Gemeinsamer Bundesausschuss

Anlage zur Zusammenfassenden Dokumentation

Beratungsverfahren nach § 137e SGB V über eine Richtlinie zur Erprobung

Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Stand: 6. Juni 2023

Unterausschuss Methodenbewertung
des Gemeinsamen Bundesausschusses

Korrespondenzadresse:
Gemeinsamer Bundesausschuss
Abteilung Methodenbewertung und Veranlasste Leistungen
Postfach 12 06 06
10596 Berlin

Tel.: +49 (0)30 – 275 838 - 0
Internet: www.g-ba.de

Inhaltsverzeichnis

Bundesanzeiger
Herausgegeben vom
Bundesministerium der Justiz
www.bundesanzeiger.de

Bekanntmachung
Veröffentlicht am Montag, 22. August 2022
BAnz AT 22.08.2022 B4

Bundesministerium für Gesundheit

Bekanntmachung eines Beschlusses des Gemeinsamen Bundesausschusses 1. über die Aufnahme von Beratungen über eine Richtlinie zur Erprobung der magnetischen Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit sowie 2. zur Ermittlung der stellungnahmeberechtigten Medizinprodukthersteller zu Beratungen des Gemeinsamen Bundesausschusses über die magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit – Aufforderung zur Meldung –

Vom 11. August 2022

1. Aufnahme von Beratungen zur Erprobungs-Richtlinie

Im Rahmen der Bescheidung eines Antrags auf Erprobung gemäß § 137e Absatz 7 des Fünften Buches Sozialgesetzbuch (SGB V) ist der Gemeinsame Bundesausschuss (G-BA) zu der Auffassung gelangt, dass der Nutzen der

- magnetischen Ösophagus-Sphinkter-Augmentation zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit, die für eine laparoskopische Fundoplicatio geeignet sind,

zwar noch nicht als hinreichend belegt anzusehen ist, die Methode aber das Potenzial einer erforderlichen Behandlungsalternative bietet. In seiner Sitzung am 18. März 2022 hat der G-BA beschlossen, die Beratungen über eine Richtlinie zur Erprobung dieser Methode gemäß § 137e SGB V aufzunehmen. In dieser Richtlinie wird die Studie konkretisiert, die die Bewertung des Nutzens dieser Methode auf einem für eine spätere Richtlinienentscheidung ausreichend sicheren Erkenntnisniveau erlauben soll.

Mit dieser Veröffentlichung soll insbesondere Sachverständigen der medizinischen Wissenschaft und Praxis, Spitzenverbänden der Selbsthilfegruppen und Patientenvertretungen sowie Verbänden von Leistungserbringern und Medizinprodukteherstellern und den jeweils betroffenen Herstellern von Medizinprodukten Gelegenheit gegeben werden, durch Beantwortung eines Fragebogens eine Ersteinschätzung zum angekündigten Beratungsgegenstand abzugeben.

Die Einschätzungen zu dem oben genannten Beratungsthema sind in deutscher Sprache anhand des Fragebogens innerhalb einer Frist von einem Monat nach der Veröffentlichung im Bundesanzeiger (bis zum 22. September 2022) möglichst in elektronischer Form an folgende E-Mail-Adresse zu senden: erprobung137e@g-ba.de.

Den Fragebogen sowie weitere Erläuterungen finden Sie auf der Internetseite des G-BA unter

https://www.g-ba.de/bewertungsverfahren/methodenbewertung/263/.

Stellungnahmeberechtigte gemäß § 91 Absatz 5 SGB V (Bundesärztekammer) und § 92 Absatz 7d SGB V (einschlägige wissenschaftliche Fachgesellschaften, Spitzenorganisationen der Medizinprodukthersteller, betroffene Medizinprodukthersteller), die eine Ersteinschätzung abgegeben haben, erhalten zudem die Gelegenheit zur Abgabe einer mündlichen Einschätzung im Rahmen einer Anhörung zum Einschätzungsverfahren. Die Terminierung der Anhörung und die Einladung übermitteln wir Ihnen in einem gesonderten Schreiben.

2. Ermittlung der stellungnahmeberechtigten Medizinprodukthersteller

– Aufforderung zur Meldung –

Der G-BA hat vor Entscheidungen über die Richtlinien nach den § 135 Absatz 1, §§ 137c und 137e SGB V zu Methoden, deren technische Anwendung maßgeblich auf dem Einsatz eines Medizinprodukts beruht, den jeweils betroffenen Medizinproduktherstellern (im Folgenden: Hersteller) Gelegenheit zur Stellungnahme zu geben. Die technische Anwendung einer Methode beruht maßgeblich auf einem Medizinprodukt, wenn ohne dessen Einbeziehung (technische Anwendung) die Methode bei der jeweiligen Indikation ihr, sie von anderen Vorgehensweisen unterscheidendes, theoretisch-wissenschaftliches Konzept verlieren würde.

Hiermit sind solche Hersteller aufgefordert sich beim G-BA zu melden, die der Auffassung sind, dass sie von Entscheidungen des G-BA zu der in Nummer 1 genannten Methode im oben genannten Sinne betroffen sind. Der G-BA prüft dann auf der Grundlage der von ihnen eingereichten Unterlagen, ob die gesetzlichen Voraussetzungen der Stellungnahmeberechtigung vorliegen.

Hierzu sind aussagekräftige Unterlagen einzureichen. Diese umfassen Ausführungen in deutscher Sprache zur

- Bezeichnung und Beschreibung des Medizinprodukts,
- Beschreibung der Einbindung des Medizinprodukts in die Methode und
- Zweckbestimmung, für die das Medizinprodukt in Verkehr gebracht wurde.

Es sind außerdem

- die medizinprodukterechtliche Konformitätserklärung beziehungsweise das Konformitätszertifikat des Medizinprodukts für das Inverkehrbringen in der Bundesrepublik Deutschland sowie
- die technische Gebrauchsanweisung

beizufügen.

Die Unterlagen sind bis zum 22. September 2022 der Geschäftsstelle des G-BA – nach Möglichkeit in elektronischer Form (zum Beispiel als Word- oder PDF-Dokumente) per E-Mail – zu übermitteln. Bitte teilen Sie uns Ihre Korrespondenz-Post- und E-Mail-Adresse unter Angabe einer Kontaktperson mit.

Sofern der G-BA in der Folge feststellen wird, dass Sie von geplanten Entscheidungen des G-BA zur oben genannten Methode betroffen sind, erhalten Sie Gelegenheit zur Abgabe einer mündlichen Einschätzung im Rahmen der Anhörung zum Einschätzungsverfahren und zu gegebenem Zeitpunkt zur Abgabe einer Stellungnahme zu Beschlussentwürfen.

Korrespondenzadresse:

Gemeinsamer Bundesausschuss
Abteilung Methodenbewertung & Veranlasste Leistungen
Postfach 12 06 06
10596 Berlin

E-Mail: erprobung137e@g-ba.de

Nachmeldungen zur Ermittlung der stellungnahmeberechtigten Medizinprodukthersteller sind zulässig. Insoweit ist zu beachten, dass bis zu der Entscheidung über die Nachmeldung die Wahrnehmung des Stellungnahmerechts nicht möglich ist.

Berlin, den 11. August 2022

Gemeinsamer Bundesausschuss
Unterausschuss Methodenbewertung

Die Vorsitzende
Lelgemann

Stand 20.07.2022

Gemeinsamer Bundesausschuss

Gelegenheit zur Abgabe erster Einschätzungen

zu Beratungen des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung: Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Am 18. März 2022 hat der Gemeinsame Bundesausschuss (G-BA) beschlossen, Beratungen über eine Richtlinie gemäß § 137e Absatz 1 des Fünften Buches Sozialgesetzbuch (SGB V) zur Erprobung der

magnetischen Ösophagus-Sphinkter-Augmentation zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit, die für eine laparoskopischen Fundoplicatio geeignet sind,

aufzunehmen.

Um den G-BA in die Lage zu versetzen, eine abschließende Bewertung des Nutzens der vorgenannten Methode durchzuführen, sollen im Wege der Erprobung die hierfür nach den §§ 135 und 137c SGB V i. V. m. den Vorgaben der Verfahrensordnung des G-BA (VerfO) notwendigen Erkenntnisse für die Bewertung des Nutzens der Methode gewonnen werden. Die zu diesem Zweck notwendige Studie soll durch eine unabhängige wissenschaftliche Institution (UWI) nach Maßgabe dieser Richtlinie entworfen, durchgeführt und ausgewertet werden. Die Ausgestaltung des Studiendesigns ist – soweit nicht im Folgenden näher bestimmt – von der UWI auf der Basis des Standes der wissenschaftlichen Erkenntnisse vorzunehmen und zu begründen.

Gemäß 2. Kapitel § 6 VerfO erhalten Sie Gelegenheit zur Abgabe einer ersten Einschätzung zum angekündigten Beratungsgegenstand. Bitte verwenden Sie zur Abgabe Ihrer Einschätzung den nachfolgenden Fragebogen.

Bitte belegen Sie Ihre Ausführungen jeweils durch Angabe von Quellen unter Nutzung der beigefügten Literaturliste (siehe Anlage). Bitte fügen Sie die Publikationen – soweit möglich – in Kopie bei.

Wir bitten Sie, den Fragebogen als Word-Dokument und alle weiteren Unterlagen als PDF-Dokumente per E-Mail an **erprobung137e@g-ba.de** zu übersenden. Die Frist zur Abgabe Ihrer Einschätzung endet am TT. Monat 2022.

Mit der Abgabe Ihrer Einschätzung erklären Sie sich damit einverstanden, dass diese, auch auszugsweise, in einem Bericht des G-BA wiedergegeben werden kann, der mit Abschluss der Beratung zu jedem Thema erstellt und der Öffentlichkeit via Internet zugänglich gemacht wird.

Gemeinsamer Bundesausschuss

Fragebogen

Funktion des Einschätzenden

Bitte geben Sie an, in welcher Funktion Sie diese Einschätzung abgeben (z. B. Verband, Institution, Hersteller, Leistungserbringer, Privatperson).

Gemeinsamer Bundesausschuss

Mit der Erprobungsstudie soll nachgewiesen werden, dass bei erwachsenen Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen Gastroösophagealer Refluxkrankheit (GERD) die Magnetische Ösophagus-Sphinkter-Augmentation (MSA) im Vergleich zur laparoskopischen Fundoplicatio (LF) bezüglich der gesundheitsbezogenen Lebensqualität nicht unterlegen ist.

Überlegungen des G-BA	**Fragen des G-BA**	**Einschätzung**
Population		
In die Erprobungsstudie einzuschließen sind erwachsene Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen GERD infolge einer Schwäche oder Insuffizienz des unteren ösophagealen Sphinkters, die trotz maximaler konservativ-medikamentöser Refluxtherapie entweder weiterhin unter GERD-Symptomen leiden oder bereits eine GERD-assoziierte Komplikation im oberen Gastrointestinaltrakt erlitten haben. Auszuschließen sind Patientinnen und Patienten mit Kontraindikationen für die Prüf- oder Vergleichsintervention. Die genauen Ein- und Ausschlusskriterien sind im Rahmen der konkreten Studienplanung festzulegen.	Ist dies die aus Ihrer Sicht treffende Beschreibung der Studienpopulation? Falls nein, wie würden Sie die Population definieren?	Klicken Sie hier, um einen Text einzugeben.

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
Intervention		
Die Prüfintervention besteht in der Magnetischen Ösophagus-Sphinkter-Augmentation: Über einen laparoskopischen Eingriff wird ein der Größe des unteren Ösophagus-Sphinkters (UÖS) entsprechendes, flexibles, ringförmiges Implantat um den UÖS gelegt, welches über seine magnetischen Anziehungskräfte die Verschlussfunktion des UÖS in physiologischer Weise unterstützt.	Stimmen Sie mit der Überlegung des G-BA zur Intervention überein? Falls nein, wie würden Sie die Intervention definieren?	Klicken Sie hier, um einen Text einzugeben.
Vergleichsintervention/Kontrolle		
Die angemessene Vergleichsintervention ist die laparoskopische Fundoplicatio.	Stimmen Sie mit der Überlegung des G-BA zur Vergleichsintervention überein? Falls nein, wie würden Sie diese definieren?	Klicken Sie hier, um einen Text einzugeben.

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
Endpunkte		
Primärer Endpunkt: • gesundheitsbezogene Lebensqualität (mittels eines krankheitsspezifischen, validierten Instruments zu messen, z. B. *GERD-HRQL*) Die genaue Operationalisierung des Endpunkts ist im Rahmen der konkreten Studienplanung festzulegen.	Stimmen Sie mit der Überlegung des G-BA zum primären Endpunkt überein? Falls nein, was ist aus Ihrer Sicht ein angemessener primärer Endpunkt für die Erprobungsstudie und welche validierten Erhebungsinstrumente gibt es nach Ihrer Kenntnis für diesen von Ihnen vorgeschlagenen Endpunkt?	Klicken Sie hier, um einen Text einzugeben.
Als sekundäre Endpunkte sind insbesondere zu erheben: • GERD-bezogene Symptome (insbesondere Sodbrennen und Regurgitationen) • postoperative Morbidität (insbesondere die Fähigkeit wieder essen zu können und die Rückkehr zu normalen Alltagsaktivitäten) • Re-Hospitalisierungen • unerwünschte Ereignisse (insbesondere die Unfähigkeit aufzustoßen oder zu erbrechen)	Stimmen Sie mit der Überlegung des G-BA zu den sekundären Endpunkten überein? Welche validierten Erhebungsinstrumente zu diesen Endpunkten halten Sie für geeignet? Sollten Ihrer Meinung nach weitere bzw. andere sekundäre Endpunkte ergänzend in der Erprobungsstudie untersucht werden? In diesem Fall benennen Sie bitte die entsprechenden validierten Erhebungsinstrumente.	Klicken Sie hier, um einen Text einzugeben.

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
Studientyp und Beobachtungszeitraum		
Die Erprobungsstudie ist als randomisierte, kontrollierte Studie (RCT) multizentrisch durchzuführen.	Stimmen Sie mit der Überlegung des G-BA zum Studientyp überein? Falls nein, welche Vorgaben zum Studientyp sollten definiert werden?	Klicken Sie hier, um einen Text einzugeben.
Die Endpunkterhebung ist zu verblinden. Auf eine Verblindung der Patientinnen und Patienten könnte verzichtet werden, weil eine eventuelle Durchführung von Diagnostik per Magnetresonanztomografie nach LF problemlos, nach MSA jedoch nur eingeschränkt möglich ist, sodass eine Entblindung für Notfallsituationen ermöglicht werden müsste, was organisatorisch und rechtlich zu erheblichem Mehraufwand führen würde.	Stimmen Sie mit der Überlegung des G-BA zur Verblindung überein? Falls nein, welche Einwände oder Vorschläge haben Sie gegen diese Einschätzung?	Klicken Sie hier, um einen Text einzugeben.
Die patientenindividuelle Beobachtungszeit soll mindestens 12 Monate ab Randomisierung betragen.	Eine Beobachtungszeit von mindestens 12 Monaten nach Randomisierung wird als angemessen angesehen, um auch die Dauerhaftigkeit der Veränderungen abschätzen zu können. Stimmen Sie mit dieser Überlegung überein? Falls nein, welche Einwände oder Vorschläge haben Sie bzgl. dieser Vorgabe?	Klicken Sie hier, um einen Text einzugeben.

Erfassung und Dokumentation bestimmter Parameter		
Die Art und Anzahl weiterer therapeutischer Interventionen mit Bezug zur Grunderkrankung oder mit möglichem Einfluss auf die zu erfassenden Endpunkte (beispielsweise die Einnahme von Protonenpumpeninhibitoren) sollten dokumentiert werden.	Stimmen Sie mit der Überlegung des G-BA überein? Falls nein, welche Einwände oder Vorschläge haben Sie bzgl. dieser Vorgabe?	Klicken Sie hier, um einen Text einzugeben.

Ergänzende Fragen	
Wie viele Studienzentren in Deutschland kämen für die Studie in Frage?	Klicken Sie hier, um einen Text einzugeben.
Wie viele Studienzentren sollten initiiert werden, um die Studie in angemessener Zeit abzuschließen?	Klicken Sie hier, um einen Text einzugeben.
Welche Maßnahmen wären erforderlich, um eine zügige Rekrutierung zu gewährleisten?	Klicken Sie hier, um einen Text einzugeben.
Gibt es aus Ihrer Sicht Aspekte zu berücksichtigen, welche die geplante Studiendurchführung erschweren könnten? (Beispielsweise geplante oder laufende Studien mit Rekrutierung derselben Patientengruppen im Indikationsgebiet der Erprobungsstudie)	Klicken Sie hier, um einen Text einzugeben.
Welche Anforderungen, insbesondere hinsichtlich der personellen, technischen und räumlichen Ausstattung, sind aus Ihrer Sicht zur Erbringung der Methode im Rahmen einer Studie zu stellen? Bitte berücksichtigen Sie hierbei auch mögliche periprozedurale Risiken ihrer Anwendung.	Klicken Sie hier, um einen Text einzugeben.
Wird bei den genannten Eckpunkten die Versorgungsrealität in Hinblick auf die Durchführbarkeit der Erprobung und der Leistungserbringung angemessen berücksichtigt?	Klicken Sie hier, um einen Text einzugeben.

Ergänzende Fragen	
Bitte benennen Sie ggf. zusätzliche Aspekte, die im Rahmen der Erstellung der Erprobungs-Richtlinie berücksichtigt werden sollten.	Klicken Sie hier, um einen Text einzugeben.

Überlegungen des G-BA zur näherungsweisen Fallzahlschätzung	**Wie lautet Ihre Einschätzung?**
Die folgenden Ausführungen zur Fallzahlschätzung sind nicht als verbindliche Kalkulation, sondern als näherungsweise Schätzung der benötigten Fallzahlen zu verstehen. Zur Schätzung der Fallzahl ist die Nichtunterlegenheitsfragestellung zum Endpunkt gesundheitsbezogene Lebensqualität maßgeblich und hierbei insbesondere die Wahl der Nichtunterlegenheitsgrenze. Es ist davon auszugehen, dass eine Studiengröße, die zum Nachweis einer zwischen MSA und LF vergleichbaren gesundheitsbezogenen Lebensqualität ausreicht, auch hinreichend sicher zeigen kann, dass die MSA gegenüber der LF Vorteile in Bezug auf die postoperative Morbidität und die unerwünschten Ereignisse (insbesondere die Unfähigkeit aufzustoßen oder zu erbrechen) hat. Für die Prüfung auf Nichtunterlegenheit wird die standardisierte Mittelwertdifferenz Hedges'g angewendet. Die Nichtunterlegenheitsgrenze von 0,25 wird für den Endpunkt gesundheitsbezogene Lebensqualität (gemessen mittels GERD-HRQL) als adäquat betrachtet. Hieraus ergibt sich als grobe Approximation eine Fallzahl von ca. 400 Patientinnen und Patienten, was der Kategorie einer mittleren Studiengröße (100 bis < 500 Patientinnen und Patienten) entspricht. Eine exakte Fallzahlkalkulation muss im Rahmen der konkreten Studienplanung erfolgen.	Klicken Sie hier, um einen Text einzugeben.

Überlegungen des G-BA zur näherungsweisen Fallzahlschätzung	**Wie lautet Ihre Einschätzung?**
Schätzung der Overheadkosten der Erprobungsstudie (Beispiel)	**Wie lautet Ihre Einschätzung?**
Für Studien mit mittlerer Fallzahl und mittlerem Aufwand lässt sich ein studienspezifischer Aufwand in Höhe von etwa 5.500 € je Teilnehmerin oder Teilnehmer beziffern. Auf der Basis dieser Annahmen lassen sich geschätzte Studienkosten von 2,2 Million € berechnen.	Klicken Sie hier, um einen Text einzugeben.

Stand 20.07.2022

Gemeinsamer Bundesausschuss

Gelegenheit zur Abgabe erster Einschätzungen

zu Beratungen des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung: Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Am 18. März 2022 hat der Gemeinsame Bundesausschuss (G-BA) beschlossen, Beratungen über eine Richtlinie gemäß § 137e Absatz 1 des Fünften Buches Sozialgesetzbuch (SGB V) zur Erprobung der

magnetischen Ösophagus-Sphinkter-Augmentation zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit, die für eine laparoskopischen Fundoplicatio geeignet sind,

aufzunehmen.

Um den G-BA in die Lage zu versetzen, eine abschließende Bewertung des Nutzens der vorgenannten Methode durchzuführen, sollen im Wege der Erprobung die hierfür nach den §§ 135 und 137c SGB V i. V. m. den Vorgaben der Verfahrensordnung des G-BA (VerfO) notwendigen Erkenntnisse für die Bewertung des Nutzens der Methode gewonnen werden. Die zu diesem Zweck notwendige Studie soll durch eine unabhängige wissenschaftliche Institution (UWI) nach Maßgabe dieser Richtlinie entworfen, durchgeführt und ausgewertet werden. Die Ausgestaltung des Studiendesigns ist – soweit nicht im Folgenden näher bestimmt – von der UWI auf der Basis des Standes der wissenschaftlichen Erkenntnisse vorzunehmen und zu begründen.

Gemäß 2. Kapitel § 6 VerfO erhalten Sie Gelegenheit zur Abgabe einer ersten Einschätzung zum angekündigten Beratungsgegenstand. Bitte verwenden Sie zur Abgabe Ihrer Einschätzung den nachfolgenden Fragebogen.

Bitte belegen Sie Ihre Ausführungen jeweils durch Angabe von Quellen unter Nutzung der beigefügten Literaturliste (siehe Anlage). Bitte fügen Sie die Publikationen – soweit möglich – in Kopie bei.

Wir bitten Sie, den Fragebogen als Word-Dokument und alle weiteren Unterlagen als PDF-Dokumente per E-Mail an **erprobung137e@g-ba.de** zu übersenden. Die Frist zur Abgabe Ihrer Einschätzung endet am TT. Monat 2022.

Mit der Abgabe Ihrer Einschätzung erklären Sie sich damit einverstanden, dass diese, auch auszugsweise, in einem Bericht des G-BA wiedergegeben werden kann, der mit Abschluss der Beratung zu jedem Thema erstellt und der Öffentlichkeit via Internet zugänglich gemacht wird.

Gemeinsamer Bundesausschuss

Fragebogen

Funktion des Einschätzenden

Bitte geben Sie an, in welcher Funktion Sie diese Einschätzung abgeben (z. B. Verband, Institution, Hersteller, Leistungserbringer, Privatperson).

Hersteller (für Torax Medical) bzw. Antragsteller

Gemeinsamer Bundesausschuss

Mit der Erprobungsstudie soll nachgewiesen werden, dass bei erwachsenen Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen Gastroösophagealer Refluxkrankheit (GERD) die Magnetische Ösophagus-Sphinkter-Augmentation (MSA) im Vergleich zur laparoskopischen Fundoplicatio (LF) bezüglich der gesundheitsbezogenen Lebensqualität nicht unterlegen ist.

Überlegungen des G-BA	**Fragen des G-BA**	**Einschätzung**
Population		
In die Erprobungsstudie einzuschließen sind erwachsene Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen GERD infolge einer Schwäche oder Insuffizienz des unteren ösophagealen Sphinkters, die trotz maximaler konservativ-medikamentöser Refluxtherapie entweder weiterhin unter GERD-Symptomen leiden oder bereits eine GERD-assoziierte Komplikation im oberen Gastrointestinaltrakt erlitten haben. Auszuschließen sind Patientinnen und Patienten mit Kontraindikationen für die Prüf- oder Vergleichsintervention. Die genauen Ein- und Ausschlusskriterien sind im Rahmen der konkreten Studienplanung festzulegen.	Ist dies die aus Ihrer Sicht treffende Beschreibung der Studienpopulation? Falls nein, wie würden Sie die Population definieren?	Ja, dies ist eine treffende Beschreibung der Studienpopulation

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
Intervention		
Die Prüfintervention besteht in der Magnetischen Ösophagus-Sphinkter-Augmentation: Über einen laparoskopischen Eingriff wird ein der Größe des unteren Ösophagus-Sphinkters (UÖS) entsprechendes, flexibles, ringförmiges Implantat um den UÖS gelegt, welches über seine magnetischen Anziehungskräfte die Verschlussfunktion des UÖS in physiologischer Weise unterstützt.	Stimmen Sie mit der Überlegung des G-BA zur Intervention überein? Falls nein, wie würden Sie die Intervention definieren?	Ja, dies ist eine treffende Beschreibung der Intervention
Vergleichsintervention/Kontrolle		
Die angemessene Vergleichsintervention ist die laparoskopische Fundoplicatio.	Stimmen Sie mit der Überlegung des G-BA zur Vergleichsintervention überein? Falls nein, wie würden Sie diese definieren?	Nach Versagen der konservativ-medikamentösen Refluxtherapie ist die laparoskopische Fundoplicatio heute das häufigste Operationsverfahren zur Behandlung einer GERD, wobei diese in einer Reihe verschiedener Varianten, z.B. nach Nissen oder Toupet durchgeführt wird. Im Sinne einer klaren Vergleichbarkeit wäre zu überlegen die Vergleichsintervention näher einzugrenzen, z.B. als laparoskopische Nissen Fundoplicatio. Auf der anderen Seite spiegelt die Zulassung der gängigen Varianten die Versorgungsrealität wider und käme jenen Operateuren entgegen, die eine Präferenz für die eine oder andere Variante der laparoskopischen

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
		Fundoplicatio haben, was auch einer zügigen Rekrutierung dienlich wäre.

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
Endpunkte		
Primärer Endpunkt: • gesundheitsbezogene Lebensqualität (mittels eines krankheitsspezifischen, validierten Instruments zu messen, z. B. *GERD-HRQL*) Die genaue Operationalisierung des Endpunkts ist im Rahmen der konkreten Studienplanung festzulegen.	Stimmen Sie mit der Überlegung des G-BA zum primären Endpunkt überein? Falls nein, was ist aus Ihrer Sicht ein angemessener primärer Endpunkt für die Erprobungsstudie und welche validierten Erhebungsinstrumente gibt es nach Ihrer Kenntnis für diesen von Ihnen vorgeschlagenen Endpunkt?	Ja, die gesundheitsbezogene Lebensqualität ist ein geeigneter Primärer Endpunkt und der GERD-HRQL ist ein geeignetes und von uns bevorzugtes Instrument diese zu erfassen [1]. Der GERD-HRQL ist relativ leicht zu erheben und in der Literatur als Referenz weit verbreitet, so dass die Ergebnisse dieser Studie vergleichbar werden und in zukünftige Meta-Analysen einfließen können. In den bisher publizierten Studien zum Vergleich der Magnetischen Ösophagus-Sphinkter-Augmentation und der laparoskopische Fundoplicatio konnte kein Unterschied im Hinblick auf den mittels GERD-HRQL gemessenen Endpunkt gesundheitsbezogene Lebensqualität festgestellt werden. Soweit angegeben, wird die Signifikanz für den Unterschied (p-Wert) im Bereich 0,9 angegeben, womit eine hohe Anzahl von Studienteilnehmern zu erwarten ist [2]–[9]. Eine Erweiterung zu einem ko-primären Endpunkt unter Einbeziehung der Verweildauer im Krankenhaus und/oder der Rückkehr zur normalen Aktivität würde die Aussagekraft der Studie verstärken und gegen ein indifferentes Ergebnis mangels Studienteilnehmern absichern.

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
		Die letzten verfügbaren Daten des InEK zu mittlerer Verweildauer vom Januar bis Mai 2022 besagen für MSA 3,7 Tage und LF 6,6 Tage. Hierbei ist zu berücksichtigen, dass die Verweildauer für MSA höchstwahrscheinlich durch die GKV-Vergütung verzerrt ist, da diese bei weniger als drei Tagen einen erheblichen Vergütungsabschlag vorsieht. Diese Verzerrungsmöglichkeit sollte bei Einbeziehung der Verweildauer als Endpunkt in die Studie durch eine verweildauerunabhängige Vergütung berücksichtigt werden.
Als sekundäre Endpunkte sind insbesondere zu erheben: • GERD-bezogene Symptome (insbesondere Sodbrennen und Regurgitationen) • postoperative Morbidität (insbesondere die Fähigkeit wieder essen zu können und die Rückkehr zu normalen Alltagsaktivitäten) • Re-Hospitalisierungen • unerwünschte Ereignisse (insbesondere die Unfähigkeit aufzustoßen oder zu erbrechen)	Stimmen Sie mit der Überlegung des G-BA zu den sekundären Endpunkten überein? Welche validierten Erhebungsinstrumente zu diesen Endpunkten halten Sie für geeignet? Sollten Ihrer Meinung nach weitere bzw. andere sekundäre Endpunkte ergänzend in der Erprobungsstudie untersucht werden? In diesem Fall benennen Sie bitte die entsprechenden validierten Erhebungsinstrumente.	Dies sind die wesentlichen, geeigneten sekundären Endpunkte. Dem hinzuzufügen wäre lediglich die post-operative Einnahme von Protonenpumpeninhibitoren, der Standardmedikation bei GERD. Sodbrennen wird mit den Fragen 1-6 des GERD-HRQL hinreichen erfasst [1]. Regurgitationen werden mit dem Foregut Symptom Questionnaire (FSQ) besser erfasst [10]. Fähigkeit wieder essen zu können sollte besser als Schluckbeschwerden/ Dysphagie verstanden werden. Dies wird mit den Fragen 7 und 8 des GERD-HRQL und dem FSQ erfasst.

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
		Re-Hospitalisierungen sollte erweitert werden auf post-operative therapeutische Interventionen, insbesondere • Revision einer Fundoplicatio • Entfernung eines MSA-Implantats • Konversion MSA nach LF oder LF nach MSA • Dilatation
Studientyp und Beobachtungszeitraum		
Die Erprobungsstudie ist als randomisierte, kontrollierte Studie (RCT) multizentrisch durchzuführen.	Stimmen Sie mit der Überlegung des G-BA zum Studientyp überein? Falls nein, welche Vorgaben zum Studientyp sollten definiert werden?	Methodisch ist ein RCT der Goldstandard und sollte, wenn möglich angestrebt werden. Jedoch hat sich gezeigt, dass eine sinnvolle Durchführung – hinreichende Anzahl Patienten in einem vertretbaren Zeitraum rekrutieren zu können – bei chirurgischen Verfahren zumindest schwierig, wenn nicht unmöglich ist. Dies wird gravierender, je unterschiedlicher die beiden Verfahren sind und die Patientenpräferenz eine immer größere Rolle spielt. Bei den beiden hier infrage kommenden Verfahren könnte sich – nach den Erfahrungen in der weltweiten Anwendung der Methode - dieser Effekt als gravierend erweisen [11], [12]. Von daher sollte von vornherein über parallele Registerarme nachgedacht werden, bei denen die Ergebnisse von Patienten mit einer klaren Präferenz für das eine oder andere

Überlegungen des G-BA	Fragen des G-BA	Einschätzung
		Verfahren erfasst werden. Siehe hierzu Grant 2013 [13].
Die Endpunkterhebung ist zu verblinden. Auf eine Verblindung der Patientinnen und Patienten könnte verzichtet werden, weil eine eventuelle Durchführung von Diagnostik per Magnetresonanztomografie nach LF problemlos, nach MSA jedoch nur eingeschränkt möglich ist, sodass eine Entblindung für Notfallsituationen ermöglicht werden müsste, was organisatorisch und rechtlich zu erheblichem Mehraufwand führen würde.	Stimmen Sie mit der Überlegung des G-BA zur Verblindung überein? Falls nein, welche Einwände oder Vorschläge haben Sie gegen diese Einschätzung?	Die vorgeschlagenen Endpunkte werden alle über von den Patienten selbständig auszufüllenden Fragebogen erhoben. Es finden keine Untersuchungen statt, bei denen der subjektive Eindruck eines Untersuchers eine Rolle spielen würde. Von daher ist der Effekt der Verblindung der Endpunkterhebung bestenfalls marginal, wenn es überhaupt einen Effekt geben sollte. Dagegen könnte die nicht-Verblindung der Patientinnen und Patienten einen wesentlich größeren Einfluss auf das Ergebnis haben, wäre aber aus angegebenen Gründen schwer darstellbar, wenn nicht sogar unmöglich.
Die patientenindividuelle Beobachtungszeit soll mindestens 12 Monate ab Randomisierung betragen.	Eine Beobachtungszeit von mindestens 12 Monaten nach Randomisierung wird als angemessen angesehen, um auch die Dauerhaftigkeit der Veränderungen abschätzen zu können. Stimmen Sie mit dieser Überlegung überein? Falls nein, welche Einwände oder Vorschläge haben Sie bzgl. dieser Vorgabe?	Wir stimmen zu, wobei die Wirksamkeit beider Verfahren über 5 Jahre hinaus sehr gut nachgewiesen wurde – für LF siehe z.B. [13], [14], [15], für MSA [3], [16], [17], [18] – was eine Beobachtungszeit von 6 Monaten nach Behandlung rechtfertigen würde.

Erfassung und Dokumentation bestimmter Parameter		
Die Art und Anzahl weiterer therapeutischer Interventionen mit Bezug zur Grunderkrankung oder mit möglichem Einfluss auf die zu erfassenden Endpunkte (beispielsweise die Einnahme von Protonenpumpeninhibitoren) sollten dokumentiert werden.	Stimmen Sie mit der Überlegung des G-BA überein? Falls nein, welche Einwände oder Vorschläge haben Sie bzgl. dieser Vorgabe?	Die Einnahme von Protonenpumpeninhibitoren sollte als sekundärer Endpunkt aufgenommen werden.

Ergänzende Fragen	
Wie viele Studienzentren in Deutschland kämen für die Studie in Frage?	In 2019 haben 24 Zentren in Deutschland 10 oder mehr Patienten oder Patientinnen mit der Magnetischen Ösophagus-Sphinkter-Augmentation versorgt, zusammen 432 Patienten oder Patientinnen. Im gleichen Zeitraum haben diese Zentren 471 Patienten oder Patientinnen mit Fundoplicatio versorgt, wobei lediglich 13 dieser Zentren jeweils mehr als 10 solcher Eingriffe vorgenommen hatten.
Wie viele Studienzentren sollten initiiert werden, um die Studie in angemessener Zeit abzuschließen?	Mit den oben genannten Zahlen sollte eine Studie mit 15 ausgesuchten Zentren innerhalb von 3 Jahren durchgeführt werden können – „vertretbare“ Anzahl von erforderlichen Studienteilnehmern vorausgesetzt.
Welche Maßnahmen wären erforderlich, um eine zügige Rekrutierung zu gewährleisten?	Adäquate Vergütung der Magnetischen Ösophagus-Sphinkter-Augmentation.
Gibt es aus Ihrer Sicht Aspekte zu berücksichtigen, welche die geplante Studiendurchführung erschweren könnten? (Beispielsweise geplante oder laufende Studien mit Rekrutierung derselben Patientengruppen im Indikationsgebiet der Erprobungsstudie)	Es werden zurzeit für die RETHINK REFLUX Registry (ClinicalTrials.gov Identifikationsnummer: NCT04253392) auch an zwei deutschen Kliniken Patienten eingeschlossen (zusammen 500 Patienten an 37 Zentren in den USA und Europa). Die Patientenrekrutierung für diese Registerstudie sollte aber bis zum Beginn der hier geplanten RCT-Studie abgeschlossen sein und einer Rekrutierung für letztgenannte nicht im Wege stehen.

Ergänzende Fragen	
Welche Anforderungen, insbesondere hinsichtlich der personellen, technischen und räumlichen Ausstattung, sind aus Ihrer Sicht zur Erbringung der Methode im Rahmen einer Studie zu stellen? Bitte berücksichtigen Sie hierbei auch mögliche periprozedurale Risiken ihrer Anwendung.	Es wird ein laparoskopischer Operationsplatz als Basisausstattung und die Möglichkeit einer anschließenden stationären Überwachung benötigt. Die Implantation sollte nur von Ärzten durchgeführt werden, die über hinreichend Erfahrung mit laparoskopischen Operationen gegen Reflux verfügen, insbesondere der LF und MSA.
Wird bei den genannten Eckpunkten die Versorgungsrealität in Hinblick auf die Durchführbarkeit der Erprobung und der Leistungserbringung angemessen berücksichtigt?	Ja, die vorgeschlagenen Studienparameter, mit unseren Ergänzungen spiegeln die Versorgungsrealität wider.
Bitte benennen Sie ggf. zusätzliche Aspekte, die im Rahmen der Erstellung der Erprobungs-Richtlinie berücksichtigt werden sollten.	Klicken Sie hier, um einen Text einzugeben.

Überlegungen des G-BA zur näherungsweisen Fallzahlschätzung	**Wie lautet Ihre Einschätzung?**
Die folgenden Ausführungen zur Fallzahlschätzung sind nicht als verbindliche Kalkulation, sondern als näherungsweise Schätzung der benötigten Fallzahlen zu verstehen. Zur Schätzung der Fallzahl ist die Nichtunterlegenheitsfragestellung zum Endpunkt gesundheitsbezogene Lebensqualität maßgeblich und hierbei insbesondere die Wahl der Nichtunterlegenheitsgrenze. Es ist davon auszugehen, dass eine Studiengröße, die zum Nachweis einer zwischen MSA und LF vergleichbaren gesundheitsbezogenen Lebensqualität ausreicht, auch hinreichend sicher zeigen kann, dass die MSA gegenüber der LF Vorteile in Bezug auf die postoperative Morbidität und die unerwünschten Ereignisse (insbesondere die Unfähigkeit aufzustoßen oder zu erbrechen) hat.	In den bisher publizierten Studien zum Vergleich der Magnetischen Ösophagus-Sphinkter-Augmentation und der laparoskopische Fundoplicatio konnte kein Unterschied im Hinblick auf den mittels GERD-HRQL gemessenen Endpunkt gesundheitsbezogene Lebensqualität festgestellt werden. Soweit angegeben, wird die Signifikanz für den Unterschied (p-Wert) im Bereich 0,9 angegeben, womit eine sehr hohe Anzahl von Studienteilnehmern zu erwarten ist, die möglicherweise sehr schwer oder unrealistisch zu erreichen ist. Von daher wird die Kombination mit anderen, oben genannten Endpunkten in einen ko-primären Endpunkt empfohlen, um einen erfolgreichen Abschluss der Studie zu ermöglichen.

Überlegungen des G-BA zur näherungsweisen Fallzahlschätzung	**Wie lautet Ihre Einschätzung?**
Für die Prüfung auf Nichtunterlegenheit wird die standardisierte Mittelwertdifferenz Hedges'g angewendet. Die Nichtunterlegenheitsgrenze von 0,25 wird für den Endpunkt gesundheitsbezogene Lebensqualität (gemessen mittels GERD-HRQL) als adäquat betrachtet. Hieraus ergibt sich als grobe Approximation eine Fallzahl von ca. 400 Patientinnen und Patienten, was der Kategorie einer mittleren Studiengröße (100 bis < 500 Patientinnen und Patienten) entspricht. Eine exakte Fallzahlkalkulation muss im Rahmen der konkreten Studienplanung erfolgen.	
Schätzung der Overheadkosten der Erprobungsstudie (Beispiel)	**Wie lautet Ihre Einschätzung?**
Für Studien mit mittlerer Fallzahl und mittlerem Aufwand lässt sich ein studienspezifischer Aufwand in Höhe von etwa 5.500 € je Teilnehmerin oder Teilnehmer beziffern. Auf der Basis dieser Annahmen lassen sich geschätzte Studienkosten von 2,2 Million € berechnen.	

Literaturliste

[1] V. Velanovich, “The development of the GERD-HRQL symptom severity instrument,” *Dis. Esophagus*, vol. 20, no. 2, pp. 130–134, Apr. 2007.

[2] L. Bonavina, T. Horbach, S. F. Schoppmann, and J. DeMarchi, “Three-year clinical experience with magnetic sphincter augmentation and laparoscopic fundoplication.,” *Surg. Endosc.*, Jul. 2020.

[3] E. Asti, G. Bonitta, A. Lovece, V. Lazzari, and L. Bonavina, “Longitudinal comparison of quality of life in patients undergoing laparoscopic Toupet fundoplication versus magnetic sphincter augmentation: Observational cohort study with propensity score analysis.,” *Medicine (Baltimore).*, vol. 95, no. 30, p. e4366, Jul. 2016.

[4] B. E. Louie, A. S. Farivar, D. Shultz, C. Brennan, E. Vallières, and R. W. Aye, “Short-term outcomes using magnetic sphincter augmentation versus Nissen fundoplication for medically resistant gastroesophageal reflux disease.,” *Ann. Thorac. Surg.*, vol. 98, no. 2, pp. 495–498, Aug. 2014.

[5] J. L. Reynolds, J. Zehetner, P. Wu, S. Shah, N. Bildzukewicz, and J. C. Lipham, “Laparoscopic Magnetic Sphincter Augmentation vs Laparoscopic Nissen Fundoplication: A Matched-Pair Analysis of 100 Patients.,” *J. Am. Coll. Surg.*, vol. 221, no. 1, pp. 123–128, Jul. 2015.

[6] W. O. Richards and C. McRae, “Comparative Analysis of Laparoscopic Fundoplication and Magnetic Sphincter Augmentation for the Treatment of Medically Refractory GERD.,” *Am. Surg.*, vol. 84, no. 11, pp. 1762–1767, Nov. 2018.

[7] E. G. Sheu, P. Nau, B. Nath, B. Kuo, and D. W. Rattner, “A comparative trial of laparoscopic magnetic sphincter augmentation and Nissen fundoplication.,” *Surg. Endosc.*, vol. 29, no. 3, pp. 505–509, Mar. 2015.

[8] H. F. Warren *et al.*, “Multi-institutional outcomes using magnetic sphincter augmentation versus Nissen fundoplication for chronic gastroesophageal reflux disease.,” *Surg. Endosc.*, vol. 30, no. 8, pp. 3289–3296, Aug. 2016.

[9] J. L. Reynolds *et al.*, “Charges, outcomes, and complications: a comparison of magnetic sphincter augmentation versus laparoscopic Nissen fundoplication for the treatment of GERD.,” *Surg. Endosc.*, vol. 30, no. 8, pp. 3225–3230, Aug. 2016.

[10] M. Costantini *et al.*, “Value of physiologic assessment of foregut symptoms in a surgical practice,” *Surgery*, vol. 114, no. 4, pp. 780–787, 1993.

[11] S. J. Spechler *et al.*, “Randomized Trial of Medical versus Surgical Treatment for Refractory Heartburn,” *N. Engl. J. Med.*, vol. 381, no. 16, pp. 1513–1523, Oct. 2019.

[12] J. G. Hunter *et al.*, “Efficacy of transoral fundoplication vs omeprazole for treatment of regurgitation in a randomized controlled trial.,” *Gastroenterology*, vol. 148, no. 2, pp. 324-333.e5, Feb. 2015.

[13] A. M. Grant *et al.*, “Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial),” *Health Technol. Assess. (Rockv).*, vol. 17, no. 22, pp. 1–167, 2013.

[14] J. P. Galmiche *et al.*, “Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD: The LOTUS randomized clinical trial,” *JAMA - J. Am. Med. Assoc.*, vol. 305, no. 19, pp. 1969–1977, May 2011.

[15] M. Anvari *et al.*, “A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for the treatment of patients with chronic gastroesophageal reflux disease (GERD): 3-year outcomes.,” *Surg. Endosc.*, vol. 25, no. 8, pp. 2547–2554, Aug. 2011.

[16] D. Ferrari, E. Asti, V. Lazzari, S. Siboni, D. Bernardi, and L. Bonavina, “Six to 12-year outcomes of magnetic sphincter augmentation for gastroesophageal reflux disease.,” *Sci. Rep.*, vol. 10, no. 1, p. 13753, Aug. 2020.

[17] R. A. Ganz *et al.*, “Long-term Outcomes of Patients Receiving a Magnetic Sphincter Augmentation Device for Gastroesophageal Reflux.,” *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.*, vol. 14, no. 5, pp. 671–677, May 2016.

[18] S. M. O’Neill, A. D. Jalilvand, J. S. Colvin, K. R. Haisley, and K. A. Perry, “S148: Long-term patient-reported outcomes of laparoscopic magnetic sphincter augmentation versus Nissen fundoplication: a 5-year follow-up study.,” *Surg. Endosc.*, Jan. 2022.

Diseases of the Esophagus (2007) **20**, 130–134
DOI: 10.1111/j.1442-2050.2007.00658.x

DISEASES OF THE ESOPHAGUS

Original article

The development of the GERD-HRQL symptom severity instrument

V. Velanovich

Division of General Surgery, Henry Ford Hospital, Detroit, Michigan, USA

SUMMARY. **The Gastroesophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) instrument was introduced approximately 10 years ago to provide a quantitative method of measuring symptom severity in gastroesophageal reflux disease (GERD). Since that time the instrument has been used to assess treatment response to medication, endoscopic procedures, and surgery for GERD. However, the development of the instrument has progressed over the course of several years, and there is no one source which reviews this progress. The purpose of this article is to summarize the development and testing of the GERD-HRQL. The GERD-HRQL was initially developed to measure the typical symptoms of GERD. It was initially determined to have face validity and subsequent studies assessed its content validity, criterion validity, concurrent validity, predictive validity and construct validity. Reliability was determined by the test-retest method. Responsiveness was determined by the effects of treatment. This instrument is practical, with little administrative burden. There are few missing responses. Because there are 51 possible scores, the instrument has a high level of precision; and because of the response anchors, cannot have a floor effect, and only 4/372 patients reached the highest score of 50, implying little ceiling effect. The instrument has been translated into several languages, and appears valid, reliable and practical in each.**

KEY WORDS: **gastroesophageal reflux disease, quality of life instruments, symptom assessment, the GERD-HRQL.**

INTRODUCTION

The early 1990s saw the development and dissemination of laparoscopic antireflux surgery for the treatment of gastroesophageal reflux disease.[1,2] The outcomes were generally measured with qualitative scales of 'poor, fair, good, or excellent', or some derivation thereof. At that time, there were very few instruments specifically designed to measure symptom severity in GERD. This lack of a good instrument inhibited progress of GERD research due to the inability to quantitatively compare the magnitude of symptomatic improvement. Specifically, good symptom severity instruments allow the clinician or researcher to characterize the impact of GERD or its treatment in terms that are of value to the patient, may be used as independent predictors of surgical outcomes, may be indicators of the severity of disease, and can provide information on the quality of care.[3] In order to meet this need, the Gastroesophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) instrument was development to assess symptomatic outcomes for the typical symptoms of GERD. This instrument is one of the most frequently used of the symptom severity instruments, and has been recommended for use by the European Association for Endoscopic Surgery.[4] Nevertheless, the entire development of the GERD-HRQL has not been documented in one source, hence the purpose of this article is to trace its development with special attention to the important attributes of a quality of life instrument.

OBJECTIVE OF THE INSTRUMENT

The Scientific Advisory Committee of the Medical Outcomes Trust has put forward recommendations to assess health status and quality of life instruments.[5] One of the key attributes is the 'Conceptual and measurement model.' This is the rationale for and the description of the concept and the populations that a measure is intended to assess. The review

Address correspondence to: Vic Velanovich, MD, Division of General Surgery, K-8; Henry Ford Hospital; 2799 West Grand Blvd., Detroit, MI 48202-2689 USA. Email: vvelano1@hfhs.org

criteria include the concept to be measured, conceptual and empiric basis for item content, target population, information on dimensionality, evidence of scale variability, intended level of measurement, and rationale for deriving a scale score. Not all of these are needed for each instrument.

For the GERD-HRQL, the primary purpose was to measure symptomatic change as a result of medical or surgical treatment of GERD. The theoretical basis of the instrument was the quantification of the 'typical' symptoms of GERD with the target population being patients with GERD-like symptoms seeking medical attention. In the mid-twentieth century, objectively determining the presence of pathologic reflux was problematic. Not all patients with GERD-like symptoms in fact have GERD.[6] Therefore, failure to achieve good symptomatic results after antireflux surgery may have been due to inappropriate patient selection. This led to the development of physiologic testing for GERD with endoscopy, esophageal manometry, and 24 h esophageal pH testing.[7] And, in fact, the combination of the 'typical' symptoms of GERD with abnormal 24 h esophageal pH monitoring has been the best predictor of symptomatic improvement after antireflux surgery.[8] Given this physiologic 'gold standard' for the diagnosis and outcome prediction of GERD, it was felt that the symptom severity questionnaire should incorporate the 24-h esophageal pH monitoring criteria.

As stated, the primary purpose in the development of the GERD-HRQL was to measure symptomatic improvement of both medical and surgical treatment for GERD. However, there were secondary considerations used in instrument development. Specifically, practicality as reflected by low administrative burden and simplicity in scoring, and appropriateness as reflected by responsiveness and interpretability were considered highly desirable. In addition, another goal was to keep the instrument short, essentially to one page, to allow for ease of use by patients, and self-explanatory to the patient so that the use of research assistance was not necessary.

METHODS OF TESTING THE INSTRUMENT

The typical symptoms of GERD include heartburn and regurgitation, occurring both during the night, frequently waking the patient up from sleep, and also occurring during the day, frequently associated with meals. As the disease progresses, strictures can form leading to dysphagia. These symptoms have a great impact on a patient's quality of life.[9,10] The 24-h pH probe measures the total number of reflux episodes, the longest reflux episode, supine reflux, upright reflux, the total time the intraesophageal pH is < 4, and the longest reflux episode. Initially, nine items were chosen for the GERD-HRQL. These items were chosen by the author based on clinical interviews of dozens of patients with GERD to reflect the progressive severity of typical GERD. In this sense, the instrument as initially designed had *face validity*.[11] That is, the instrument appears to cover the issues of the disease as determined by those familiar with the disease. To this, an additional item related to bloating was added, as this is one of the side-effects of antireflux surgery and a measure of pre-existing gastroparesis, which commonly occurs in GERD patients. All subsequent studies have revalidated the GERD-HRQL with this additional item. In addition to the items reflecting the symptoms, an additional item was added with regard to use of medication as a measure of the effect of medication usage on quality of life, which has often been overlooked by other investigators. Table 1 presents the GERD-HRQL questionnaire. The final instrument contains a total of 10 scaled items which are scored, and a patient-reported global satisfaction assessment which is not added to the total GERD-HRQL score.[11]

Once the items of the questionnaire were chosen, a scaling system was needed to be devised to allow for increase in severity of symptoms to be appropriately quantified across patients. A main concern was that of floor and ceiling effects. The 'floor' effect is when a patient reports that he/she is at the lowest score possible as measured by the instrument,

Table 1 The Gastroesophageal Reflux Disease-Health Related Quality of Life instrument

• **Scale**: No symptoms = **0**; Symptoms noticeable, but not bothersome = **1**; Symptoms noticeable and bothersome, but not every day = **2**; Symptoms bothersome every day = **3**; Symptoms affect daily activities = **4**; Symptoms are incapacitating, unable to do daily activities = **5**

• **Questions**

__ 1. How bad is your heartburn?	0 1 2 3 4 5
__ 2. Heartburn when lying down?	0 1 2 3 4 5
__ 3. Heartburn when standing up?	0 1 2 3 4 5
__ 4. Heartburn after meals?	0 1 2 3 4 5
__ 5. Does heartburn change your diet?	0 1 2 3 4 5
__ 6. Does heartburn wake you from sleep?	0 1 2 3 4 5
__ 7. Do you have difficulty swallowing?	0 1 2 3 4 5
__ 8. Do you have pain with swallowing?	0 1 2 3 4 5
__ 9. Do you have bloating or gassy feelings?	0 1 2 3 4 5
__ 10. If you take medication, does this affect your daily life?	0 1 2 3 4 5
__ How satisfied are you with your present condition? Satisfied __ Neutral __ Dissatisfied __	

but later he/she reports symptoms beyond this lowest point. The 'ceiling' effect is at the opposite end of the scale. As an example of a ceiling effect is in the visual analog scale (VAS) for pain. In this instrument the worst possible score is 10 (from a scale of 0 to 10). Yet, the patient reports that his pain is 'an 11 out of 10.' This response cannot be measured by the VAS. Only four of 372 patients who completed the instrument scored 50, implying very little ceiling effect. In addition, it was important to insure that patients could understand the response scale. This was approached by having the numerical Likert-type responses attached to an anchor, whereby each patient can assess the severity of his or her own symptoms on an ordinal scale. Because the severity would be defined in laymen's terms, the responses would be standardized from patient to patient. The scale and anchors were defined in such a manner that zero was defined as no symptoms; (and therefore since patients could not be better than asymptomatic, the problem with the floor effect is avoided) and 5 was defined as 'incapacitating, unable to do daily activities' (and this was felt to be an adequate ceiling, since it is unlikely that patients could be any worse than completely incapacitated from the reflux). The figure also shows the scale with the anchors. The total GERD-HRQL score is derived by simply adding the individual item scores. No transformation of the raw scores to a scaled score is required, thereby insuring practicality. Therefore, the best possible total GERD-HRQL score is 0 (asymptomatic in all items) and the worst possible score is 50 (incapacitated in all items). Because the total GERD-HRQL score has 51 possible scores, it has a high level of precision,[12] especially compared to other GERD instruments. Item 11 pertaining to satisfaction has no numerical score and it is not reflected in the total GERD-HRQL score. This item should be interpreted as a patient-reported 'global' assessment of his or her present condition with respect to GERD.

VALIDITY OF THE GERD-HRQL

Whether the GERD-HRQL has validity or not has been questioned.[13] Let us look carefully at the types of validity used in quality of life research. Fayers and Machin[12] have defined three broad types of validity, and within these, subtypes. 'Content validity' relates to the adequacy of the content of the instrument to the quality of life characteristics it intends to measure. An aspect of content validity is 'face validity'; that is, whether the instrument appears to cover the issues of the disease as determined by those familiar with the disease (as mentioned above). The GERD-HRQL was intended to measure the typical symptoms of reflux; therefore, patients had to feel that it measured the symptoms they were experiencing. Patients were asked to assess the GERD-HRQL and the SF-36 with the following questions:

1. Which questionnaire do you like best?
2. Which questionnaire was easier to understand?
3. Which questionnaire was more reflective with your problems of reflux?
4. Given the choice, which questionnaire would you rather fill out?

The GERD-HRQL was chosen more often by patients for all of these questions and particularly for question #3, 85% of patients felt that the GERD-HRQL better reflected their problems with reflux than the SF-36 and 68% of patients would rather fill out the GERD-HRQL rather than the SF-36.[14] From the standpoint of patient preference, the GERD-HRQL was a better questionnaire. More importantly, patients felt that the GERD-HRQL better reflected their problems with GERD and this supports the instrument's content and face validity.

'Criterion validity' involves measuring the instrument against a 'gold standard.' At the time the instrument was developed, there was no gold standard questionnaire for GERD. It was felt that the gold standard was physiologic assessment. Therefore, the instrument incorporates aspects of the physiologic goal standard of the 24-h pH probe; hence, it does have criterion validity in this sense. In addition, the instrument was assessed by comparing it to other physiologic standards such as endoscopically demonstrated esophagitis, results of the 24-h pH probe and esophageal manometry. Triadalfilopoulos[15] has shown that there is a correlation between question #1 of the GERD-HRQL (How bad is your heartburn) and the percentage of time that the pH < 4 by 24 h esophageal pH monitoring. In another study, it has been shown that as esophagitis grade increases, so does the total GERD-HRQL score.[16] This correlation with esophagitis grade and 24 h esophageal pH monitoring meet the criteria of criterion validity subtype of 'concurrent validity.' The total GERD-HRQL score did not correlate with the DeMeester score.[16] It is believed that this reflects the fact that only three of the items (#1–3) directly correlate to the aspects recorded by 24-h pH monitoring. Another subtype of criterion validity is 'predictive validity.' The GERD-HRQL has been shown to predict which patients would chose antireflux surgery and which patients would continue with medical management.[11]

Lastly, 'construct validity' is an assessment of the degree to which an instrument measures the theoretical construct that it was designed to measure. A subtype of construct validity is 'known-groups' validity; that is, it would be expected that similar groups would have similar scores and differing groups would have different scores. In the case of

GERD, symptomatic improvement is a primary outcome endpoint. This is why patients seek medical attention. Therefore, an instrument measuring GERD symptoms must be able to differentiate patients who are satisfied with their present level of symptoms and those who are not. The GERD-HRQL has been shown to do this.[11,14] In addition, we would expect patients who have had treatment for GERD to have better scores, which it true for the GERD-HRQL for both medical and surgical therapy,[11] and patients who have concomitant esophageal disorders to have worse scores, as was demonstrated with patients with non-specific esophageal motility disorders.[17]

'Concurrent validity' is agreement with the 'true' value. As this is not possible with most quality of life instruments because the 'true' value is not discernable, another way to address this is to compare them with other instruments. Instruments that measure the same phenomenon should have similar results. The GERD-HRQL was compared with another instrument which measures symptom severity, the quality of life questionnaire for patients undergoing antireflux surgery (QOLARS), and was found to correlate.[18]

Another subtype of construct validity is 'discriminant validity', in which instruments which do not measure the same aspects of quality of life would have scores which poorly correlate. When comparing the responses to the GERD-HRQL to the SF-36, it was shown using both univariate and multivariate analysis that the total GERD-HRQL score was a better predictor of patient satisfaction with level of reflux symptoms than the SF-36 and there was little correlation between the scores of the SF-36, and the GERD-HRQL.[14] Moreover, the range of scores had little overlap between the satisfied and the dissatisfied groups. Therefore, given these findings, validity of the GERD-HRQL has been assessed.

RELIABILITY OF THE GERD-HRQL

Reliability is the degree to which an instrument is free from random error.[5,12] Another way of stating this is that the instrument should give the same score at the same level of symptoms. Reliability of the instrument was assessed with the test-retest standard.[11] When patients at the same level of their reported symptom severity had retaken the test in two consecutive visits, the average difference of the scores was less than seven points. This difference was less than the difference between the total scores between the satisfied and dissatisfied patients. Therefore, there is stability in patient scores from test-to-test at the same level of patient-perceived symptoms.

RESPONSIVENESS OF THE GERD-HRQL

The strength of the GERD-HRQL is its sensitivity to change (responsiveness) to the effect of treatment. This attribute is an instrument's ability to detect change over time.[5] This characteristic is important in assessing the efficacy of treatments,[19] particularly surgical treatments.[20] The GERD-HRQL total score does improve (that is, reduces in score) with both medical and surgical treatment.[11] In addition, the magnitude of improvement reflects the initial severity of the score. The score shows that there is similar improvement in patients who have undergone laparoscopic versus open antireflux surgery both for the total score and for the first six items of the instrument individually. With respect to laparoscopy, there was also improvements in items number 8, 9, and 10. So for both medical, laparoscopic surgical treatment, and open surgical treatment, the total GERD-HRQL score and most of the individual item scores were responsive to improvements in patients' symptoms.[21] In addition, we see that when patients are less satisfied with antireflux surgery, such as those with chronic pain syndromes or psychoemotional problems,[22,23] the magnitude of the change is less. Also, the GERD-HRQL has been used in a number of studies evaluating new endoscopic treatments with similar responsiveness as with surgery.[24]

PRACTICALITY OF THE GERD-HRQL

Another strength in the GERD-HRQL is its practicality. The instrument has a total of 11 items, 10 of which are related to the scale and are included in assessing the total GERD-HRQL. Item number 11 is a global item related to patient satisfaction. The instrument is generally administered by simply handing it to the patient during an office visit or can be easily given over the phone in less than 2 minutes. Patients find it easy to understand and there are relatively few unanswered points when assessing the questionnaire. It has been my experience that the number of unanswered items is in the 1–2% range (unpubl. data). Few self-administered questionnaires have such a low unanswered question rate.

LIMITATIONS OF THE GERD-HRQL

Although the GERD-HRQL is an appropriate instrument to measure the severity of the typical symptoms of GERD, it does have important limitations. The GERD-HRQL is not appropriate for measuring the atypical symptoms of GERD. Specifically, there are no items for respiratory or laryngeal symptoms and none for chest pain as an

independent symptom from heartburn. Also, the GERD-HRQL does not measure the symptoms or effects of laryngopharyngeal reflux as a separate clinical entity. Other instruments have been developed for this purpose.[25] In addition, the GERD-HRQL is not an appropriate instrument for the measurement of the effects of GERD on lifestyle or other activities of daily living. There does exit another instrument which measures such problems.[26]

As the GERD-HRQL focuses on the typical symptoms of GERD, investigators may need to supplement its use with other quality of life (QoL) instruments. For example, to assess the effects of GERD on other aspects of QoL or to be able to assess the QoL effects of GERD as compared to other diseases, a generic instrument would be most appropriate. Such instruments as the SF-36, the Psychological General Well-Being, or the Sickness Impact Profile have been used in GERD and many other disease processes.[3] Therefore, whether to use the GERD-HRQL alone or in combination with other instruments will depend entirely on the purpose of the investigator or clinician.

CONCLUSION

In conclusion, the GERD-HRQL has found a place in the assessment of symptom severity in gastroesophageal reflux disease. It is reliable, valid, and practical for this purpose. Further areas of research include additional comparisons with other instruments as well as further studies in the areas of physiologic testing.

References

1 Dallemagne B, Weerts J M, Jehaes C, Markiewicz S, Lombard R. Laparoscopic nissen fundoplication: preliminary report. Surg Laparosc Endosc 1991; 1: 138–43.
2 Hinder R A, Filipi C J, Wetscher G, Neary P, DeMeester T R, Perdikis G. Laparoscopic Nissen fundoplication is an effective treatment for gastroesophageal reflux disease. Ann Surg 1994; 220: 472–81.
3 Wood-Dauphinee S, Korolija D. Symptoms, health-related quality of life and patient satisfaction: using these patient-reported outcomes in people with gastroesophageal reflux disease. In: Granderath F A, Kamolz T, Pointner R, (eds). Gastroesophageal Reflux Disease: Principles of Disease, Diagnosis, and Treatment. Vienna: Springer, 2006; 269–86.
4 Korolija D, Sauerland S, Wood-Dauphinee S *et al.* Evaluation of quality of life after laparoscopic surgery: evidence-based guidelines of the European Association of Endoscopic Surgery. Surg Endosc 2004; 18: 879–97.
5 Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality of life instruments: attributes and review criteria. Qual Life Res 2002; 11: 193–205.
6 Costantini M, Crookes P F, Bremner R M *et al.* Value of physiologic assessment of foregut symptoms in a surgical practice. Surgery 1993; 114: 780–6.
7 Stein H J, DeMeester T R, Hinder R A. Outpatient physiologic testing and surgical management of foregut motility disorders. Curr Prob Surg 1992; 24: 415–555.
8 Campos G M, Peters J H, DeMeester T R *et al.* Multivariate analysis of factors predicting outcome after laparoscopic Nissen fundoplication. J Gastrointest Surg 1999; 3: 292–300.
9 Kamolz T, Pointner R, Velanovich V. The impact of gastroesophageal reflux disease on quality of life. Surg Endosc 2003; 17: 1193–9.
10 Kamolz T, Velanovich V. Quality of life in gastro-oesophageal reflux disease. Digest Liv Dis 2001; 33: 737–9.
11 Velanovich V, Vallance S R, Gusz J R, Tapice F V, Hurkabus M A. Quality of life scale for gastroesophageal reflux disease. J AM Coll Surg 1996; 183: 217–24.
12 Fayers P M, Machin D. Quality of life: assessment, analysis, and interpretation. Chichester: John Wiley & Sons, Ltd, 2000.
13 Borganokor M R, Irvine E J. Quality of life measurement in gastrointestinal and live disorders. Gus 2000; 47: 444–54.
14 Velanovich V. Comparison of generic (SF-36) vs. disease-specific (GERD-HRQL) quality-of-life scales for gastroesophageal reflux disease. J Gastrointest Surg 1998; 2: 141–5.
15 Triadalfilopoulos G. Changes in GERD symptom scores correlate with improvement in esophageal acid exposure after the Stretta procedure. Surg Endosc 2004; 18: 1038–44.
16 Velanovich V, Karmy-Jones R. Measuring gastroesophageal reflux disease: relationship between healthy-related quality of life scores and physiologic parameters. Am Surg 1998; 64: 649–53.
17 Velanovich V, Mahatme A. Effects of manometrically discovered nonspecific motility disorders of the esophagus on the outcomes of antireflux surgery. J Gastrointest Surg 2004; 8: 335–41.
18 Zeman Z, Rozsa S, Tarko E, Tihanyi T. Psychometric documentation of a quality of life questionnaire for patients undergoing antireflux surgery (QOLARS). Surg Endosc 2005; 19: 257–61.
19 Fitzpatrick R, Ziebland S, Jenkinson C *et al.* Importance of sensitivity to change as a criterion for selecting health status measures. Qual Health Care 1992; 1: 88–93.
20 Velanovich V. Using quality of life instruments to assess surgical outcomes. Surgery 1999; 126: 1–4.
21 Velanovich V. Comparison of symptomatic and quality of life outcomes of laparoscopic versus open antireflux surgery. Surgery 1999; 126: 782–9.
22 Velanovich V. The effect of chronic pain syndromes and psychoemotional disorders on symptomatic and quality of life outcomes of antireflux surgery. J Gastrointest Surg 2003; 7: 53–8.
23 Velanovich V. Using quality of life measurements to predict patient satisfaction outcomes for antireflux surgery. Arch Surg 2004; 139: 621–6.
24 Kamolz T, Velanovich V. The impact of disease and treatment on health-related quality of life in patients suffering from GERD. In: Granderath F A, Kamolz T, Pointner R, (eds). Gastroesophageal Reflux Disease: Principles of Disease, Diagnosis, and Treatment. Vienna: Springer, 2006; 287–98.
25 Westcott C J, Hopkins M B, Bach K *et al.* Fundoplication for laryngopharyngeal reflux. J Am Coll Surg 2004; 199: 23–30.
26 Liu J Y, Woloshin S, Laycock W S *et al.* Symptoms and treatment burden of gastroesophageal reflux disease: validating the GERD Assessment Scales. Arch Intern Med 2004; 164: 2058–64.

Surgical Endoscopy
https://doi.org/10.1007/s00464-020-07792-1

Three-year clinical experience with magnetic sphincter augmentation and laparoscopic fundoplication

Luigi Bonavina[1] · Thomas Horbach[2] · Sebastian F. Schoppmann[3] · Janet DeMarchi[4,5]

Received: 18 February 2020 / Accepted: 7 July 2020

Abstract

Background Magnetic sphincter augmentation (MSA) is a surgical intervention for gastroesophageal reflux disease (GERD) which has been evaluated in numerous studies and has shown beneficial effects. Long-term effectiveness data for MSA as well as laparoscopic fundoplication (LF) in patients with GERD are needed.

Objective The objective of this study was to evaluate the 3-year outcomes for MSA and LF in patients with GERD.

Methods This prospective, multi-center, observational registry study evaluated MSA and LF in clinical practice over 3 years (ClinicalTrials.gov identifier: NCT01624506). Data collection included baseline characteristics, reflux symptoms, medication use, satisfaction and complications. Post-surgical evaluations were collected at yearly intervals.

Results Between December 2009 and December 2014, 631 patients (465 MSA and 166 LF) were enrolled in the registry. Both MSA and LF resulted in improvements in total GERD-HRQL score (mean reduction in GERD-HRQL from baseline to 3 years post-surgery: MSA 22.0 to 4.6 and LF 23.6 to 4.9) and in satisfaction (GERD-HRQL satisfaction increase from baseline to 3 years: MSA 4.6% to 78.2% and LF 3.7% to 76.5%). Most patients were able to belch as needed with both therapies (MSA 97.6% and LF 91.7% at 3 years). MSA allowed a higher percentage of patients the ability to vomit as needed (MSA 91.2% and LF 68.0% at 3 years). PPI usage declined from baseline to 3 years for both groups after surgery (MSA 97.8% to 24.2% and LF 95.8% to 19.5%). The mean procedure time was shorter for MSA than for LF. Intraoperative and procedure-related complication rates ($\leq 2\%$) were low for both therapies.

Conclusions This 3-year prospective observational registry study contributes to the mounting evidence for the effectiveness of MSA and LF. Despite the more severe nature of GERD in the LF group, the clinical outcomes for MSA and LF were favorable from an effectiveness and safety standpoint.

Keywords Anti-reflux surgery · GERD · LINX · Fundoplication · Regurgitation · Proton pump inhibitors

Gastroesophageal reflux disease (GERD) is defined as symptoms or complications resulting from the reflux of gastric contents into the esophagus or beyond, into the oral cavity (including larynx) or lung [1]. Persistent and intense GERD symptoms have a significant and negative impact on both health-related quality of life (HRQL) and healthcare resource utilization [2]. Treatment options for people who suffer from GERD vary widely, depending on the severity and symptoms of their disease. There are currently three primary means of treating GERD: lifestyle changes, medical therapy, and surgical intervention [3]. Reasons to refer GERD patients for surgery may include: proton pump inhibitor (PPI) non-compliance, side effects associated with medical therapy, presence of a large hiatal hernia, esophagitis refractory to medical therapy, or persistent symptoms caused by refractory GERD [1, 4, 5].

The traditional surgical treatment for GERD is a laparoscopic fundoplication [6]. Nissen fundoplication involves wrapping a portion of the stomach around the entire

✉ Janet DeMarchi
jdemarch@its.jnj.com

1 Division of General and Foregut Surgery, Department of Biomedical Sciences for Health, University of Milano, IRCCS Policlinico San Donato, Milano, Italy

2 Department of Surgery, Schön Klinik Nürnberg Fürth, 90763 Fürth, Germany

3 Department of Surgery, Medical University Vienna, 1090 Vienna, Austria

4 Ethicon, Inc., Cincinnati, OH, USA

5 Ethicon, Inc., 15803 Sunset Rd, Minnetonka, MN 55345, USA

Published online: 16 July 2020

esophagus(360°) to reinforce the weakened lower esophageal sphincter (LES). Another variation of fundoplication is the Toupet procedure which wraps a portion of the stomach around 270° of the esophagus. Yet a third variation is the Dor fundoplication which consists of wrapping a portion of the stomach around 180° of the esophagus. The fundoplication procedure has several shortcomings that have limited its use: (1) it results in anatomical and physiological alteration of the fundus; (2) potentially debilitating side effects including gas bloat and an inability to belch or vomit [7, 8]; (3) the procedure is not standardized, resulting in variable effectiveness [8]; and (4) it is a not a fully reversible procedure [9]. For the purpose of this publication, laparoscopic fundoplication (LF) includes Nissen, Toupet or Other/Unspecified fundoplication procedures. Magnetic sphincter augmentation (MSA) is an alternative to fundoplication which has been studied in diverse patient populations and has shown beneficial effects [5, 9–23]. MSA augments the weak lower esophageal sphincter (LES), restoring the body's natural barrier to reflux [24]. The LINX System is a small flexible band of interlinked titanium beads with magnetic cores. The magnetic attraction between the beads is intended to help the LES resist opening to gastric pressures, preventing reflux from the stomach into the esophagus.

The outcomes associated with laparoscopic fundoplication (LF) and MSA have previously been evaluated, including a prospective, multi-center 1-year real-world prospective registry evaluating MSA and LF by Reigler et al. [23]. The 1-year registry study findings showed that both MSA and LF resulted in significant improvements in reflux control, with similar safety and reoperation rates [23]. This same patient population was followed an additional 2 years to demonstrate the long-term safety and effectiveness outcomes of MSA and LF and is the basis of this publication.

Methods

This was a prospective, multi-center, observational registry evaluating MSA and LF in clinical practice (ClinicalTrials.gov identifier: NCT01624506). The study was conducted under a common protocol and approved by each center's Ethics Committee. Informed consent was obtained before enrolling patients into the study.

Patient enrollment and study population

Twenty-two medical centers in four countries (Austria, Germany, Italy, and the United Kingdom) enrolled all patients who were candidates for a surgical anti-reflux procedure. All patients had a diagnosis of GERD confirmed by abnormal esophageal acid exposure on a prolonged pH or pH impedance study and chronic reflux symptoms despite the daily use of medical therapy with PPIs. Patients with severe GERD, defined as meeting one or more of the following conditions, were also included: large hiatal hernia (> 3 cm diameter of the esophageal hiatus), Barrett's esophagus, motility disorder, and Grade C or D esophagitis by Los Angeles (LA) classification. Patients without advanced GERD characteristics were considered to have moderate GERD (abnormal esophageal pH, reflux symptoms despite medication). Patients were excluded if they had known conditions that would make it unlikely for them to complete the 3-year follow-up (e.g., life expectancy < 3 years).

Study procedures and outcomes

The type of anti-reflux procedure performed (MSA or LF [Nissen, Toupet or Other/Unspecified]) was provisionally agreed upon by the surgeon in close consultation with the patient. If a patient met the labeling requirements for MSA (hiatal hernia ≤ 3 cm, esophagitis less than Grade C, absence of Barrett's esophagus, absence of motility disorders), MSA was recommended. Motility disorders were defined as having distal esophageal pressure less than 35 mm/Hg (or HRM equivalent of a DCI < 450), less than 70% effective swallows, or known disorders such as achalasia, nutcracker esophagus, diffuse esophageal spasm or a hypertensive lower esophageal sphincter. The MSA device was placed utilizing the minimal dissection technique. For those patients with larger hiatal hernias, more severe esophagitis, Barrett's esophagus or some types of dysmotility, LF was recommended. The final choice of procedure was made by the surgeon at the time of laparoscopy. A patient may have been a MSA candidate based upon the preoperative screening but upon entering the abdomen, a surgeon may have encountered a larger than anticipated hiatal hernia or was unable to obtain enough intra-abdominal esophageal length to safely implant the MSA device. In this instance, a fundoplication would be performed. The majority of the LF group underwent Nissen fundoplication (62%). A Toupet fundoplication was performed in 31% of patients in the LF group with the remaining 7% undergoing an Other/Unspecified procedure. Postoperative care was directed by the surgeon based on the patient's clinical condition and practices of the institution.

Data collection included baseline characteristics and pre- and post-surgical HRQL (evaluated with Gastroesophageal Reflux Disease-Health-Related Quality of Life (GERD-HRQL) symptom severity instrument) [25], patient satisfaction (GERD-HRQL satisfaction with present condition [proportion of satisfied, neutral, or dissatisfied]), symptoms, and use of PPIs (proportion of patients regularly taking PPIs), (GERD-HRQL PPI medication status) as well as the duration of surgery, length of stay, complications, and healthcare resource use. The GERD-HRQL was developed to measure symptomatic outcomes and therapeutic effects in patients

with GERD [26]. The scale has 11 items which focus on heartburn symptoms, dysphagia, medication effects, and the patient's present health condition. Each item is scored from 0 to 5, with a higher score indicating a better quality of life.

Study-related questionnaires were administered to all patients prior to and 1, 2, and 3 years after surgery. The hospital staff administering the standardized questionnaires were unaware of the procedure that was performed. Pre- and post-surgery questionnaires were completed at the appropriate time points (i.e., preoperatively, at time of surgery, from 6 weeks to 6 months, at 12 months, at 24 months, and at 36 months). All registry follow-up visits had a ± 60-day visit window and could be conducted over the telephone. Both incorporated the GERD-HRQL, satisfaction with current condition, other GERD-related symptoms, reason for choosing anti-reflux surgery and perceived benefits of anti-reflux surgery. Use of PPIs was tracked before and after surgery. In addition, healthcare utilization related to procedural complaints or complications was tracked after surgery. Patients were also asked about their willingness to undergo surgery again.

Study analyses

The registry was not a clinical study and therefore it was not powered to evaluate a study hypothesis. Statistical analyses were primarily descriptive. Means and standard deviations (SD) and/or medians were used to describe continuous variables and frequency counts and/or percentages summarized categorical variables. Statistical tests were only conducted for evaluating differences in baseline characteristics and for evaluating within-group changes over time. Two-tailed unpaired Student's t test or the Wilcoxon rank sum test were used to compare continuous baseline characteristic values for MSA and LF. Chi-square tests or the Mann–Whitney were used to compare categorical baseline characteristic values for MSA and LF. Differences were considered to be significant at the 0.05 level. Analyses were performed using SAS version 9.4.

Results

Patient baseline demographic and clinical characteristics

Between Dec 2009 and Dec 2014, 631 patients (465 MSA and 166 LF) had enrolled in the registry. Patient baseline demographic and clinical characteristics for MSA and LF are presented in Table 1.

The particular patient baseline characteristics that were statistically significantly different between patients with MSA vs. LF (all $p < 0.0001$) were patient age (LF 56.3 years vs. MSA 46.6 years), BMI (LF 27.8 vs MSA 25.7), frequency of large hiatal hernias (LF 48.1% vs. MSA 1.4%), and the presence of Barrett's esophagus at the time of surgery (LF 12.7% vs MSA 1.7%). Also, a greater proportion of patients with MSA had no esophagitis ($p = 0.0130$). It is notable that approximately 91% of patients in the MSA group met the definition of moderate GERD (hiatal hernia absent or ≤ 3 cm, normal motility, no Barrett's esophagus and esophagitis ≤ Grade B by LA Classification), while just 18% of the LF group were categorized as having moderate GERD. On the other hand, approximately 83% of patients in the LF group met the criteria for severe GERD (one of more of the following: hiatal hernia > 3 cm, Barrett's esophagus, motility disorder and/or esophagitis Grade C or D by LA Classification), while 9% of the MSA group met the

Table 1 Patient baseline demographic and clinical characteristics for MSA and LF

Measure	MSA $N = 465$	LF $N = 166$	p value
Age, years (mean ± SD)	46.6 ± 13.6	56.3 ± 12.6	< 0.0001
Gender, % of patients			0.866
Male	63.7%	49.4%	
Female	36.3%	50.6%	
BMI (kg/m^2) (mean ± SD)	25.7 ± 3.7	27.81 ± 4.0	< 0.0001
Years with GERD (mean ± SD)	9.0 ± 7.7	9.2 ± 8.6	0.7950
Years of PPI Use (mean ± SD)	6.1 ± 5.3	5.7 ± 6.0	0.5184
Esophagitis, % of patients			0.0130
None	53.0%	40.9%	
Grade A	31.7%	29.6%	
Grade B	13.5%	16.4%	
Grade C	1.1%	8.2%	
Grade D	0.7%	5.0%	
Barrett's Esophagus, % of patients	1.7%	12.7%	< 0.0001
Hiatal Hernia Size, % of patients			< 0.0001
None	19.7%	7.5%	
1–3 cm	78.9%	44.4%	
> 3 cm	1.4%	48.1%	
Total % Time pH < 4 (mean ± SD)	12.2 ± 11.4	13.0 ± 14.7	0.5830
Moderate GERD, % of patients[a]	90.8	18.1	
Severe GERD, % of patients[b]	9.2	81.9	

BMI body mass index, *GERD* gastroesophageal reflux disease, *LF* laparoscopic fundoplication, *MSA* magnetic sphincter augmentation, *PPI* proton pump inhibitor

[a]Moderate GERD defined as hiatal hernia ≤ 3 cm, no Barrett's esophagus, no motility disorder, and esophagitis ≤ Grade B by LA Classification

[b]Severe GERD defined as one or more of: hiatal hernia > 3 cm, Barrett's esophagus, motility disorder and/or Grade C or D esophagitis by LA Classification

criteria for severe GERD. Given this difference in severity between the two groups, the median GERD-HRQL scores were similar with 22.0 for the MSA group and 23.0 for the LF group but not statistically significant with a *p* value of 0.0620. Patients did not differ in other evaluated baseline characteristics.

Clinical effectiveness and HRQL of MSA and LF over time

Table 2 presents the clinical effectiveness of MSA and LF pre- and post-surgery.

Both MSA and LF resulted in an improvement in satisfaction (satisfaction score from GERD-HRQL) over the 1, 2, and 3 years after surgery. The proportion of patients whose sleep was affected by GERD declined upon surgery for both MSA and LF. At 24 months, the proportion of MSA patients with sleep problems was twice that of LF patients; however, the confidence intervals overlapped and these differences were not observed at 36 months. Despite the one-way valve of LF and the two-way physiologic valve of MSA, patients appear to be able to belch as needed with both therapies. It is notable that at each postoperative time point, MSA allowed a higher percentage of patients the ability to vomit if needed with 91.2% of patients noting the ability to vomit at 36 months. At the same timepoint, 68% of the LF patients were able to vomit if needed. PPI usage also declined for both groups after surgery. Most patients reported a willingness to have the surgery again throughout the 3-year follow-up period.

Both MSA and LF resulted in substantial improvements in total GERD-HRQL score (a lower GERD-HRQL score signifies higher quality of life) from baseline over the 1, 2, and 3 years after surgery. This reduction, seen in both

Table 2 Clinical effectiveness of MSA and LF pre- and post-surgery

Measure	MSA % (*n*/*N*) [95% CI]	LF % (*n*/*N*) [95% CI]
Satisfaction with current condition (from GERD-HRQL)		
Baseline	4.6% (21/460) [2.7%, 6.5%]	3.7% (6/164) [0.8%, 6.5%]
12 M	75.3% (326/433) [71.2%, 79.4%]	77.2% (122/158) [70.7%, 83.8%]
24 M	78.9% (254/322) [74.4%, 83.3%]	83.3% (90/108) [76.3%, 90.4%]
36 M	78.2% (230/294) [73.5%, 82.9%]	76.5% (65/85) [67.5%, 85.5%]
GERD interfering with sleep		
Baseline	73.3% (333/454) [69.3%, 77.4%]	78.0% (128/164) [71.7%, 84.4%]
12 M	11.9% (50/419) [8.8%, 15.0%]	9.6% (15/157) [5.0%, 14.2%]
24 M	11.7% (37/315) [8.2%, 15.3%]	5.5% (6/109) [1.2%, 9.8%]
36 M	9.0% (26/290) [5.7%, 12.3%]	10.7% (9/84) [4.1%, 17.3%]
Ability to belch		
Baseline	96.7% (441/456) [95.1%, 98.3%]	93.9% (154/164) [90.2%, 97.6%]
12 M	96.7% (406/420) [94.9%, 98.4%]	88.5% (138/156) [83.4%, 93.5%]
24 M	97.2% (308/317) [95.3%, 99.0%]	92.5% (99/107) [87.5%, 97.5%]
36 M	97.6% (284/291) [95.8%, 99.4%]	91.7% (77/84) [85.8%, 97.6%]
Ability to vomit[a]		
Baseline	96.6% (343/355) [94.8%, 98.4%]	92.0% (115/125) [87.2%, 96.8%]
12 M	89.7% (191/213) [85.6%, 93.8%]	55.8% (29/52) [42.3%, 69.3%]
24 M	85.8% (133/155) [80.3%, 91.3%]	52.6% (20/38) [36.7%, 68.5%]
36 M	91.2% (134/147) [86.6%, 95.8%]	68.0% (17/25) [49.8%, 86.2%]
Use of PPIs		
Baseline	97.8% (453/463) [95.6%,100%]	95.8% (158/165) [91.6%, 100%]
12 M	18.9% (81/428) [15.2%, 22.6%]	19.7% (31/157) [13.5%, 26.0%]
24 M	21.4% (74/346) [17.1%, 25.7%]	18.1% (21/116) [11.1%, 25.1%]
36 M	24.2% (76/314) [19.5%, 28.9%]	19.5% (17/87) [11.2%, 27.9%]
Willingness to have surgery again		
12 M	89.5% (366/409) [86.5%, 92.5%]	91.1% (143/157) [86.6%, 95.5%]
24 M	90.6% (281/310) [87.4%, 93.9%]	94.4% (102/108) [90.1%, 98.8%]
36 M	93.1% (270/290) [90.2%, 96.0%]	94.0% (79/84) [89.0%, 99.1%]

CI confidence interval, *GERD* gastroesophageal reflux disease, *HRQL* health-related quality of life, *LF* laparoscopic fundoplication, *M* months, *MSA* magnetic sphincter augmentation, *PPI* proton pump inhibitor

[a]Data adjusted to include only those patients who found the need to vomit

groups, was statistically significant and maintained over the entire study period which shows appreciable durability of the therapeutic effect of both anti-reflux surgical options. The results of this analysis are summarized in Table 3.

With respect to preoperative and postoperative dysphagia, both MSA and LF groups reported daily, bothersome dysphagia at baseline with 15.7% and 24.4%, respectively. Both groups showed a decrease postoperatively, summarized in Table 4.

When looking at the LF group and dysphagia, the percentage of patients reporting daily bothersome dysphagia was similar across the three procedure options, as outlined in Table 5.

Procedure time, duration of hospital stay, and clinical safety of MSA and LF

The procedure time, duration of hospital stay, and clinical safety of MSA and LF are presented in Table 6.

Table 4 Dysphagia results for MSA and Fundoplication over study duration

Timepoint	MSA	Fundo	Q7 *p* value
Baseline			
Score	1.0 ± 1.3	1.3 ± 1.5	0.0227
% Q7 > 3.0	15.7%	24.4%	0.0174
12 months			
Score	0.8 ± 1.1	0.6 ± 1.1	–
% Q7 > 3.0	8.8%	7.6%	
24 months			
Score	0.6 ± 0.9	0.4 ± 0.9	–
% Q7 > 3.0	4.4%	4.6%	
36 months			
Score	0.5 ± 0.9	0.4 ± 1.1	–
% Q7 > 3.0	3.8%	4.8%	

Table 3 GERD-HRQL scores and change from baseline

Measure	MSA		LF	
	Mean GERD-HRQL ± SD Median (Min, Max)	Mean Change from Baseline ± SD Median (Min, Max) [95% CL] *p* value	Mean GERD-HRQL ± SD Median (Min, Max)	Mean Change from Baseline ± SD Median (Min, Max) [95% CL] *p* value
Baseline	*n* = 457 22.0 ± 9.1 22.0 (0.0, 47.0)		*n* = 163 23.6 ± 9.8 23.0 (3.0, 47.0)	
Paired Baseline/Month 12	*n* = 414 21.9 ± 9.0 22.5 (0.0, 46.0)		*n* = 152 23.4 ± 9.9 23.0 (3.0, 47.0)	
Month 12	*n* = 418 5.2 ± 6.4 3.0 (0.0, 42.0)	*n* = 414 − 16.7 ± 10.0 − 17.0 (− 41.0, 21.0) [− 17.6, − 15.7] < 0.001	*n* = 154 4.9 ± 7.2 3.0 (0.0, 48.0)	*n* = 152 − 18.5 ± 11.5 − 19.5 (− 45.0, 20.0) [− 20.3, − 16.6] < 0.001
Paired Baseline/Month 24	*n* = 296 21.6 ± 9.2 22.0 (0.0, 41.0)		*n* = 103 23.9 ± 10.1 24.0 (3.0, 47.0)	
Month 24	*n* = 300 4.9 ± 6.1 2.0 (0.0, 35.0)	*n* = 296 − 16.7 ± 10.6 − 17.0 (− 39.0, 28.0) [− 17.9, − 15.5] < 0.001	*n* = 105 3.9 ± 4.4 3.0 (0.0, 19.0)	*n* = 103 − 20.0 ± 10.0 − 20.0 (− 45.0, 0.0) [− 22.0, − 18.1] < 0.001
Paired Baseline/Month 36	*n* = 278 21.3 ± 9.3 22.0 (0.0, 41.0)		*n* = 80 22.5 ± 9.7 22.5 (3.0, 47.0)	
Month 36	*n* = 283 4.6 ± 6.0 3.0 (0.0, 39.0)	*n* = 278 − 16.6 ± 10.2 − 18.0 (− 41.0, 12.0) [− 17.8, − 15.4] < 0.001	*n* = 82 4.9 ± 7.1 3.0 (0.0, 45.0)	*n* = 80 − 17.8 ± 10.6 − 18.0 (− 39.0, 17.0) [− 20.1, − 15.4] < 0.001

CL confidence limit, *GERD* gastroesophageal reflux disease, *HRQL* health-related quality of life, *LF* laparoscopic fundoplication, *max* maximum, *min* minimum, *MSA* magnetic sphincter augmentation, *SD* standard deviation

Table 5 Dysphagia results for Nissen, Toupet and Other over study duration

Timepoint	Nissen	Toupet	Other
Baseline			
Score	1.3 ± 1.6	1.3 ± 1.4	1.7 ± 1.5
% Q7 > 3.0	26.7% (27/101)	19.6% (10/51)	25.0% (3/12)
12 months			
Score	0.7 ± 1.1	0.6 ± 1.1	0.8 ± 1.3
% Q7 > 3.0	7.4% (7/95)	8.0% (4/50)	8.3% (1/12)
24 months			
Score	0.5 ± 0.9	0.4 ± 0.9	0.5 ± 0.9
% Q7 > 3.0	4.4% (3/68)	6.3% (2/32)	0.0% (0/8)
36 months			
Score	0.4 ± 1.0	0.5 ± 1.1	0.4 ± 0.9
% Q7 > 3.0	5.9% (3/51)	3.7% (1/27)	0.0% (0/5)

Table 6 Procedure time, duration of hospital stay, and clinical safety of MSA and LF

Measure	MSA ($n = 459$)	LF ($n = 163$)
Mean procedure time (min)	43.2 ± 19.7	79.7 ± 47.7
Length of Stay < 24 h	36.1%	11.4%
Length of Stay > 48 h[a]	50.8%	72.3%
Intraoperative complication rate	1.8%	1.2%
Procedure-related complication rate	2.0%	1.8%

LF laparoscopic fundoplication, *MSA* magnetic sphincter augmentation

[a]238 of 465 were German patients, have longer stay built into reimbursement

The mean procedure time was 43.2 min for MSA compared with 79.7 min with LF. More than one-third of patients with MSA had a length of stay < 24 h versus 11.4% of patients with LF. Nearly three-quarters of patients with LF had a length of stay > 48 h versus 50.8% of patients with MSA. The intraoperative and procedure-related complication rates were low (≤ 2%) and similar for MSA and LF.

Healthcare resource use with MSA and LF

The healthcare resource use for patients treated with MSA and LF is presented in Table 7.

The proportion of patients with outpatient clinic visits for GERD symptoms or due to procedural complaints or complications was similar for MSA and LF over the 3 years. Most outpatient clinic visits were due to GERD symptoms. In the early postoperative period, approximately 40% of outpatient visits in both MSA and LF groups were for procedure complaints; however, this rate trended downwards as expected over time. The surgical intervention rates and the device removal rate remained low over the 3 years. The

Table 7 Healthcare resource use with MSA and LF

Measure	MSA ($n = 459$)	LF ($n = 163$)
Outpatient clinic visits		
12 M	18.9%	15.3%
24 M	14.7%	12.9%
36 M	10.5%	8.0%
Return to clinic for GERD symptoms		
12 M	58.5%	54.2%
24 M	80.4%	86.7%
36 M	87.9%	100%
Return to clinic due to procedural complaint/complication		
12 M	39.3%	41.7%
24 M	19.2%	20.0%
36 M	15.2%	0.0%
Surgical intervention		
12 M	1.6%	1.9%
24 M	1.2%	0.0%
36 M	0.6%	0.0%
Device removal*		
12 M	1.5% (7/459)	NA
24 M	2.0% (9/459)	NA
36 M	2.4% (11/459)	NA

GERD gastroesophageal reflux disease, *LF* laparoscopic fundoplication, *MSA* magnetic sphincter augmentation

*Removal rates are cumulative across the 3 years

surgical intervention rate for the MSA group at 3 years was 2.4% (11/459). The LF group had a surgical intervention rate of 1.9% (3/157). The intervention for the MSA group was the removal of the device for dysphagia (45%), ongoing GERD (18%), vomiting/regurgitation (18%), gastric pain (9.5%) and need for MRI (9.5%). There were no complications noted during the removal procedures. Two patients underwent fundoplication at the time of the device removal. The interventions for the LF group were revision of a Nissen wrap due to ongoing GERD, re-herniation and a sigmoid resection secondary to diverticulitis. No complications or ongoing sequelae were reported.

Discussion

This is the first long-term prospective study that looks specifically at the safety and effectiveness outcomes of MSA and LF. This 3-year prospective, observational registry study contributes to the mounting real-world evidence for the effectiveness of MSA and LF. Randomized controlled trials (RCTs) remain the gold standard for assessing efficacy; however, they are inadequate for addressing questions about real-world effectiveness and safety of interventions [27, 28].

The high degree of internal validity with RCTs comes at the price of reduced external validity because study populations, protocols, and circumstances may not be relevant to the 'real-world' or diverse populations [29]. Specific to the GERD treatment options of a type of fundoplication and magnetic sphincter augmentation, randomization may not be appropriate due to the inherent variability in the fundoplication procedure as opposed to the standardized magnetic sphincter augmentation procedure. Patients who have researched the options often request one treatment over the other and would not accept leaving the treatment option to chance. There is also an inherent bias built into the patient selection for the two treatments. This study concluded that the patients who received MSA were less severe by definition. The Instructions for Use (IFU) for MSA and therefore inclusion criteria at the time of this trial, excluded hiatal hernias > 3 cm, Barrett's esophagus, motility disorders and more significant esophagitis. These are common findings in the population affected by GERD. The surgeon, in following the IFU for MSA, would generally choose the less severe patient who would fit the criteria. It would be difficult to randomize this group because there are patients who would not be appropriate for MSA as outlined in the current labeling, such as those with large hiatal hernias, Barrett's esophagus or motility disorders, who may benefit from a type of fundoplication instead. Even so, given these challenges, a randomized controlled study would contribute to the evidence base for these two surgical procedures. Unfortunately, such a study was proposed to the United Kingdom's Medical Research Council in 2012 but the funding was denied [30]. Comparative effectiveness research (CER) focuses on real-world evidence to assist consumers, clinicians, purchasers, and policymakers to make informed decisions that will improve healthcare at the individual and population levels [31]. Real-world observational studies can address many of the limitations of RCTs because they leverage data originating from clinical practice and better reflect real-world conditions, a central concept of Comparative Effectiveness Research [32].

The findings from this study showed that the clinical outcomes for MSA and LF were favorable in regards to effectiveness, quality of life, satisfaction, PPI use, and safety, with low rates of complications and similar rates of clinic visits over 3 years. These results are in alignment with previously published effectiveness data for MSA and LF which showed that MSA and LF both lead to a decrease in HRQL score and an increase in patient satisfaction when compared with patient's preoperative symptoms [33, 34]. The proportion of patients with the ability to belch with LF in our study increased over the first 2 years and then stayed relatively steady, possibly due to the loosening of the fundoplication over time (ie, transthoracic migration of the wrap) [35, 36]. At 3 years, in the MSA cohort, 91.2% of patients reported the ability to vomit with just 68.0% of the LF group having the ability to do so.

The previously published 1-year registry results by Riegler et al., [23] which used unadjusted data, showed that the median GERD-HRQL score improved from 20.0 to 3.0 after MSA and 23.0 to 3.5 after LF. Moderate or severe regurgitation improved from 58.2% to 3.1% after MSA and 60.0% to 13.0% after LF ($p = 0.014$). Discontinuation of PPIs was achieved by 81.8% of patients after MSA and 63.0% after LF ($p = 0.009$). Excessive gas and abdominal bloating were reported by 10.0% of patients after MSA and 31.9% following LF ($p < 0.001$). Following MSA, 91.3% of patients were able to vomit if needed, compared with 44.4% of those undergoing LF ($p < 0.001$).

Aiolfi and colleagues performed a systematic review and meta-analysis of patients undergoing MSA or LF (Nissen or Toupet) [19]. Seven studies, published between 2014 and 2017, met the criteria for the analysis, providing a total of 1211 patients. The MSA cohort consisted of 686 patients (56%) and the LF cohort, those undergoing either laparoscopic full (Nissen) or partial (Toupet) fundoplication, included 525 patients (44%). The postoperative follow-up ranged from six to twelve months. In the entire cohort, the mean age ranged from 39.3 to 54 years. The mean hernia size ranged from 1 to 2 cm, esophagitis ≥ Grade B was present in 15.4%. Barrett's esophagus was diagnosed in 16.2% of the patients. At up to 1 year, improvement in the quality of life scores, assessed by GERD-HRQL and the rate of PPI suspension were similar between the two groups. There were 13 reoperations in the MSA group (1.9%), consisting of 12 device removals and one crural release. In the LF group, there were 11 reoperations (2.1%) including herniation of the fundic wrap (5), persistent GERD (3), retro-esophageal abscess (2), crural release (1). The authors concluded that patients with GERD may benefit from both LF and MSA in terms of safety, postoperative quality of life, and PPI suspension at 1-year follow-up.

These findings are also consistent with long-term single-arm data for MSA. Saino et al. [13] conducted a prospective multi-center study at four clinical sites in the US and Europe and found that mean total GERD-HRQL score improved significantly from 25.7 to 2.9 ($p < 0.001$) with MSA when comparing baseline and 5 years ($n = 33$), and that 93.9% of patients had at least a 50% reduction in GERD-HRQL total score compared with baseline. Complete discontinuation of PPIs was achieved by 87.8% of patients. No complications occurred in the long-term, including no device erosions or migrations at any point. Similarly, Ganz and colleagues [37] conducted a prospective study of MSA in 100 adults who were partially responsive to daily PPIs at 14 centers in the US and The Netherlands. Median baseline GERD-HRQL scores were 27 in patients not taking PPIs and 11 in patients on PPIs; 5 years after MSA GERD-HRQL scores decreased

to 4. All patients used PPIs at baseline; however, only 15.3% used them at 5 years. Moderate or severe regurgitation occurred in 57% of subjects at baseline, but only 1.2% at 5 years. All patients reported the ability to belch and vomit if needed. Bothersome dysphagia was present in 5% at baseline and in 6% at 5 years. Bothersome gas bloat was present in 52% at baseline and decreased to 8.3% at 5 years. A recent study by Ayazi et al. [38] on 553 patients at a single institution, looked specifically at identifying patient characteristics that may predict a positive outcome with MSA. Four of the independent predictors identified are age less than 45 years, male gender, GERD-HRQL preoperative score greater than 15 and an abnormal DeMeester score. These predictors align more closely with the MSA cohort in this study.

There are limitations associated with this analysis. MSA and LF effectiveness estimates were derived from 22 medical centers in four European countries and are not necessarily representative of all settings of care. All investigators performed fundoplication routinely and in addition, were trained specifically on the implantation of MSA which is only available in select centers. Within the LF group, there were different procedures performed at the discretion of the surgeon. While this individualizes the treatment to the patient, it makes the analysis more complex. This was a non-randomized study and was not intended to detect statistically significant clinical outcomes between MSA and LF. This registry was intended to capture everyday clinical practice and patients were screened based upon the current Instructions for Use (IFU) for MSA as those who met the MSA criteria. The final decision for the type of surgery was made by the surgeon at the time of laparoscopy, based upon a variety of patient-specific factors. Comparisons of baseline characteristics indicated that patients in the LF group were older, had a higher BMI, and had a greater frequency of large hiatal hernias and Barrett's esophagus. It is possible that results were confounded and observed differences between groups were due to these and other variables. Due to the noted differences in the two populations, the results must be carefully considered. Furthermore, the procedure to implant MSA has since evolved to include full crural and gastroesophageal junction dissection as opposed to the minimal dissection utilized in this investigation. The timeframe for this study will determine if the procedural modifications are relevant to the outcomes in this study population. The current procedure theoretically may provide better outcomes for patients as compared to those done under the "minimal dissection" protocol as the hiatal hernia is often addressed.

An unmet need exists for the effective management of some refractory and/or chronic patients with GERD. In the current climate of tighter budgets and stricter management guidelines, it is important for health systems to assess both the clinical and economic value of surgical interventions and technologies. While they constitute a small proportion of the overall GERD population, GERD patients who are eligible for surgery present a significant clinical and socioeconomic burden. It is important to note that patients in both groups, on average, had suffered with GERD for 9 years and had been medicating with PPIs for 6 years. Persistent reflux symptoms, despite PPI therapy, can cause mental health disorders, sleep disorders, and psychological distress to patients [9, 39, 40]. Persistent and intense GERD symptoms have a negative impact on HRQL and healthcare resource utilization, including lower work productivity, greater activity impairment, more hours missed from work due to health problems, more visits to both primary-care physicians and specialists, and more emergency room visits [40].

Conclusions

This 3-year prospective, observational registry study contributes to the mounting evidence for the effectiveness of MSA and LF. Despite the more severe nature of GERD in the LF group, the clinical outcomes for both MSA and LF were favorable from an effectiveness (GERD-HRQL), satisfaction, PPI use, and safety standpoint, with similar rates of complications and clinic visits. The use of MSA or LF is associated with improved patient clinical and QOL outcomes and lower overall burden of GERD disease. The analyses demonstrated that the evidence supports the use of MSA and LF as effective treatment strategies for patients with GERD eligible for surgery.

Acknowledgments The authors would like to acknowledge the work of fellow investigators who contributed to the data for this publication: Martin Riegler, Medical University Vienna, Austria; David Ashton, London, United Kingdom; Rishi Singhal, Healthier Weight and the Reflux Center, Birmingham United Kingdom; Matthias Kemen, Evangelisches Krankenhaus, Herne, Germany. The authors also wish to acknowledge the professional medical writing services of Natalie Edwards.

Compliance with ethical standards

Disclosures Drs. Luigi Bonavina and Thomas Horbach have received consulting fees from Torax Medical Inc. in the past. Dr. Sebastian Schoppmann has received a research grant from Torax Medical Inc. Ms. DeMarchi is an employee of Ethicon. Drs. Luigi Bonavina, Sebastian Schoppmann and Thomas Horbach declare no current conflict of interest. This work was sponsored and partially funded by Torax Medical, Inc. (ClinicalTrials.gov identifier: NCT01624506).

References

1. Katz PO, Gerson LB, Vela MF (2013) Guidelines for the diagnosis and management of gastroesophageal reflux disease. Am J Gastroenterol 3:308–329
2. Toghanian S, Johnson DA, Stalhammar NO, Zerbib F (2011) Burden of gastro-oesophageal reflux disease in patients with persistent and intense symptoms despite proton pump inhibitor therapy: A post hoc analysis of the 2007 national health and wellness survey. Clin Drug Investig 1:703–715
3. Yadlapati R, Dakhoul L, Pandolfino JE, Keswani RN (2017) The Quality of Care for Gastroesophageal Reflux Disease. Dig Dis Sc 3:569–576
4. Stefanidis D, Hope WW, Kohn GP, Reardon PR, Richardson WS, Fanelli RD (2010) Guidelines for surgical treatment of gastroesophageal reflux disease. Surg Endosc 11:2647–2669
5. Ganz R (2016) A Review of New Surgical and Endoscopic Therapies for Gastroesophageal Reflux Disease. Gastroenterol Hepatol 7:424–431
6. Moore M, Afaneh C, Benhuri D, Antonacci C, Abelson J, Zarnegar R (2016) Gastroesophageal reflux disease: a review of surgical decision making. World J Gastrointest Surg 1:77–83
7. Patti M (2016) An evidence-based approach to the treatment of gastroesophageal reflux disease. JAMA Surg 1:73–78
8. Richter JE (2013) Gastroesophageal reflux disease treatment: side effects and complications of fundoplication. Clin Gastroenterol Hepato 5:465–471
9. Ganz RA (2017) A modern magnetic implant for gastroesophageal reflux disease. Clin Gastroenterol Hepatol 9:1326–1337
10. Ganz RA, Edmundowicz SA, Taiganides PA, Lipham JC, Smith CD, DeVault KR, Horgan S, Jacobsen G, Luketich JD, Smith CC, Schlack-Haerer SC, Kothari SN, Dunst CM, Watson TJ, Peters J, Oelschlager BK, Perry KA, Melvin S, Bemelman WA, Smout AJ, Dunn D (2016) Long-term outcomes of patients receiving a magnetic sphincter augmentation device for gastroesophageal reflux. Clin Gastroenterol Hepatol 5:671–677
11. Lipham JC, DeMeester TR, Ganz RA, Bonavina L, Saino G, Dunn DH, Fockens P, Bemelmen W (2012) The LINX® reflux management system: confirmed safety and efficacy now at 4 years. Surg Endosc 10:2944–2949
12. Warren HF, Reynolds JL, Lipham JC, Zehetner J, Bildzukewicz NA, Taiganides PA, Mickley J, Aye RW, Farivar AS, Louie BE (2016) Multi-institutional outcomes using magnetic sphincter augmentation versus Nissen fundoplication for chronic gastroesophageal reflux disease. Surg Endosc 8:3289–3296
13. Saino G, Bonavina L, Lipham JC, Dunn D, Ganz RA (2015) Magnetic Sphincter augmentation for gastroesophageal reflux at 5 years: final results of a pilot study show long-term acid reduction and symptom improvement. J Laparoendosc Adv Surg Tech A 10:787–792
14. Smith CD, DeVault KR, Buchanan M (2014) Introduction of mechanical sphincter augmentation for gastroesophageal reflux disease into practice: early clinical outcomes and keys to successful adoption. J Am Coll Surg 4:776–781
15. Skubleny D, Switzer NJ, Dang J, Gill RS, Shi X, de Gara C, Birch DW, Wong C, Hutter MM, Karmali S (2017) LINX® magnetic esophageal sphincter augmentation versus Nissen fundoplication for gastroesophageal reflux disease: a systematic review and meta-analysis. Surg Endosc 8:3078–3084
16. Reynolds JL, Zehetner J, Nieh A, Bildzukewicz N, Sandhu K, Katkhouda K, Lipham JC (2016) Charges, outcomes, and complications: a comparison of magnetic sphincter augmentation versus laparoscopic Nissen fundoplication for the treatment of GERD. Surg Endos 8:3225–3230
17. Ganz RA (2013) The esophageal sphincter device for treatment of GERD. Gastroenterol Hepatol (NY) 10:661–663
18. Prakash D, Campbell B, Wajed S (2018) Introduction into the NHS of magnetic sphincter augmentation: an innovative surgical therapy for reflux—results and challenges. Ann R Coll Surg Engl 4:251–256
19. Aiolfi A, Asti E, Bernardi D, Bonitta G, Rausa E, Siboni S, Bonavina L (2018) Early results of magnetic sphincter augmentation versus fundoplication for gastroesophageal reflux disease: systematic review and meta-analysis. Int J Surg 52:82–88
20. Bonavina L, Saino G, Bona D, Sironi A, Lazzari V (2013) One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 4:577–585
21. Bonavina L, Saino G, Lipham JC, Demeester TR (2013) LINX® Reflux Management System in chronic gastroesophageal reflux: a novel effective technology for restoring the natural barrier to reflux. Ther Adv Gastroenterol 4:261–268
22. Bell R, Lipham J, Louie B, Williams V, Luketich J, Hill M, Richards W, Dunst C, Lister D, McDowell-Jacobs L, Reardon P, Woods K, Gould J, Buckley FP III, Kothari S, Khaitan L, Smith CD, Park A, Smith C, Jacobsen G, Abbas G, Katz P (2019) Laparoscopic magnetic sphincter augmentation versus double-dose proton pump inhibitors for management of moderate-to-severe regurgitation in GERD: a randomized controlled trial. Gastrointest Endosc 1:14–22
23. Riegler M, Schoppman SF, Bonavina L, Ashton D, Horbach T, Kemen M (2015) Magnetic sphincter augmentation and fundoplication for GERD in clinical practice: one-year results of a multicenter, prospective observational study. Surg Endosc 5:1123–1129
24. Warren HF, Louie BE, Farivar AS, Wilshire C, Aye RW (2017) Manometric changes to the lower esophageal sphincter after magnetic sphincter augmentation in patients with chronic gastroesophageal reflux disease. Ann Surg 1:99–104
25. Velanovich V (2007) The development of the GERD-HRQL symptom severity instrument. Dis Esophagus 2:130–134
26. Velanovich V, Vallance SR, Gusz JR, Tapia FV, Harkabus MA (1996) Quality of life scale for gastroesophageal reflux disease. J Am Coll Surg 3:217–224
27. Dreyer NA, Schneeweiss S, McNeil BJ, Berger ML, Walker AM, Ollendorf DA, Gliklich RE (2010) GRACE principles: recognizing high-quality observational studies of comparative effectiveness. Am J Manag Care 6:467–471
28. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA (2011) Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. Clin Pharmacol Ther 6:777–790
29. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC (1992) Economic analysis in randomized control trials. Med Care 3:231–243
30. Bonavina L, Boyle N, Dunn C, Horbach T, Knowles TB, Lipham JC, Louie BE, Markar S, Schoppmann SF, Zehetner J (2020) Comment on Systematic review of the introduction and evaluation of magnetic augmentation of the lower oesophageal sphincter for gastro-oesophageal reflux disease. Br J Surg. https://doi.org/10.1002/bjs.11567
31. Sox HC (2010) Defining comparative effectiveness research: the importance of getting it right. Med Care 6:S7–8

32. Chang TI, Winkelmayer WC (2012) Comparative effectiveness research: what is it and why do we need it in nephrology? Nephrol Dial Transplant 6:2156–2161
33. Richards WO, McRae C (2018) Comparative analysis of laparoscopic fundoplication and magnetic sphincter augmentation for the treatment of medically refractory GERD. Am Surg 11:1762–1767
34. Guidozzi N, Wiggins T, Ahmed AR, Hanna GB, Markar SR (2019) Laparoscopic magnetic sphincter augmentation versus fundoplication for gastroesophageal reflux disease: systematic review and pooled analysis. Dis Esophagus. https://doi.org/10.1093/dote/doz031
35. Smith CD, McClusky DA, Rajad MA, Lederman AB, Hunter JG (2005) When fundoplication fails: redo? Ann Surg 6:861–869
36. Bell RC, Hufford RJ, Fearon J, Freeman KD (2013) Revision of failed traditional fundoplication using EsophyX transoral fundoplication. Surg Endosc 3:761–767
37. Ganz RA, Edmundowicz SA, Taiganides PA, Lipham JC, Smith CD, DeVault KR, Horgan S, Jacobsen G, Luketich JD, Smith CC, Schlack-Haerer SC, Kothari SN, Dunst CM, Watson TJ, Peters J, Oelschlager BK, Perry KA, Melvin S, Bemelmen WA, Smout AJ, Dunn D (2016) Long-term outcomes of patients receiving a magnetic sphincter augmentation device for gastroesophageal reflux. Clin Gastroenterol Hepatol 5:671–677
38. Ayazi S, Zheng P, Zaidi AH, Chovanec K, Salvitti M, Newhams K, Hoppo T, Jobe BA (2020) Clinical outcomes and predictors of favorable result after laparoscopic magnetic sphincter augmentation: single-institution experience with more than 500 patients. J Am Coll Surg 230(5):733–743
39. Gerson LB, Fass R (2009) A systematic review of the definitions, prevalence, and response to treatment of nocturnal gastroesophageal reflux disease. Clin Gastroenterol Hepatol 4:372–378
40. Mody R, Bolge SC, Kannan H, Fass R (2009) Effects of gastroesophageal reflux disease on sleep and outcomes. Clin Gastroenterol Hepatol 9:953–959

Observational Study

Medicine®

OPEN

Longitudinal comparison of quality of life in patients undergoing laparoscopic Toupet fundoplication versus magnetic sphincter augmentation

Observational cohort study with propensity score analysis

Emanuele Asti, MD, Gianluca Bonitta, MSc, Andrea Lovece, MD, Veronica Lazzari, MD, Luigi Bonavina, MD, PhD, FACS*

Abstract

Only a minority of patients with gastro-esophageal reflux disease (GERD) are offered a surgical option. This is mostly due to the fear of potential side effects, the variable success rate, and the extreme alteration of gastric anatomy with the current gold standard, the laparoscopic Nissen fundoplication. It has been reported that laparoscopic Toupet fundoplication (LTF) and laparoscopic sphincter augmentation using a magnetic device (LINX) can treat reflux more physiologically and with a lower incidence of side-effects and reoperation rate. We present the first comparing quality of life in patients undergoing LTF versus LINX.

Observational cohort study. Consecutive patients undergoing LTF or LINX over the same time period were compared by using the propensity score full matching method and generalized estimating equation. Criteria of exclusion were >3 cm hiatal hernia, grade C–D esophagitis, ineffective esophageal motility, body mass index >35, and previous upper abdominal surgery. The primary study outcome was quality of life measured with the Gastro-Esophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) questionnaire. Secondary outcomes were proton pump inhibitors (PPI) use, presence of gas-related symptoms or dysphagia, and reoperation-free probability.

Between March 2007 and July 2014, 238 patients with GERD met the criteria of inclusion in the study. Of these, 103 underwent an LTF and 135 a LINX procedure. All patients had a minimum 1-year follow-up. Over time, patients in both groups had similar GERD-HRQL scores (odds ratio [OR] 1.04, confidence interval [CI] 0.89–1.27; $P=0.578$), PPI use (OR 1.18, CI 0.81–1.70; $P=0.388$), gas-related symptoms (OR 0.69, CI 0.21–2.28; $P=0.542$), dysphagia (OR 0.62, CI 0.26–1.30; $P=0.241$), and reoperation-free probability (stratified log-rank test = 0.556).

In 2 concurrent cohorts of patients with early stage GERD undergoing LTF or LINX and matched by propensity score analysis, health-related quality of life significantly improved and GERD-HRQL scores had a similar decreasing trend over time up to 7 years of follow-up. We conclude that LTF and LINX provide similar disease-specific quality of life over time in patients with early stage GERD.

Abbreviations: GEE = generalized estimating equation, GERD = gastro-esophageal reflux disease, GERD-HRQL = Gastro-Esophageal Reflux Disease-Health Related Quality of Life, LTF = laparoscopic Toupet fundoplication, PPI = proton pump inhibitors, PS = propensity score.

Keywords: gastroesophageal reflux, GERD-HRQL score, laparoscopic Nissen fundoplication, laparoscopic Toupet fundoplication, lower esophageal sphincter, magnetic LES augmentation, propensity score

1. Introduction

Gastro-esophageal reflux disease (GERD) is a highly prevalent disease affecting up to 30% of the population in Western countries. The burden of the disease in the general population is probably underestimated because many people with symptoms do not consult a physician. The diagnosis of GERD has increased more than 200% from 1998 to 2005, and it is now the most common reason for access to gastroenterology outpatient clinics.[1] The impact of GERD on quality of life is worse than other common chronic conditions such as diabetes, arthritis, and congestive heart failure. Gastroesophageal reflux interferes with physical activity and social functioning, disturbs sleep, reduces productivity at work, and leads to increased healthcare resource utilization; therefore, the primary goal of therapy in uncomplicated GERD is to improve patient's symptoms and quality of life over time.[2]

Proton pump inhibitors (PPI) represent the first-line therapy approach in GERD; however, nearly 40% of patients have

Editor: Papavramidis Theodossis.

LB has been consultant for Torax Medical Inc.

The authors have no funding and conflicts of interest to disclose.

Department of Biomedical Sciences for Health, Division of General Surgery, University of Milan Medical School, IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy.

* Correspondence: Luigi Bonavina, Dipartimento di Chirurgia, IRCCS Policlinico San Donato, Piazza Malan 2, 20097 San Donato Milanese, Milan, Italy (e-mail: luigi.bonavina@unimi.it).

Medicine (2016) 95:30(e4366)

Received: 9 March 2016 / Received in final form: 21 April 2016 / Accepted: 5 July 2016

http://dx.doi.org/10.1097/MD.0000000000004366

inadequate symptom relief despite high dose medication. This is mainly because the therapeutic gain for the relief of regurgitation is modest and considerably lower than that for heartburn.[3]

Laparoscopic Nissen fundoplication is generally recognized as the gold standard of antireflux surgery worldwide. On the other hand, the Nissen procedure is highly operator-dependent, has a variable success rate, can lead to potential side effects, and is regarded by some a sort of overtreatment for patients with mild to moderate GERD. As a consequence, the number of Nissen fundoplications has steadily declined over the recent years.[4]

The debate about the choice of the most appropriate surgical technique to provide optimal reflux control while minimizing the side effects is still ongoing. It has been assumed that the laparoscopic Toupet fundoplication (LTF) would provide less outflow resistance, thereby lowering the dysphagia and the bloating rate, and some surgeons favor this operation arguing that a partial fundoplication is more physiological and effective at least in patients with "mild" disease.[5]

Over the past 15 years we have been performing LTF in patients with early stage GERD and in those with large hiatal hernia or ineffective esophageal motility. In 2007, we started to perform laparoscopic magnetic sphincter augmentation with the LINX device as part of a feasibility trial,[6] and this is still an option we offer to patients with early stage GERD. It is a simple standardized laparoscopic procedure that does not alter gastric anatomy, provides relief of reflux-related symptoms without impeding the ability to belch or vomit, and is reversible if necessary.[7] The LINX device is FDA approved and is currently available in the market.

The aim of this study was to assess and compare health-related quality of life over time in 2 concurrent cohorts of patients undergoing LTF or LINX matched by propensity score (PS).

2. Methods

2.1. Study design

Observational cohort study. All patients undergoing an LTF or LINX procedure between March 2007 and July 2014 were identified from a prospectively collected data base. The time frame was chosen to include patients undergoing both surgical operations during the same period and to allow for a minimum 1-year follow-up. Ethical approval was waived and written consent was not given to the patients because all data were analyzed anonymously. We adopted the STROBE criteria for reporting an observational study.[8]

Inclusion criteria were age >18 years, chronic GERD symptoms despite PPI use for at least 6 months, objective evidence of reflux at the pH study, and normal esophageal motility documented by manometry. Exclusion criteria were the following: hiatal hernia >3 cm, esophagitis grade C–D, ineffective esophageal motility, body mass index >35, and previous esophago-gastric surgery. The primary outcome was postoperative quality of life measured with the validated and disease-specific Gastro-Esophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) questionnaire.[9] Questionnaires were administered in the outpatient clinic to all patients. Secondary outcomes were PPI use, presence of gas-related symptoms or dysphagia, and reoperation-free probability. Patients in both groups were evaluated at 3 to 12 months, and then every 12 months with the GERD-HRQL survey plus questions about PPI use, gas-related symptoms and dysphagia.

2.2. Surgical procedure

Both the LTF and LINX procedures were performed according to a standard institutional protocol. In the LTF procedure, the right and left diaphragmatic crus were dissected in order to create a large retroesophageal window and encircle the esophagus en bloc with the posterior vagus nerve. The vagal branch to the gallbladder was routinely preserved. After division of the phrenoesophageal ligament, the esophagus was pulled down and dissection carried out in the mediastinum in order to obtain a 4 cm tension-free intra-abdominal esophageal segment. A posterior hiatoplasty was routinely performed. The gastrophrenic ligament along with the proximal short gastric vessels was divided and the fundus was wrapped behind the esophagus. Four cardinal nonabsorbable stitches were placed to held the fundoplication in place by tailoring the fundus symmetrically on both sides and without tension. The 2 proximal stitches were also fixed to the diaphragm. Finally, both edges of the fundic wrap were sutured to the right and left side of the esophageal wall using nonabsorbable running sutures.

In the LINX procedure, the peritoneal reflection overlying the esophago-gastric junction was divided, and the mediastinal cavity was not routinely entered in an attempt to preserve the phrenoesophageal ligament. A posterior hiatoplasty was added at discretion of the surgeon only when the hiatus appeared to be markedly enlarged. The vagal branch to the gallbladder was routinely preserved. The posterior vagus nerve and the area corresponding to the Z line were identified and a tunnel was created between the posterior vagus nerve and the esophagus. A special sizing instrument was used to measure the circumference of the esophagus and an appropriately sized device was placed through the tunnel. The ends of the device were then secured anteriorly.

2.3. Statistical analysis

Continuous data are presented as median and interquartile range (IQR). Categorical variables are shown as numbers and percentages. Wilcoxon signed-rank, Wilcoxon matched pairs signed-rank, Fisher exact, or χ^2 tests and conditional logistic regression were performed as appropriate. Confidence intervals (CIs) at 95% confidence level, 2-sided statistical test with type I error = 0.05, and q-value (q) for multiple test with false discovery rate = 0.05. All analyses were carried out using R software package version 3.2.2.[10]

To reduce the impact of treatment selection bias inherent to an observational study, we compared postoperative GERD-HRQL of patients with LTF and LINX using the PS matching method. The PS, defined as the conditional probability of assignment to a treatment given a vector of particular observed covariates, is designed to mimic some of the particular characteristics of a randomized clinical trial in the context of an observational study. As appropriate and with caution, PS analysis allows estimation of relative risk in binary outcomes.[11]

We computed a PS for individual patients with logistic regression using demographic and clinical variables,[12] and evaluated the interaction among all preoperative covariates and square terms without time-dependent variables. The generalized additive model was used to check linear assumption in PS model. The 2 patient groups were then matched using the PS full-matching method.[13] The balance of baseline covariates after matching was assessed using the standardized difference of mean,[14] and the overlap degree of PS distribution. We also performed sensitivity analysis[15] to assess possible hidden bias

due to unobserved confounders. We analyzed correlated repeated measure of GERD-HRQL, PPI use, presence of gas-related symptoms, and dysphagia over time with generalized estimating equation (GEE) in PS full-matched dataset; we used sandwich estimator and autoregressive of order 1 working correlation matrix.[16] In particular, to account for skewed and doubly bounded nature of GERD-HRQL scores, we performed linear transformation of GERD-HRQL values. Then, we specified mean and variance function of beta distribution with logit as mean function.[17] The linear predictor for means includes preoperative transformed GERD-HRQL, treatment, time, and time–treatment interaction. GEE takes into account that the PS method allows for estimation of marginal treatment effect and the matched nature of data. For models selection we used quasi-likelihood information criterion. Wald test and Wald CI were computed in GEE analysis. We established a 1.5 clinical effect-size threshold for the odds ratio (OR), a value that is compatible with clinical experience and published indices.[18]

The Kaplan–Meier reoperation-free probability curves were estimated separately for LTF and LINX in the PS full-matched sample. Stratified log-rank test was used to compare the curves. Hazard ratio was estimated with univariate marginal survival model with robust standard errors in PS full-matched data set.[14] Proportional hazard assumption was tested. Reference group was LTF for all models.

3. Results

All 238 patients were successfully treated via a laparoscopic approach. The duration of the surgical procedure was 87 minutes (IQR 28) in the LTF group and 42 minutes (IQR 34) in the LINX group ($P < 0.001$). One patient in the LINX group had a respiratory arrest within the 1st hour postoperatively and was successfully resuscitated without consequences. Postoperative morbidity consisted of atrial fibrillation (n = 1), urinary retention (n = 1), and bleeding from a trocar site (n = 1), all occurring in the LTF group.

3.1. Preoperative patient characteristics

The preoperative patients' characteristics are reported in Table 1. There were statistically significant differences in 6 of the 19 covariates. Thus, as typically occurs in observational studies, there were systematic differences in preoperative characteristics between the 2 treatments.

In the PS full-matched data, 16 of the variables had absolute standardized differences of mean after matching that exceeded 0.10 (Table 2). The absolute standardized differences of mean ranged from 0.06 to 0.1, with a median of 0.044 (25th and 75th percentiles 0.023–0.071, respectively), indicating that the means and prevalences of continuous and dichotomous variables were very similar between treatment groups. The variance ratios for continuous variables ranged from 0.83 to 1.10, indicating that the variance variables were similar between the 2 treatment groups.

3.2. Postoperative follow-up

All patients had a minimum 1-year follow-up. The mean postoperative follow-up was 42 and 44 months in LTF and LINX groups, respectively.

3.3. Quality of life

The GERD-HRQL score significantly decreased within normal values to a similar extent after both procedures (Fig. 1A and B).

Table 1

Baseline characteristics of the study sample.

	Toupet (n = 103)	LINX (n = 135)	*P*
Sex, M/F	61/42	44/91	<0.001
Age, y	50 (24)	44 (20)	ns
BMI, kg/m^2	25.10 (4.47)	23.94 (4.54)	ns
Comorbidities, n (%)			
Anxiety/depression	4 (4)	20 (15)	0.008
Asthma	5 (5)	9 (7)	ns
Barrett esophagus	7 (7)	6 (4)	ns
Hypertension	24 (23)	20 (15)	ns
Symptoms, n (%)			
Typical	56 (54)	43 (32)	<0.001
Atypical	1 (1)	7 (5)	ns
Mixed	46 (45)	85 (63)	0.008
GERD-HRQL score	19.70 (11.00)	21.00 (9.00)	ns
Disease duration, y	6.0 (7.0)	5.0 (7.0)	ns
Duration of PPI therapy, y	5.0 (5.0)	4.0 (5.5)	ns
%pH < 4 (total)	8.30 (7.70)	8.00 (6.15)	ns
%pH < 4 (upright)	6.50 (7.30)	7.90 (7.25)	0.04
%pH < 4 (supine)	11.00 (14.20)	5.80 (9.40)	0.002
DeMeester score	37.6 (26.4)	31.4 (25.3)	ns
Hernia size, cm	2.00 (1.00)	2.00 (1.00)	ns
Hospital stay, d	2.0 (1.0)	2.0 (0.0)	ns

Continuous values are expressed as median and interquartile range.
BMI = body mass index, GERD-HRQL = Gastro-Esophageal Reflux Disease-Health Related Quality of Life, PPI = proton pump inhibitors.

Parameters estimate by the beta GEE model for GERD-HRQL are shown in Table 3. Over time, there was no statistical difference in the GERD-HRQL scores between the LTF and LINX groups as indicated by the time–treatment interaction term

Table 2

Standardized difference of mean before and after full propensity score matching with relative balance improvement (%) for each baseline characteristic.

	Standardized difference of mean	
Variables	**Before matching**	**After matching**
Propensity score	0.494	0.010
Sex	0.556	0.012
Age	0.230	0.024
BMI	−0.319	0.050
Comorbidities		
Anxiety/depression	−0.299	0.022
Asthma	0.060	0.039
Barrett esophagus	−0.134	0.090
Hypertension	−0.221	0.057
Symptoms		
Typical	0.465	0.044
Atypical	−0.187	0.057
Mixed	−0.363	−0.070
GERD-HRQL	0.150	0.090
Disease duration	−0.252	−0.089
PPI therapy duration	−0.159	0.006
%pH < 4 (total)	0.049	0.010
%pH < 4 (upright)	0.191	0.042
%pH < 4 (supine)	−0.189	0.033
DeMeester score	−0.029	0.014
Hernia size	0.103	0.100
Hospital stay	1.304	0.072

BMI = body mass index, GERD-HRQL = Gastro Esophageal Reflux Disease-Health Related Quality of Life, PPI = proton pump inhibitors.

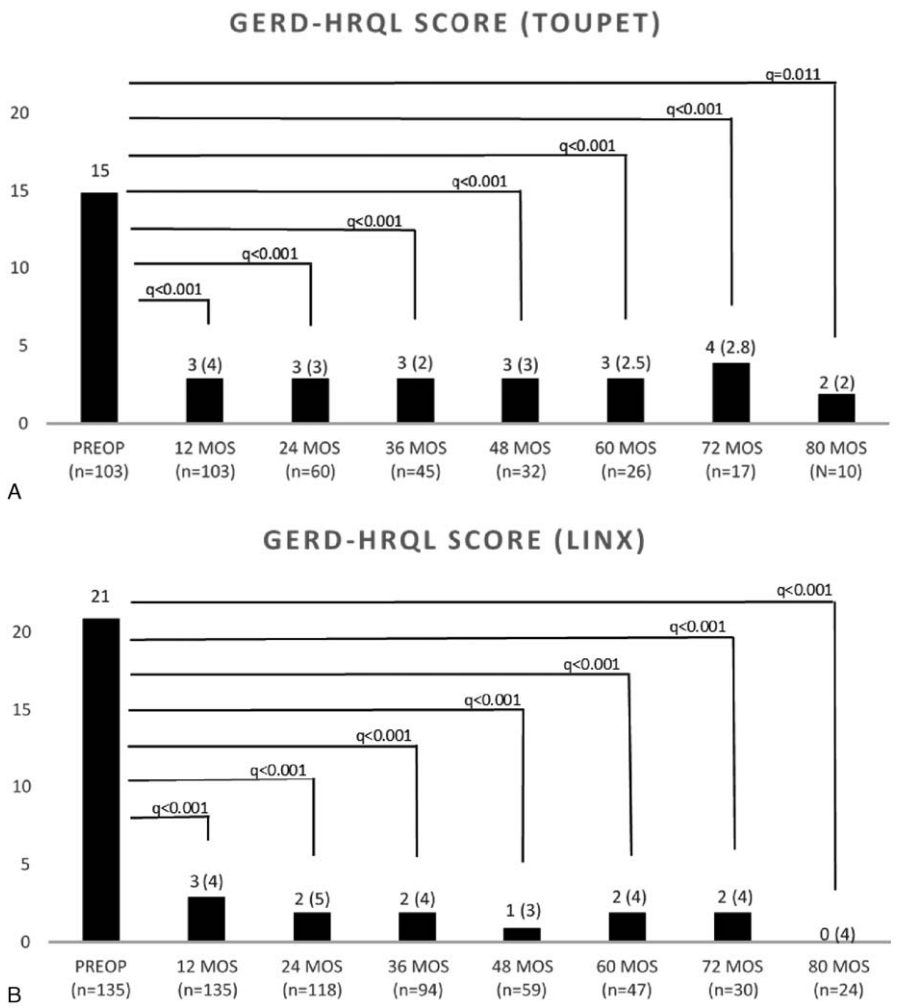


Figure 1. (A and B) Comparison of pre and postoperative GERD-HRQL scores in patients with LTF and LINX at each measured follow-up time. Wilcoxon matched pairs signed-rank, q values (q), and false discovery rate = 0.05 level. Values are expressed as median and interquartile range. GERD-HRQL = Gastro-Esophageal Reflux Disease-Health Related Quality of Life, LTF = laparoscopic Toupet fundoplication.

in the GEE model (OR 1.04, CI 0.89–1.27; $P = 0.578$). Furthermore, the OR CI did not encompass the clinical significance previously established at 1.5 threshold. The raw analysis showed similar results (OR 0.99, CI 0.94–1.03; $P = 0.528$). Due to the fact that the pattern of change was the same over time in both groups, we assumed a linear trend and refitted the model by excluding the interaction term to investigate the trend of the logit of mean of GERD-HRQL. This showed a decreasing linear trend over time (GEE model time parameter −0.069, CI −0.104 to −0.032; $P < 0.001$) (Fig. 2).

3.4. PPI use, gas-related symptoms, dysphagia, and reoperation rate

As indicated by the time–treatment interaction term in the GEE model, over time there was no statistical difference in PPI use (OR 1.18, CI 0.81–1.70; $P = 0.388$), gas-related symptoms (OR 0.69, CI 0.21–2.28; $P = 0.532$), and dysphagia (OR 0.62, CI 0.26–1.30; $P = 0.241$) between the LTF and LINX groups. As expected, the prevalence of dysphagia was significantly greater in patients with LINX at 3-month follow up (OR 9.42, CI 2.22–20.10; $P < 0.001$). At 1-year follow-up, there was no difference in the prevalence of dysphagia (interaction time term not significant, treatment term 0.78, CI −0.02 to 1.57; $P = 0.357$). There was no statistical difference in the reoperation-free probability between patients with LTF and LINX (stratified log-

Table 3

Results of beta GEE model for GERD-HRQL outcomes.

	Beta GEE		
Parameter coefficient	**Estimate**	**95% Confidence interval**	***P***
Intercept	−1.735	−2.359/−1.111	<0.001
GERD-HRQL pre	0.318	−0.770/1.406	0.365
Time	−0.130	−0.261/−0.002	0.049
Procedure	−0.309	−0.666/0.047	0.090
Procedure × Time	0.043	−0.110/0.197	0.578
Covariance estimates			
Correlation parameters AR(1)	0.565	—	—
Scale	0.267	—	—

AR = auto regressive, GEE = generalized estimating equation, GERD-HRQL = Gastro-Esophageal Reflux Disease-Health Related Quality of Life.

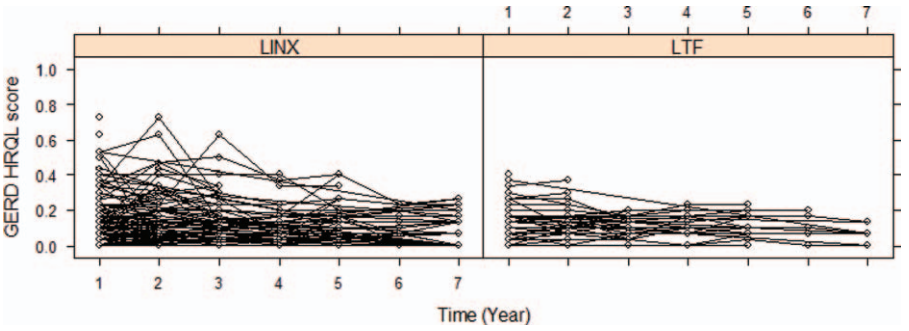


Figure 2. Spaghetti plot: trajectories over time of linearly transformed GERD-HRQL scores for each individual according to the surgical technique. GERD-HRQL = Gastro-Esophageal Reflux Disease-Health Related Quality of Life.

rank test, $P=0.556$, HR 0.77, CI 0.234–2.57; $P=0.687$). At 80 months, the estimated Kaplan–Meier reoperation-free probability survival was 0.97 (CI 0.88–0.99) in patients with LTF and 0.94 (CI 0.89–0.98) in patients with LINX (Fig. 3).

All study outcomes are summarized in Table 4.

4. Discussion

The main finding of this observational study is that long-term health-related quality of life was similar in patients undergoing LTF and LINX. We also found that over time the GERD-HRQL scores had a similar decreasing trend in both groups. Longitudinal data, analyzed with appropriate statistical techniques such as GEE, can depict the temporal evolution of the outcome and take into account intra-individual variability. Since patient perception and satisfaction are reasonable and measurable indicators of the success of a surgical procedure, we assume that longitudinal quality of life data play an important role in the process of decision-making and counseling patients with GERD who are candidates for a laparoscopic antireflux procedure.

The long-held dogma that Nissen fundoplication is the ideal antireflux operation has recently been challenged. Systematic review and meta-analyses of randomized controlled studies have shown that LTF is equally effective to improve quality of life and is associated with less postoperative dysphagia and gas-related symptoms compared to Nissen fundoplication.[19,20]

Laparoscopic magnetic sphincter augmentation with the LINX device is an emerging surgical option for the treatment of GERD. A single-center study[21] and a multicenter single-arm study,[22] enrolling 100 patients each, evaluated the long-term results of magnetic augmentation and showed that the procedure provides significant and sustained control of reflux with minimal side-effects or complications up to 6 years of follow-up. However, no randomized trials exist that can validate the effectiveness of the LINX and reliably compare its results with other established surgical therapies. Interestingly, the patient profile for the LINX procedure is very similar to that required for the Toupet fundoplication, which has been successfully employed in patients with "mild" GERD.[5] Three recent observational studies have compared LINX and laparoscopic Nissen fundoplication. Louie et al[23] performed a retrospective case–control study comparing 66 patients undergoing LINX or Nissen fundoplication; at a mean follow-up of 6 and 10 months, respectively, scores on the GERD-HRQL scale significantly improved in both groups. Reynolds et al[24] conducted a retrospective analysis of 1-year outcomes of 100 patients matched by PS; although the GERD-HRQL scores were similar in both groups, there were 10.6% of patients in the Nissen group complaining of severe gas-bloat symptoms compared with 0% in the LINX group. Finally, a smaller case-matched study by Sheu et al[25] including 24 patients followed for an average of 7 months showed that severe dysphagia requiring endoscopic dilation was more frequent in the LINX group.

In the present study, for the first time, the long-term outcomes of a sizeable number of patients with LTF and LINX having a

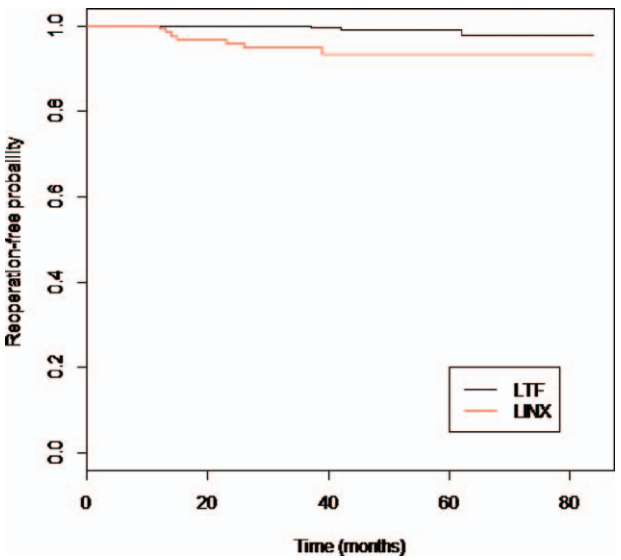


Figure 3. Reoperation-free probability in patients with LTF and LINX (stratified log-rank test for comparison of Kaplan–Meier curves, $P=0.556$). LTF = laparoscopic Toupet fundoplication.

Table 4

Results of primary and secondary study outcomes over time.

Outcome	Measure	Estimate	CI	P
HRQL	OR	1.04	0.89–1.27	0.578
PPI use	OR	1.18	0.81–1.70	0.388
Gas-related symptoms	OR	0.69	0.21–2.28	0.532
Dysphagia	OR	0.62	0.26–1.30	0.241
Reoperation rate	HR	0.77	0.234–2.57	0.687

OR values are relative to time procedure interaction of GEE.

CI = confidence interval, GEE = generalized estimating equation, HR = hazards ratio, HRQL = Health Related Quality of Life, OR = odds ratio, PPI = proton pump inhibitors.

minimum 1-year follow-up and sharing similar characteristics of disease severity were compared in a longitudinal manner. We confirm that these surgical procedures equally normalize GERD-HRQL scores, and that the results are maintained over time for up to 7 years with a similar trend in both groups. As expected, dysphagia occurred more frequently at 3 months in patients undergoing the LINX procedure but this difference disappeared at 1 year. The reason for reoperation among patients with LINX was persistent dysphagia in 3, recurrent heartburn/regurgitation in 3, and erosion in 1. In all these individuals, the LINX device was safely removed through laparoscopy and a standard fundoplication was performed. Conversely, all 4 patients with LTF who required a reoperation complained of recurrent heartburn/regurgitation and underwent Nissen fundoplication.

A strength of this study is that we used a longitudinal model to analyze health-related quality of life scores over time. To our knowledge, this is the first time that the GERD-HRQL questionnaire is employed in a longitudinal study using beta GEE analysis. As opposed to pure cross-sectional studies, longitudinal studies with repeated measurements taken over time are more reliable in establishing causality.[26] Waiting for a randomized clinical trial, the present study may provide some indications to surgeons and patients for selecting the most suitable laparoscopic antireflux procedure. The study sample size was sufficient to produce a narrow OR CI that did not encompass the clinical significance. The possible residual selection bias of our propensity-matched analysis was further mitigated by the fact that we have compared LTF to the LINX and not to the Nissen, a procedure that we usually reserve to patients with advanced GERD. In addition, we used similar criteria of exclusion from the study for both procedures.

Limitations of this study are that the GERD-HRQL is a validated, but still subjective test, and the LINX procedures were not standardized regarding crural repair. Hidden bias typical of an observational study cannot be excluded due to unmeasured and unmeasurable confounding factors. The PS model could be biased, and we did not consider possible measurable time-dependant confounders. Despite the sensitivity analysis showed negligible residual bias, we need to be cautious in interpreting the overall study results.

5. Conclusions

Both LTF and LINX can be safely offered as a first choice surgical option in patients with early stage GERD. However, a randomized clinical trial would be required to demonstrate the equivalence of the 2 procedures. Compared to fundoplication, LINX appears to be a simple, standardized, and easily reversible procedure that does not alter gastric anatomy; operative time is shorter but the cost of the device should be considered. Further research is needed to investigate correlation between longitudinal quality of life data with objective long-term outcome of these procedures.

References

[1] Gastroesophageal Reflux Disease (GERD) Hospitalizations in 1998 and 2005. HCUP-US Home Page. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb44.jsp.

[2] Wicklund L. Review of the quality of life and burden of illness in gastroesophageal reflux disease. Dig Dis 2004;22:108–14.

[3] Kahrilas PJ, Howden CW, Hughes N. Response of regurgitation to proton pump inhibitor therapy in clinical trials of gastroesophageal reflux disease. Am J Gastroenterol 2011;106:1419–25.

[4] Finks JF, Wei Y, Brinkmeyer JD. The rise and fall of antireflux surgery in the United States. Surg Endosc 2006;20:1698–701.

[5] Bell RCW, Hanna P, Mills MR, et al. Patterns of success and failure with laparoscopic Toupet fundoplication. Surg Endosc 1999;13:1189–94.

[6] Bonavina L, Saino G, Bona D, et al. Magnetic augmentation of the lower esophageal sphincter: results of a feasibility clinical trial. J Gastrointest Surg 2008;12:2133–40.

[7] Bonavina L, DeMeester TR, Fockens P, et al. Laparoscopic sphincter augmentation device eliminates reflux symptoms and normalizes esophageal acid exposure: one and 2-year results of a feasibility trial. Ann Surg 2010;252:857–62.

[8] von Elm E, Altman DG, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573–7.

[9] Velanovich V. Comparison of generic (SF-36) vs disease-specific quality of life (GERD-HRQL) scales for gastroesophageal reflux disease. J Gastrointest Surg 1998;2:141–5.

[10] R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015. ISBN 3-900051-07-0. http://www.R-project.org. Accessed February 29, 2016.

[11] Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res 2011;46:399–424.

[12] Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. Am J Epidemiol 2006;163:1149–56.

[13] Stuart EA, Green KM. Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes. Dev Psychol 2008;44:395–406.

[14] Austin P. A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. Multivar Behav Res 2011;46:119–51.

[15] Liu W, Kuramoto SJ, Stuart EA. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. Prev Sci 2013;14:570–80.

[16] Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. 2nd ed.New York:Wiley; 2011.

[17] Hunger M, Döring A, Holle R. Longitudinal beta regression models for analyzing health-related quality of life scores over time. BMC Med Res Methodol 2012;12:144.

[18] Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. J Grad Med Educ 2012;9:279–82.

[19] Broeders JAJL, Mauritz FA, Ahmed Ali U, et al. Systematic review and meta-analysis of laparoscopic Nissen (posterior total) versus Toupet (posterior partial) fundoplication for gastro-oesophageal reflux disease. Br J Surg 2010;97:1318–30.

[20] Tian XC, Wang B, Shan CX, et al. A meta-analysis of randomized controlled trials to compare long-term outcomes of Nissen and Toupet fundoplication for gastroesophageal reflux disease. PLoS ONE 2015;10:e0127627.

[21] Bonavina L, Saino G, Bona D, et al. One-hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 2013;17:577–85.

[22] Ganz RA, Edmundowicz SA, Taiganides P, et al. Long-term outcomes of patients receiving a magnetic sphincter augmentation device for gastroesophageal reflux. Clin Gastroenterol Hepatol 2016;14:671–7. doi:10.1016/j.cgh.2015.05.028 [Epub ahead of print].

[23] Louie BE, Farivar A, Shultz D, et al. Short-term outcomes using magnetic sphincter augmentation versus Nissen fundoplication for medically resistant gastroesophageal reflux disease. Ann Thorac Surg 2014;98:498–504.

[24] Reynolds J, Zehetner J, Wu P, et al. Laparoscopic magnetic sphincter augmentation vs laparoscopic Nissen fundoplication: a matched-pair analysis of 100 patients. J Am Coll Surg 2015;221:123–8.

[25] Sheu EG, Nau P, Nath B, et al. A comparative trial of laparoscopic magnetic sphincter augmentation and Nissen fundoplication. Surg Endosc 2015;29:505–9.

[26] Cambridge University Press, Frees EW. Longitudinal and Panel Data: Analysis and Applications in the Social Sciences. 2004.

Short-Term Outcomes Using Magnetic Sphincter Augmentation Versus Nissen Fundoplication for Medically Resistant Gastroesophageal Reflux Disease

Brian E. Louie, MD, Alexander S. Farivar, MD, Dale Shultz, BS, Christina Brennan, CCRP, Eric Vallières, MD, and Ralph W. Aye, MD

Division of Thoracic Surgery, Swedish Cancer Institute and Medical Center, Seattle, Washington

Background. In 2012 the United States Food and Drug Administration approved implantation of a magnetic sphincter to augment the native reflux barrier based on single-series data. We sought to compare our initial experience with magnetic sphincter augmentation (MSA) with laparoscopic Nissen fundoplication (LNF).

Methods. A retrospective case-control study was performed of consecutive patients undergoing either procedure who had chronic gastrointestinal esophageal disease (GERD) and a hiatal hernia of less than 3 cm.

Results. Sixty-six patients underwent operations (34 MSA and 32 LNF). The groups were similar in reflux characteristics and hernia size. Operative time was longer for LNF (118 vs 73 min) and resulted in 1 return to the operating room and 1 readmission. Preoperative symptoms were abolished in both groups. At 6 months or longer postoperatively, scores on the Gastroesophageal Reflux Disease Health Related Quality of Life scale improved from 20.6 to 5.0 for MSA vs 22.8 to 5.1 for LNF. Postoperative DeMeester scores (14.2 vs 5.1, $p = 0.0001$) and the percentage of time pH was less than 4 (4.6 vs 1.1; $p = 0.0001$) were normalized in both groups but statistically different. MSA resulted in improved gassy and bloated feelings (1.32 vs 2.36; $p = 0.59$) and enabled belching in 67% compared with none of the LNFs.

Conclusions. MSA results in similar objective control of GERD, symptom resolution, and improved quality of life compared with LNF. MSA seems to restore a more physiologic sphincter that allows physiologic reflux, facilitates belching, and creates less bloating and flatulence. This device has the potential to allow individualized treatment of patients with GERD and increase the surgical treatment of GERD.

(Ann Thorac Surg 2014;98:498–505)

Chronic gastroesophageal reflux disease (GERD) occurs in patients as a spectrum of disease that ranges from the endoscopically normal esophagus to erosive esophagitis and to Barrett's esophagus [1]. This is influenced by a hiatal hernia of varying sizes, a stricture of varying degrees (rings to fibrotic), and the potential for a foreshortened esophagus. Despite the wide variation, only two dominant therapies have been used to treat the entire spectrum of GERD during the past 70 years: Nissen fundoplication and proton pump inhibitors (PPIs). Both treatments are effective at controlling GERD, with a slight advantage toward operative treatment based on two randomized control trials [2, 3].

Despite similar outcomes, there is a large gap in the use of both treatments. Using current rates of antireflux operations, it is estimated that surgical repair is used in less than 1% of patients [4]. Even though PPIs are the dominant therapy, only 60% of patients are satisfied with their treatment [5]. This leaves a therapy gap of at least 40% of patients who are taking PPIs with ongoing GERD symptoms. These patients are either not being referred for an equally effective therapy or have chosen not to undergo surgical treatment. The reasons for this include concerns about the ability to belch or vomit and the development of hyperflatulence or bloating [2, 6]. Furthermore, there are concerns about the perceived invasiveness and durability of the surgical outcomes because upwards of 25% of repairs will deteriorate over time [2]. This creates an opportunity for the development of new treatments.

In March 2012, the United States Food and Drug Administration approved a novel device to control GERD composed of a series of magnets set in a titanium casing and connected by titanium wires interconnected with a hollow housing in the configuration of a Roman arch

Accepted for publication April 14, 2014.

Presented at the Fiftieth Annual Meeting of The Society of Thoracic Surgeons, Orlando, FL, Jan 25–29, 2014.

Address correspondence to Dr Louie, Swedish Cancer Institute and Medical Center, Division of Thoracic Surgery, 1101 Madison St, Ste 900, Seattle, WA 98105; e-mail: brian.louie@swedish.org.

Dr Louie discloses a financial relationship with Torax Medical, Inc.

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2014.04.074

(Fig 1). This "sphincter," by virtue of the magnets, is potentially durable for the life of the patient and would augment the lower esophageal sphincter by limiting lower esophageal shortening and relaxations during gastric distension but open to gastric pressure to allow belching, and thus prevent hyperflatulence and bloating. One small pilot trial [7] and two single-series trials [8, 9] comprising 244 patients have demonstrated its initial efficacy. However, no comparison with standard treatments has been performed and is necessary. We evaluated our experience with magnetic sphincter augmentation (MSA) and compared it with laparoscopic Nissen fundoplication (LNF) at 6 months.

Material and Methods

We retrospectively reviewed prospectively collected data on consecutive patients who underwent laparoscopic implantation of a magnetic sphincter at Swedish Medical Center from September 2012 to December 2013. The Institutional Review Board of Swedish Medical Center approved this study and waived the need to implant the devices under a research protocol. Magnetic sphincters were placed as part of clinical care, and patient consent was provided for implantation; however, individual patient consent for this study was waived because of the study's retrospective nature.

For comparison, we reviewed 427 primary antireflux repairs from a prospectively maintained benign esophageal surgical database from January 2010 to July 2013 to identify consecutive patients undergoing LNF. We excluded patients based on the following criteria: age younger than 18 years, body mass index exceeding 36 kg/m^2, hiatal hernia exceeding 3 cm in axial length without a paraesophageal component, Barrett's esophagus exceeding 1 cm, and named motility disorders.

We identified 98 patients, with 50 excluded because they were part of another clinical trial and a further 16 patients excluded when preoperative video esophagograms and endoscopic photos showed the hiatal hernias were too large. Ten of the 32 Nissen patients were considered for MSA before the Nissen but chose not to proceed with MSA, were denied by insurance, or were excluded due to the need for magnetic resonance imaging. Additional exclusion criteria included allergy to metal, delayed gastric emptying, prior esophageal or gastric operations, and an esophageal stricture.

All patients underwent preoperative evaluation, including video esophagogram, esophagogastroduodenoscopy, pH analysis with a 48-hour wireless probe or a 24-hour impedance-pH catheter, and high-resolution manometry. Patients underwent postoperative clinical follow-up at approximately 2 weeks, 6 weeks, and 6 months. Quality of life and symptom severity were assessed with the Gastroesophageal Reflux Disease Health Related Quality of Life (GERD-HRQL) scale, Quality of Life in Reflux and Dyspepsia (QOLRAD), and a modified Dakkak Dysphagia Severity Score preoperatively and at each clinic follow-up appointment. At 6 months, patients were requested and encouraged to repeat preoperative studies.

At endoscopy, the gastroesophageal junction was evaluated using the Hill Classification, and the presence or absence of esophagitis was graded according to the Los Angeles (LA) Classification system. The presence or absence of a hiatal hernia was noted and the size determined by the distance from the top of the rugal folds and the diaphragmatic impression. The patient's use of PPIs was stopped 7 days before pH analysis. The highest score during a 48-hour wireless probe evaluation was used for the DeMeester score and the percentage of time the pH was less than 4.

Operative Techniques

LNF was performed using 5 ports. The esophageal hiatus was completely dissected and mediastinal dissection carried to the level of the inferior pulmonary veins. The upper gastric fundus was mobilized by dividing the proximal short gastric vessels and the retrogastric pancreatic attachments. Once 3 cm of intraabdominal esophagus was established, the esophageal hiatus was closed with single "0" polyester sutures (Ethicon, Cincinnati, OH) and a Ti-Knot (LSI Solutions, Victor, NY) suture-securing device.

After hiatal closure, a 2-0 silk marking suture was placed on the posterior fundus 6 cm down the greater

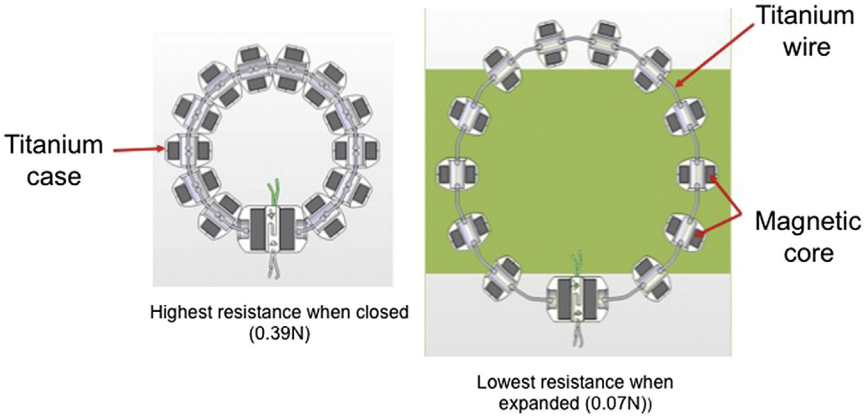


Fig 1. Magnetic sphincter device (left) closed and (right) open. (Images courtesy of Torax Medical Inc, Shoreview, Minnesota.)

Ann Thorac Surg
2014;98:498–505

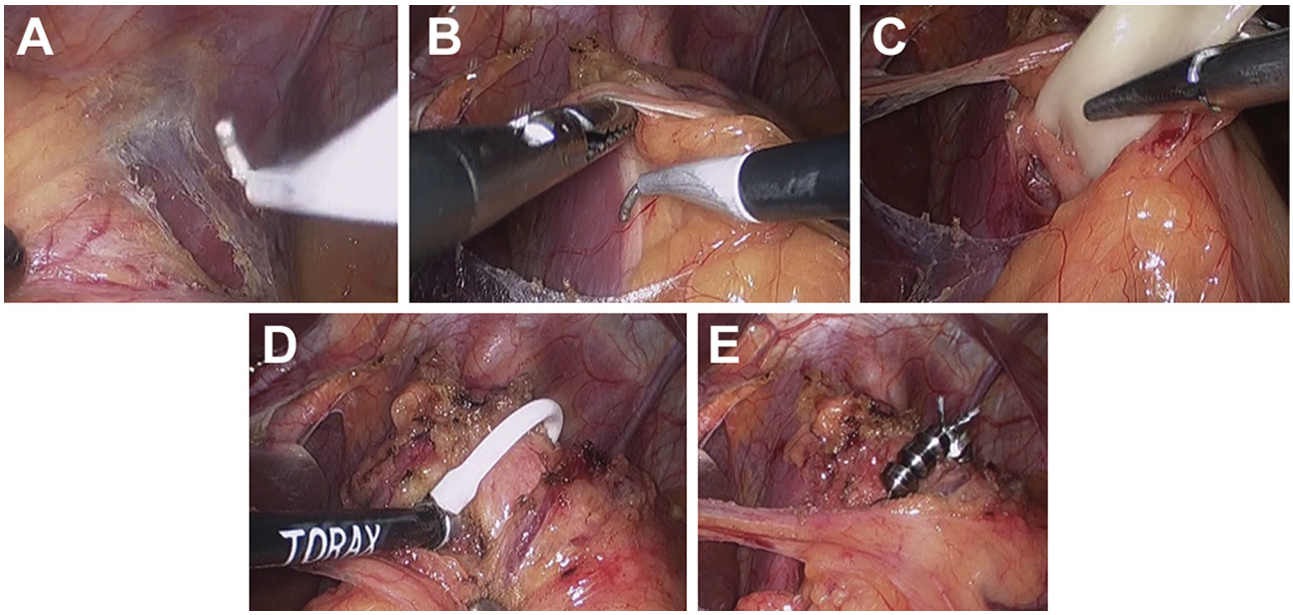


Fig 2. Implantation of magnetic sphincter augmentation (MSA) device: (A) Landing zone on left crus. (B) Site of MSA placement under the hepatic branch of the vagal nerve. (C) Creation of tunnel with isolation of posterior vagal nerve. (D) Determination of MSA size. (E) Implanted MSA device in anterior trench.

curve and one-third of the distance perpendicular to the lesser curve. The posterior fundus was brought through the retroesophageal window, and a "shoe-shine" maneuver was performed to ensure a 1:1 relationship between anterior and posterior fundus. A 58F or 60F bougie was advanced and the shoe-shine maneuver repeated. The fundoplication was created by placing a 2-0 polypropylene suture with a 1-cm pledget in a horizontal mattress formation through the anterior fundus—right lateral wall of esophagus—posterior fundus, followed by the second pledget. The fundoplication was lengthened by placing 2-0 silk sutures from anterior to posterior fundus above and below the pledget to create a 2-cm to 2.5-cm length wrap.

The magnetic sphincter was implanted laparoscopically using 5 ports in a similar configuration to fundoplication (Fig 2). The location of device placement was identified on the patient's right side, opening the gastrohepatic ligament above and below the hepatic branch of the anterior vagus nerve. Directly opposite to this on the left crus, a "landing zone" was created by incising the peritoneum between the lateral left crus and the posterior fundus. A tunnel was created behind the esophagus from the edge of the right crus under the vagal branch and directed toward the landing zone. A 0.25-inch Penrose drain was placed through the tunnel behind the esophagus. The posterior vagal nerve was identified and isolated by placing the Penrose inside of the nerve and around the gastroesophageal junction.

On the anterior surface, the peritoneum and fat were cleared with monopolar cautery, taking care to preserve the anterior vagal nerve and the phrenoesophageal ligament; thus, creating a "trench" for the device to sit in on the surface of the esophagus. If the posterior esophageal hiatus showed a "gap," it was closed with 1 or 2 "0" polyester sutures.

To determine the correct size of device, the outer circumference of the esophagus is measured using a sizing device provided by the manufacturer. The device is placed in the tunnel between the posterior vagus and the esophagus and wrapped around the esophagus. The device is tightened till it approximates the circumference of the esophagus without indenting the tissue. At this point, the device identifies the recommended size. The appropriately sized device was situated in the tunnel and around the esophagus, and the sutures were secured using a suture-securing device.

Postoperatively, patients who underwent LNF were kept nothing by mouth with the nasogastric tube to low intermittent suction. After an overnight stay, the nasogastric was removed, and a barium swallow was obtained. If satisfactory, the patient was initiated on clear to full liquids and discharged after tolerating oral intake. Patient-controlled analgesia was used for all LNFs.

Patients who received a magnetic sphincter were initiated on clear liquids immediately after the procedure. A barium swallow was obtained the next day and a regular diet started the morning after the procedure. For most patients, only oral analgesics were used. The most recent 5 patients were discharge home the same day.

The tests used for statistical comparisons were the *t* test for continuous variables and the Pearson χ^2 test for categoric variables. Symptom improvement was assessed using the McNemar paired change test. All *p* values were two-tailed, and no adjustments were made for multiple comparisons.

Table 1. Patient Baseline Demographics and Gastroesophageal Reflux Disease Characteristics

Characteristic	MSA (n = 34)	Nissen (n = 32)	*p* Value
Age, mean ± SD y	54 ± 11.8	47 ± 12.2	0.007
Gender, No.			0.32
Female	16	19	
Male	18	13	
BMI, kg/m^2 ± SD	27 ± 5.1	30 ± 4.4	0.03
GERD duration, y ± SD	12 ± 11.6	12 ± 8.5	0.88
Hernia size, cm ± SD	1.4 ± 1.1	1.5 ± 0.8	0.62
Hill grade, No.			
I	4	2	0.44
II	11	8	
III	14	18	
IV	5	5	
Esophagitis grade, No.			
No esophagitis	14	15	0.78
A	14	10	
B	4	6	
C	1	1	

BMI = body mass index; GERD = gastroesophageal reflux disease.

Results

Of the 34 patients who underwent MSA, 24 completed the 6-month follow-up. For comparison, 32 patients underwent LNF. The baseline demographic and GERD characteristics of both groups were similar, except MSA patients were older, and LNF patients had a higher body mass index (Table 1).

The operative time for MSA was 73 minutes compared with 118 minutes for LNF ($p = 0.001$). There were no operative deaths. In the MSA group, there were no major morbidities. Minor morbidities included symptomatic bradycardia in 1 patient and corneal abrasion in 1 patient. Two major morbidities occurred in the Nissen group. One patient was readmitted 5 days after discharge with dehydration and nausea, and 1 patient had symptoms of esophageal obstruction, which was confirmed on barium swallow. The patient returned to the operating room so a suture could be removed from the hiatal closure and was discharged without further sequelae. Two minor morbidities in the Nissen group included a postoperative seizure and a urinary tract infection.

At a mean follow-up of 6 months for MSA patients and 10 months for LNF patients, the symptoms of heartburn, regurgitation, cough, aspiration, chest pain, and ear, nose, and throat symptoms, such as throat clearing and hoarseness, were significantly improved compared with baseline (Table 2). The quality of life improved in both MSA (n = 23) and Nissen (n = 17) from baseline to 6 weeks postoperatively and to 6 months, and there was no difference between the groups for the QOLRAD (4.4, 6.0, and 6.6 vs 4.3, 5.8, and 6.6; $p = 0.77$, 0.57, and 0.91, respectively) or for the GERD-HRQL (20.6, 8.8, and 5.0 vs 22.8, 10.0, and 5.1; $p = 0.51$, 0.43, and 0.93, respectively). Swallowing ability worsened in both groups at 6 weeks (37.7 to 33.2 vs 37.1 to 26.3) but was significantly worse in the Nissen group ($p = 0.023$). Swallowing returned to baseline at 6 months in both groups (40.2 vs 36.9; $p = 0.24$; Fig 3). One component of the GERD-HRQL evaluates bloating and gassy feelings and showed a trend (1.32 vs 2.36; $p = 0.059$) in favor of MSA patients. Similarly, 16 of 24 MSA patients (67%) reported the

Table 2. Symptom Resolution by Magnetic Sphincter Augmentation and Nissen Procedure

	Magnetic Sphincter Augmentation				Laparoscopic Nissen Fundoplication			
		Symptom Present Postoperatively?				Symptom Present Postoperatively?		
Symptoms	Preoperative	No (No.)	Yes (No.)	*p* Value[a]	Preoperative	No (No.)	Yes (No.)	*p*-Value[a]
Heartburn	No	31	3	0.000	No	28	1	0.001
	Yes	31	2		Yes	31	4	
Regurgitation	No	34	6	0.000	No	30	3	0.001
	Yes	28	0		Yes	29	2	
Cough	No	31	27	0.000	No	31	22	0.001
	Yes	7	2		Yes	10	1	
Aspiration	No	32	28	0.000	No	31	29	0.001
	Yes	6	0		Yes	3	1	
Chest pain	No	32	25	0.000	No	32	26	0.001
	Yes	9	2		Yes	6	0	
ENT symptoms	No	32	15	0.000	No	31	24	0.001
	Yes	19	1		Yes	8	1	

[a] Related samples McNemar's change test.

ENT = ear, nose, and throat.

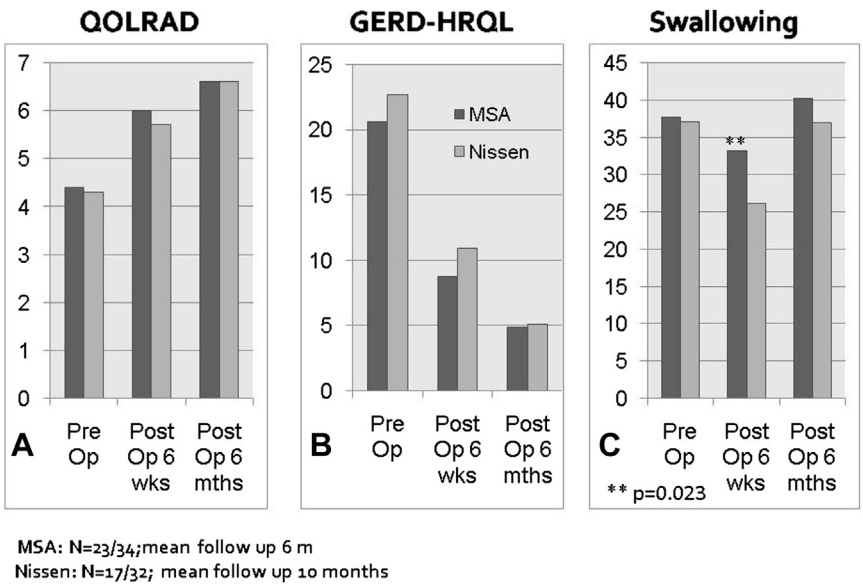


Fig 3. Quality of life scores are shown for (A) Quality of Life in Reflux and Dyspepsia questionnaire, (B) Gastroesophageal Reflux Disease-Health Related Quality of Life scale, and (C) swallowing for 23 of 24 MSA patients (mean follow-up, 6 months), and 17 of 32 Nissen patients (mean follow-up, 10 months).

ability to belch, whereas none of the Nissen patients could belch ($p = 0.0001$).

Postoperative pH testing was performed in 18 of 34 MSA patients and in 22 of 32 Nissen patients between 6 and 10 months postoperatively. Both procedures normalized the DeMeester score, with the MSA group dropping from 49.5 to 14.2 and the Nissen group from 49.0 to 5.1. Similarly, the percentage of time pH was less than 4 was normalized, with the MSA group going from 14.8% to 4.6% and the Nissen group going from 13.5% to 1.1%. Despite normalization, there was still a significant difference between the postoperative DeMeester scores ($p = 0.0001$) and the percentage of time the pH was less than 4 ($p = 0.0001$) when MSA and LNF were compared.

In absolute terms, 10 of 18 MSA patients (56%) had a DeMeester score below the 14.7 threshold and 10 of 18 had a percentage of time the pH was below the 4.9 threshold. Comparatively, only 1 Nissen patient had a DeMeester score above 14.7. Furthermore, when two components of the DeMeester score were reviewed—total number of refluxes and the number of postprandial refluxes—the total number of refluxes were below the normal of 104, with MSA having 60.1 refluxes compared with 21.5 for Nissen ($p = 0.002$). The number of postprandial refluxes was 35.1 for MSA and 8.4 for Nissen ($p = 0.001$).

At their respective mean follow-up assessments, all of the MSA patients (0 of 24) remained off PPI therapy whereas 1 of the 32 Nissen patients was on a PPI despite having normal postoperative testing. In the MSA group, 1 patient had an episode of a food bolus impaction requiring evaluation, but no invasive treatment, and 1 patient underwent endoscopic balloon dilation for dysphagia early in our experience. Comparatively, gas bloat occurred in 2 Nissen patients, symptomatic esophageal spasms occurred in 2 requiring medical therapy, and 1 patient had new-onset diarrhea related to fundoplication. Endoscopically, esophagitis occurred in 4 MSA patients (LA class A in 3; LA class B in 1), with each having an elevated DeMeester score. Comparatively, 1 Nissen patient had LA class A esophagitis with a normal DeMeester score. There were no identified erosions, device migrations, or removals in the MSA group. A recurrent hiatal hernia developed in 1 Nissen patient at 1 year, but the Nissen was intact.

Comment

The main finding in this study is that patients with GERD, with or without a hiatal hernia smaller than 3 cm, undergoing MSA with the LINX device (Torax Medical Inc, Shoreview, MN) have equivalent outcomes compared with patients with similar characteristics undergoing LNF. MSA alleviates typical and atypical symptoms of GERD, improves quality of life, and normalizes distal esophageal acid exposure. Our MSA results are similar compared with previous published studies and add to the growing experience with this device [7–9].

Although MSA results in normalization of distal esophageal acid exposure overall, our mean DeMeester approaches the normal of 14.7, and only 56% have a normalized score. These findings are similar to the results of Ganz and colleagues [8], who reported postoperative DeMeester scores at 1 year of 13.5 and an absolute normalization in 58%. Comparatively, Bonavina and colleagues [9] reported a median composite DeMeester score of 11.2 and absolute normalization of 80% but had a longer median follow-up of 4.2 years.

There are several possibilities for these findings. First, it is possible that with time, there is "maturation" of the device with scarring around the gastroesophageal junction leading to continued improvement in GERD control. Second, endoscopy and pH testing are done at a point in time and the test results may reflect only what has occurred during a short period of time before testing. Lastly, and most likely, it may depend on the grade of esophagitis before MSA because the relative

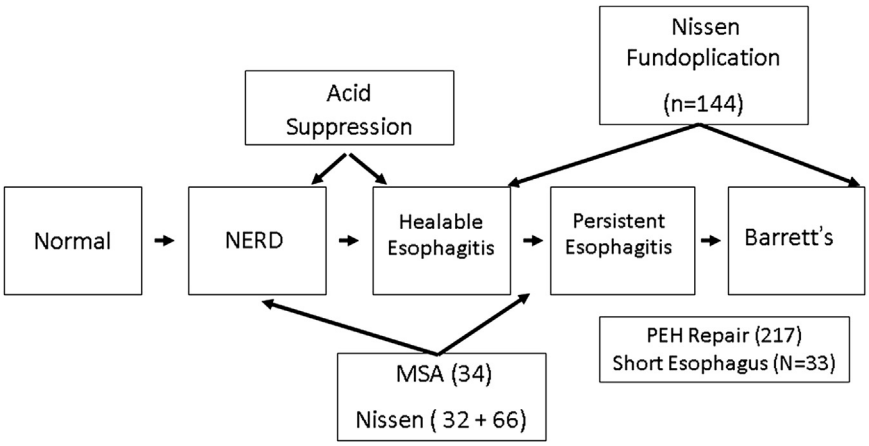


Fig 4. Individualized treatment for the spectrum of gastroesophageal reflux disease. (MSA = magnetic sphincter augmentation; NERD = nonerosive reflux disease; PEH = paraesophageal hernia). (Adapted from Lord and colleagues [1].)

proportions of no esophagitis vs LA class A vs LA class B were 60:20:20 in the Ganz trial, 83:10:6 in the Bonavina trial, and 41:41:12 in our study. This may reflect subtle differences in disease severity and lower esophageal sphincter dysfunction that may not be augmented as well with the device.

The outcome data presented here suggest that there are differences between MSA and Nissen that address patient and referring physician concerns about Nissen fundoplication. First, the ability to belch is substantially improved by MSA. Second, the trend toward less bloating and gassy feelings also favors MSA. Third, the side effects experienced by the Nissen patients, such as gas bloat, spasm, and diarrhea/dumping, which can last more than 5 years, did not occur after MSA [3]. We hypothesize that these differences may be explained by restoration of a more normal sphincter when MSA is used.

Asymptomatic normal patients experienced a mean number of total refluxes of 43.8 in 48 hours [10]. In MSA patients, 60 total refluxes occurred, whereas Nissen patients experienced only 21 refluxes. Furthermore, in the postprandial period, air that has been entrained during ingestion of food is often vented. But because Nissen allows for only 8 reflux episodes compared with 35 with MSA, the ability to vent is less effective with Nissen, giving way to the side effects of bloating and flatulence. Thus, MSA likely results in a more normal sphincter, whereas Nissen may be considered "super-normal" because little to no reflux is not physiologic.

Given the findings in this study, MSA may allow GERD treatment to be further individualized because it offers patients an option if PPIs are not effective (Fig 4). Of the 427 patients we screened, 217 were for type II to IV paraesophageal hernias, including 33 with a short esophagus requiring Collis gastroplasty. Another 144 Nissen procedures were performed for refractory esophagitis and hiatal hernias sized between 3 and 5 cm. In the spectrum of disease where MSA is indicated, there were only 98 patients representing the group of patients who previously would have remained on acid suppression with incomplete control but might never be referred for or considered surgical intervention.

This study has several limitations. First, it represents a small series with very short follow-up. As such, these results may not be indicative of future outcomes, although the longer-term data have shown durability out to 6 years [9].

Second, the short follow-up precludes definitive comments about the issues of erosion, migration, and removal. However, an analysis of the first 1,048 MSA implants showed no migrations and removal of 36 devices [11]. One erosion was reported, but recent reports have identified 4 erosions in nearly 1,600 implants, which is significantly less than Angelchik and lap bands, which were considerably larger and exerted pressure on the esophagus (Torax Medical Data).

Lastly, the study is retrospective and thus subject to biases, but the patients in the comparison group were carefully evaluated and all would have qualified for MSA, thus making the conclusions perhaps more meaningful. A randomized controlled trial would be ideal to compare MSA with existing therapies; however, because there are objective data points, such as pH, to allow comparison, the need for such a trial may not be as great.

In conclusion, MSA in patients with chronic GERD and a hiatal hernia of less than 3 cm in size results in similar objective control of GERD, symptom resolution, and improved quality of life compared with Nissen fundoplication. MSA seems to restore a more physiologic sphincter that allows physiologic reflux in patients with earlier reflux disease that facilitates belching and creates less bloating and flatulence by allowing total reflux events to move toward the mean and maintaining postprandial reflux events. This device has the potential to allow individualized treatment of patients with GERD and increase the surgical treatment of GERD.

References

1. Lord R, DeMeester S, Peters J, et al. Hiatal hernia, lower esophageal sphincter incompetence, and effectiveness of Nissen fundoplication in the spectrum of gastroesophageal reflux disease. J Gastrointest Surg 2009;13:602–10.
2. Lundell L, Miettinen P, Myrvold HE, et al. Comparison of outcomes twelve years after antireflux surgery or

omeprazole maintenance therapy for reflux esophagitis. Clin Gastroenterol Hepatol 2009;7:1292–8; quiz: 1260.
3. Galmiche JP, Hatlebakk J, Attwood S, et al. Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD: the LOTUS randomized clinical trial. JAMA 2011;305:1969–77.
4. Finks J, Wei Y, Birkmeyer J. The rise and fall of antireflux surgery in the United States. Surg Endosc 2006;20:1698–701.
5. Kahrilas P, Jonsson A, Denison H, et al. Regurgitation is less responsive to acid suppression than heartburn in patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol 2012;10:612–9.
6. Humphries L, Hernandez J, Clark W, et al. Causes of dissatisfaction after laparoscopic fundoplication: the impact of new symptoms, recurrent symptoms, and the patient experience. Surg Endosc 2013;27:1537–45.
7. Ganz RA, Gostout CJ, Grudem J, Swanson W, Berg T, DeMeester TR. Use of a magnetic sphincter for the treatment of GERD: a feasibility study. Gastrointest Endosc 2008;67:287–94.
8. Ganz RA, Peters JH, Horgan S, et al. Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 2013;368:719–27.
9. Bonavina L, Saino G, Bona D, Sironi A, Lazzari V. One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 2013;17:577–85.
10. Ayazi S, Lipham J, Portale G, et al. Bravo catheter-free pH monitoring: normal values, concordance, optimal diagnostic thresholds, and accuracy. Clin Gastroenterol Hepatol 2009;7:60–7.
11. Lipham J, Taiganides P, Louie BE, et al. Safety analysis of first 1,000 patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease. Dis Esophagus 2014 Mar 11. http://dx.doi.org/10.1111/dote.12199 [Epub ahead of print].

DISCUSSION

DR STEVEN DeMEESTER (Los Angeles, CA): Great presentation. A lot of times with these new devices and so forth, we learn more from the failures than we do from the successes. We expect it to work, but the failures are where we can really dig in and understand the mechanism and where it might be best applicable. So in light of that, I didn't really see that you presented the absolute percent of normalization. You showed us mean or median data for pH.

Can you tell us what percentage of patients that had the LINX (Torax Medical Inc, Shoreview, MN) and the Nissen were normalized, and then tell us about the patients that weren't normalized with the LINX, particularly the 3 with esophagitis. What can we learn from them? Did they have esophagitis before the surgery and it persisted, or is this new esophagitis? Can you also describe how you managed these patients? Thanks.

DR LOUIE: Certainly. So the percent normalization for the Nissen group, 100% of those patients had normalization of their DeMeester scores all the way down below, which I think accounts for such a low number.

There, the percent normalization for the LINX device is about 60%, and those values are just above the 14.7 threshold. So we looked at those patients specifically. And you look at them, they are like 16, 18, 19, and almost all of those are postprandial reflux events. When we looked at the 3 with esophagitis, none of those patients had any symptoms. They had some esophagitis preoperatively. It was better postoperatively, but we never got rid of it. But they were 1-mm breaks, and we were very strict about our assessment of esophagitis. Whether those episodes or those findings endoscopically, at that one point in time, will persist, or they come and go is unclear. None were treated with proton pump inhibitors (PPIs).

You know, one of the anecdotal stories is on one of those patients, right before he was tested, he was an Air Force pilot. He was a very smart guy. He was interested in figuring out what the burst pressure for his LINX device was. So before surgery he could normally eat about 2 bowls of chili, and then he would have horrendous reflux disease. So he figured he would do the same, and he did it, unfortunately, right before we did his reflux testing. And he figured that at $2^1/_2$ bowls of chili, he was good. When he got to the top of the third bowl, he had horrendous reflux disease. So he's, like: Doc, my burst pressure is this. This is when I get reflux disease. So I think, unfortunately, the patients select themselves for LINX in this early period, and some of them are very interested in sort of understanding the physiology.

But I think it is correct. I think that you have a little bit of postprandial reflux, and if those patients continue their habits, which are big meals once or twice a day, they might get some reflux or esophagitis because of it.

DR THOMAS WATSON (Rochester, NY): Brian, I really enjoyed your talk. I agree with you; there is a place for LINX in the marketplace. The device seems to be as efficacious as a Nissen in controlling gastroesophageal reflux disease (GERD), and perhaps with a slightly better side effect profile. I think also that the reversibility aspect of it has appeal to patients.

We have been putting in the LINX device in Rochester, and I would say we have had three major barriers to utilizing it more frequently. I would be curious how you have dealt with these issues:

One is insurance reimbursement. We are having a hard time getting our insurers to pay for it. And while I hope that will get easier in time, I am curious if you have any insights into how you have been dealing with that problem in Seattle.

Number two, we still do not have good long-term data about potential complications, such as erosions, from these devices left in for decades or more. What are you telling patients about such possibilities?

And thirdly, the fact that having an internal magnet precludes the patient from ever having an magnetic resonance imaging (MRI) scan is a big turnoff, particularly for younger patients looking at long life ahead of them. Will this problem ever be overcome?

Thanks for your insights and a great presentation.

DR LOUIE: Sure. Let us talk about insurance. Insurance continues to be a battle. I think the company has been very good about that. Since Swedish is self-insured, we went to the medical director, and for our internal employees, we have made a deal with them and we are covered. We have used the two companies that the Torax Medical is engaged in. We have been reasonably successful, but still, we have some insurers that absolutely will not cover it, deeming it experimental. We have gone to peer-to-peer review, we have done the whole scheme of things. I think it is better. The company has statistics showing that they are getting closer to 30%, 40%, 50% approval over

the last year, with the process that they are using, and it still remains early.

The MRI issue, I think, is a concern for younger patients, although they did receive conditional Food and Drug Administration approval about 3 weeks ago for a 0.7 Tesla MRI, so the LINX can go in a 0.7 Tesla MRI. I have an e-mail out to the radiologist to find out what a 0.7 Tesla MRI can do, but as I understand it, that includes most bone and joint procedures, but will not allow you to have a sort of a spinal MRI.

And then the third point, long-term foreign body, I think that remains a concern when you implant anything around the esophagus. We have seen that with permanent mesh around the hiatus. We do know, and it is available online, that there have been 4 erosions in the world. These erosions, in total, about 1,550 cases worldwide, 3 of them in Europe, 1 of them in the United States. And I think, at least my interpretation of that data is that if you are having difficulty getting in that tunnel underneath, that is where most of the erosions are occurring. I think probably, at the time of surgery, maybe that is surgeon judgment. We are not putting them in properly or we are having some difficulty back there doing stuff, when you look at that data. So I think that is still a far cry from other implantable devices, but I think it is a concern.

We have told all our patients that, look, this is a concern. We have told them what the data are. We have also engaged them that if you are going to have LINX at this point in time, that your follow-up has to be complete, so they are getting an endoscopy once a year, or they are being surveyed so we have an idea what is going on, because it is such a new device. And most patients who are interested in the new device will continue to come back, because they are interested in the device, as well.

DR MARK B. ORRINGER (Ann Arbor, MI): As a "gray beard," I would like to add historical perspective to this discussion. Years ago, a thoracic surgeon named Mr Ronald Belsey developed the Belsey Mark IV hiatal hernia repair. And the Belsey Mark IV operation was named that because there had been a Mark I, a Mark II, and a Mark III before the Mark IV.

It took Dr David Skinner going over to Bristol, England, to convince Mr Belsey to allow him to review and report his series with 10-year follow-up, because Belsey felt that a minimum of 5 to 10 years of follow-up was necessary before one could authoritatively talk about the efficacy of an antireflux operation, which in his case he modified four times in search of better results. This principle of long-term follow-up has guided thoracic surgeons whose practice focuses upon esophageal disease.

So I would submit that a new antireflux operation with only a few months of follow-up has little proven efficacy justifying clinical adoption. This paper is more or less a "proof of concept" discussion. The LINX device appears to control reflux in the short-term, but how it will fare in long-term follow-up is unknown.

Historically, we can draw upon the experience with the Angelchik prosthesis of a few decades ago. With this device, there was not any question of "tunneling too deep" or "getting too near" the esophageal wall as discussed with the LINX device today. One took this beautiful, soft, spongy silicone ring and secured it around the esophagus. What could be safer than that? Except that in long-term follow-up, the eventual migrations up through the hiatus and down onto the stomach, and erosions through the esophageal wall, at times with passage of the ring per rectum, were disasters which led to the ring being pulled off the market. So the concept of placing a semirigid ring around a part of the body that is constantly exposed to the motion of the diaphragm and moving up and down is not exactly new. The face may be a little different, but it remains to be seen if history will be repeated.

I would like to ask one question: We always show on the slides of the indications for antireflux surgery "failure of PPI therapy." The decision to undertake an antireflux operation should not be made lightly, as multiple failed repairs may ultimately result in an esophagectomy and the physical adjustments that have to be made to it afterward.

And the quality of life after an esophagectomy is generally not as good as it is with putting up with a little reflux and modifying lifestyle: getting on a weight reduction program, limiting carbohydrate intake, and walking 3 miles every day.

So I would submit that "failure of PPI therapy" alone is not a sufficient indication for antireflux surgery as it does not constitute "failure of medical management," but rather just one aspect of medical management. I would like to know how you counsel patients being considered for a LINX procedure. Do you make them lose weight? Or, do they come to you saying, "I heard about this new antireflux device you are putting in, and I would like one of these?"

DR LOUIE: Well, you can see that we chose people who were under a body mass index of 35, and the LINX people were actually much lower statistically. But yes, these patients, we counseled them. I counseled them extensively about diet, exercise, and weight, because we know that that is the biggest significant contributor to recurrent GERD after repair.

And I think your comments about the longer-term effects are key. I think this is clearly early, which we have labeled it early because we do not have those long-term data. And I think it is important that we are going to follow these people out so that we avoid potentially the issues with Angelchik.

DR BLACKMON: Thank you.

Laparoscopic Magnetic Sphincter Augmentation vs Laparoscopic Nissen Fundoplication: A Matched-Pair Analysis of 100 Patients

Jessica L Reynolds, MD, Joerg Zehetner, MD, FACS, Phil Wu, BS, Shawn Shah, BS, Nikolai Bildzukewicz, MD, FACS, John C Lipham, MD, FACS

BACKGROUND: The efficacy and safety of magnetic sphincter augmentation (MSA) with the LINX device (Torax Medical) has been reported in several short-and long-term studies, rivaling historic results of laparoscopic Nissen fundoplication (LNF), but with fewer side effects. However, there have been no studies comparing patients with similar disease to validate these results.

STUDY DESIGN: We conducted a retrospective analysis of 1-year outcomes of patients undergoing MSA and LNF from June 2010 to June 2013. Patients were matched using propensity scores incorporating multiple preoperative variables. Outcomes were measured by GERD Health Related Quality of Life scores, proton-pump inhibitor use, satisfaction, and complications.

RESULTS: One hundred and seventy-nine patients met inclusion criteria, 62 MSA and 117 LNF. Propensity score matching identified 50 patients in both groups using the "best-fit" model with a caliper of 0.5 SD. At 1 year after surgery, both groups had similar GERD Health Related Quality of Life scores (4.2 MSA and 4.3 LNF; $p = 0.897$) and proton-pump inhibitor use (17% of MSA and 8.5% of LNF; $p = 0.355$). Although there was no difference in the number of patients reporting mild gas and bloating (27.6% MSA and 27.6% LNF; $p = 1.000$), there were no patients with severe gas and bloating in the MSA group compared with 10.6% in the LNF group ($p = 0.022$). More LNF patients were unable to belch (8.5% of MSA and 25.5% of LNF; $p = 0.028$) or vomit (4.3% of MSA and 21.3% of LNF; $p = 0.004$). The incidence of postoperative dysphagia was similar between the groups (46.8% MSA and 44.7% LNF; $p = 0.766$).

CONCLUSIONS: Analogous GERD patients had similar control of reflux symptoms after both MSA and LNF. The inabilities to belch and vomit were significantly fewer with MSA, along with a significantly lower incidence of severe gas–bloat symptoms. These results support the use of MSA as first-line therapy in patients with mild to moderate GERD. (J Am Coll Surg 2015;221:123–128. © 2015 by the American College of Surgeons)

Laparoscopic Nissen fundoplication (LNF) is the gold-standard surgical treatment for GERD. Despite this, it is estimated that <1% of patients use or are offered this option.[1] Ninety-nine percent of patients with GERD are treated with acid-suppression therapy, the majority of which are proton-pump inhibitors (PPIs). These agents reduce the acidic symptoms of reflux by increasing the pH of the refluxed gastric juice without reducing the incidence of reflux episodes. Despite their ubiquitous use, it is estimated that approximately 60% of patients on PPIs continue to have symptoms or are unable to tolerate the medication.[2,3] Consequently, there is a considerable portion of GERD patients who remain poorly controlled on PPI therapy, and one-fourth have endoscopic evidence of progressive disease, such as esophagitis

Disclosure Information: Dr Lipham is a consultant for Torax Medical, the manufacturer of the LINX device. All other authors have nothing to disclose.

Support: Supported by the James & Pamela Muzzy Endowed Chair in GI Cancer.

Presented at the Western Surgical Association 122nd Scientific Session, Indian Wells, CA, November 2014.

Received December 1, 2014; Revised February 1, 2015; Accepted February 1, 2015.

From the Department of General Surgery, Keck Medical Center of the University of Southern California, Los Angeles, CA.

Correspondence address: John C Lipham, MD, FACS, Department of General Surgery, Keck Medical Center of the University of Southern California, 1510 San Pablo St, Suite 514, Los Angeles, CA 90033. email: john.lipham@med.usc.edu

http://dx.doi.org/10.1016/j.jamcollsurg.2015.02.025
ISSN 1072-7515/15

Abbreviations and Acronyms

HRQL = Health Related Quality of Life survey
LES = lower esophageal sphincter
LNF = laparoscopic Nissen fundoplication
MSA = magnetic sphincter augmentation
PPI = proton pump inhibitor

or Barrett's esophagus.[4] Magnetic sphincter augmentation (MSA) was specifically designed for such patients.

Magnetic sphincter augmentation with the LINX device (Torax Medical) was approved by the FDA in 2012 for patients with mild to moderate GERD. Labeling of the device excludes patients with moderate dysphagia, severe esophagitis, and large hiatal hernias, exactly the population that often present for LNF. There have been several studies looking at the efficacy of MSA, and all find both short- and long-term efficacy to be similar to that reported for LNF.[5-9] However, these studies are likely reporting on a population with different disease severity than that undergoing LNF.

The aim of this study was to compare the outcomes of MSA with LNF in a group of patients matched by propensity scores calculated from disease-specific preoperative symptoms and endoscopic findings to evaluate the 2 procedures in patients with similar disease severity.

METHODS

All patients undergoing MSA or LNF between April 2007 and October 2013 were identified from a prospectively collected database. This range was chosen to include patients undergoing MSA and LNF during the same time period as the first MSA was performed at our institution in 2007 as part of the FDA Feasibility trial. Inclusion criteria included objective evidence of GERD, defined as an abnormal pH study, presence of biopsy-proven Barrett's esophagus, or esophagitis grade B or greater; documented history of PPI therapy for a minimum of 6 months; and normal esophageal motility documented by videoesophagram or esophageal manometry.

Surgical procedure

Magnetic sphincter augmentation was performed by 3 surgeons (JL, NB, and JZ) at 2 institutions according to a standard protocol (Table 1). The decision to repair unsuspected hiatal hernias was made intraoperatively by the operating surgeon. In general, crural closure was performed if a hiatal hernia was visible after a posterior dissection of the hiatus that kept the phrenoesophageal membrane intact anteriorly and laterally. Laparoscopic Nissen fundoplication was performed by the same 3 surgeons and at the same 2 institutions according to the standard protocol (Table 2).

Table 1. Magnetic Sphincter Augmentation Protocol

1.	Patient is placed in low lithotomy with the surgeon standing between the patient's legs.
2.	A 12-mm camera port is placed at the umbilicus. A 5-mm working port is placed in the right upper quadrant and a Nathanson liver retractor is placed in the right upper quadrant. An 8-mm working port in placed in the left upper quadrant to allow passage of the MSA device.
3.	The hepatic branch of the vagus nerve is identified and preserved. The right and left crura are identified and minimally dissected to create a tunnel behind the esophagus.
4.	If the posterior vagus nerve can be easily identified, it is dissected posterior. However, this step is often omitted in favor of keeping dissection to a minimum.
5.	Tissue on the anterior esophagus is removed so the MSA device can lie flush to the esophagus.
6.	The provided sizing device is used to determine the number of beads on the device.
7.	The MSA device is passed through the 8-mm port and pulled through the retro-esophageal tunnel. If the vagus nerve was dissected, it is placed anterior to the vagus.
8.	The MSA device is secured using the clasp on the device.
9.	Ports are removed and the abdomen desufflated.

MSA, magnetic sphincter augmentation.

At 1 year after surgery, patients were evaluated with a GERD-Health Related Quality of Life (HRQL) survey with questions about gas—bloat, their ability to belch, their

Table 2. Laparoscopic Nissen Fundoplication Protocol

1.	Patient is placed in low lithotomy with the surgeon standing between the patient's legs.
2.	A 12-mm camera port is placed above and to the left of the umbilicus. Two 12-mm working ports are placed at the bilateral subcostal margins in the mid-clavicular line. A Nathanson liver retractor is placed in the right upper quadrant. A 12-mm port in the left anterior axillary line at the level of the camera port is placed as an assistant port.
3.	The plane between the right crus and esophagus is developed and extended to the left crus until complete circumferential dissection of the esophagus is obtained with a large retro-esophageal window.
4.	Both vagus nerves are identified and preserved.
5.	The esophagus is encircled with a Penrose and the crura are closed.
6.	The short gastric vessels are divided to mobilize the fundus.
7.	A 52 to 56F esophageal bougie is passed and the fundus passed behind the esophagus creating a loose, floppy wrap.
8.	The fundus is sutured to itself to size the wrap and then the bougie is removed.
9.	Additional sutures are placed in the fundus, including the anterior esophageal wall to secure the wrap around the esophagus.
10.	Ports are removed and the abdomen desufflated.

ability to vomit, dysphagia, and overall satisfaction with surgical treatment. The GERD-HRQL is a validated disease-specific questionnaire consisting of 10 questions focused on heartburn, dysphagia, and gas–bloat, with each question rated 0 (least severe) to 5 (most severe), for a total score ranging from 0 to 50.[10] Dysphagia was recorded as mild if patients reported occasional food sticking less than once a week, moderate if they experienced symptoms more than once a week without regurgitation of undigested food or vomiting or made dietary modifications to accommodate symptoms, and severe if they experienced symptoms more than once a week that included regurgitation of undigested food or vomiting. Patients were also specifically asked if their preoperative symptoms improved, resolved, or did not change. The medical record was also searched for any complications or interventions that occurred within the first postoperative year.

Statistical analysis

Matching was performed by calculating propensity scores for MSA and LNF patients using the following predictors: age, sex, BMI, duration of GERD symptoms, esophagitis grade (Los Angeles classification) as described on endoscopy report, size of a hiatal hernia as measured on endoscopy, Hill valve grade, laryngeal–pharyngeal reflux symptoms, and dysphagia. Per inclusion criteria, all patients had objective evidence of pathologic reflux defined as an abnormal pH study, biopsy-proven Barrett's esophagus, or esophagitis grade B or greater. Also, all patients had normal motility assessed by videoesophagram or manometry and were on PPIs for at least 6 months before the procedure. Patients were matched using the "best-fit" model with a caliper of 0.5 SD and a limit of 100 cases.

Means were compared using independent *t*-test and categorical variables were compared using either Fisher's exact test for variables with 2 categories or Pearson's chi-square for variable with >2 categories. A p value <0.05 was considered significant.

All statistical analysis was performed with SPSS software, version 22 (IBM SPSS).

RESULTS

There were 179 patients that met the inclusion criteria for the study, 62 had MSA and 117 had a LNF. Propensity score matching with a 0.5-SD caliper identified 51 matches and the 50 best matches were included for analysis. Comparison of the preoperative characteristics between the patients who had MSA with those who had LNF showed no significant difference in any of the included variables (Tables 3 and 4). The presence of a hiatal hernia and the size of the hiatal hernia were distributed similarly in the MSA and LNF patients (Table 4). All LNF patients underwent crural closure as part of the standard LNF procedure (Table 2). In MSA patients, crural closure was performed if a hiatal hernia was visible after a posterior dissection of the hiatus that kept the phrenoesophageal membrane intact anteriorly and laterally. Twenty-two percent (11 of 50) of MSA patients underwent crural closure, including 33.3% (6 of 18) of patients with a 2-cm hernia, 44.4% (4 of 9) of patients with a 3-cm hernia, and 100% (1 of 1) of patients with a 4-cm hernia.

Table 3. Preoperative Demographics and GERD Charecteristics

Characteristics	MSA (n = 50)	LNF (n = 50)	p Value
Age, y	53.0	54.0	0.748
Sex, n			0.686
Male	30	27	
Female	20	23	
BMI, kg/m^2	26.4	26.7	0.741
GERD duration, mo	146.9	144.5	0.932
GERD-HRQL score	19.7	18.8	0.596
Esophagitis, n			0.711
None	35	36	
A	9	7	
B	5	4	
C	1	3	
D	0	0	
Hiatal hernia size, cm	1.5	1.6	0.735
Hill grade valve, n			0.482
1	1	0	
2	5	5	
3	17	12	
4	27	33	
LPR, n	20	20	1.000
Dysphagia, n	5	7	0.760
Barrett's esophagus, n	10	11	1.000

HRQL, Health Related Quality of Life survey; LNF, laparoscopic Nissen fundoplication; LPR, laryngeal–pharyngeal reflux; MSA, magnetic sphincter augmentation.

One-year follow-up data were available for 47 of 50 (94%) patients who had MSA and 47 of 50 (94%) LNF patients. Among MSA patients, 46 of 47 (97.8%) stated that their GERD symptoms had improved, and 24 of 47 (51.1%) reported complete resolution. Similarly, 46 of 47 (97.8%) LNF patients reported improvement of their GERD symptoms at 1 year, and 23 of 47 (48.9%) reporting complete resolution of symptoms (p = 0.978).

Mean GERD-HRQL scores at 1 year after surgery were similar, with 4.2 for the MSA patients and 4.3 for the LNF patients (p = 0.879). At 1 year after surgery, 39 of 47 (83.0%) MSA patients were off PPI therapy, compared with 43 of 47 (91.5%) LNF patients (p = 0.355).

Table 4. Hiatal Hernia Characteristics

Characteristics	LNF group (n = 50)	MSA group (n = 50)
Hiatal hernia size, n		
None	15	15
1 cm	4	7
2 cm	20	18
3 cm	9	9
4 cm	2	1
Any, n (%)	35 (70)	35 (70)

LNF, laparoscopic Nissen fundoplication; MSA, magnetic sphincter augmentation.

One year after surgery, 17 of 47 (36.2%) patients with MSA had mild dysphagia, characterized as food sticking less than once a week and 5 of 47 (10.6%) had moderate to severe dysphagia, characterized as symptoms more than once a week, regurgitation of undigested food, vomiting, or requiring dietary modifications. This was similar to patients who had LNF, with 15 of 47 (31.9%) reporting mild dysphagia and 6 of 47 (12.8%) reporting moderate to severe dysphagia ($p = 0.766$). Endoscopic dilation was performed for symptom of dysphagia in 8 of 50 (16%) patients who had MSA compared with 5 of 50 (10%) who had LNF ($p = 0.554$).

Of the patients who had MSA, 13 of 47 (27.7%) reported symptoms of gas–bloat, all characterized it as "mild." Eighteen of 47 (38.3%) patients reported gas–bloat after LNF and 5 (10.6%) characterized it as "severe" ($p = 0.067$). No MSA patients reported severe gas and bloating symptoms ($p = 0.022$). After MSA, 4 of 47 (8.5%) patients were unable to belch and 2 of 47 (4.3%) were unable to vomit when necessary. After LNF, 12 of 47 (25.5%) patients were unable to belch ($p = 0.028$) and 10 of 47 (21.3%) were unable to vomit when necessary ($p = 0.004$).

Overall, 42 of 47 (89.4%) MSA patients and 43 of 47 (92.5%) LNF patients were satisfied with the procedure ($p = 0.603$), and 44 of 47 (93.6%) MSA and 38 of 47 (80.9%) LNF patients would elect to have the procedure again ($p = 0.053$).

Complications within 30 days of surgery did not occur in patients who had MSA. In contrast, 2 patients who had an LNF had complications; 1 was readmitted for intractable nausea and oral intake intolerance on postoperative day 1, and another patient required an esophagogastroduodenoscopy for food impaction. One year after surgery, there were no MSA device removals or erosions and no LNF patients required reoperation. Endoscopic dilation for dysphagia during the first postoperative year occurred in 8 of 50 (16.0%) MSA patients and in 5 of 50 (10.0%) LNF patients ($p = 0.554$).

DISCUSSION

The era of minimally invasive surgical treatment of GERD began in the early 1990s with LNF. During the latter part of the intervening 25 years, several endoscopic procedures were developed to provide a less invasive surgical treatment of GERD. These included an attempt to form a fundoplication around the lower esophageal sphincter (LES) with a variety of endoscopic suturing devices, bulk and stiffen the LES with foreign materials, or decrease the compliance of the LES by producing escharotic lesions with radiofrequency ablation. However, none of these have been able to provide long-term control of acid reflux comparable with LNF with acceptable side effects.[11-13] The concept of sphincter augmentation was developed to prevent transient sphincter relaxation due to effacement and shortening of the sphincter's length secondary to postprandial gastric distention or gastric dilation caused by adaptive relaxation. In this way, MSA is fundamentally different from other anti-reflux procedures. It was designed specifically to augment a near-normal LES, the length of which is starting to shorten from reflux-induced inflammatory injury and to provide additional support during transient failures of the LES, such as during postprandial gastric distention or dilation. Magnetic sphincter augmentation provides a surgical alternative to patients with mild to moderate disease who have evidence of progressive disease, such as esophagitis on baseline endoscopy, failure of esophagitis to heal with acid-suppression therapy, the need to escalate the dose of acid-suppression therapy to achieve symptomatic relief, and a compulsive dependency on daily PPIs to control symptoms. The impetus to identify and counsel patients with progressive disease about the need for surgical therapy is critical if progression to inflammatory and metaplastic complications of the disease are to be prevented.

In this matched-pair analysis between MSA and LNF in patients with early disease, we found that those who had MSA obtained the same efficacy in symptomatic reflux control with less gas–bloat symptoms and less restriction in their ability to belch and vomit. Louie and colleagues[14] showed similar short-term results in a smaller group of patients controlled for hernia size and GERD symptoms.

A criticism of comparative studies to date is that the patients undergoing MSA have less severe disease than patients undergoing LNF. The current study is the first to test this criticism by matching patients for disease severity. By using propensity scores and matching on 9 variables known to affect disease severity, we were able to compare similar patients. Analysis of these characteristics confirms

that the majority of patients in both groups, those who had MSA and those who had LNF, had mild to early disease (ie, no Barrett's esophagus, no dysphagia, mild or no esophagitis, and small hiatal hernias). Analysis of 1-year outcomes confirmed that MSA is comparable with LNF in terms of efficacy, safety, and patient satisfaction. There was a significant difference favoring MSA in both severe gas–bloat symptoms and ability to belch and vomit, with twice as many LNF patients as MSA patients not being able to belch normally and 5 times as many being unable to vomit.

These results substantiate that MSA is as effective as LNF in controlling reflux, with the benefit of also having fewer side effects. Consequently, there is the potential that more patients with early evidence of progressive disease will be amenable to surgical therapy and potentially more gastroenterologists will be willing to refer patients with progressive disease earlier for MSA. This is particularly important because medical therapy with PPIs does not prevent reflux into the esophagus of the alkalized gastric juice, allowing the potential for progressive disease, despite silencing of the patient's symptoms. Treating such patients with MSA can prevent progressive disease and decrease the incidence of severe reflux complications and development of chronic inflammation, metaplasia, dysplasia, and esophageal adenocarcinoma.

The main shortcoming of this study is that it is a retrospective study and not a randomized controlled trial. Therefore, there is an inherent selection bias. We used propensity score analysis to limit the effect of this selection bias, but it does still exist. Early in the study, MSA was done as part of the FDA pre-approval trials and, therefore, was only available to a small subset of patients who met strict inclusion criteria and were willing to undergo the yearly invasive testing for 5 years, which was part of the study. Once MSA was approved by the FDA, the lack of insurance coverage for MSA was a considerable barrier for patients who wanted to undergo the procedure. Many of these patients elected to continue PPI therapy in lieu of undergoing LNF, however, some patients proceeded with LNF. Although infrequent, there were patients who were offered MSA but preferred to proceed with LNF for various personal reasons. The only way to remove all selection bias would be to perform a randomized controlled trial. However, given that current data already show that MSA is as effective as LNF and with a shorter operative time, shorter length of stay, and less-severe side effects, it seems unlikely that such a trial would be able to accrue the necessary number of patients.

CONCLUSIONS

Analogous GERD patients had similar control of reflux symptoms after both MSA and LNF. Annoying inabilities to belch and vomit and severe gas–bloat symptoms were significantly fewer with MSA. The more favorable side effect profile of MSA supports its use early in the course of GERD.

Author Contributions

Study conception and design: Reynolds, Zehetner, Bildzukewicz, Lipham
Acquisition of data: Reynolds, Wu, Shah
Analysis and interpretation of data: Reynolds, Zehetner, Lipham
Drafting of manuscript: Reynolds, Zehetner, Lipham
Critical revision: Reynolds, Zehetner, Lipham

REFERENCES

1. Finks JF, Wei Y, Birkmeyer JD. The rise and fall of antireflux surgery in the United States. Surg Endosc 2006;20:1698–1701.
2. Kahrilas PJ, Jonsson A, Denison H, et al. Regurgitation is less responsive to acid suppression than heartburn in patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol 2012;10:612–619.
3. Kahrilas PJ. Clinical practice. Gastroesophageal reflux disease. N Engl J Med 2008;359:1700–1707.
4. Malfertheiner P, Nocon M, Vieth M, et al. Evolution of gastro-oesophageal reflux disease over 5 years under routine medical care—the ProGERD study. Aliment Pharmacol Ther 2012;35:154–164.
5. Bonavina L, DeMeester T, Fockens P, et al. Laparoscopic sphincter augmentation device eliminates reflux symptoms and normalizes esophageal acid exposure: one- and 2-year results of a feasibility trial. Ann Surg 2010;252:857–862.
6. Lipham JC, DeMeester TR, Ganz RA, et al. The LINX(R) reflux management system: confirmed safety and efficacy now at 4 years. Surg Endosc 2012;26:2944–2949.
7. Ganz RA, Peters JH, Horgan S, et al. Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 2013;368:719–727.
8. Lipham JC, Taiganides PA, Louie BE, et al. Safety analysis of first 1000 patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease. Dis Esophagus [Epub ahead of print], http://dx.doi.org/10.1111/dote.12199; 2014.
9. Riegler M, Schoppman SF, Bonavina L, et al. Magnetic sphincter augmentation and fundoplication for GERD in clinical practice: one-year results of a multicenter, prospective observational study. Surg Endosc 2014 [Epub ahead of print].
10. Velanovich V. The development of the GERD-HRQL symptom severity instrument. Dis Esophagus 2007;20:130–134.
11. Testoni PA, Testoni S, Mazzoleni G, et al. Long-term efficacy of transoral incisionless fundoplication with Esophyx (Tif 2.0) and factors affecting outcomes in GERD patients followed for up to 6 years: a prospective single-center study. Surg Endosc 2014 [Epub ahead of print].

12. Hogan WJ. Clinical trials evaluating endoscopic GERD treatments: is it time for a moratorium on the clinical use of these procedures? Am J Gastroenterol 2006;101:437–439.
13. Lipka S, Kumar A, Richter JE. No evidence for efficacy of radiofrequency ablation for treatment of gastroesophageal reflux disease: a systematic review and meta-analysis. Clin Gastroenterol Hepatol Clin Gastroenterol Hepatol 2014. pii: S1542–3565(14)01504-3. http://dx.doi.org/10.1016/j.cgh.2014.10.013. [Epub ahead of print] Review.
14. Louie BE, Farivar AS, Shultz D, et al. Short-term outcomes using magnetic sphincter augmentation versus Nissen fundoplication for medically resistant gastroesophageal reflux disease. Ann Thorac Surg 2014;98:498–504; discussion 504–505.

Discussion

DR MICHAEL UJIKI (Evanston, IL): Drs Reynolds, Lipham, and colleagues describe their experience with the magnetic sphincter augmentation through a matched pair analysis with patients who have undergone laparoscopic Nissen fundoplication, and correctly state one of the issues with the data to date, which is that patients having magnetic sphincter augmentation have been compared with groups of patients undergoing Nissen fundoplication who have more severe gastroesophageal reflux disease. They have analyzed 1-year outcomes in 100 patients who were identified through a propensity "best fit" analysis. Both groups reported improved or resolved symptoms of GERD, and most were off acid suppression. Dysphagia and satisfaction rates were similar. Complications and need for dilation were similar. Patients who underwent magnetic sphincter augmentation were more able to belch and vomit.

Up to what size hiatal hernia and body mass index are you currently willing to still perform magnetic sphincter augmentation? Or, in other words, have inclusion criteria expanded in your series of patients? This seems to be a limiting factor in the expansion of use with this device because most patients in whom surgery is indicated tend to have large hernias and are obese.

Has your group experienced any patients with Barrett's esophagus that progressed to dysplastic disease after sphincter augmentation? If so, how did you manage them? If not, how would you handle such a patient, as I would assume that ablation is contraindicated with these patients?

Given that magnetic sphincter augmentation tends to be aimed at a select group of thin patients with small hiatal hernias, mild reflux, and normal esophageal motility—a group that is not common in the surgeon's office—would it be more appropriate to compare magnetic sphincter augmentation with acid suppression rather than fundoplication? Does your group have any experience in comparing these 2 groups? And if so, can you share your results?

DR JOHN C LIPHAM (Los Angeles, CA): In regard to the question about Barrett's and progression, since the clinical trials, we have been placing the device in patients with short segment Barrett's. All patients come back yearly for an endoscopy, given the fact that it is a new procedure. And we have not seen progression of the Barrett's with patients now out 3 to 3½ years.

Specifically in regard to ablation, we have done it with the LINX in place using the HALO90 device. Ken Wang's group at the University of California-Irvine also did an animal study, looking at ablation with LINX in a pig model, and showed that the energy delivered by the ablation did not reach the level of the device, so there were no untoward effects of it. If we see a patient with dysplasia that develops afterwards, we will use the HALO90 to ablate that. So I do not think it precludes ablation in those. I would be a little nervous, however, using the 360, which is the balloon device, given the fact that the balloon there is quite large, and the diameter of the device runs from somewhere around 21 to 26 mm.

In regard to your question about hiatal hernia size, since the clinical trials, we have loosened our requirements in regard to hiatal hernias. We have operated on patients with hiatal hernia size up to 7 cm with seemingly similar results. Now, that is our own case series on those patients. We have recently started to enroll patients in a multicenter trial looking at patients with hiatal hernia size 4 to 7 cm. So hopefully within the next year, I will be able to give you an answer to that question.

Although BMI was an exclusion criterion within the clinical trials, since that time, we have really gone away from excluding patients solely based on their BMI. Even in this study, there were patients in that trial that got LINX with BMIs greater than 35 kg/m^2.

In regard to the MRI and erosion question or comment, initially, when the LINX came out, MRI was a contraindication to the device. Since that time, they have loosened their contraindications somewhat. Currently, a low powered MRI, or a 0.7-Tesla MRI, is allowable with the device in place. There has also been a redesign of the magnets that is currently at a level that seemingly can tolerate a 1.5-Tesla MRI, and the company expects approval from the FDA by January or February. The concern is not that the device will come shooting out like little BBs; the concern has been the repolarization of the magnetic beads, given the high power of the magnet.

In regard to erosion, there have been 5 erosions out of about 3,000 patients being operated on worldwide, giving it about a 0.2% erosion rate. Most of the erosions, at least 3 of the 5, seem to be infectious in origin, meaning the device, at the time of implant or shortly thereafter, got infected, and that seemed to be the reason for its erosion. In the 5 patients in whom it did erode, it was fairly easy to manage. They went down endoscopically. There were usually 2 or 3 or 4 little beads that were intraluminal. The link on each side of those beads was cut and those 3 or 4 beads were then removed endoscopically. Later, they went in laparoscopically and took the remaining part of the device out. No patients required esophagectomy, gastrectomy, or anything more major than that.

In regard to your last question, comparison with proton pump inhibitors (PPI), I agree. I think that is something that needs to be done. I think what we have shown, and others have shown here, is that the device is very effective for that gap population, which admittedly is the milder to moderate reflux patient. I think there are really 2 questions here. One is its comparison to PPI therapy alone, but I think we also need to look at a comparison with the more advanced degrees of reflux. The first step in that is going to be looking at the patients with the larger hiatal hernia size.

DR MARK A TALAMINI (Stony Brook, NY): The context or framework here, I think, is really fascinating, because there is the backdrop of the Angelchik prosthesis story, which is considered

Comparative Analysis of Laparoscopic Fundoplication and Magnetic Sphincter Augmentation for the Treatment of Medically Refractory GERD

WILLIAM O. RICHARDS, M.D., CARLY McRAE, B.S.

From the Department of Surgery, University of South Alabama Health System, Mobile, Alabama

We have recently introduced laparoscopic magnetic sphincter augmentation (MSA) combined with hiatal hernia repair for treatment of patients with medically refractory gastroesophageal reflux disease (GERD). MSA is a novel surgical approach to the treatment of severe GERD, in which magnetic beads are secured around the lower esophageal sphincter, augmenting the lower esophageal sphincter function as an anti-reflux barrier. We hypothesize that patients undergoing MSA will achieve GERD relief, equal to that obtained after laparoscopic Nissen fundoplication. The GERD Health Related Quality of Life (GERD HRQL) Questionnaire is a validated clinical tool that was used to quantify patient outcomes in terms of GERD-related symptoms both on and off proton pump inhibitors and after acute radiation syndrome. We retrospectively reviewed data from patients at our institution enrolled in a prospective institutional review board–approved database "Registry Outcomes Anti-Reflux Surgery" that applies objective and subjective information about patients undergoing anti-reflux surgery. Information from both the database and patient HRQL scores were used to compare the effectiveness of medical intervention with acute radiation syndrome (laparoscopic Nissen fundoplication and MSA) in decreasing GERD-related symptoms in patients. Results are expressed as mean ± SE, and single-factor ANOVA test was used to compare groups. We found that MSA and laparoscopic fundoplication both lead to a comparable decrease in HRQL score and an increase in patient satisfaction when compared with patient's preoperative symptoms with maximum proton pump inhibitor use. In addition, our study shows that MSA is a safe minimally invasive anti-reflux procedure without the negative side-effects, such as gas bloat, inability to belch, and inability to vomit, commonly associated with NF.

MAGNETIC SPHINCTER AUGMENTATION (MSA) using the LINX® (Torax Medical, Saint Paul, MN) device was Food and Drug Administration (FDA) approved in 2012 after the pivotal trial showed that implantation of the magnetic beads around the lower esophageal sphincter (LES) without repair of the hiatal hernia, in patients with smaller than 3-cm hiatal hernias, produced good results.[1] Subsequent studies have shown that MSA has long-term five year success in improving gastroesophageal reflux disease (GERD) symptoms, and reduction of esophageal acid exposure.[2–4] Other long-term studies have shown that MSA is safe and has a lower rate of negative side-effects such as gas bloat and inability to vomit compared with laparoscopic fundoplication (LF).[5]

Although LF was increasingly utilized to treat patients with refractory GERD in the 1990's, many studies identified a marked reduction in the utilization of LF as an option from 2000 to 2010.[6] Most authors attributed this decline in the performance of LF to the negative side-effects after LF, including gas bloat, inability to vomit, and gradual loss of efficacy.[2] At the University of South Alabama introduction of MSA has led to an influx of patients requesting evaluation for the MSA procedure as a perceived improvement over laparoscopic anti-reflux procedures, in particular, the laparoscopic Nissen fundoplication (LNF). Furthermore, it became apparent that gastroenterologists who treat most patients with severe GERD symptoms that do not

Presented at the Annual Scientific Meeting Southeastern Surgical Congress, February 13, 2018, Tampa, FL.

Address correspondence and reprint requests to William O. Richards, M.D., Department of Surgery, University of South Alabama Medical Center, 2451 USA Medical Center Drive, Mastin 719, Mobile, AL 36617-2293. E-mail: brichards@health.southalabama.edu.

Disclosure: William Richards has served as the principal investigator at University of South Alabama Health System during the industry funded trial using the LINX device. This was a multi-institutional trial comparing PPI therapy with MSA as a surgical anti-reflux operation.

respond to proton pump inhibitor (PPI) therapy[7] have been more willing to refer patients for a surgical evaluation if MSA was an option.

This study is a retrospective review of our prospective collected data on patients undergoing anti-reflux surgery at the University of South Alabama, before and after introduction of MSA into surgical practice. Thus, it reflects the change in the types of procedures and the tailoring of the surgical procedure to the anatomy, pathophysiology, and the esophageal motility of the individual patient. The primary outcome objective of the study was to measure the GERD symptom outcomes between LF and the MSA procedure using the patient's own reported standardized validated GERD Health Related Quality of Life (GERD HRQL) questionnaire pre- and postsurgery.

In this study, we found that MSA and LF both lead to a comparable decrease in HRQL score and an increase in patient satisfaction when compared with the patient's preoperative symptoms with maximum PPI use. In addition, our study shows that MSA is a safe minimally invasive anti-reflux procedure without the negative side-effects, such as gas bloat, inability to belch, and inability to vomit, commonly associated with NF.

Methods

Patients undergoing anti-reflux surgery at the University of South Alabama by a single surgeon (W.O.R.) were enrolled in a prospectively collected database and administered the GERD Quality of Life questionnaires over a 12-month period (June 2016–June 2017). The database is an institutional review board–approved prospective database and forms the basis of this retrospective analysis of patients who underwent anti-reflux surgery at the University of South Alabama Health System by a single attending surgeon (W.O.R.). The GERD HRQL is a standardized, validated instrument previously well documented in the literature, and has been shown to accurately reflect GERD symptoms in a numerical fashion.[8] The GERD HRQL survey was administered to the patients preoperatively and at 3 to 6 months postoperatively (Fig. 1).[8]

The results are expressed as mean ± SE and subjected to statistical analysis using ANOVA (Single factor). Statistical significance was met when F > F crit. Dysphagia scores before and after MSA and LF were evaluated using a paired Student's *t* test, which was considered statistically significant when $P < 0.05$.

Both LF and MSA procedures were performed by the attending surgeon (W.O.R.) working with residents on the GI surgical service in a standard fashion. All patients underwent complete dissection of the esophageal hiatus, resection of hernia sac, reduction of the stomach into the abdomen with a minimum of 2 cm of abdominal esophagus, with suture closure of the left and right crus of the diaphragm posterior to the esophagus using figure-of-eight 0 polyester braided sutures. LNF was performed over a 54 to 58 French bougie to create a loose floppy wrap approximately 2 cm in size. The laparoscopic Dor fundoplication was performed over a standard flexible video endoscope to create a 180° anterior fundoplication using 0 braided polyester simple sutures in a technique described previously. (W.O.R.) MSA was performed after complete dissection of the esophageal hiatus, reduction of the stomach into the abdomen and crural closure in the exact same fashion as the Nissen or Dor fundoplication. The MSA was performed by dissecting the posterior vagus nerve away from the esophagus at the level of the gastroesophageal junction and then using the esophageal sizing device to determine the size of the

***Taken from:** Velanovich, V. The development of the GERD-HRQL symptom severity instrument. Diseases of the Esophagus 2007; 20:130-134.

• **Scale**: No symptoms = **0**; symptoms noticeable, but not bothersome = **1**; symptoms noticeable and bothersome, but not every day = **2**; symptoms bothersome every day = **3**; symptoms affect daily activities = **4**; symptoms are incapacitating, unable to do daily activities = **5**

• **Questions**

__ 1. How bad is your heartburn?	0 1 2 3 4 5
__ 2. Heartburn when lying down?	0 1 2 3 4 5
__ 3. Heartburn when standing up?	0 1 2 3 4 5
__ 4. Heartburn after meals?	0 1 2 3 4 5
__ 5. Does heartburn change your diet?	0 1 2 3 4 5
__ 6. Does heartburn wake you from sleep?	0 1 2 3 4 5
__ 7. Do you have difficulty swallowing?	0 1 2 3 4 5
__ 8. Do you have pain with swallowing?	0 1 2 3 4 5
__ 9. Do you have bloating or gassy feelings?	0 1 2 3 4 5
__ 10. If you take medication, does this affect your daily life?	0 1 2 3 4 5

__ How satisfied are you with your present condition? Satisfied __ Neutral __ Dissatisfied __

FIG. 1. The GERD HRQL instrument used in the study to assess patient GERD symptoms.[8]

LINX® device to be implanted. The LINX® magnetic beads are then placed around the esophagus between the posterior Vagus and esophagus and clasped together, which positions the device at the GE junction to augment the LES and prevent reflux.[9]

Postoperative care typically consisted of overnight stay and a mechanical soft diet for four weeks after LF and resumption of a normal diet with frequent (Q2H) snacks for patients undergoing MSA. The regular diet for the MSA patients is used to ensure that a constricting scar does not form around the magnetic beads causing long-term dysphagia. A short course of steroids is given to patients who develop dysphagia after MSA to prevent intense scar formation inhibiting the free motion of the magnetic beads.

Results

Thirty-eight patients underwent laparoscopic anti-reflux surgery. Six of the patients underwent fundoplication and were slightly older in age, had larger hiatal hernias, and had a longer duration of surgery than did the 32 patients who underwent MSA. The patients undergoing LF were chosen because they either had significant dysphagia or esophageal dysmotility on high resolution manometry, or had obstructive symptoms from their large hiatal hernias. Thirty-two patients underwent MSA with concomitant repair of their hiatal hernias. Operative time for the MSA procedure was shorter than that of the LF as shown in Table 1.

As shown in table one, preoperatively, the patients on single- and double-dose PPI were uniformly dissatisfied with their quality of life and had a high GERD HRQL (28.0 ± 9.9), which signified their continued symptoms of heartburn, dysphagia, sleep disturbances, and regurgitation. GERD HRQL scores were significantly reduced (ANOVA F = 43.8 > F crit = 3.12) 3 to 6 months after LF (8.6 ± 15.7) and MSA (6.4 ± 8.2). Patient satisfaction scores were improved in 89 per cent of the patients undergoing MSA, whereas 89 per cent of the patients undergoing LF were satisfied.

Preoperatively, 92 per cent of the patients reported the ability to belch, whereas postoperatively 93, per cent of MSA and 100 per cent of LF patients reported the ability to belch.

Questions seven and eight from the GERD HRQL survey specifically target dysphagia and were used to measure dysphagia symptoms before and after LF and MSA. Dysphagia was slightly more severe in the LF group (4.7 + 2.9) before surgery compared with the MSA group (4.2 ± 3.4). Dysphagia was significantly reduced after both LF (0.7 ± 1.2) and MSA (2.0 ± 2.7) [$P < 0.05$, paired Student's t test].

One patient in the MSA group underwent removal of the LINX® beads approximately one year after the original implantation because of severe dysphagia not responsive to endoscopic dilation. The patient had evidence of scar formation around the LINX® beads, with mild esophageal dilation that was unresponsive to endoscopic dilation. Her dysphasia resolved after explantation of the beads, but she continues to have GERD symptoms which are treated medically.

Two patients underwent laparoscopic capsulotomy of the LINX® device for treatment of severe dysphagia. Both patients responded and experienced improvement of their dysphagia after operation, and are presently satisfied with their condition.

We assessed practice patterns in the US before and after the initiation of MSA (August 2015) by retrospectively comparing operative notes from the time of the study (June 2016–June 2017) to operative notes from a similar time frame before the utilization of MSA (June 2014–June 2015). As shown in Table 2, there has been a 4-fold increase in the number of primary anti-reflux procedures performed after the introduction of the MSA procedure.

Discussion

As a referral center for foregut disease we have seen a steady stream of patients referred for takedown of their previously performed Nissen fundoplication secondary to gas bloat, dysphagia, regurgitation, and development of

TABLE 1. *Demographics and GERD HRQL, for Patients Undergoing MSA, LF Pre- and Post-surgery*

Intervention Tvpe	GERD HRQL	Patient Satisfaction	Ability to Belch	Age	Duration of surgery (minutes)	Size of Hiatal Hernia (cm)
Preoperative on PPI (n = 37)	28.0 ± 9.9	3% Neutral 97% Dissatisfied	92%	55.6 ± 13.2	N/A	3.6 ± 2.4
Post MSA (n = 32)	6.4 ± 8.2	74% Satisfied 15% Neutral 11% Dissatisfied	93%	54.5 ± 12.8	90.9 ± 16.4*	3.2 ± 1.9
Post LF (n = 6)	8.6 ± 13.7	89% Satisfied 11% Neutral	100%	61.7 ± 14.4	131.8 ± 34.6	5.7 ± 3.7

* Four outliers who underwent extensive lysis of adhesions >60 minutes were removed from the duration of surgery analysis in the MSA group.

TABLE 2. *Comparison of Anti-Reflux Surgery before and after Institution of MSA*

Procedure	June 2014 to June 2015	June 2016 to June 2017
LF	8 (Seven Nissen, one Toupet)	6 (Two Nissen, four Dor)
MSA	0	32
Total	8	38

secondary achalasia. These problems that arise from the inherent nature of the Nissen creating a one-way valve might be avoided by the performance of the MSA, which only augments the LES as an anti-reflux barrier. Advantages of MSA over LF include the ability to belch and vomit, while demonstrating equal improvement in GERD symptoms, reduction in PPI use, and in esophageal acid exposure.[10] Our study found that MSA was just as effective as LF in improving GERD HRQL and in patient satisfaction.

The patients undergoing LF in this study were not candidates for MSA because their preoperative workup identified esophageal dysmotility, or they had giant hiatal hernias with obstructive symptoms, which contraindicated performance of the MSA procedure. Our current practice for patients with esophageal dysmotility or severe obstructive symptoms includes hiatal hernia repair combined with partial 180° partial fundoplication, to prevent side-effects of the 360° Nissen fundoplication such as gas, bloat, and the inability to belch.[5, 11, 12] The decision to perform the Dor partial fundoplication was made because of our favorable results when using it on patients undergoing Heller myotomy[13] and from the results of a trial of Dor fundoplication versus LNF which demonstrated equal improvement in GERD symptoms after surgery, but a marked improved ability to belch in the patients who underwent Dor fundoplication.[12] In our study, patients undergoing MSA overwhelmingly reported their ability to belch in the peri-operative period.

Postoperative care and diet are dramatically different for patients undergoing MSA compared with patients undergoing LF. In our experience, approximately 70 per cent of the patients exhibit some levels of dysphagia starting around 10 days postoperatively. Postoperative diet for MSA patients concentrates on eating at regular intervals to exercise the esophagus and move the beads, to prevent circatrix formation around the beads, causing long-term dysphagia. Our diet instructions include eating a regular diet on postoperative day one with small snacks of solid food every two hours while awake during the first eight weeks after surgery. For patients who develop dysphagia after MSA, institution of steroids and/or esophageal antispasmodic medications are used. This study also shows that preoperative dysphagia is quite common in GERD patients despite maximal PPI use; however, our objective measurements showed that dysphagia is significantly reduced at three months after both LF and MSA compared with their preoperative condition results that are similar to that of the literature.[10, 14]

Since the introduction of MSA into our surgical practice, there has been a 4-fold increase in the number of patients undergoing anti-reflux surgery to treat GERD symptoms refractory to PPI use, reflecting a growing consensus by patients and referring gastroenterologists that the benefits of surgical therapy outweigh the risks. Previous studies have shown improved patient outcomes if the hiatal hernia is repaired in conjunction with MSA.[15] In this study, all 32 patients undergoing MSA also underwent a concomitant hiatal hernia repair, which reflects the expanding indications for MSA combined with hiatal hernia repair.

Our study is in concordance with other studies and indicates that operation times for MSA are shorter than for LF. The difference in operative time reflects both the smaller size of the hiatal hernias in the MSA group and the reduction in time to suture the fundoplication compared with placement of the LINX® device in patients undergoing MSA. Operative times for our cohort of patients undergoing MSA are longer than for patients undergoing MSA in the literature, which may reflect the fact that all of our patients underwent complete dissection of larger hiatal hernias with repair of the hiatal hernia.

There was one patient who had persistent dysphasia beyond six months who ultimately underwent laparoscopic removal of the LINX® device. Her rapid return to her preoperative condition is an advantage for the MSA procedure. Performance of the LF procedure requires takedown of the short gastric vessels and wrapping the fundus of the stomach around the esophagus, resulting in anatomical distortion of the stomach and lower esophagus. Reversal of the Nissen is a more difficult procedure which ultimately does not restore the original anatomy, and frequently leaves the patient with continued foregut symptoms.[16, 17] The relative ease of explantation of the LINX® device is a positive for our gastroenterology colleagues and for the patients at large.

Two patients underwent laparoscopic capsulotomy more than three months after MSA for dysphagia unresponsive to endoscopic dilation. In this procedure, the capsule around the magnetic beads was incised using monopolar electrocautery, exposing the magnetic beads and insuring they had total freedom of motion. Postoperatively, patients were placed on a 21-day steroid taper and managed with a vigorous diet of regular food and Q2H snacks to exercise the magnetic beads and prevent recurrence of the cicatrix that limited motion of the beads after original implantation. In short term follow-up, both patients improved although it is too

early to determine whether this course of action will prevent further formation of scar around the magnetic beads.

One of the striking advantages of the MSA procedure is the constant force exerted by the magnetic beads which translates to a constant augmentation of the LES and restoration of the anti-reflux barrier.[2] Long-term studies of MSA demonstrate very stable GERD symptoms over the five years, which is attributable to the lack of deterioration of the magnetic force in the LINX® device.[18, 19] It is well recognized that the LF is a tissue repair and loses efficacy each year in its anti-reflux barrier function.[17] Although this study did not evaluate long-term results, many of the patients referred themselves for the MSA procedure based on their perception that it is equally efficacious compared with LF, but without troubling side-effects.

An additional advantage of the MSA procedure is that the surgeon utilizes an esophageal measuring device that precisely identifies the size of the LINX®[20] beads required for each patient, allowing for a more standardized procedure that is less dependent on surgeon's identification of the proper size of the fundoplication.

All patients undergoing MSA and LF underwent a battery of tests that are considered essential before anti-reflux surgery including: endoscopic evaluation, barium esophagram, high resolution manometry, and 24-hour esophageal Ph study while off PPI therapy for at least seven days, to demonstrate the presence of reflux. In our experience, approximately 20 per cent of patients we work-up have identified conditions and other diagnoses that contraindicate the performance of MSA. In the first 20 patients undergoing workup for MSA, we identified two patients with achalasia, one patient with severe esophageal dysmotility secondary to scleroderma, and one patient with esophagogastric outlet obstruction. Patients are all highly screened before undergoing MSA to insure that they will do well postoperatively in terms of the pathophysiology and severity of disease and adequate esophageal motility to overcome the increased LES pressure from the LINX® device.[21]

Although some studies indicate that the MSA procedure is much simpler than the Nissen,[2] we would argue that, in our experience, 85 per cent of the MSA procedure is identical to LF. The procedure includes the complete dissection of the phrenoesopageal membrane and hernia sac, reduction of the gastroesophageal junction into the abdomen through complete dissection into the mediastinum, identification and preservation of the anterior and posterior vagal nerves, and then suture repair of the esophageal hiatus. The surgeon must dissect precisely between the posterior vagus and the esophagus and precisely place the magnetic beads on the esophagus at the level of the LES.

In conclusion, our study shows that MSA is a safe, minimally invasive anti-reflux procedure without the negative side-effects such as gas bloat, inability to belch, and inability to vomit commonly associated with the Nissen. MSA and LF significantly improve GERD symptoms and dysphagia in patients who have PPI-resistant GERD symptoms to the same degree. Our experience has shown a remarkable increase in the number of patients referred for anti-reflux surgery which reflects the view of the referring gastroenterologists and patients who conclude that MSA is a better long-term procedure with less side-effects than laparoscopic Nissen. Despite the need for re-operation in a small percentage of MSA patients in this study, our procedure of choice in patients who have medically refractory GERD symptoms is the MSA with hiatal hernia repair. We conclude this based on the advantages of MSA previously discussed including standardization of bead sizing, decreased operation times, ability to belch and vomit postoperatively, retention of normal anatomy and easy explanation of the LINX® device, and the ability of the magnetic beads to last much longer than the tissue wrap used in LF.

REFERENCES

1. Ganz RA, Peters JH, Horgan S, et al. Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 2013; 368:719–27.

2. Zak Y, Rattner DW. The use of LINX for gastroesophageal reflux. Adv Surg 2016;50:41–8.

3. Bonavina L, Attwood S. Laparoscopic alternatives to fundoplication for gastroesophageal reflux: the role of magnetic augmentation and electrical stimulation of the lower esophageal sphincter. Dis Esophagus 2016;29:996–1001.

4. Saino G, Bonavina L, Lipham JC, et al. Magnetic sphincter augmentation for gastroesophageal reflux at 5 years: final results of a pilot study show long-term acid reduction and symptom improvement. J Laparoendosc Adv Surg Tech A 2015;25: 787–92.

5. Reynolds JL, Zehetner J, Wu P, et al. Laparoscopic magnetic sphincter augmentation vs laparoscopic nissen fundoplication: a matched-pair analysis of 100 patients. J Am Coll Surg 2015;221: 123–8.

6. Worrell SG, Greene CL, DeMeester TR. The state of surgical treatment of gastroesophageal reflux disease after five decades. J Am Coll Surg 2014;219:819–30.

7. Brockmeyer JR, Connolly EE, Wittchow RJ, et al. Resolution of fundic gland polyposis following laparoscopic magnetic sphincter augmentation and subsequent cessation of proton pump inhibitors [internet]. Case reports in gastrointestinal medicine. 2015. Available at: https://www.hindawi.com/journals/crigm/2015/576263/. Accessed January 28, 2018.

8. Velanovich V. The development of the GERD-HRQL symptom severity instrument. Dis Esophagus 2007;20:130–4.

9. Kuckelman JP, Barron MR, Martin MJ. "The missing LINX" for gastroesophageal reflux disease: operative techniques video for the Linx magnetic sphincter augmentation procedure. Am J Surg 2017;213:984–7.

10. Czosnyka NM, Buckley FP, Doggett SL, et al. Outcomes of magnetic sphincter augmentation—a community hospital perspective. Am J Surg 2017;213:1019–23.

11. Sheu EG, Nau P, Nath B, et al. A comparative trial of laparoscopic magnetic sphincter augmentation and Nissen fundoplication. Surg Endosc 2015;29:505–9.

12. Cao Z, Cai W, Qin M, et al. Randomized clinical trial of laparoscopic anterior 180° partial versus 360° Nissen fundoplication: 5-year results. Dis Esophagus 2012;25:114–20.

13. Richards WO, Torquati A, Holzman MD, et al. Heller myotomy versus heller myotomy with dor fundoplication for achalasia. Ann Surg 2004;240:405–15.

14. Spechler SJ, Lee E, Ahnen D, et al. Long-term outcome of medical and surgical therapies for gastroesophageal reflux disease: follow-up of a randomized controlled trial. JAMA 2001;285: 2331–8.

15. Rona KA, Reynolds J, Schwameis K, et al. Efficacy of magnetic sphincter augmentation in patients with large hiatal hernias. Surg Endosc 2017;31:2096–102.

16. Wang YR, Dempsey DT, Richter JE. Trends and perioperative outcomes of inpatient antireflux surgery in the United States, 1993–2006. Dis Esophagus 2011;24:215–23.

17. Zhou T, Harnsberger C, Broderick R, et al. Reoperation rates after laparoscopic fundoplication. Surg Endosc 2015;29: 510–4.

18. Zhang H, Dong D, Liu Z, et al. Revaluation of the efficacy of magnetic sphincter augmentation for treating gastroesophageal reflux disease. Surg Endosc 2016;30:3684–90.

19. Magnetic sphincter augmentation for gastroesophageal reflux at 5 years: final results of a pilot study show long-term acid reduction and symptom improvement [internet]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4624249/. Accessed January 28, 2018.

20. Desart K, Rossidis G, Michel M, et al. Gastroesophageal reflux management with the LINX® system for gastroesophageal reflux disease following laparoscopic sleeve gastrectomy. J Gastrointest Surg 2015;19:1782–6.

21. Revaluation-of-the-efficacy-of-magnetic-sphincter-augmentation-for-treat….pdf [internet]. Available at: http://17dxz634tcba1yaiqj33jt8t.wpengine.netdna-cdn.com/wp-content/uploads/2015/12/Revaluation-of-the-efficacy-of-magnetic-sphincter-augmentation-for-treat….pdf. Accessed January 28, 2018.

Surg Endosc (2015) 29:505–509
DOI 10.1007/s00464-014-3704-6

ORIGINAL ARTICLES

A comparative trial of laparoscopic magnetic sphincter augmentation and Nissen fundoplication

Eric G. Sheu · Peter Nau · Barbara Nath · Braden Kuo · David W. Rattner

Received: 5 April 2014 / Accepted: 22 June 2014 / Published online: 11 July 2014

Abstract
Background Laparoscopic magnetic sphincter augmentation (MSA) with the LINX device is a promising new therapy for the treatment of gastroesophageal reflux disease (GERD). Initial studies have demonstrated MSA to be safe and effective. However, no direct comparison between MSA and laparoscopic Nissen fundoplication (LNF), the gold standard surgical therapy for GERD, has been performed.
Methods A single institution, case–control study was conducted of MSA performed from 2012 to 2013 and a cohort of LNF matched for age, gender, and hiatal hernia size.
Results MSA and LNF were both effective treatments for reflux with 75 and 83 % of patients, respectively, reporting resolution of GERD at short-term follow-up. Dysphagia was common following both MSA and LNF, but severe dysphagia requiring endoscopic dilation was more frequent after MSA (50 vs. 0 %, $p = 0.01$). Need for dilation did not correlate with size of the LINX device or any other examined patient factors. A trend toward decreased adverse GI symptoms of bloating, flatulence, and diarrhea was seen after MSA compared to LNF (0 vs. 33 %). MSA had a shorter operative time (64 vs. 90 min, $p < 0.01$) but other peri-operative outcomes, including pain, morbidity, and re-admissions were equivalent to LNF. MSA patients were more likely to be self-referred (58 vs. 0 %, $p < 0.001$).
Conclusions MSA and LNF are both effective and safe treatments for GERD; however, severe dysphagia requiring endoscopic intervention is more common with MSA. Other adverse GI side effects may be less frequent after MSA. Consideration should be paid to these distinct post-operative symptom profiles when selecting a surgical therapy for reflux disease.

Keywords Gastroesophageal reflux disease · Fundoplication · LINX · Magnetic sphincter augmentation · Dysphagia · Endoscopic dilation

Presented at the SAGES 2014 Annual Meeting, April 2–5, 2014, Salt Lake City, Utah.

E. G. Sheu · D. W. Rattner (✉)
Department of Surgery, Massachusetts General Hospital, Boston, MA, USA
e-mail: drattner@mgh.harvard.edu

E. G. Sheu
e-mail: esheu@partners.org

E. G. Sheu · D. W. Rattner
Division of General & Gastrointestinal Surgery, Massachusetts General Hospital, 15 Parkman Street, Boston, MA, USA

P. Nau
Department of Surgery, University of Iowa, Iowa City, IA, USA

B. Nath · B. Kuo
Department of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA

Laparoscopic magnetic sphincter augmentation (MSA) is a recently introduced operation for the treatment of gastroesophageal reflux disease (GERD) [1–4]. MSA involves a limited crural dissection to place a necklace of magnetized beads (LINX®, Torax Medical) around the esophagus at the gastroesophageal junction [1]. The magnetic force between the beads augments the lower esophageal sphincter to prevent GERD. The device, however, will open when enough luminal pressure is generated to disrupt the magnetic force between the beads. Thus, the device is meant to recreate a physiologic LES, preventing GERD while theoretically allowing appropriate transit of food, liquid, and gas through the GE junction.

Early published experience with MSA has shown excellent efficacy with high levels of resolution of both symptomatic GERD and physiologic reflux as documented by pH testing. The two case series of MSA that have been published did not report any serious peri-operative complications although a high rate of post-operative dysphagia has been seen [3, 4]. In their multi-center trial, Ganz et al. found that 68 % of patients experienced early dysphagia, and 19 % required endoscopic dilation. Bonavina, in their single center case series, had a lower endoscopic dilation rate—2 %—although they did not report the overall frequency of dysphagia. Indeed, refractory dysphagia has been the most common indication for device removal in early experience with MSA [5].

No direct comparison has been made with laparoscopic fundoplication—the standard surgical therapy for GERD [6]. We therefore compared our initial experience with MSA against a contemporaneous, case-matched series of laparoscopic Nissen fundoplication (LNF). We compared safety, efficacy, and side effect profiles in an effort to better define optimal selection criteria for the two procedures.

Methods

Institutional review board approval for this study was obtained from the Partners Human Research Committee. A retrospective study of all laparoscopic MSAs performed at the Massachusetts General Hospital between 2012 and 2013 was performed ($n = 12$). A case-matched cohort of patients ($n = 12$) undergoing primary LNF during the same time period was identified by matching for age, gender, GERD symptoms, and hiatal hernia size.

Pre-operative work-up for all patients included upper endoscopy and esophageal manometry. All MSA patients underwent standard pH or pH/impedance testing to document pathologic reflux. For LNF, pH testing was omitted if there was documented evidence of reflux on upper endoscopy (e.g., Barrett's esophagus or erosive esophagitis confirmed by pathology) and a classic symptom complex, which occurred in five patients. Upper GI fluoroscopic and gastric emptying studies were performed selectively. Exclusion criteria for MSA included prior anti-reflux surgery, hiatal hernia >2 cm, esophageal dysmotility, allergy to device materials, and need for future MRI imaging. MSA and LNF were otherwise both offered to all appropriate candidates based on this work-up.

All MSA and LNF were performed by a single surgeon (D. W. R.) using standard techniques. For MSA, the gastrohepatic ligament was opened with preservation of the hepatic branch of the vagus nerve. A limited dissection of the phrenoesophageal ligament was performed to allow development of a retroesophageal window at the decussation of the right and left crus. The posterior vagus was identified, a window was created between the posterior vagus and esophagus, and this space was marked with a vessel loop. The LINX® device (Torax Medical, Minnesota, USA) was sized according to the manufacturer's instructions, and the device was subsequently placed to encircle the esophagus at the GE junction, excluding the posterior vagus, and secured with the TI-Knot device (LSI Solutions, New York, USA).

LNF was performed as described [7]. A complete crural dissection to mobilize >3 cm intra-abdominal esophagus was performed. The short gastric vessels supplying the fundus were divided, and a 360° floppy fundoplication was fashioned over a 56F bougie using three pledgeted Ethibond sutures, each incorporating esophageal wall. A posterior crural repair was performed with pledgeted 0 Ethibond sutures.

Post-operatively, patients were kept overnight for observation. MSA patients were started on a liquid diet immediately and most were discharged the following day on a soft diet. LNF patients were maintained on a liquid diet for 1 week post-op and then advanced to a soft solid diet. GERD medications were discontinued at the patients' first 2–3 week post-op office visit. Indication for dilation was persistent intolerance to a soft solid diet for greater than 6 weeks with ongoing weight loss. Patients with dysphagia who were able to maintain their weight did not undergo dilation.

Pre-operatively, patients were asked to describe their reflux symptom(s), and the primary presenting symptom(s) was recorded in a prospective fashion. Symptoms were classified as heartburn, regurgitation, laryngopharyngeal, or pulmonary symptoms. Post-operatively, patients were assessed at 1 and 3 weeks by the same examiner in a non-blinded fashion. Patients were questioned about changes in their major presenting reflux symptoms and development of any other UGI symptoms. Each patient was specifically asked about dysphagia to liquids or solids, bloating, flatulence, and the ability to belch. Subsequent follow-up visits were with primary care providers and/or gastroenterologist with surgeon follow-up as needed; these records were reviewed retrospectively. GERD treatment success was defined as resolution of pre-operative symptoms on decreased acid blocking medication and/or a normal pH study off-medication. Operative time was defined as time from first incision to closure of all port sites.

Statistical analysis was performed using 2-tailed Student's t test or 2-tailed Fisher's exact test as appropriate. Results are reported as mean ± standard deviation or median values as indicated.

Table 1 Demographic characteristics of MSA and LNF patients are shown

	LINX	Fundoplication	*p*
Age	39.3 ± 12.9	43.8 ± 9.2	0.34
Sex (% female)	41.7	50	0.99
BMI	26.8 ± 4.4	26.8 ± 3.6	0.96
Symptoms (%)			0.99
Heartburn	42	58	
Regurgitation	25	17	
Atypical (LPR, pulm)	33	25	
Self-referred (%)	58	0	0.004

Values shown in each column represent mean values ± standard deviation. Calculated *p* values for two tailed Student's and Fisher's exact test are shown

Table 2 Peri-operative outcomes after MSA and LNF

	LINX	Fundoplication	*p*
OR time (min)	63.7 ± 11.6	90.3 ± 18.0	<0.001
Length of stay (days)	1.0 ± 0	1.1 ± 0.3	0.32
Pain meds (mg)	17.8 ± 16.5	27.1 ± 29.2	0.36

Mean operating time, length of stay, and inpatient narcotic usage (expressed in equivalent mg of oxycodone) are shown with variance reported as standard deviation

Results

Demographic data of patients who underwent MSA and LNF are presented in Table 1. The two groups were well matched for age, gender, and BMI. The proportion of patients who reported atypical symptoms—i.e., laryngopharyngeal (LPR) or pulmonary symptoms—was also similar. The only significant difference in the groups pre-operatively was in referral patterns. 58 % of MSA patients were self-referred, most specifically seeking LINX placement. In contrast, none of the LNF were self-referred ($p < 0.004$).

Early peri-operative outcomes were similar between MSA and LNF (Table 2). Hospital length of stay and need for post-operative narcotic medication were equivalent. Both procedures were safe, with no 30-day morbidity, mortality, or re-admissions occurring in either group. Average operative time was slightly but significantly shorter for MSA (64 vs. 90 min, $p < 0.001$).

With an average follow-up time of 7 months, both operations were effective in treating GERD (Table 3). 75 % of MSA and 83 % of LNF reported resolution of their pre-operative symptoms. However, the side effect profiles of the procedures differed. 33 % of LNF experienced post-operative symptoms of bloating, flatulence, or diarrhea, whereas no MSA patients reported these adverse GI side effects. However, this difference did not reach statistical significance.

Table 3 Comparison of efficacy and side effect profiles of MSA and LNF

	LINX	Fundoplication	*p*
GERD resolution (%)	75	83	0.99
Any dysphagia (%)	83	58	0.37
Dysphagia requiring endoscopic dilation (%)	50	0	0.014
Other GI symptoms (%)	0	33	0.21

Percentage of patients reporting resolution of pre-operative reflux symptoms, adverse GI symptoms of bloating, flatulence, and/or irritable bowel/diarrhea symptoms, and dysphagia to liquids or solids are shown

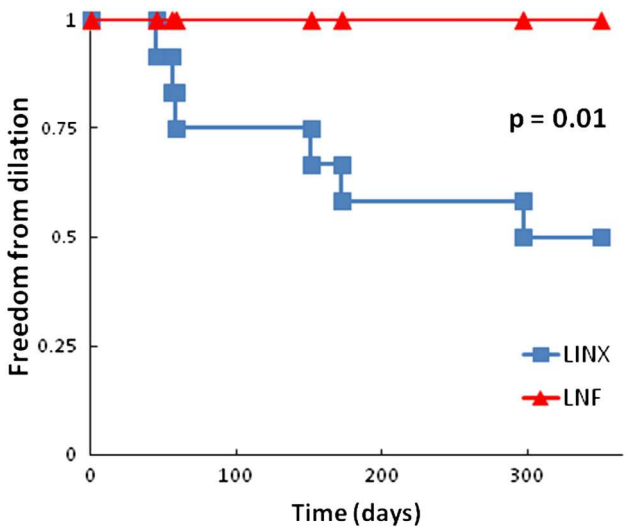


Fig. 1 Kaplan–Meier curve showing time to endoscopic dilation after MSA and LNF. Median time to dilation after MSA was 130 days

Dysphagia was a significant adverse effect of both operations, with 83 % of LINX and 58 % of LNF patients reporting some degree of dysphagia post-operatively. Dysphagia after LNF occurred within the first 30 days post-op, and all cases resolved with expectant management. In contrast, dysphagia after MSA did not develop in the first 7 days post-op, but once it occurred it was more persistent and severe. 50 % of MSA patients were eventually requiring endoscopic dilation. Dysphagia resolved after dilation in all but one patient. Mean time to dilation after MSA was 130 days, with 50 % of dilations occurring within the first 2 months post-operatively (Fig. 1). Need for dilation after MSA did not correlate with the presence of a hiatal hernia, pre-operative manometry, sex, BMI, or size of the LINX device placed.

Routine post-operative studies were not performed in this series. Two patients in the MSA group did undergo post-op pH testing—one for persistent LPR symptoms and another for acute chest pain. While both patients had positive pH studies pre-op their post-op studies showed that the MSA successfully abolished pathologic acid reflux.

Discussion

This study is the first, to our knowledge, to directly compare MSA with LNF. We find that both operations have similar short-term effectiveness for the treatment of reflux disease. In the immediate peri-operative period, both operations were safe, with no mortality, complications, or re-admissions, and a short inpatient length of stay. Our GERD response rates with both therapies are similar to those previously reported with longer follow-up [3, 4, 7–9]. However, direct comparison of LNF and MSA with long-term follow-up is needed to determine which is the more durable therapy for GERD.

While dysphagia is common after both MSA and LNF, we find that post-operative dysphagia after MSA is more severe, longer lasting, and more frequently requires intervention with endoscopic dilation compared to LNF. In our experience, post-fundoplication dysphagia occurs early and usually resolves in the first several weeks as post-operative edema at the GE junction resolves [10]. In contrast, we find that dysphagia post-MSA often begins later and can be progressive, as reflected in our median time to dilation of 130 days post-op, with half of dilations occurring >2 months post-operatively. Post-MSA dysphagia is speculated to occur due to scarring around the device impairing opening of the magnetic sphincter [3]. In our experience, endoscopic dilation is highly effective in relieving this dysphagia, presumably by disrupting the scar that has formed around the implant and restoring normal function of the LINX device.

It is not clear to us why our dilation rate after MSA is higher than those previously reported [3–5]. We do not believe technical factors explain this difference. We did not perform crural repairs in this series, and we sized the LINX device with a manufacturer's representative present. There was no association of device size with dysphagia. We selected patients for MSA using the same criteria as the initial trial, and all our MSA patients had appropriate esophageal manometry studies and no significant hiatal hernias. It is possible that we had a lower threshold for proceeding to dilation than others. Our criteria for intervention were persistent intolerance of a soft diet with ongoing weight loss. Given the effectiveness of dilation for treatment of dysphagia, we felt the benefit of intervention outweighed the relatively low risks.

Although the MSA and LNF groups were well matched for variables such as pre-operative symptoms, age, and anatomic factors (e.g., hiatal hernia), it is likely that differences in the two patient populations exist that are not measurable. Suggestive of underlying differences in the groups is our finding that more MSA patients were self-referred, usually specifically for the LINX device. In our initial experience, we have found that the MSA population is highly educated about GERD and perhaps more vigilant about their symptoms both pre- and post-operatively.

An advantage of MSA is that MSA patients did not report classic post-fundoplication GI side effects of bloating, excess flatulence, and inability to belch or vomit, although these differences did not reach statistical significance in our small series [10, 11]. Others series have also shown a low frequency of adverse GI side effects post-MSA [3]. Should head to head trials of MSA and LNF confirm this finding, such an improved outcome would be a significant advantage for MSA that would be important to patients and surgeon in choosing among surgical reflux therapies.

Our study is limited by its retrospective, non-blinded, single institution design and lack of long-term follow-up. We did not quantify changes in GERD symptoms and GI adverse effects with validated questionnaires, which we hope to do in the future [12]. As is our usual practice, we did not perform objective post-operative pH or manometry testing in all patients. Nevertheless, we believe our endpoint of endoscopic dilation to be objective, and the observed differences between MSA and LNF striking. Clearly, our findings require confirmation in larger and preferably randomized controlled studies comparing fundoplication and MSA.

In conclusion, we find that MSA and LNF are both effective treatments for GERD. Severe dysphagia requiring intervention occurs more frequently following MSA than LNF. The higher rate of significant dysphagia must be weighed against the possibility of fewer 'windy' GI side effects when the patient and surgeon choose an anti-reflux therapy.

Disclosures Eric G. Sheu, Peter Nau, Barbara Nath, and Braden Kuo have no conflicts of interest to disclose. David W. Rattner received consulting fees from Olympus and honoraria from TransEnterix.

References

1. Bonavina L, Saino G, Bona D, Lipham J, Ganz RA, Dunn D, Demeester T (2008) Magnetic sphincter augmentation of the lower esophageal sphincter: results of a feasibility clinical trial. J Gastrointest Surg 12:2133–2140

2. Bonavina L, deMeester TR, Fockens P et al (2010) Laparoscopic sphincter augmentation device eliminates reflux symptoms and normalizes esophageal acid exposure: one- and 2-year results of a feasibility trial. Ann Surg 23:325–337
3. Ganz RA, Peters JH, Horgan S, Bemelman WA, Dunst CM, Edmundowicz SA, Lipham JC, Luketich JD, Melvin WS, Oelschlager BK, Schlack-Haerer SC, Smith CD, Smith CC, Dunn D, Taiganides PA (2013) Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 368:719–727
4. Bonavina L, Saino G, Bona D, Sironi A, Lazzari V (2013) One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 217:577–585
5. Lipham JC, Taiganides PA, Louie BE, Ganz RA, Demeester T (2014) Safety analysis of first 1000 patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease. Dis Esophagus. doi:10.1111/dote.12199
6. Stefanidis D, Hope WW, Kohn GP, Reardon PR, Richardson WS, Fanelli RD (2010) Guidelines for surgical treatment of gastroesophageal reflux disease. Surg Endosc 24:2467–2469
7. Gee DW, Andreoli MT, Rattner DW (2008) Measuring the effectiveness of laparoscopic antireflux surgery. Arch Surg 143:482–487
8. Oelschlager BK, Ma KC, Soares RV et al (2012) A broad assessment of clinical outcomes after laparoscopic antireflux surgery. Ann Surg 256:87–94
9. Galmiche JP, Hatlebakk J, Attwood S et al (2011) Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD: the LOTUS randomized clinical trial. JAMA 305:1969–1977
10. Richter JE (2013) Gastroesophageal reflux disease treatment: side effects and complications of fundoplication. Clin Gastroenterol Hepatol 11:465–471
11. Humphries LA, Hernandez JM, Whale C et al (2013) Causes of dissatisfaction after laparoscopic fundoplication: the impact of new symptoms, recurrent symptoms, and the patient experience. Surg Endosc 11:465–471
12. Velanovich V (1998) Comparison of generic (SF-36) versus disease specific quality-of-life (GERD-HRQL) scales for gastroesophageal disease. J Gastrointest Surg 2:141–145

Surg Endosc (2016) 30:3289–3296
DOI 10.1007/s00464-015-4659-y

Multi-institutional outcomes using magnetic sphincter augmentation versus Nissen fundoplication for chronic gastroesophageal reflux disease

Heather F. Warren[1] · Jessica L. Reynolds[2] · John C. Lipham[2] · Joerg Zehetner[2] · Nikolai A. Bildzukewicz[2] · Paul A. Taiganides[3] · Jody Mickley[3] · Ralph W. Aye[1] · Alexander S. Farivar[1] · Brian E. Louie[1]

Received: 23 April 2015 / Accepted: 28 October 2015 / Published online: 5 November 2015

Abstract
Background Magnetic sphincter augmentation (MSA) has emerged as an alternative surgical treatment of gastroesophageal reflux disease (GERD). The safety and efficacy of MSA has been previously demonstrated, although adequate comparison to Nissen fundoplication (NF) is lacking, and required to validate the role of MSA in GERD management.
Methods A multi-institutional retrospective cohort study of patients with GERD undergoing either MSA or NF. Comparisons were made at 1 year for the overall group and for a propensity-matched group.
Results A total of 415 patients (201 MSA and 214 NF) underwent surgery. The groups were similar in age, gender, and GERD-HRQL scores but significantly different in preoperative obesity (32 vs. 40 %), dysphagia (27 vs. 39 %), DeMeester scores (34 vs. 39), presence of microscopic Barrett's (18 vs. 31 %) and hiatal hernia (55 vs. 69 %). At a minimum of 1-year follow-up, 354 patients (169 MSA and 185 NF) had significant improvement in GERD-HRQL scores (pre to post: 21–3 and 19–4). MSA patients had greater ability to belch (96 vs. 69 %) and vomit (95 vs. 43 %) with less gas bloat (47 vs. 59 %). Propensity-matched cases showed similar GERD-HRQL scores and the differences in ability to belch or vomit, and gas bloat persisted in favor of MSA. Mild dysphagia was higher for MSA (44 vs. 32 %). Resumption of daily PPIs was higher for MSA (24 vs. 12, $p = 0.02$) with similar patient-reported satisfaction rates.
Conclusions MSA for uncomplicated GERD achieves similar improvements in quality of life and symptomatic relief, with fewer side effects, but lower PPI elimination rates when compared to propensity-matched NF cases. In appropriate candidates, MSA is a valid alternative surgical treatment for GERD management.

Keywords Gastroesophageal reflux disease · Anti-reflux surgery · Multi-institutional · Nissen fundoplication · Outcomes

Chronic gastroesophageal reflux disease (GERD) is a common disease affecting up to 25 % of the US population [1]. Despite a wide clinical and physiologic spectrum ranging from mild reflux symptoms to severe regurgitation and aspiration, Barrett's esophagus and adenocarcinoma [1–6], therapeutic options have been predominately limited to: proton pump inhibitors (PPIs) and anti-reflux surgery, predominately Nissen fundoplication (NF) [7–9]. While PPIs have been shown to effectively control GERD symptoms in most patients, up to 40 % of patients are not completely controlled by maximal medical therapy and continue to experience breakthrough symptoms [2]. Despite this significant percentage of patients experiencing inadequate control of their reflux symptoms, less than 1 % will opt for surgery to treat their symptoms of GERD. Consequently, there is significant therapy gap between

Presented at the SAGES 2015 Annual Meeting, April 15–18, 2015, Nashville, Tennessee

✉ Brian E. Louie
Brian.Louie@swedish.org

[1] Division of Thoracic Surgery, Swedish Medical Center and Cancer Institute, 1101 Madison Street Suite 900, Seattle, WA 98104, USA

[2] Division of Upper GI and General Surgery, Keck Medical Center at University of Southern California, Los Angeles, CA, USA

[3] Knox Community Hospital, Mount Vernon, OH, USA

those completely satisfied with their medical management and those seeking surgical treatment for GERD [10–15].

This therapy gap persists even though NF has been shown to be more effective than PPIs at controlling reflux disease, particularly when performed at specialized centers [7, 16–19]. Underutilization and delayed employment of anti-reflux surgery is likely due, in part, to patient and referring provider concerns about long-term durability, potential for intraoperative complications, and postoperative NF side effects including dysphagia, the inability to belch or vomit, and resulting hyperflatulence and bloating [16, 20–29]. More recently less invasive anti-reflux surgery options have emerged to address this treatment gap and target patients earlier in the GERD disease process.

In March 2012, the United States Food and Drug Administration approved the use of magnetic sphincter augmentation (MSA), as an alternative surgical intervention in the management of GERD based on two single-arm studies involving 144 patients [16, 30]. Several additional studies have confirmed the safety and efficacy of MSA in the treatment of chronic GERD [16, 24, 30–34]. However, there is a lack of comparative data evaluating MSA against NF and comparisons that have been completed have been predominately limited to single-center retrospective reviews with limited sample sizes [25, 35–37]. Larger, multi-center comparative data are necessary in order to validate MSA as an alternative surgical treatment in the management of chronic GERD.

The aim of the present study was to validate MSA as an anti-reflux procedure through comparison of perioperative and clinical outcomes following MSA to NF in clinical practice, in a standardized fashion, for a larger number of patients than previously published.

Materials and methods

From April 6, 2007, to December 12, 2014, three high-volume esophageal centers participated in a retrospective case–control review of prospectively collected data on patients who underwent either MSA or NF for the treatment of chronic GERD. The institutional review board at each center approved the study. Since the magnetic sphincters were placed as part of clinical care, the need to implant the devices under a research protocol was waived. Standard informed consent was provided for surgical intervention; however, individual patient consent for this study was waived because of the study's retrospective nature.

Patients were identified in each center's database and included if they met the inclusion and exclusion criteria and the patient was eligible for either MSA or NF during the study time period. Patients included in the NF group were often eligible for MSA implantation, but excluded from MSA because of insurance denial, MRI requirements, patient preference, or known allergy to titanium, stainless steel, nickel. Patients were included if they were greater than 18 years and less than 85 years, had a documented history of GERD at least partially responsive to PPI treatment, and positive pH testing. Patients were excluded if they had prior gastric or esophageal surgery, a hiatal hernia greater than 3 cm in size, esophageal dysmotility (as defined by manometry findings demonstrating effective swallows <70 % and/or, distal esophageal amplitude of <35 mm Hg), and the presence of endoscopically visible Barrett's or esophageal stricture.

A total of 455 patients (222 MSA and 233 NF) were identified as having undergone surgical intervention for GERD. From this, 21 MSA and 19 NF patients were excluded from analysis as a result of inadequate follow-up data. Ultimately, 415 patients (201 MSA and 214 NF) were selected for comparative analysis, and a subgroup of 354 patients (169 MSA and 185 NF) with a minimum of 1-year follow-up were used to validate clinical outcomes. When the preliminary analysis demonstrated differences between the groups, a propensity analysis was completed matching patients based upon preoperative esophagitis, presence of microscopic Barrett's, hiatal hernia size, Hill grade and BMI (Fig. 1).

Data collected from the medical record included: patient demographic information including age, gender, and body mass index (BMI). Results from the preoperative evaluation included: barium swallow; endoscopic findings including Hill classification, the absence or presence and grade of esophagitis according to the Los Angeles (LA) classification, the absence or presence and size of hiatal hernia, measured from the top of the rugal folds to the diaphragmatic impressions; pH analysis with a 48-h wireless probe or 24-h impedance pH catheter, with the highest score during a 48-h period evaluation used for the DeMeester score and percentage of time that pH was less than 4; high-resolution manometry; and clinical symptom severity as measured with the Gastroesophageal Reflux Disease Health-Related Quality of Life (GERD HRQL) scale. Surgical outcomes included operative time and hiatal closure. Postsurgical data included length of stay, 30 day major and minor complications and need for explant, revision or postoperative dilation. The evaluation of treatment effect was done by comparing postsurgical to presurgical GERD-HRQL scores, postoperative ability to belch or vomit, PPI use, and patient satisfaction, when available.

Procedure

MSA implantation and NF was completed laparoscopically for all patients. MSA implantation was completed using the LINX Reflux Management System (Torax Medical,

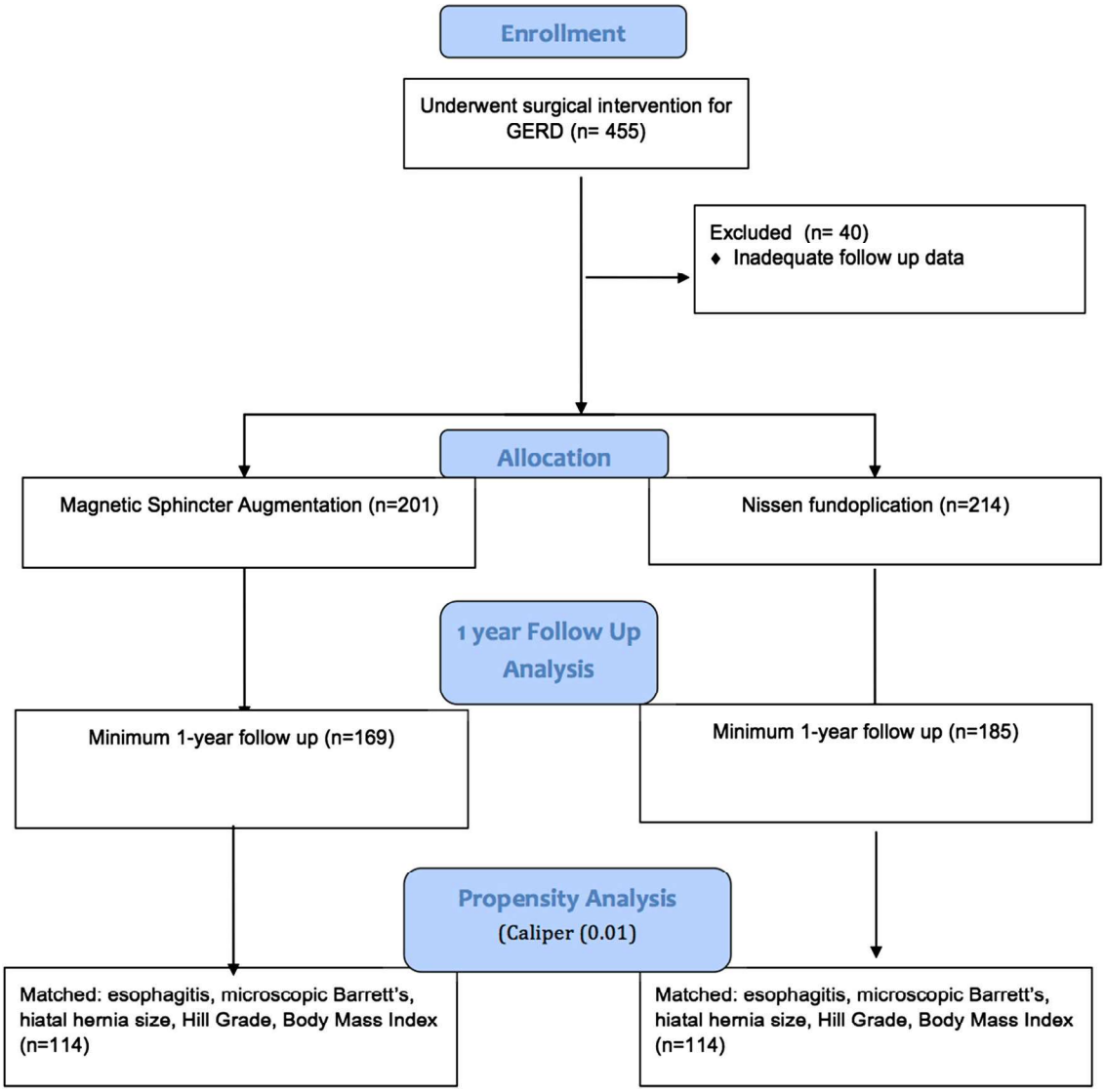


Fig. 1 Patient enrollment distribution

Shoreview Minnesota). The magnetic sphincter is composed of a series of magnets set in a titanium casing and connected by titanium wires, which augments the lower esophageal sphincter and controls reflux by limiting lower esophageal shortening and relaxation during gastric distention [35]. Implantation is completed using 5 ports in a similar configuration traditionally used for NF, with minimal hiatal dissection, posterior closure of the crura with 1–2 sutures if indicated, and preservation of the gastroesophageal junction, specifically the phrenoesophageal ligament, and gastric anatomy. The complete procedure has been previously described and published in detail [32, 35]. The surgical approach for NF was left to the discretion of the individual surgeon; however, all surgeries included the basic tenets of an anti-reflux repair including hiatal dissection and closure if indicated, reestablishment of at least 2 cm of intra-abdominal esophageal length, fundus mobilization with division of the short gastric vessels, and creation of a symmetrical wrap over an appropriately sized Bougie (58–60 Fr.) at the gastroesophageal junction.

Data between groups were compared with the Student's *t* test for continuous variables and the Pearson X [2] test for categorical variables. Differences were considered significant at the $p \leq 0.05$ level. Propensity matching was completed using a caliper of 0.01. Statistical analysis was performed using the SPSS 19 statistical software package.

Results

A total of 415 patients (201 MSA and 214 NF) underwent comparative analysis (Table 1). The groups were similar with respect to age, gender, and GERD-HRQL scores. Patients undergoing MSA had a significantly smaller body

mass index than those undergoing NF and were less likely to report preoperative dysphagia.

During preoperative evaluation (Table 1), MSA patients were found to have lower DeMeester scores, a lower incidence of microscopically identified Barrett's esophagus, but similar rates of esophagitis and percent time pH less than 4. MSA patients were also less likely to have a hiatal hernia, and when present it was likely to be smaller than patients who underwent NF. When the hiatus was assessed by Hill classification, more grade I and 3 valves were present in patients who underwent MSA, while a significantly greater percentage of patients who underwent NF had preoperative Hill grade 4 valves.

The operating time and length of stay were significantly shorter in patients undergoing MSA versus NF (60 vs. 76 min; and 13 vs. 32 h respectively; $p < 0.001$). Statistically significantly more patients had a hiatal closure with NF (19 % MSA vs. 83 % NF, $p < 0.001$).

There were no mortalities, and overall there were no significant differences in 30-day postoperative minor and major morbidities. There was one major complication in a MSA patient, which involved GEJ obstruction and required a return to the operating room for removal of a crural stitch. Three major complications were noted in the NF group including one GEJ obstruction requiring a return to the operating room for fundoplication wrap revision, and two retroesophageal abscesses, associated with biologic mesh placement and bioglue, necessitating surgical drainage (Table 2).

Two patients had removal of their magnetic sphincter, and two patients had revision of their NF. The two fundoplication revisions were for recurrence of hiatal hernia with symptomatic GERD. With respect to the MSA explants, one patient was converted from MSA to NF at 13 months postoperatively for failure to control reflux, as evidenced by positive pH testing and persistent clinical symptoms; and one patient had a device erosion. This patient presented with dysphagia 20 months after implantation. An initial esophagogastroduodenoscopy (EGD) showed no evidence of erosion or other abnormality. A second EGD, performed 30 days later for persistent dysphagia, demonstrated a portion of the magnetic sphincter within the esophageal lumen. This was removed by cutting the exposed magnetic beads with an endoscopic Endoloop Cutter (Olympus Medical Systems, Center Valley, PA, USA). Serial endoscopies demonstrated complete healing of the erosion without any further complication. The patient elected to have the remainder of the device removed laparoscopically 90 days later and has since had a full recovery with no significant complications.

A total of 354 patients (169 MSA and 185 NF) had a minimum of 1-year follow-up, with a 12-month median

Table 1 Patient demographics and preoperative characteristics[a]

	MSA ($n = 201$)	NF ($n = 214$)	P value
Age (years)	54 (42–64)	52 (43–64)	0.76
Gender	52 %M, 48 %F	43 %M, 57 %F	0.06
BMI > 30 (%)	32	40	0.05
GERD-HRQL	21 (15–25)	19 (15–25)	0.56
Preoperative dysphagia (%)	27	39	0.008
% time pH < 4	10 (6–15)	11 (7–16)	0.20
DeMeester Score	34 (21–51)	39 (27–56)	0.03
Esophagitis LA class (%)			
None	63	59	0.48
A	18	15	0.42
B	13	14	0.72
C	4	8	0.14
D	2	4	0.3
Barrett's esophagus (%)	18	31	0.001
Hiatal Hernia present (%)	55	69	0.002
Hiatal Hernia size (cm)	1 (0–2)	2 (0–2)	<0.001
Preoperative hill grade (%)			
I	7	1	0.001
II	19	19	0.09
III	42	29	0.02
IV	32	51	<0.001

[a] All data, unless otherwise specified, expressed as median values with (interquartile range)

Table 2 Adverse events

	MSA (%) ($n = 201$)	NF (%) ($n = 214$)	P value
Major complications (30 days)	0.5	1.4	0.34
Minor complications (30 days)	7	9	0.49
Explant/revision	1	0.9	0.66

duration of follow-up for both MSA and NF. Both groups reported significant improvement in GERD-HRQL scores with no significant difference in the postoperative GERD-HRQL scores between the MSA and NF patients (3 MSA vs. 4 NF, $p = 0.17$)(Fig. 2).

Patients who underwent MSA were significantly more likely, than those undergoing NF, to retain the ability for eructation and emesis. Additionally, the patients who underwent MSA were less likely to experience gas bloat. The incidence of moderate or severe dysphagia was similar between the 2 groups, while MSA patients had significantly higher incidence of mild dysphagia ($p = 0.02$)(Table 3).

An equal distribution of MSA and NF patients resumed daily PPI use, reported satisfaction with the procedure (85 % MSA vs. 91 % NF, $p = 0.09$) and likelihood of undergoing the procedure again (90 % MSA vs. 89 % NF, $p = 0.75$).

Propensity analysis (Table 4) identified 114 matched pairs with similar preoperative esophagitis, presence of microscopic Barrett's, hiatal hernia size, Hill grade and BMI. The mean follow-up was 11 months for MSA and 16 months for NF ($p < 0.001$). There was no significant difference between the matched pairs with respect to postoperative GERD-HRQL scores, but postoperative daily PPI use was higher in the MSA group. Patients who underwent MSA were more likely to have mild postoperative dysphagia, less likely to experience gas bloat, and had greater retention of the ability for postoperative vomiting and eructation. Although patient-reported satisfactions with the procedures were similar, patients undergoing MSA were more likely to report they would undergo the same procedure again.

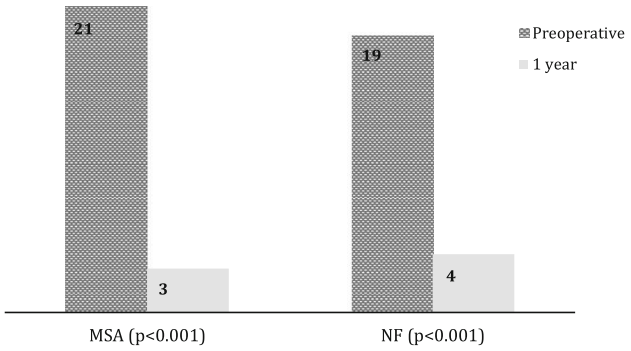


Fig. 2 Preoperative and postoperative 1-year HRQL scores in patients who underwent magnetic sphincter augmentation (MSA) and Nissen fundoplication (NF)

Discussion

The primary finding in this large multi-institutional study is that MSA achieves excellent symptom resolution as measured by the GERD-HRQL questionnaire. When compared to patients undergoing NF, the GERD-HRQL is similar in patients who have achieved at least 1-year follow-up and in propensity-matched patients. MSA also appears to maintain normal physiologic function of the LES as evidenced by the ability to belch and vomit with less gas/bloat symptoms. The incidence of dysphagia is also similar between the two groups. These results add to an increasing knowledge base and experience with MSA, while the comparative data against NF provide validation of MSA as an addition to the surgical management of GERD. Lastly, these results also provide insight into the role MSA will play in the management of GERD.

The results of MSA in this study are similar to those from previously published studies [16, 24, 25, 30–34]. Under strict inclusion criteria, a median GERD-HRQL score of 2 was achieved after MSA in the initial report on outcomes [31]. The minor differences in GERD-HRQL scores in the current series, when compared to previous studies, are likely due to the slightly relaxed inclusion criteria (e.g., patients included with grade C + esophagitis) as more centers are implanting the device and more experienced LINX surgeons are becoming comfortable with the outcomes of the device. Despite this expansion, there has been very consistent symptom relief measured with MSA.

The comparative data illustrate the significant differences in side effect profiles of these two procedures. The inability to belch and vomit is often raised by patients as a concern when contemplating NF [7, 8]. MSA appears to retain that ability as well as having less gassy or bloating sensations which is theoretically due to the dynamic nature of the device allowing gastric venting by way of transient relaxations. Some degree of dysphagia is expected with a NF and is anticipated with post MSA about 2 weeks into the healing process. We were surprised to see that at 1 year, MSA patients appeared to experience a slightly higher rate of "mild" dysphagia compared to propensity-

Table 3 Quality of life at 1 year

	MSA (%) ($n = 169$)	NF (%) ($n = 185$)	P value
Ability for eructation	96	69	<0.001
Ability for emesis	95	43	<0.001
Gas bloat			
None	53	41	0.03
Mild	27	40	0.02
Moderate	14	16	0.65
Severe	5	3	0.24
Dysphagia			
None	42	53	0.04
Mild	44	32	0.03
Moderate	13	11	0.57
Severe	1	5	0.24
Postoperative PPI	19	14	0.18

Table 4 Quality of life: propensity-matched analysis

	MSA (%) ($n = 114$)	NF (%) ($n = 114$)	P value
GERD-HRQL	6	5	0.54
Postoperative PPI	24	12	0.02
Ability for eructation	97	66	<0.001
Dysphagia			
None	42	53	0.31
Mild	44	32	0.04
Moderate	13	11	0.15
Severe	1	5	0.55
Gas bloat			
None	59	41	0.008
Mild	30	42	0.08
Moderate	9	16	0.09
Severe	2	1	0.58
Ability for emesis	88	40	<0.001
Ability for eructation	97	66	<0.001
Satisfaction	88	89	0.61
Would undergo procedure again	93	83	0.01

matched NFs. Our experience suggests this difference is related to several postoperative differences. Patients undergoing NF learn to eat over time with a graduated diet and learn to slow down ingestion early on to adjust to new anatomy, whereas MSA patients eat regular food immediately after surgery and have yet to modify ingestion but when instructed to slow down their eating pattern immediately obtain relief of this mild dysphagia.

The PPI elimination rate of 76 % is slightly lower than reported by previous studies which has shown PPI elimination rates of 81–85 % in multiple studies some of which have long-term follow-up out to 5 years or more [16, 24, 25, 31, 32, 37]. Comparatively, studies with longer-term follow-up with NF show a steady resumption of PPIs beginning at 1 year [38]. What is interesting is that despite a higher need for PPIs postoperatively, patients who underwent MSA were equally satisfied with their outcome but more likely to report that they would undergo the procedure again. Presumably, in this group of uncomplicated GERD patients, MSA has struck a reasonable balance between symptom control and the need for PPIs while lessening the troublesome side effects related to NF that patients and referring MDs are concerned about. The result of this balance is that it likely will increase the number of patients undergoing anti-reflux surgery since 40 % of patients on PPIs still have troublesome symptoms but because of the side effects of NF were waiting on the sidelines until something "better" came along.

Even though entry into this trial was restricted so that patients were required to be eligible for both procedures, it

is apparent there are subtle preoperative differences in these two populations with more Barrett's, as well as more and larger hiatal hernias in patients undergoing NF. This suggests that there may have been a subtle selection bias in which MSA is being applied versus in whom NF is being offered. For most surgeons, the patients with symptomatic GERD without a hernia or esophagitis represents a departure from the typical patient referred to most practices for NF who usually have Barrett's, a moderate to large hiatal hernia, are overweight and a defective LES with bipositional reflux. Along the spectrum of GERD, the patients with complicated GERD, a larger hiatal hernia, erosive esophagitis, esophageal strictures, endoscopically visible Barrett's esophagus, or motility disorders may be better served by NF [4, 9, 25]. In these patients with advanced disease, the LES is likely to be severely defective or absent and may require reconstruction rather than augmentation. The addition of MSA increases the treatment options for GERD by targeting patients with uncomplicated disease earlier in the disease spectrum that are inadequately managed with medical therapy, but not so severe they are willing to accept the side effects of NF.

This study has several limitations. First, its retrospective nature is subject to inherent biases. Second, it should be noted that all of the centers were high-volume esophageal centers, and high-volume surgeons with extended proficiency in dissection of the diaphragmatic hiatus, performing laparoscopic fundoplication, and early adopters of MSA technology. This may limit the applicability of the results outside of high-volume esophageal centers. Lastly, the lack of an objective postoperative GERD control measure such as postoperative pH is acknowledged. However, prior series including a small comparative trial have documented the ability of MSA to normalize pH and with the consistent results seen across trials would be expected to be similar in the current study [31, 35].

In conclusion, MSA in patients with uncomplicated GERD results in equivalent symptom control, improved quality of life but lower PPI elimination rates when compared to propensity-matched NF cases. MSA had significantly less side effects and patients retained their ability to belch and vomit. MSA offers patients an effective procedure with a better side effect profile and is an alternative surgical treatment option for patients with GERD that is not adequately controlled by PPI's and has not progressed to the point of needing a Nissen fundoplication.

Acknowledgments Dr. Heather Warren's Advanced Gastrointestinal Surgery Fellowship at Swedish Medical Center was supported in part by the Ryan Hill Foundation.

Compliance with ethical standards

Disclosures Drs. Lipham, Taiganides and Louie have received consulting fees from Torax Medical. Drs. Warren, Reynolds, Zehetner, Bildzukewicz, Aye, and Farivar, and Ms. Mickley have no conflicts of interest or financial ties to disclose.

References

1. Dent J, El-Serag HB, Wallander MA, Johansson S (2005) Epidemiology of gastroesophageal reflux disease: a systematic review. Gut 54:710–717
2. Toghanian S, Johnson DA, Stalhammar NO, Zerbib R (2011) Burden of gastroesophageal reflux disease in patients with persistent and intense symptoms despite proton pump inhibitor therapy; a post hoc analysis of the 2007 National Health and Wellness Survey. Clin Drug Investig 31:703–715
3. Ying C (2015) Gastroesophageal reflux disease and non-digestive tract diseases. Expert Rev Gastroenterol Hepatol 9:685–692
4. Lord RVN, DeMeester SR, Peters JH, Hagen JH, Elyssnia D, Sheth CT, DeMeester TR (2009) Hiatal hernia, lower esophageal sphincter incompetence and effectiveness of Nissen fundoplication in the spectrum of gastroesophageal reflux disease. J Gastrointest Surg 13:602–610
5. El-Serag HB (2007) Time trends of gastroesophageal reflux disease: a systematic review. Clin Gastroenterol Hepatol 5:17–26
6. Lagergren J, Bergstrom R, Lindgren A, Nyren O (1999) Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. NEJM 340:825–831
7. Lundell L, Miettinen P, Myrvold HE, Hatlebakk JG, Wallin L, Engstrom C, Julkunen R, Montgomery M, Malm A, Lind T, Walan A, Nordic GERD Study Group (2009) Comparison of outcomes twelve years after antireflux surgery or omeprazole maintenance therapy for reflux esophagitis. Clin Gastroenterol Hepatol 7:1292–1298
8. Galmiche JP, Hatlebakk J, Attwood S, Ell C, Fiocca R, Eklund S, Langstrom G, Lind T, Lundell L (2011) Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD: the LOTUS randomized clinical trial. JAMA 305:1969–1977
9. Stein HJ, Barlow AP, DeMeester TR, Hinder RA (1992) Complications of gastroesophageal reflux disease. Role of the lower esophageal sphincter, esophageal acid and acid/alkaline exposure, and duodenogastric reflux. Ann Surg 216:35–43
10. Finks J, Wei Y, Birkmeyer J (2006) The rise and fall of antireflux surgery in the United States. Surg Endosc 20:1698–1701
11. Katz PO, Zavala S (2010) Proton pump inhibitors in the management of GERD. J Gastrointest Surg 14(S1):S62–S66
12. Kahrilas P, Jonsson A, Denison H, Wernersson B, Hughes N, Howden C (2012) Regurgitation is less responsive to acid suppression than heartburn in patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol 10:612–619
13. Kahrilas PJ (2008) Gastroesophageal reflux disease. NEJM 359:1700–1707
14. Castell DO, Kahrilas PJ, Richer JE, Vakil NB, Johnson DA, Zuckerman S, Skammer W, Levine JG (2002) Esomeprazole (40 mg) compared with lansoprazole (30 mg) in the treatment of erosive esophagitis. Am J Gastroenterol 97:575–583
15. Tauseef A, David RN, William TM (2009) Long term safety concerns with proton pump inhibitors. Am J Med 122:896–903

16. Bonavina L, Saino GI, Bona D, Lipham J, Ganz RA, Dunn D, DeMeester T (2008) Magnetic augmentation of the lower esophageal sphincter: results of a feasibility clinical trial. J Gastrointest Surg 12:2133–2140
17. Wileman SM, McCann S, Grant AM, Krukowski ZH (2010) Medical versus surgical management for gastro-oesophageal reflux disease (GORD) in adults. Cochrane Database Syst Rev 13:332–341
18. Varban OA, McCoy TP, Westcott C (2011) A comparison of preoperative comorbidities and post operative outcomes among patients undergoing laparoscopic nissen fundoplication at high and low volume centers. J Gastrointest Surg 15:1121–1127
19. Peters JH, DeMeester TR, Crookes P, Oberg S, Vos Shoop M, Hagen JA, Bremner CG (1998) The treatment of gastroesophageal reflux disease with laparoscopic Nissen fundoplication. A prospective evaluation of 100 patients with typical symptoms. Ann Surg 228(1):40–50
20. Dellemagne B, Weerts J, Markiewics S, Dewandre JM, Wahlen C, Monami B, Jehaes C (2006) Clinical results of laparoscopic fundoplication at ten years after surgery Surg. Endoscopy 20:159–165
21. Kelly J, Watson D, Chin K, Devitt P, Game P, Jamieson G (2007) Laparoscopic Nissen fundoplication: clinical outcomes at 10 years. J Am Coll Surg 205:570–575
22. Morgenthal C, Shane M, Steval A, Gletsu N, Milam G, Milam G, Swafford V, Hunter JG, Smith CD (2007) The durability of laparoscopic Nissen fundoplication: 11 year outcomes. J Gastrointest Surg 11:693–700
23. Hunter J, Smith C, Branum G, Waring J, Trus T, Cornwell M, Galloway K (1999) Laparoscopic fundoplication failures: patterns of failure and response to fundoplication. Ann Surg 230:595–606
24. Lipham JC, Taiganides PA, Louie BE, Ganz RA, DeMeester TR (2014) Safety analysis of first 1000 patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease. Dis Esophagus. doi:10.1111/dote.12199
25. Riegler M, Sebastian FS, Bonavina L, Ashton D, Horbach T, Kemen M (2014) Magnetic sphincter augmentation and fundoplication for GERD in clinical practice: one year results of a multicenter, prospective observational study. Surg Endosc 29(5):1123–1129
26. Catarci M, Gentileschi P, Papi C, Carrara A, Marrese R, Gaspari L, Grassi GB (2004) Evidence based appraisal of antireflux fundoplication. Ann Surg 23:325–337
27. Wang YR, Dempsey DT, Richter JE (2011) Trends and perioperative outcomes in inpatient antireflux surgery in the United States, 1993–2006. Dis Esophagus 24:215–223
28. Richter JE (2013) Gastroesophageal reflux disease treatment: side effects and complications of fundoplication. Clin Gastroeneterol Hepatol 11:465–471
29. Carlson MA, Frantzides CT (2001) Complications and results of primary minimally invasive antireflux procedures: a review of 10735 reported cases. J Am Coll Surg 193:428–439
30. Ganz RA, Gostout CJ, Grudem J, Swanson W, Berg T, DeMeester TR (2008) Use of a magnetic sphincter for the treatment of GERD: a feasibility study. Gastrointest Endosc 67(2):287–294
31. Ganz RA, Peters JH, Horgan S, Bermelman WA, Dunst CM, Edmondowicz SA, Lipham JC, Lukitich JD, Melvin WS, Oelschlager BK, Schlack-Haerere SC, Smith CD, Smith CC, Dunn D, Taiganides PA (2013) Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 368:719–727
32. Bonavina L, Saino G, Bona D, Sironi A, Lazzari V (2013) One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 17:577–585
33. Bonavina L, DeMeester T, Fockens P, Dunn D, Saino G, Bona D, Lipham J, Bemelman W, Ganz R (2010) Laparoscopic sphincter augmentation device eliminates reflux symptoms and normalizes esophageal acid exposure. One and 2 year results of a feasibility trail. Ann Surg 252(5):857–862
34. Lipham JC, DeMeester TR, Ganz RA, Bonavina L, Saino G, Dunn DH, Fockens P, Bemelman W (2012) The LINX reflux management system: confirmed safety and efficacy now at 4 years. Surg Endosc 26:2944–2949
35. Louie BE, Farivar AS, Shultz D, Brennan C, Vallieres E, Aye RW (2014) Short term outcomes using magnetic sphincter augmentation versus nissen fundoplication for medically resistant gastroesophageal reflux disease. Ann Thorac Surg 98(2):498–504
36. Sheu EG, Nau P, Nath B, Kuo B, Rattner DW (2015) A comparative trial of laparoscopic magnetic sphincter augmentation and nissen fundoplication. Surg Endosc 29:505–509
37. Reynolds J, Zehetner J, Wu P, Shah S, Bildzukewicz Lipham J (2015) Laparoscopic magnetic sphincter augmentation versus laparoscopic Nissen fundoplication: a matched pair analysis of 100 patients. J Am Coll Surg 221(1):123–128
38. Simorov A, Ranade A, Jones R, Tadaki C, Shostrom V, Boilesen E (2014) Long term patient outcomes after laparoscopic anti reflux procedures. J Gastrointest Surg 18(1):157–162

Surg Endosc (2016) 30:3225–3230
DOI 10.1007/s00464-015-4635-6

Charges, outcomes, and complications: a comparison of magnetic sphincter augmentation versus laparoscopic Nissen fundoplication for the treatment of GERD

Jessica L. Reynolds[1] · Joerg Zehetner[1] · Angela Nieh[1] · Nikolai Bildzukewicz[1] · Kulmeet Sandhu[1] · Namir Katkhouda[1] · John C. Lipham[1]

Received: 8 May 2015 / Accepted: 17 October 2015 / Published online: 5 November 2015

Abstract
Background Magnetic sphincter augmentation (MSA) is approved for uncomplicated GERD. Multiple studies have shown MSA to compare favorably to laparoscopic Nissen fundoplication (LNF) in terms of symptom control with results out to 5 years. The MSA device itself, however, is an added cost to an anti-reflux surgery, and direct cost comparison studies have not been done between MSA and LNF. The aim of the study was to compare charges, complications, and outcome of MSA versus LNF at 1 year.
Methods This is a retrospective analysis of all patients who underwent MSA or LNF for the treatment of GERD between January 2010 and June 2013. Patient charges were collected for the surgical admission. We also collected data on 30-day complications and symptom control at 1 year assessed by GERD-HRQL score and PPI use.
Results There were 119 patients included in the study, 52 MSA and 67 LNF. There was no significant difference between the mean charges for MSA and LNF ($48,491 vs. $50,111, $p = 0.506$). There were significant differences in OR time (66 min MSA vs. 82 min LNF, $p < 0.01$) and LOS (17 h MSA vs. 38 h LNF, $p < 0.01$). At 1-year follow-up, mean GERD-HRQL was 4.3 for MSA versus 5.1 for LNF ($p = 0.47$) and 85 % of MSA patients versus 92 % of LNF patients were free from PPIs ($p = 0.37$). MSA patients reported less gas bloat symptoms (23 vs. 53 %, $p \leq 0.01$) and inability to belch (10 vs. 36 %, $p \leq 0.01$) and vomit (4 vs. 19 %, $p \leq 0.01$).
Conclusion The side effect profile of MSA is better than LNF as evidenced by less gas bloat and increase ability to belch and vomit. LNF and MSA are comparable in symptom control, safety, and overall hospital charges. The charge for the MSA device is offset by less charges in other categories as a result of the shorter operative time and LOS.

Keywords Gastroesophageal reflux disease · Magnetic sphincter augmentation · Fundoplication · Charges

Gastroesophageal reflux disease (GERD) is a common problem that affects up to 25 % of the population. Proton pump inhibitors (PPIs) are the standard medical therapy, but they do not prevent reflux and only change the character of the refluxate from acidic to non-acidic or weakly acidic. Furthermore, up to 60 % of patients on appropriate PPI therapy will continue to have breakthrough symptoms. For these patients, the current gold standard surgical treatment is laparoscopic Nissen fundoplication (LNF). However, <1 % of eligible patients pursue surgical treatment due to concerns over the side effects, such as gas bloat, inability to vomit and belch, and concerns about the durability of LNF.

MSA was developed as an additional surgical treatment option specifically for patients with uncomplicated GERD who still have some lower esophageal sphincter (LES) function. A flexible band of interlinked magnetic titanium beads is placed laparoscopically around the gastroesophageal junction (GEJ) with minimal dissection of the hiatus in order to preserve the native LES. The MSA device augments a weak LES, preventing reflux while allowing normal functions such as swallowing, belching, and vomiting.

Presented at the SAGES 2015 Annual Meeting, April 15–18, 2015, Nashville, Tennessee.

✉ John C. Lipham
john.lipham@med.usc.edu

[1] Department of Surgery, Keck Medical Center, University of Southern California, 1510 San Pablo Street, Suite 514, Los Angeles, CA 90033, USA

Previous studies have shown MSA to be safe and effective with PPI elimination rates, pH normalization, and symptom control similar to that reported for LNF [1–7]. However, the device has only been approved for a few years and many payers are hesitant to cover the cost of the procedure over continued concerns about long-term outcomes, potential side effects, and the cost of the device, which averages $5000. However, MSA has been reported to have shorter operative times and shorter lengths of stay than LNF [8], and the difference in total charges between the two procedures has not been studied. Therefore, the aim of our study was to compare the charges incurred for the index admission between MSA and LNF and to compare outcomes, side effects, and complications in this same group of patients.

Methods

This study was approved by the University of Southern California Keck Medical Center Institutional Review Board. We performed a retrospective review of a prospectively collected database of all patients who underwent MSA or LNF from January 2010 to June 2013 at two institutions by a single surgeon (JL). All hospital charges were obtained for the index admission and divided into six broad categories for comparison: (1) total billable supplies, (2) pharmacy and drugs, (3) laboratories, test, and radiology, (4) OR services, (5) anesthesia, and (6) room and board. The MSA device was included under the category of total billable supplies. Physician charges were not included as they are a fixed cost and do not vary between procedure and patient and are billed separately at our institution. Length of stay (LOS) information was obtained from hospital record and was calculated from the time the patient was admitted to the pre-op area until the time the discharge was recorded in the computer. Operative time was obtained from the anesthesia record and was calculated from the recorded surgery start time and surgery end time.

All patients completed a gastroesophageal reflux disease health-related quality of life (GERD-HRQL) questionnaire at 1 year and answered questions regarding PPI use, gas bloat symptoms, ability to belch and vomit, and satisfaction. Charts were reviewed for 30-day complications and for any reoperations in the first year following the procedure.

Surgical techniques

MSA was performed by a single surgeon (JL) at two institutions according to a standard protocol (Table 1). The decision to repair hiatal hernias was made intra-operatively by the operating surgeon. In general, crural closure was performed if a hiatal hernia was visible after a posterior dissection of the hiatus that kept the phrenoesophageal membrane intact anteriorly and laterally. LNF was performed by the same surgeon and at the same two institutions according to the standard protocol (Table 2).

Mesh was not used to reinforce the hiatus as part of our standard LNF.

Postoperative care

MSA patients were transferred to PACU and were immediately given a soft diet in the PACU. Patients were discharged from PACU when able to tolerate a soft diet and instructed to eat at least six small regular meals.

LNF patients were admitted to the surgical ward and started on a clear liquid diet in the morning after surgery. Patients were discharged when tolerating enough clear liquids to no longer need intravenous fluids, usually postoperative day (POD) #1 or #2. They are discharged with instruction to continue a clear liquid diet and to advance to a full liquid diet on POD #4 and continue for 1 week. They

Table 1 MSA protocol

1. Patient is placed in low lithotomy with the surgeon standing between the patient's legs
2. A 5-mm camera port is placed at the umbilicus. A 5-mm working port is placed in the right upper quadrant and a Nathanson liver retractor is placed in the right upper quadrant. An 8-mm working port in placed in the left upper quadrant to allow passage of the MSA device
3. The hepatic branch of the vagus nerve is identified and preserved. The right and left crura are identified and minimally dissected to create a tunnel behind the esophagus
4. If the posterior vagus nerve can be easily identified, it is dissected posterior. However, this step is often omitted in favor of keeping dissection to a minimum
5. Tissue on the anterior esophagus is removed so the MSA device can lie flush to the esophagus
6. The provided sizing device is used to determine the number of beads on the device
7. The MSA device is passed through the 8-mm port and pulled through the retro-esophageal tunnel. If the vagus nerve was dissected, it is placed anterior to the vagus
8. The MSA device is secured using the clasp on the device
9. Ports are removed and the abdomen desufflated

Table 2 LNF protocol

1.	Patient is placed in low lithotomy with the surgeon standing between the patient's legs
2.	A 12-mm camera port is placed above and to the left of the umbilicus. Two 12-mm working port are placed at the bilateral subcostal margins in the mid-clavicular line. A Nathanson liver retractor is placed in the right upper quadrant. A 5-mm port in the left anterior axillary line at the level of the camera port is placed as an assistant port
3.	The plane between the right crus and esophagus is developed and extended to the left crus until complete circumferential dissection of the esophagus is obtained with a large retro-esophageal window
4.	Both vagus nerves are identified and preserved
5.	The esophagus is encircled with a Penrose and the crura are closed
6.	The short gastric vessels are divided to mobilize the fundus
7.	A 52- to 56-F esophageal bougie is passed and the fundus passed behind the esophagus creating a loose, floppy wrap
8.	The fundus is sutured to itself to size the wrap, and then the bougie is removed
9.	Additional sutures are placed in the fundus including the anterior esophageal wall to secure the wrap around the esophagus
10.	Ports are removed and the abdomen desufflated

then advance to a soft diet for 1 week and thereafter switch to a regular diet if tolerated.

Symptom scores

The GERD-HRQL is a disease specific, validated questionnaire consisting of ten questions about reflux symptoms each graded from 0 (least severe) to 5 (most severe) for a total score ranging from 0 to 50 [9]. Patients were instructed to stop taking all acid-reducing medications for at least 5 days prior to completing the questionnaire.

Statistical analysis

All statistical analyses were performed using SPSS v22.0 (SPSS, Inc., Chicago, IL, USA) software. Continuous variables were compared with an unpaired *t* test, categorical values were compared with a Chi-square, and *p* values <0.05 were considered significant.

Results

We identified 119 patients, 52 underwent MSA and 67 underwent LNF. There were more males in the MSA group, but no differences in mean age, mean BMI, mean preoperative GERD-HRQL, presence of hiatal hernia, laryngeal pharyngeal reflux symptoms, Barrett's, or esophagitis LA class B or less between the two groups (Table 3).

There was no significant difference in the total charges with a mean charge of $48,491 for MSA and $50,111 for LNF ($p = 0.506$) (Table 4). Total billable supplies were significantly higher for MSA, $24,552 versus $17,118 for LNF ($p < 0.01$), but this was offset by significantly higher charges for LNF in pharmacy/drugs, laboratories/tests/ radiology, OR services, and room and board (Table 4). MSA had a significantly shorter operative time than LNF, 66 versus 82 min ($p < 0.01$), and shorter LOS, 17 h (range 6–37 h with a median of 13 h and a mode of 7 h) versus 38 h (range 25–83 h with a median of 32 h and a mode of 31 h) ($p < 0.01$).

Table 3 Preoperative characteristics

	MSA $N = 52$	LNF $N = 67$	*p* value
Mean age in years	53	53	0.95
Male gender in %	61.5	46.2	0.04
Mean BMI in kg/m^2	26	27	0.56
Mean preoperative GERD-HRQL	17	19	0.10
+Hiatal Hernia	35 (67 %)	50 (75 %)	0.84
+LPR	22 (42 %)	25 (37 %)	0.34
+Pre-op dysphagia	7 (13 %)	16 (23 %)	0.35
+Barrett's esophagus	16 (31 %)	18 (27 %)	0.41
Esophagitis $\leq$ LA class B	50 (96 %)	62 (93 %)	0.93

We were able to obtain 1-year follow-up data for 48/52 (92 %) of the MSA patients and 59/67 (88 %) of the LNF patients (Table 5). There was no difference in postoperative GERD-HRQL with a mean score of 4 in the MSA group and 5 in the LNF group ($p = 0.47$). Both groups had a significant decrease in PPI use with 85 % of MSA patients and 92 % of LNF patients off PPIs at 1 year ($p = 0.37$). Mild-to-moderate dysphagia was reported by 46 % of the MSA patients and 56 % of the LNF patients, while severe dysphagia was not reported by any MSA patients and 5 % of LNF patients ($p = 0.25$). There were nine MSA patients who required dilation for persistent dysphagia in the MSA group and eight in the LNF group ($p = 0.22$). Significantly fewer MSA patients reported gas bloat symptoms (23 vs. 53 %, $p \leq 0.01$) or an inability to

Table 4 Charges and operative details

	MSA $N = 52$	LNF $N = 67$	p value
Total charges (mean ± SD)	$48,491 ± $16,481	$50,111 ± $17,376	0.50
Total billable supplies	$24,552 ± $8143	$17,118 ± $6187	<0.01
Pharmacy/drugs	$2243 ± $2791	$5453 ± $4196	<0.01
Laboratories/tests/radiology	$1358 ± $1891	$2758 ± $2744	0.01
OR services	$15,849 ± $4398	$18,664 ± $5170	0.01
Anesthesia	$2782 ± $1429	$3224 ± $1519	0.10
Room and board	$2619 ± $2226	$4235 ± $3006	<0.01
OR time (min ± SD)	66 ± 23	82 ± 18	<0.01
Length of stay (h ± SD)	17 ± 10	38 ± 14	<0.01

Table 5 1-year outcomes

	MSA $N = 48/52$ (92 %)	LNF $N = 59/67$ (88 %)	p value
GERD-HRQL (mean ± SD)	4 ± 6	5 ± 5	0.47
Off PPIs	41 (85 %)	54 (92 %)	0.37
Dysphagia			0.25
Mild/moderate	22 (46 %)	30 (56 %)	
Severe	0 (0 %)	3 (5 %)	
Dilation	9 (19 %)	8 (14 %)	0.22
Gas/bloating	11 (23 %)	31 (53 %)	<0.01
Inability to belch	5 (10 %)	21 (36 %)	<0.01
Inability to vomit	2 (4 %)	11 (19 %)	<0.01
Satisfied with procedure	43 (90 %)	54 (92 %)	0.4
Would have procedure again	45 (94 %)	50 (85 %)	0.21

belch (10 vs. 36 %, $p \leq 0.01$) or vomit (4 vs. 19 %, $p \leq 0.01$) than LNF patients. Patients were equally satisfied with their procedure, and 94 % MSA patients and 85 % of LNF patients would have the procedure again ($p = 0.21$).

There were only two 30-day complications, both in the LNF group. One LNF patient was readmitted on POD #3 with intractable vomiting and received intravenous fluid hydration. He was able to tolerate clear liquids and be discharged on hospital day 2. Another LNF patient was readmitted with a food bolus impacted at the GEJ requiring endoscopic removal.

Discussion

This is the first study comparing overall charges for MSA versus LNF. Our study shows that the additional charge for the MSA device itself is completely offset by shorter operative times and LOS for MSA. The shorter LOS by about a day results in fewer laboratory tests, less medication usage of narcotics and anti-emetics, and decreased charges for room and board as most patients are discharged from the recovery room, whereas LNF patients are admitted and stay one midnight. This results in overall charges for MSA being equal to LNF. Previous studies have already shown LNF to be cost-effective compared to lifetime PPIs [10, 11]. As this study shows equivalency in charges between the MSA and the LNF, it is assumable that the MSA is also more cost-effective than lifetime PPI.

MSA was approved by the FDA in 2012, and several studies have confirmed both efficacy, safety, and durability with results out to 6 years, including a safety analysis with more than 1000 patients [1–4, 8]. Despite this, insurance carriers have been hesitant to cover MSA as a surgical option for GERD, either due to concerns over perceived additional costs of the procedure or due to the lack of comparative studies to the traditional surgical treatment for GERD. Several recent comparative studies including this one have now definitively shown that MSA is equally effective as LNF in symptom control and PPI elimination rates [5, 6, 12]. Additionally, in this analysis of MSA versus LNF, we found that MSA patients had less gas bloat symptoms and increased ability to belch and vomit. Similar results were published recently from our group in a matched-pair propensity analysis of MSA versus LNF [12].

The MSA device is meant to augment a near normal LES and provide additional support during temporary failures of the LES, such as during post-prandial gastric distention or transient LES relaxation. MSA is not meant to replace LNF which has been typically used for patients with complicated GERD and complete LES failure. MSA, rather, provides a surgical alternative for patients earlier in the disease progression that have not developed complications and that are not completely controlled by medications.

With up to 60 % of patients on PPIs still having residual symptoms and <1 % of eligible patients seeking surgical treatment, there is a large "gap" population of patients that are seeking an alternative treatment for their GERD. It appears that patient's hesitation in pursuing surgical treatment for their GERD is mainly due to the perceived side effects and complexity of Nissen fundoplication including gas bloat, inability to vomit or belch, and inconsistent results and complications of the procedure in the community. Additionally, these "gap" patients tend to have uncomplicated disease without esophageal strictures, Barrett's esophagus, or higher grades of esophagitis. Their symptoms, however, tend to be only partially controlled by PPIs, but given that their disease has not progressed to the level of complications, they are hesitant to accept the side effects of LNF. MSA has been offered as a treatment option specifically for this gap population.

As there is now a proven alternative to the LNF available, potentially more patients with earlier, uncomplicated GERD will seek a surgical approach to reflux control. By treating patients earlier in the disease, we may be able to stop the progression of this disease and hopefully decrease the incidence of severe reflux complications and the subsequent development of dysplasia and esophageal cancer. MSA should be offered to all patients who have incomplete symptom control with PPIs and are hesitant to pursue LNF before they progress to complicated GERD.

There are two main limitations to this study. First, as the primary outcome was a difference in charges, this study was not adequately powered for our secondary outcomes. Specifically, it was not powered to detect a difference in PPI use <10 %. Since we only showed a 7 % difference in PPI use, we cannot say whether this is significant or not. Second, we compared charges between the two procedures which are not the same as comparing cost. Cost is determined by many factors that are often in flux including vendor contracts, payer contracts, and the mix of payers. We were unable to obtain this information from our institution. Therefore, we looked at charges which are set by the hospital and somewhat artificial, but are constant between patient and procedure such that the charge for a given medication is not affected by the type of insurance, whether it is an outpatient or inpatient procedure, or whether the contract with the vendor changed. This means that although we cannot look at the actual cost and reimbursement to the hospital for the procedures, we can compare resource utilization between them by comparing charges.

Conclusion

There is no significant difference between MSA and LNF in symptom control, PPI elimination rate, and charges. The better side effect profile of MSA should allow for its use earlier in the disease progression of GERD before complications occur. MSA is an alternative treatment option to PPI therapy in GERD patients that have not progressed to the point of having significant complications from their disease.

Compliance with ethical standards

Disclosures Dr. John Lipham is a consultant for Torax Medical. Drs. Zehetner, Bildzukewicz, Crookes, Sandhu, and Katkhouda have nothing to disclose. Angela Nieh has nothing to disclose.

References

1. Bonavina L, DeMeester T, Fockens P, Dunn D, Saino G, Bona D, Lipham J, Bemelman W, Ganz RA (2010) Laparoscopic sphincter augmentation device eliminates reflux symptoms and normalizes esophageal acid exposure: one- and 2-year results of a feasibility trial. Ann Surg 252:857–862
2. Bonavina L, Saino G, Bona D, Sironi A, Lazzari V (2013) One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 217:577–585
3. Ganz RA, Peters JH, Horgan S, Bemelman WA, Dunst CM, Edmundowicz SA, Lipham JC, Luketich JD, Melvin WS, Oelschlager BK, Schlack-Haerer SC, Smith CD, Smith CC, Dunn D, Taiganides PA (2013) Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 368:719–727
4. Lipham JC, Taiganides PA, Louie BE, Ganz RA, Demeester TR (2014) Safety analysis of first 1000 patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease. Dis Esophagus 28(4):305–311
5. Louie BE, Farivar AS, Shultz D, Brennan C, Vallieres E, Aye RW (2014) Short-term outcomes using magnetic sphincter augmentation versus Nissen fundoplication for medically resistant gastroesophageal reflux disease. Ann Thorac Surg 98:498–504 **(discussion 504–495)**
6. Riegler M, Schoppman S, Bonavina L, Ashton D, Horbach T, Kemen M (2014) Magnetic sphincter augmentation and fundoplication for GERD in clinical practice: one-year results of a multicenter, prospective observational study. Surg Endosc 29(5):1123–1129
7. Sheu EG, Nau P, Nath B, Kuo B, Rattner DW (2014) A comparative trial of laparoscopic magnetic sphincter augmentation and Nissen fundoplication. Surg Endosc 29(3):505–509
8. Reynolds JL, Zehetner J, Bildzukewicz N, Katkhouda N, Dandekar G, Lipham JC (2014) Magnetic sphincter augmentation with the LINX device for gastroesophageal reflux disease after U.S. Food and Drug Administration approval. Am Surg 80:1034–1038

9. Velanovich V (2007) The development of the GERD-HRQL symptom severity instrument. Dis Esophagus 20:130–134
10. Goeree R, Hopkins R, Marshall JK, Armstrong D, Ungar WJ, Goldsmith C, Allen C, Anvari M (2011) Cost-utility of laparoscopic Nissen fundoplication versus proton pump inhibitors for chronic and controlled gastroesophageal reflux disease: a 3-year prospective randomized controlled trial and economic evaluation. Value Health 14:263–273
11. Grant AM, Boachie C, Cotton SC, Faria R, Bojke L, Epstein DM, Ramsay CR, Corbacho B, Sculpher M, Krukowski ZH, Heading RC, Campbell MK (2013) Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial). Health Technol Assess 17:1–167
12. Reynolds JL, Zehetner J, Wu P, Shah S, Bildzukewicz N, Lipham JC (2015) Laparoscopic magnetic sphincter augmentation versus laparoscopic Nissen fundoplication: a matched-pair analysis of 100 patients. J Am Coll Surg 221(1):123–128

Value of physiologic assessment of foregut symptoms in a surgical practice

Mario Costantini, MD, Peter F. Crookes, MD, Ross M. Bremner, MD, Sebastian F. Hoeft, MD, Afshin Ehsan, BS, Jeffrey H. Peters, MD, Cedric G. Bremner, MD, *and* **Tom R. DeMeester, MD,** *Los Angeles, Calif.*

Background. *The aim of this study was to evaluate the reliability of symptoms in the diagnosis of gastroesophageal reflux disease and esophageal motility disorders as assessed by functional tests.*
Methods. *In 365 patients referred for suspected esophageal functional disease, symptomatic assessment was compared with the results of esophageal manometry and ambulatory 24-hour pH monitoring of the distal esophagus.*
Results. *Based on the patients' chief complaint, the symptomatic diagnosis was gastroesophageal reflux (44%), esophageal motor disorder (26%), chest pain of esophageal origin (9%), reflux and aspiration (8%), and abdominal pathology (12%). The symptomatic diagnosis was considerably altered by the results of the esophageal function tests: gastroesophageal reflux and motility disorders were found in all symptomatic diagnostic groups and a large number of patients in each group tested normal. The sensitivity and specificity of symptom-based diagnoses for functional disease were low.*
Conclusions. *The results of this study showed that symptoms are an unreliable guide of esophageal abnormality, illustrating the need for objective testing in these patients, particularly to avoid inappropriate medical or surgical therapy. (SURGERY 1993;114:780-7.)*

From the Department of Surgery, University of Southern California School of Medicine, Los Angeles, Calif.

A CAREFUL EVALUATION OF a patient's clinical history has traditionally been a cornerstone in the diagnostic approach to any disease. The symptoms of heartburn, acid regurgitation, and dysphagia have long been linked to esophageal abnormalities.[1] However, it has been noted that esophageal symptoms are generally nonspecific and do not always indicate the correct pathologic condition.[2] Despite this, patients are often treated on a purely symptomatic basis. Furthermore, morphologic studies such as endoscopy and barium swallow only reveal alterations that represent complications of functional disease, such as esophagitis or stricture. Moreover, they are poorly sensitive, and in only a few situations, specific.[3]

With the advent of physiologic testing techniques such as esophageal manometry and 24-hour pH monitoring of the distal esophagus, more precise diagnoses are possible. These tests have helped consistently in the understanding of the pathophysiologic alteration, allowing an objective diagnosis of a motility disorder or gastroesophageal reflux disease (GERD), a prerequisite for correct treatment. They are, however, invasive and require significant expense in terms of equipment and trained personnel.

This study was therefore undertaken to assess the reliability of symptoms in the diagnosis of GERD and esophageal motor disorders in patients referred for evaluation for surgical therapy. Second, we aimed to identify the most specific symptoms, if any, for the diagnosis of GERD and for primary esophageal motor disorders.

PATIENTS AND METHODS

Three hundred and sixty-five consecutive patients referred to our institution for suspected esophageal disease, in whom routine examinations (i.e., endoscopy or barium swallow) were unable to give a satisfactory diagnosis, form the basis of this study. Patients with previous esophageal or gastric surgery and those with malignancies were excluded.

The 205 female and 160 male patients had a median age of 53 years (range, 2 months to 89 years). All patients were studied according to the same protocol. Before esophageal functional tests, patients completed a detailed questionnaire to score esophageal symptoms (heartburn, regurgitation, dysphagia) from 0 to 3

Presented at the Fiftieth Annual Meeting of the Central Surgical Association, Cincinnati, Ohio, March 4-6, 1993.

Reprint requests: Tom R. DeMeester, MD, Department of Surgery, University of Southern California School of Medicine, 1510 San Pablo St., Suite 514, Los Angeles, CA 90033-4612.

0039-6060/93/$1.00 + .10 **11/6/49591**

Table I. Symptom scoring system

Heartburn
0, None
1, Minimal; identifiable symptom; occasional episodes; no prior medical visit
2, Moderate; primary reason for visit
3, Severe; constant marked disability in activities of daily life
Regurgitation
0, None
1, Mild; after straining or large meals
2, Moderate; predictable with position change; straining or lying down
3, Severe; constant regurgitation; presence of aspiration
Dysphagia
0, None
1, Occasionally with coarse food (meat, sandwich, hard roll); lasting for few seconds
2, Requiring clearing with liquids
3, Severe; semiliquid diet; history of meat impaction

Table II. Functional esophageal abnormalities suspected on symptomatic basis and symptom complex classification of patients

Group 1: Gastroesophageal reflux
Heartburn ≥2
Acid regurgitation ≥2 chief complaints
Dysphagia ≤2 may be associated
Group 2: Motility disorder
Dysphagia ≥2 chief complaint
Heartburn ≤2 may be associated
Group 3: Noncardiac chest pain
Chest pain chief complaint
Heartburn ≤2
Dysphagia ≤2 may be associated
Group 4: Aspiration
Asthma
Choking
Wheezing
Coughing chief complaints
Group 5: Gastric pathologic condition
Nausea
Vomiting
Epigastric pain
Bloating chief complaints
Heartburn ≤2 may be associated

according to severity (Table I). Other symptoms were recorded as present or absent. Particular importance was placed on the patients' chief complaint. On the basis of their symptom complex, patients were classified into five groups (Table II). Group 1 had typical *GERD symptoms* of heartburn and acid regurgitation. Mild to moderate dysphagia was present in some patients of this group but was not the chief complaint. Group 2 patients *(Motility disorder)* complained of severe dysphagia suggesting an esophageal motor disorder. In this group mild to moderate heartburn was also sometimes associated but was not the chief complaint. In group 3 *(Aspiration)*, patients complained of chronic respiratory symptoms and were referred to exclude underlying GERD or motor disorder as the basis for their symptoms, whereas group 4 *(Noncardiac chest pain)* patients were referred by cardiologists for investigation of a possible esophageal etiology for their chest pain. Finally, in group 5 *(Abdominal pathologic condition)*, patients' chief complaint suggested an abdominal organ abnormality.

Stationary esophageal manometry was performed according to the standard method previously described.[4] Briefly, a multilumen infused catheter, with 5 side-holes located 5 cm apart and radially oriented, was used. After confirmation of the gastric pressure pattern, the lower esophageal sphincter (LES) was evaluated by a station pull-through technique, in which the catheter was withdrawn across the cardia at 1 cm steps. The LES pressure at the respiratory inversion point in midinspiration, the overall length, and the length of the LES segment below the respiratory inversion point (i.e., the abdominal length) were measured.[4] The evaluation of the esophageal body was performed by positioning the most proximal side-hole 1 cm below the lower border of the upper esophageal sphincter, thus having the others 6, 11, 16, and 21 cm below, incorporating the entire esophageal body. Ten dry and ten wet (5 ml of room temperature water) swallows were performed at 20-second intervals. Pressure data were stored in a computer and analyzed by dedicated software. The test was completed with the evaluation of pressure, length, and function of the pharyngoesophageal sphincter. Fifty healthy volunteers studied with the same protocol formed the control group. Classification of motility patterns followed the criteria summarized in Table III.[5]

Twenty-four-hour pH monitoring of the distal esophagus was performed with a bipolar glass electrode located 5 cm above the upper border of the manometrically located LES, following our standard protocol, and gastroesophageal reflux pattern was quantitated with a composite score as previously described.[6] Patients whose composite score exceeded the 95th percentile of values obtained in the population of 50 healthy volunteers were considered to have GERD.

Statistical analysis was performed with a commercially available package (SAS 6.04; SAS Institute, Cary, N.C.). Chi-squared test was used to test proportions among groups, and the Kruskal-Wallis test was used to compare means. The sensitivity, specificity, and accuracy of symptomatic assessment were calculated.[7]

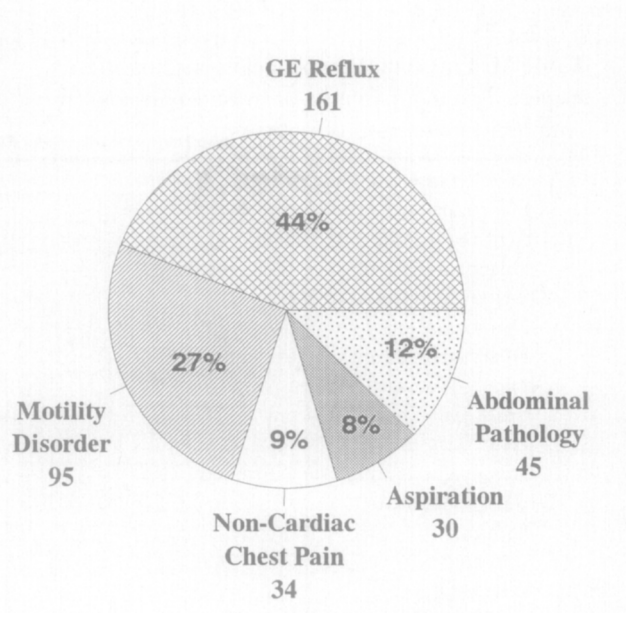


Fig. 1. Distribution of patients in five symptomatic groups.

RESULTS

Fig. 1 graphically displays the distribution of patients in the 5 groups according to their symptom complex. Table IV shows the relationship between proven esophageal function abnormalities and the abnormalities suspected on a symptomatic basis.

One hundred and sixty-one of the referred patients (44%) had symptoms suggestive of GERD. One hundred and five of them (65.2%) had objective evidence of reflux on 24-hour pH monitoring. In 11 patients whose pH monitoring showed normal esophageal exposure to gastric juice, an esophageal motor disorder was detected at manometry. One patient had diffuse esophageal spasm (DES) and 10 had nonspecific esophageal motor disorders (NEMD).

Ninety-five (26%) patients presented with dysphagia as their chief complaint, thus suggesting an esophageal motor disorder (group 2). The suspected esophageal motor disorder was confirmed by stationary manometry in only 35 (36.8%): six had classic achalasia, five had DES, six had nutcracker esophagus (NE), two had hypertensive LES, two had cricopharyngeal incoordination, and 14 had NEMD. Notably, 24-hour pH monitoring revealed the presence of GERD in a further 31 (32.6%) of the patients classified in this group.

Twenty-six patients with suspected noncardiac chest pain (76.5%) had evidence on function tests of an esophageal abnormality: 17 patients (50%) exhibited a primary esophageal motor disorder (two DES, five NE, and 10 NEMD), and nine (26.5%) had evidence of GERD on pH monitoring.

Only 12 (40%) of patients suspected of having chronic aspiration secondary to reflux had objective evidence of pathologic reflux on pH testing. An underlying esophageal motor disorder was the cause of symptoms in five of the other 18 patients (17% of the whole group).

In the group of patients with symptoms suggesting an abdominal abnormality, GERD was found on pH monitoring in 14 patients (31%). A further 12 patients (27%) in this group had an unsuspected esophageal motor disorder (one DES, three NE, eight NEMD).

Overall, in the whole group of 365 patients, 24-hour pH monitoring objectively revealed the presence of GERD in 171 (46.8%). Of this group only 105 (61.4%) presented with typical symptoms of heartburn and acid regurgitation and were correctly identified on a symptomatic basis (Fig. 2, *A*). The sensitivity of symptom-based diagnosis was therefore 61.4% and the specificity 58.6%, with an accuracy of 50.4%.

On the other hand, 80 patients (22%) had a primary esophageal motor disorder diagnosed manometrically: 52 (65%) were suspected of having the disorder on a symptomatic basis either because of dysphagia (35, 44%) or chest pain (17, 21%) (Fig. 2, *B*). The sensitivity of symptom-based diagnosis in this group of patients was therefore 65% and the specificity 73%, with an accuracy of 71.2%.

Table III. Criteria for manometric diagnosis

Achalasia
Incomplete LES relaxation (≤75% relaxation)
100% aperistalsis in esophageal body
Elevated LES pressure (≥26 mm Hg)*
Increased intraesophageal baseline pressure relative to gastric baseline*
Diffuse esophageal spasm
Simultaneous (nonperistaltic contractions) (>20% of wet swallows)
Repetitive and multipeaked contractions
Spontaneous contractions
Intermittent normal peristalsis
Contractions may be of increased amplitude and duration
Normally relaxing LES
NE
Normally relaxing LES
Normal peristalsis in esophageal body
Contractions exceeding 180 mm Hg in distal esophagus
Hypertensive LES
Elevated LES pressure (≥26 mm Hg)
Normal LES relaxation
Normal peristalsis in esophageal body
Nonspecific esophageal motility disorders
Decreased (<30 mm Hg) or absent amplitude of esophageal peristalsis
Increased number of nontransmitted contractions
Abnormal wave forms

*Not absolutely required.
(Modified from Stein HJ, DeMeester TR, Hinder RA. Curr Probl Surg 1992;24:418-555.)

Table IV. Relationship between proven esophageal function abnormalities and abnormalities suspected on symptomatic basis

		Detected abnormality		
Suspected abnormality	*n*	*GERD n (%)*	*Motility disorder n (%)*	*No esophageal disease n (%)*
GERD	161	105 (65.2)	11 (6.8)	45 (28)
Motility disorder	95	31 (32.6)	35 (36.8)	29 (31)
Noncardiac chest pain	34	9 (26)	17 (50)	8 (24)
Aspiration	30	12 (40)	5 (17)	13 (43)
Abdominal pathologic condition	45	14 (31)	12 (27)	19 (42)
TOTAL	365	171	80	114

Table V. Prevalence of separate symptoms in patients with positive 24-hour pH monitoring (GERD), in patients with manometrically detected esophageal motility disorder, and patients with no evidence of esophageal disease

	24-hour pH proven GERD %	*Manometrically detected motor disorder %*	*No esophageal disease on function tests %*
Heartburn 2-3	69	31	51*
Regurgitation 2-3	49	40	50
Dysphagia 2-3	51	58	44
Cumulative symptom score >3	68	38	58.4*
Chest pain	30	44	34
Cough	25	15	12
Nausea	8	10	11
Epigastric pain	22	15	25
Vomiting	9.4	5	10

*$p < 0.001$, chi-squared test.

The prevalence of separate symptoms related to the results of functional tests is shown in Table V. The only symptom that was significantly different in prevalence among the three groups was heartburn (moderate to severe), which was present in nearly 70% of patients with GERD (sensitivity, 69%). The symptom was also present in one third of patients with a motor disorder and in one half of the patients with no detectable abnormalities on functional tests (specificity, 56.7%). Moderate to severe dysphagia and chest pain were slightly more frequent in patients with a manometrically detectable motor disorder, but there was no statistical difference in their prevalence with respect to the other groups of patients. Other symptoms were equally present in all of the patients, irrespective of whether they had esophageal abnormalities.

Finally, by using our scoring system for grading typical esophageal symptoms (Table I), patients with proven GERD, as expected, had a higher mean score for heartburn than the other patients (1.8 ± 0.1 vs 1.0 ± 0.1 and 1.4 ± 0.1, respectively, i.e., for patients with motor disorders and patients without esophageal abnormalities; mean ± SEM; $p < 0.05$) (Fig. 3). The mean cumulative symptom score (obtained by adding the score of each symptom) was also higher in this group of patients ($p < 0.05$). Table V shows that a cumulative symptom score greater than 3 was more apt to be present in a patient with proven GERD ($p < 0.001$); however, a score >3 was also present in 38% of patients with a manometrically detected motility disorder and in 58.4% of patients without any detectable esophageal abnormality.

DISCUSSION

It is often possible to suspect GERD or an esophageal motility disorder by a careful history, but objective evidence of these disorders is essential in their management, particularly if surgery is contemplated. Most such patients are treated by the primary physician without objective tests, and only persistent symptoms lead to referral of these patients for further investigation. Failure to establish the cause of the symptoms before operation can result in an unfavorable result. There are already reports of performing an antireflux operations when the

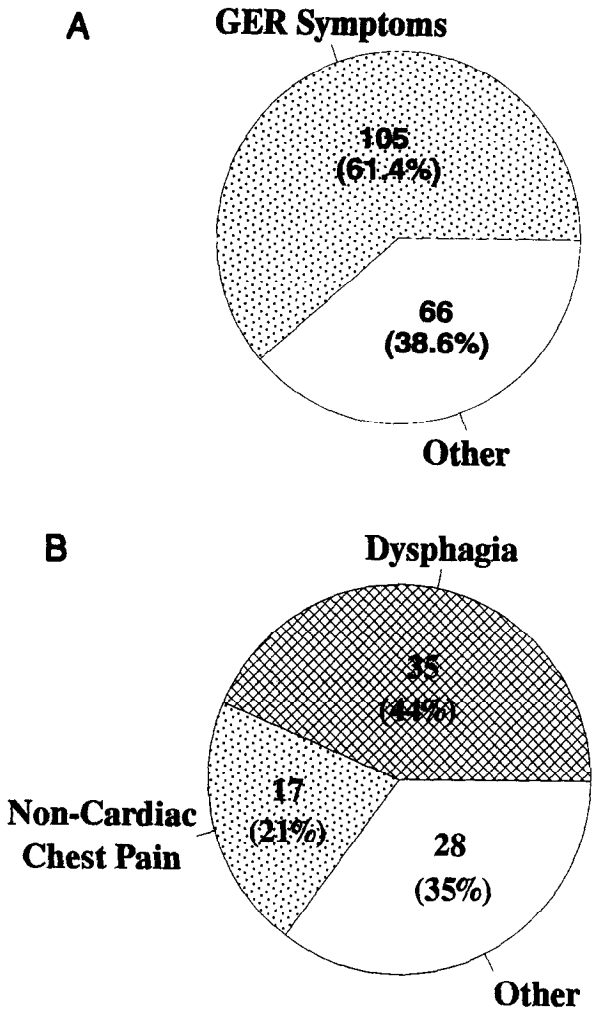


Fig. 2. A, Prevalence of GER symptoms in patients with 24-hour pH proven reflux (n = 171). **B,** Prevalence of symptoms suggestive for motility disorders and chest pain in patients with manometrically proven motility disorder (n = 80).

underlying problem was achalasia but mistakenly thought to be GERD.[8] Extensive clinical experience has shown that 24-hour esophageal pH monitoring has the highest sensitivity and specificity for the detection of increased esophageal acid exposure, the hallmark of GERD.[9] Esophageal manometry still represents the basis for the current classification of esophageal motility disorders,[10] although ambulatory motility may prove to be more accurate for this purpose.[11, 12]

Our results clearly illustrate the inadequate sensitivity and specificity of the cardinal symptoms related to esophageal disease. In particular, 114 patients (31%) who were tested had neither GERD nor a motility abnormality. This large group of patients with symptoms and negative test results require further evaluation to exclude gastric, hepatobiliary, pancreatic disease, or irritable bowel syndrome[13] before starting any medical or surgical therapy.

The results also highlight the value of physiologic tests in making a diagnosis of unsuspected esophageal motility disorders, although this has been questioned by some authors.[14] More than one third of detected esophageal motility disorders were found in patients with symptoms suggestive of GERD or abdominal disorders. In another study of 100 patients complaining of dysphagia in whom barium roentgenogram and endoscopy were normal, 89 were shown to have a motility disorder.[15] Treatment or dismissal of these patients without a motility study would have been inappropriate.

The symptoms evaluated in this study are known to be very prevalent in the general population, as demonstrated by Ruth et al.[16] in a Scandinavian population. These authors found symptoms suggestive of GERD in 25%, chest pain in 13%, and dysphagia in 10% of the interviewed subjects. Moreover, in that study chronic cough and breathing difficulties were significantly associated with symptoms suggestive of GERD. In this country it has been estimated that heartburn occurs daily in about 10% of the population and that intermittent symptoms can be present in one third of the general population.[1] Starting phase I therapy (antacids and elevation of the head of the bed) in these patients is probably expedient, but escalating therapy to phase 2 (H_2 blocker and proton pump inhibitors) without proper testing would, on the basis of our findings, be very cost ineffective.

Calculation of sensitivity and specificity of esophageal symptoms in GERD was reported by Klauser et al.[17] These authors compared symptoms of 304 patients referred for suspected GERD with the results of ambulatory pH monitoring. They found that the only symptoms appreciably different in frequency between patients with abnormal and patients with normal acid exposure of the distal esophagus were heartburn and acid regurgitation. These symptoms had a sensitivity for a correct diagnosis of GERD of 73% and 66%, respectively, but a low specificity (53% and 59%) was present also in nearly one half the patients without the disease. If these symptoms clearly dominated the patients' complaints, the specificity was higher, but the sensitivity was then much lower. Use of sophisticated statistical techniques such as discriminant analysis did not improve the relationship between symptoms and objective diagnosis. Our own results confirm the conclusions of Klauser et al., in that the only symptom significantly different in prevalence between patients with and without documented GERD was heartburn of grade 2 or more. The sensitivity (69%) and specificity (56.7%) of heartburn in the diagnosis of GERD in our own patients were also very similar to those reported by Klauser et al.

Ambulatory pH monitoring was used by Johnson et

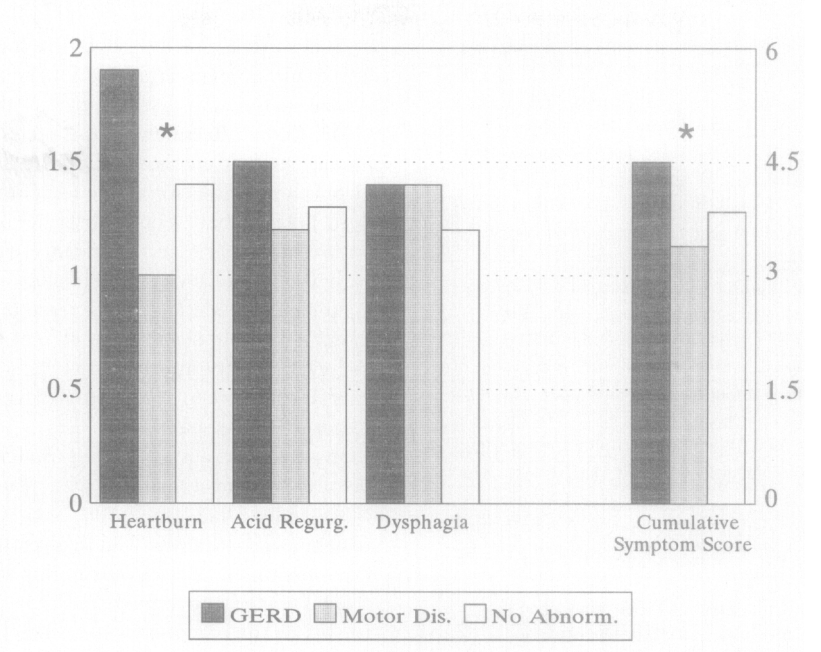


Fig. 3. Comparison of mean scores for heartburn, acid regurgitation, and dysphagia, and mean cumulative symptom score among patients who had 24-hour pH proven GERD, a manometrically proven motility disorder, or no abnormality on function tests. Significantly higher heartburn and cumulative score were found in patients with 24-hour pH proven GERD compared with other groups. $*p < 0.05$.

al.[18] in 220 patients to evaluate the ability of symptoms and endoscopic findings to establish a diagnosis of reflux disease.[18] Daily occurrence of heartburn or acid regurgitation had a positive predictive value of 59% and 66%, respectively. Their study confirms our findings that it is difficult to establish a diagnosis of GERD by patient history alone.

In an attempt to refine the diagnosis of GERD based on symptom assessment, we also graded symptoms in our patients according to severity. Patients with proven GERD had a higher mean score for heartburn and a higher mean cumulative symptom score than patients without the disease. A cumulative symptom score for heartburn regurgitation and dysphagia greater than 3 was statistically more prevalent in patients with GERD than in those without, but this approach also would have incorrectly diagnosed 50% of patients without the disease. Categorization by symptom groups also proved to be inaccurate, because 38.6% of patients with pH-proven GERD presented with symptoms that did not suggest the disease, and 35% of patients with a manometrically detected motility disorder were not suspected of having such a disorder.

It may be argued that endoscopy provides sufficient information to render physiologic tests unnecessary. It must be borne in mind that endoscopy is more expensive than esophageal function test and is useful only in detecting complications of reflux disease. In this situation it has high specificity but low sensitivity.[3] In addition, it is not of value in assessment of esophageal motor function.

In conclusion, there are no symptoms sufficiently sensitive and specific for esophageal functional disease. A symptomatic approach in these patients is therefore an unreliable guide and results in an incorrect diagnosis in 25% to 35% of patients. The value of a careful history is not primarily to form the basis for treatment but rather to guide the physician to perform functional assessment. Documentation of esophageal disease is necessary to make an accurate diagnosis before commencing phase II or surgical intervention.

We thank Suzanne Talbot for her expert assistance in the preparation of the paper.

REFERENCES

1. Nebel OT, Fornes MF, Castell DO. Symptomatic gastroesophageal reflux: incidence and precipitating factors. Am J Dig Dis 1976;21:953-6.
2. DeMeester TR, Bonavina L, Iascone C, et al. Chronic respiratory symptoms and occult gastroesophageal reflux. Ann Surg 1990;211:337-45.
3. Richter JE, Castell DO. Gastroesophageal reflux. Pathogenesis, diagnosis, and therapy. Ann Intern Med 1982;97:93-103.
4. Zaninotto G, DeMeester TR, Schwizer W, et al. The lower esophageal sphincter in health and disease. Am J Surg 1988; 155:104-11.
5. Stein HJ, DeMeester TR, Hinder RA. Outpatient physiologic testing and surgical management of functional foregut disorders. Curr Probl Surg 1992;24:418-555.
6. DeMeester TR, Wang CI, Wernly JA, et al. Technique, indi-

cations and clinical use of 24-hour esophageal pH monitoring. J Thorac Cardiovasc Surg 1980;79:656-70.
7. Vecchio TJ. Predictive value of a single diagnostic test in unselected populations. N Engl J Med 1966;274:1171-3.
8. Hocking MP, Ryckman FC, Woodward ER. Achalasia mimicking peptic esophageal stricture. Am Surg 1985;51:563-6.
9. Fuchs KH, DeMeester TR, Albertucci M. Specificity and sensitivity of objective diagnosis of gastroesophageal reflux disease. SURGERY 1987;102:575-80.
10. Castell DO, Richter JE, Dalton CB. Esophageal motility testing. New York: Elsevier Science Publishing Co Inc, 1987.
11. Stein HJ, DeMeester TR, Eypasch EP, Klingman RR. Ambulatory 24-hour esophageal manometry in the evaluation of esophageal motor disorders and noncardiac chest pain. SURGERY 1991;110:753-63.
12. Richter JE, Hackshaw BT, Wu WC. Edrophonium: a useful provocative test for esophageal chest pain. Ann Intern Med 1985;103:14-21.
13. Costantini M, Sturniolo GC, Zaninotto G, et al. Altered esophageal pain threshold in irritable bowel syndrome. Dig Dis Sci 1993;38:206-12.
14. Meshkinpour H, Glick ME, Sanchez P, Tarvin J. Esophageal manometry: a benefit and cost analysis. Dig Dis Sci 1982; 27:772-5.
15. Bremner CG, Wilson M. Value of oesophageal motility with attached pH probe in undiagnosed dysphagia. [Abstract]. S Afr J Surg 1987;25:113.
16. Ruth M, Mansson I, Sandberg N. The prevalence of symptoms suggestive of esophageal disorders. Scand J Gastroenterol 1991; 26:73-81.
17. Klauser AG, Schindlbeck NE, Müller-Lissner SA. Symptoms in gastro-oesophageal reflux disease. Lancet 1990;335:205-8.
18. Johnsson B, Joelsson K, Gudmundsson K, Greiff L. Symptoms and endoscopic findings in the diagnosis of gastroesophageal reflux disease. Scand J Gastroenterol 1987;22:714-8.

DISCUSSION

Dr. Philip E. Donahue (Chicago, Ill.). This excellent presentation amplifies and reinforces observations about esophageal disease. The take-home message that foregut disorders require a complete evaluation, especially before surgical intervention, is quite timely.

First, you did not present any data on gastric emptying studies or a systematic exclusion of patients with gastric outlet or duodenal pathologic condition, which might lead to secondary gastroesophageal reflux symptoms. It is essential that a referral clinic such as yours define such patients clearly, because once they go back to their referring clinic all of their treatments might hinge on your opinions. So, did you study and exclude patients to find gastric disorders?

Second, what about patients who had a negative pH study? When we have looked at patients with reflux we have found that 20% with documented gastroesophageal reflux have a normal pH study on a given day. One third of the patients in your two major categories had a normal pH study and were assigned to the grouping "No esophageal disease." I think that some of these patients had reflux disease that you may have missed. Would you comment on that?

Last, how frequently did you find that reflux and other motility disorders coexist? Things in the clinic are not as nicely defined as they are in these pie charts. If they did coexist, did you find the treatment of the motility disorder might aggravate the reflux?

This paper makes a very positive statement about the capabilities of a foregut clinic and for the role of experienced clinicians as the most essential part of that clinic.

Dr. Cedric Bremner (Los Angeles, Calif.). You have reminded us that endoscopists and radiologists do not always give us a diagnosis for upper foregut visceral symptoms.

In Johannesburg, South Africa, we had a similar experience with a smaller series in 100 patients referred for dysphagia but with normal endoscopy and radiology studies. We did pH and manometric studies, and our results paralleled these. We found a possible diagnosis in 89% of the 100 cases, and this included motility disorders such as achalasia in two, hypertensive sphincter in two, scleroderma, polymyositis, and diffuse esophageal spasm.

In particular, our achalasia prevalence parallels this study, because you had six patients with achalasia in more than 300 cases, and we had two in 100. Now we are seeing reports of patients with achalasia who inadvertently underwent antireflux operations, and, of course, this is an absolute disaster.

This paper also reminds us that a lot of patients being treated by the practitioner have these vague epigastric or abdominal symptoms and are perhaps being treated inadvertently with H_2 receptor blockers, and patients with respiratory symptoms are being treated inadvertently with bronchial dilators, which would aggravate a reflux situation. In these groups of patients, esophageal studies are extremely important.

Dr. Robert E. Condon (Milwaukee, Wis.). I have been listening to this message from Tom DeMeester and his colleagues for years now. I have heard essentially this data presented half a dozen times. I want to record once again my disagreement with the global conclusion.

There is no question that physiologic testing has an important diagnostic role in some patients with esophageal disease, but my objection is to the uniform application of all of these tests to every patient, because some patients do not need it.

More than a decade ago my colleagues and I reported our experience in a series of more than 400 patients. We segregated our patients into those who had typical symptoms of reflux esophagitis (i.e., they had heartburn aggravated by positional and postcibal change confirmed by the presence of 2+ or worse esophagitis). In those patients the use of physiologic testing did not alter the diagnosis, did not alter the therapy applied, and had no influence on the outcome.

On the other hand, in patients who have additional symptoms or atypical symptoms or who failed to exhibit significant esophagitis on endoscopy, physiologic testing was very helpful in sorting out the presence of significant motility disorders and other forms of esophageal disease.

The bottom line for me is that if you have heartburn, it is worse when you bend over, there are no other symptoms, and you exhibit esophagitis when you undergo endoscopy, you do not need a battery of tests to determine what you need in terms of surgical therapy. You need a fundoplication.

Dr. M. Costantini (closing). Dr. Donahue asked about gastric emptying studies in these patients. Many of our patients were outside referrals from other physicians and were referred only for esophageal tests and pH monitoring. In those patients with negative test results and suspected gastric problems, we recommended further investigation. In Dr. De-

Meester's own practice all patients undergo a complete outpatient foregut monitoring consisting of esophageal motility and esophageal and gastric pH monitoring. Moreover, patients in whom we suspect a gastric emptying problem undergo isotopic gastric emptying studies. At the moment we are also studying and evaluating two new probes, one with a micro-γ camera to monitor gastric emptying in outpatients and one based on fiberoptic technology that looks at bile in the stomach to monitor possible duodenogastric reflux.

I agree that the diagnosis of reflux may be missed in a few patients by 24-hour pH monitoring. But in these patients we have to make sure that omeprazole has been stopped for at least 2 weeks before testing because of its long-lasting action. We also have to make sure that the patients follow their own typical lifestyle during the test. We have to make sure that these patients are not taking medication that can have a caustic effect on the esophageal mucosa, because it has been shown that at least 10% of endoscopic esophagitis is caused not by reflux but by drugs, especially in elderly patients who have a poor acid clearance of the esophagus.

You asked about reflux and motility disorders. Yes, we did observe severe motility disorders associated with esophageal reflux. At least 52 of 171 patients with gastroesophageal reflux exhibited some kind of motor abnormality in the esophagus.

This leads me to the last comment by Dr. Condon. Esophageal function tests are especially necessary in patients with atypical symptoms and no endoscopic esophagitis. We also believe that it is important to understand the physiologic defect before undertaking antireflux operation. The result will warn us of a primary or coexisting motility disorder and poor acid clearance and allow us to modify the antireflux procedure; for example, a Belsey Mk IV procedure may be preferable to a Nissen fundoplication in patients with poor motility because it is an incomplete wrap and less obstructive.

I would like to thank Dr. Bremner for his comment. He also stressed the importance of these tests to make a correct diagnosis of functional disorders of the esophagus.

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 OCTOBER 17, 2019 VOL. 381 NO. 16

Randomized Trial of Medical versus Surgical Treatment for Refractory Heartburn

S.J. Spechler, J.G. Hunter, K.M. Jones, R. Lee, B.R. Smith, H. Mashimo, V.M. Sanchez, K.B. Dunbar, T.H. Pham, U.K. Murthy, T. Kim, C.S. Jackson, J.M. Wallen, E.C. von Rosenvinge, J.P. Pearl, L. Laine, A.W. Kim, A.M. Kaz, R.P. Tatum, Z.F. Gellad, S. Lagoo-Deenadayalan, J.H. Rubenstein, A.A. Ghaferi, W.-K. Lo, R.S. Fernando, B.S. Chan, S.C. Paski, D. Provenzale, D.O. Castell, D. Lieberman, R.F. Souza, W.D. Chey, S.R. Warren, A. Davis-Karim, S.D. Melton, R.M. Genta, T. Serpi,* K. Biswas, and G.D. Huang

ABSTRACT

BACKGROUND

Heartburn that persists despite proton-pump inhibitor (PPI) treatment is a frequent clinical problem with multiple potential causes. Treatments for PPI-refractory heartburn are of unproven efficacy and focus on controlling gastroesophageal reflux with reflux-reducing medication (e.g., baclofen) or antireflux surgery or on dampening visceral hypersensitivity with neuromodulators (e.g., desipramine).

METHODS

Patients who were referred to Veterans Affairs (VA) gastroenterology clinics for PPI-refractory heartburn received 20 mg of omeprazole twice daily for 2 weeks, and those with persistent heartburn underwent endoscopy, esophageal biopsy, esophageal manometry, and multichannel intraluminal impedance–pH monitoring. If patients were found to have reflux-related heartburn, we randomly assigned them to receive surgical treatment (laparoscopic Nissen fundoplication), active medical treatment (omeprazole plus baclofen, with desipramine added depending on symptoms), or control medical treatment (omeprazole plus placebo). The primary outcome was treatment success, defined as a decrease of 50% or more in the Gastroesophageal Reflux Disease (GERD)–Health Related Quality of Life score (range, 0 to 50, with higher scores indicating worse symptoms) at 1 year.

RESULTS

A total of 366 patients (mean age, 48.5 years; 280 men) were enrolled. Prerandomization procedures excluded 288 patients: 42 had relief of their heartburn during the 2-week omeprazole trial, 70 did not complete trial procedures, 54 were excluded for other reasons, 23 had non-GERD esophageal disorders, and 99 had functional heartburn (not due to GERD or other histopathologic, motility, or structural abnormality). The remaining 78 patients underwent randomization. The incidence of treatment success with surgery (18 of 27 patients, 67%) was significantly superior to that with active medical treatment (7 of 25 patients, 28%; P=0.007) or control medical treatment (3 of 26 patients, 12%; P<0.001). The difference in the incidence of treatment success between the active medical group and the control medical group was 16 percentage points (95% confidence interval, −5 to 38; P=0.17).

CONCLUSIONS

Among patients referred to VA gastroenterology clinics for PPI-refractory heartburn, systematic workup revealed truly PPI-refractory and reflux-related heartburn in a minority of patients. For that highly selected subgroup, surgery was superior to medical treatment. (Funded by the Department of Veterans Affairs Cooperative Studies Program; ClinicalTrials.gov number, NCT01265550.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Spechler at Baylor University Medical Center at Dallas, 3500 Gaston Ave., 2 Hoblitzelle, Suite 250, Dallas, TX 75246, or at sjspechler@aol.com.

*Deceased.

N Engl J Med 2019;381:1513-23.
DOI: 10.1056/NEJMoa1811424

In the United States, approximately 20% of adults regularly have symptoms of gastroesophageal reflux disease (GERD),[1] and annual costs for managing GERD exceed $12 billion.[2] Patients with heartburn, the cardinal symptom of GERD, report reduced work productivity and significant impairments in health-related quality of life.[3,4] Proton-pump inhibitors (PPIs) are highly effective for healing reflux esophagitis[5] but less effective for eliminating GERD symptoms, which persist in some 30% of patients treated with PPIs.[6] Only 58% of patients taking prescription PPIs for chronic heartburn report complete satisfaction with this treatment,[7] and "PPI-refractory GERD" is the most common reason for GERD-related referrals to gastroenterologists.[8]

Five major mechanisms might underlie PPI-refractory heartburn[9]: first, abnormal acid reflux persists despite PPI therapy; second, there is reflux hypersensitivity, in which esophageal exposure to acid is normal but "physiologic" reflux episodes (acidic or nonacidic) evoke heartburn[10]; third, heartburn is caused by esophageal disorders other than GERD (e.g., achalasia); fourth, heartburn is caused by extraesophageal disorders (e.g., heart disease); or fifth, heartburn is functional (i.e., not due to GERD or any other identifiable histopathologic, motility, or structural abnormality).[10] The frequency with which these mechanisms underlie PPI-refractory heartburn is not clear, and distinguishing among them requires systematic evaluation that includes endoscopy with esophageal biopsy, esophageal manometry, and esophageal multichannel intraluminal impedance (MII)–pH monitoring. MII-pH monitoring measures reflux episodes (acidic according to pH, and nonacidic according to MII) and their association with heartburn episodes.

For patients with PPI-refractory heartburn that is reflux-related (due to persistently abnormal acid reflux or reflux hypersensitivity), there are no medical treatment options of established long-term benefit. PPIs are often continued despite inadequate symptom relief.[11] Other options include reflux-reducing medications, such as baclofen,[12] or neuromodulators (e.g., tricyclic antidepressants) that dampen visceral hypersensitivity.[13] However, baclofen and neuromodulators often have unacceptable side effects, and studies of their efficacy for PPI-refractory heartburn are few and of short duration.[12-14] Recommendations for medical management of this condition are largely opinion-based.[6,15-17]

In principle, antireflux surgery (fundoplication), which creates a barrier to reflux of all gastric material (acidic and nonacidic), should relieve PPI-refractory heartburn that is reflux-related. In practice, however, patients with "GERD symptoms" that are unresponsive to PPIs often do not have a response to surgery either.[18] This might result from preoperative failure to document that the symptoms are truly reflux-related. Alternatively, for patients with reflux hypersensitivity, surgical reduction of reflux might not relieve symptoms generated by a hypersensitive esophagus.

Systematic evaluation including esophageal MII-pH monitoring may distinguish PPI-refractory patients with non-GERD disorders (who will not benefit from fundoplication) from those with persistently abnormal acid reflux or reflux hypersensitivity (who might have a response to surgery).[19] MII-pH monitoring is a relatively recent innovation, however, and experts disagree on its clinical usefulness.[20] Surgeons are reluctant to rely on it to select patients for fundoplication, generally preferring traditional esophageal pH monitoring for that purpose.[21] We hypothesized that if non-GERD and functional disorders were excluded by systematic workup, then antireflux surgery would be superior to medical therapy for patients with PPI-unresponsive heartburn that MII-pH monitoring identifies as being reflux-related.

METHODS

TRIAL OVERSIGHT

This trial was approved by the Veterans Affairs (VA) central institutional review board. All the patients provided written informed consent. The authors vouch for the completeness and accuracy of the data and for the fidelity of the trial to the protocol, available with the full text of this article at NEJM.org.

TRIAL DESIGN

Prerandomization Trial Procedures

All patients who were referred to VA gastroenterology clinics for heartburn refractory to PPIs were screened (Fig. 1). Eligible patients completed the GERD–Health Related Quality of Life (GERD-HRQL) questionnaire, which measures

severity of heartburn and other GERD symptoms (scores range from 0 to 50, with higher scores indicating worse symptoms).[22] Irrespective of PPI type and dose that patients were taking at trial entry, all received a 2-week trial of omeprazole at a dose of 20 mg twice daily, with instructions to take omeprazole 30 minutes before breakfast and dinner, and GERD-HRQL scoring was repeated (this score was considered the baseline score).

Patients with an improvement (decrease) of less than 50% in the GERD-HRQL score completed questionnaires (Veterans RAND 36-Item Health Survey [VR-36] measuring health-related quality of life [scores range from 0 to 100, with higher scores indicating better function], Patient Health Questionnaire 9 [PHQ-9] measuring depression [scores range from 0 to 27, with higher scores indicating worse depression], and the Generalized Anxiety Disorder 7-Item Questionnaire [GAD-7] measuring anxiety [scores range from 0 to 21, with higher scores indicating worse anxiety]) and underwent endoscopy with esophageal biopsy, esophageal manometry, and MII-pH monitoring while receiving 20 mg of omeprazole twice daily. Patients with severe reflux esophagitis, non-GERD endoscopic abnormalities, eosinophilic esophagitis, achalasia, or absent contractility were excluded.

Trial participation required trial surgeon approval and a positive symptom association probability (SAP >95%, indicating a significant [$P<0.05$] association between heartburn and reflux episodes [acidic, nonacidic, or all]), abnormal acid reflux (esophageal pH <4 for ≥4.2% of the 24-hour monitoring period), or both. The sequence of trial procedures varied owing to logistic and patient convenience issues, and further testing was not performed if any test ruled out reflux-related, PPI-refractory heartburn. Initial slow recruitment of patients resulted in intratrial changes to the protocol, including revised power calculations. (For details on prerandomization trial procedures, randomization, intratrial changes to the protocol, and the statistical analysis, see the Supplementary Appendix, available at NEJM.org.)

Randomization

Patients were assigned to receive active medical, control medical, or surgical treatment with the use of an adaptive-randomization procedure that stratified patients according to MII-pH results (positive SAP alone, abnormal acid reflux alone, or both positive SAP and abnormal acid reflux) and with the use of a "biased coin" procedure to balance treatment assignments.[23] Randomization status (determined by the adaptive-randomization and biased-coin procedures) was programmed centrally on a secured server, which ensured concealment of treatment assignments.

Medical Treatment Groups

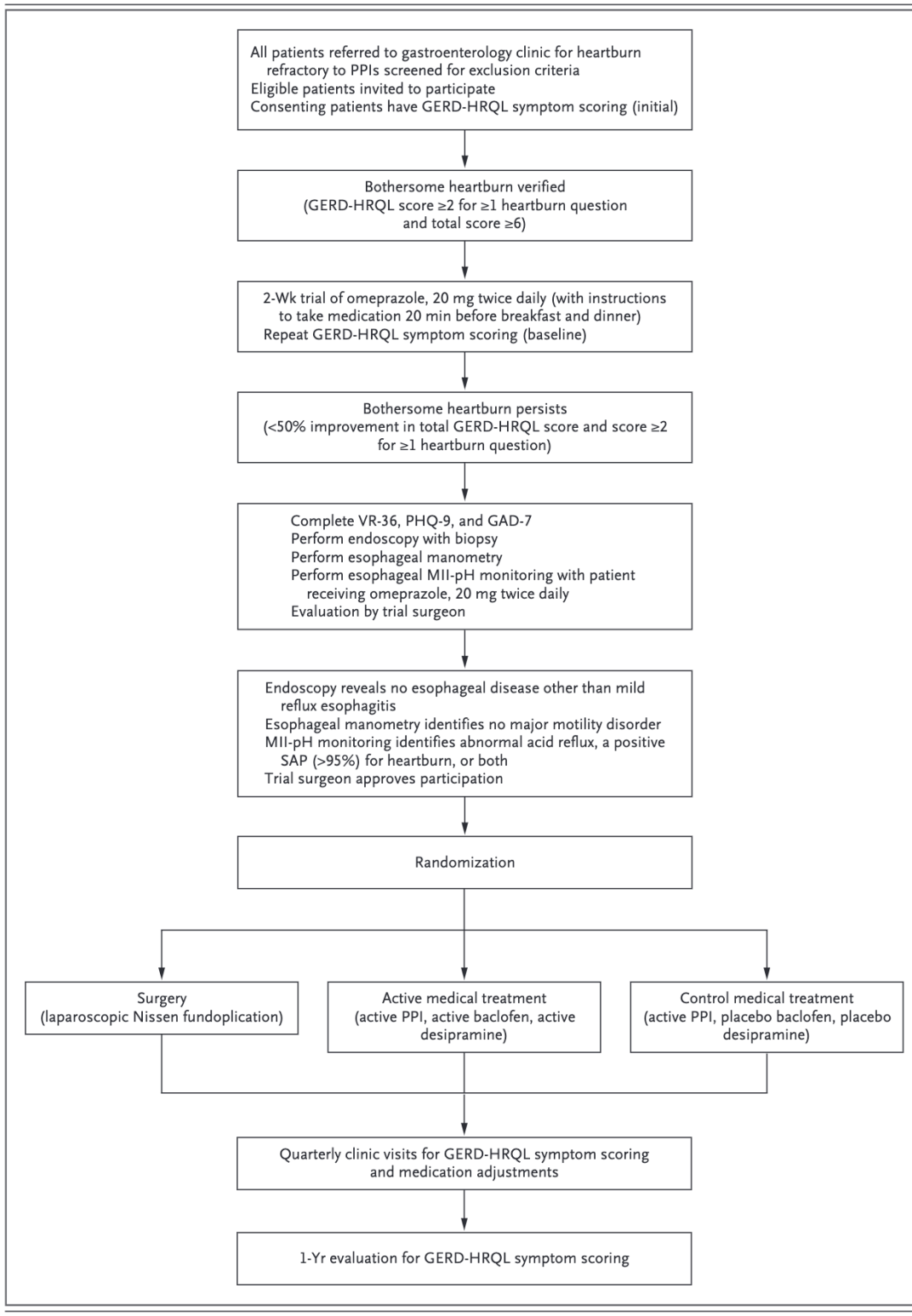
Investigators and patients were unaware of whether medical treatment was active or placebo. At all clinic visits, patients were queried about missed medication doses, and medication counts were performed. Patients in both groups received active omeprazole at a dose of 20 mg twice daily throughout the trial. Baclofen, desipramine, or identical-appearing placebos were added sequentially. All patients received baclofen (or baclofen placebo), with the dose gradually increased to 20 mg three times daily, for the trial duration unless unacceptable side effects occurred or less than 50% improvement in the GERD-HRQL score was found on any quarterly clinic visit, in which case baclofen (or baclofen placebo) was discontinued. After discontinuation, patients with contraindications to desipramine were declared to have treatment failure. Patients without contraindications to desipramine received desipramine (or desipramine placebo), with the dose gradually increased to 100 mg at bedtime, for the trial duration unless unacceptable side effects occurred or less than 50% improvement in the GERD-HRQL score was found on any quarterly clinic visit, at which time patients were declared to have treatment failure.

Surgical Treatment Group

Surgical treatment was laparoscopic Nissen fundoplication. Heartburn medications were prohibited after fundoplication. Patients with less than 50% improvement from the baseline GERD-HRQL score at any quarterly clinic visit or with heartburn severe enough to result in medication were declared to have treatment failure.

Follow-up

Patients who had undergone randomization were seen at quarterly clinic visits for GERD-HRQL scoring and medication adjustments and at 1 year for GERD-HRQL scoring.



STATISTICAL ANALYSIS

The primary outcome was treatment success, defined as an improvement (decrease) of 50% or more in the GERD-HRQL score from baseline to 12 months, a definition used in other prospective trials of antireflux procedures.[24] In the surgery group, patients were considered to have treatment failure if they did not have an improve-

Figure 1 (facing page). Trial Design.

From August 29, 2012, through December 2, 2015, medical records of all consecutive patients who were referred to gastroenterology clinics at participating Veterans Affairs (VA) medical centers for heartburn refractory to proton-pump inhibitors (PPIs) (on the basis of the referring physician's assessment) were screened for preliminary exclusion criteria (see the Supplementary Appendix), and those with no identified exclusions were invited to participate in the trial with the understanding that they would enter the randomized trial of medical or surgical treatment if trial procedures documented reflux-related heartburn. The sequence in which trial procedures were performed varied as a result of logistic and patient convenience issues. Further testing was not performed if any test excluded reflux-related, PPI-refractory heartburn (e.g., if endoscopy established an alternative diagnosis, such as eosinophilic esophagitis, then manometry and multichannel intraluminal impedance [MII]–pH monitoring were not performed if they had not already been performed). Total scores on the Gastroesophageal Reflux Disease–Health Related Quality of Life (GERD-HRQL) scale range from 0 to 50, with higher scores indicating worse symptoms. Scores on individual heartburn questions range from 0 to 5, with higher scores indicating worse symptoms. Symptom association probability (SAP) values of more than 95% indicate significant ($P<0.05$) associations between periods of reflux episodes and periods of heartburn symptoms. GAD-7 denotes the Generalized Anxiety Disorder 7-Item Questionnaire (which measures anxiety), PHQ-9 Patient Health Questionnaire 9 (which measures depression), and VR-36 the Veterans RAND 36-Item Health Survey (which measures health-related quality of life).

ment of 50% or more in the GERD-HRQL score or if they had used medication for heartburn symptoms at any of the quarterly assessments during the 12-month follow-up period; in the medical treatment groups, patients were considered to have treatment failure if they did not have an improvement of 50% or more in the GERD-HRQL score at any assessment after dose adjustment was completed (see above). Prespecified secondary outcomes included the frequency with which non-GERD disorders underlay PPI-refractory heartburn and the frequency of anxiety and depression in these patients. This article does not include all analyses and outcomes that were prespecified in the protocol and does not include any analyses or outcome measures that were not prespecified in the protocol.

Analyses were performed on an intention-to-treat basis. Fisher's exact test of two proportions was used to compare treatment success in three pairwise comparisons: surgery versus active medical treatment, surgery versus control medical treatment, and active medical treatment versus control medical treatment. To keep the overall type I error at 0.05 for the null hypothesis, the alpha level for each of the three comparisons was adjusted with the use of the Hochberg method.[25] Comparisons between groups that were excluded from the trial and randomly assigned patients (Table 1) and comparisons among treatment groups (Table 2) used chi-square analyses for categorical variables and the Kruskal–Wallis test for continuous measures. All reported P values are two-sided and unadjusted. We used SAS software, version 9.4 (SAS Institute), for all analyses.

RESULTS

PRERANDOMIZATION PROCEDURE EXCLUSIONS

Patients were recruited from August 29, 2012, through December 2, 2015; follow-up ended December 31, 2016. A total of 366 patients (280 men; mean [±SD] age, 48.5±12.2 years) were enrolled and began the prerandomization evaluation; 288 were excluded during prerandomization trial procedures (Fig. 2). A total of 70 patients were unwilling or unable to complete prerandomization procedures (Table S1 in the Supplementary Appendix); 54 were excluded for miscellaneous reasons; 42 were excluded because heartburn was relieved during the 2-week omeprazole trial; 23 were found to have non-GERD organic disorders (9 had eosinophilic esophagitis, 7 had other endoscopic or histologic abnormalities [2 had severe reflux esophagitis, 1 candida esophagitis, 2 gastric ulcer, and 2 unspecified abnormalities], and 7 had manometric abnormalities [2 had achalasia, 2 esophago-gastric junction outflow obstruction, and 3 severe ineffective esophageal motility]); and 99 had functional heartburn (MII-pH monitoring showed normal esophageal acid exposure and an SAP of ≤95%, which indicated no significant association between reflux episodes and heartburn). Thus, only 78 patients were eligible for randomization because they completed the full evaluation and were determined to have reflux-related, PPI-unresponsive heartburn documented by MII-pH monitoring. A revised power calculation called for 108 randomly assigned patients, but enrollment was capped at 78 solely because of funding limitations.

Baseline characteristics (age, sex, race, and GERD-HRQL scores) were generally similar in

Table 1. Demographic and Clinical Characteristics of the Patients.*

Characteristic	Patients with Functional Heartburn (N=99)	Patients with Reflux-Related, PPI-Refractory Heartburn Who Underwent Randomization (N=78)
Sex — no. (%)		
Male	76 (77)	64 (82)
Female	23 (23)	14 (18)
Race — no. (%)†		
White	69 (70)	54 (69)
Black	17 (17)	9 (12)
Other or mixed	13 (13)	15 (19)
Age	50.5±12.2	45.4±11.8
GERD-HRQL score‡		
Initial	23.9±7.9	25.5±8.1
Baseline	21.4±7.7	23.9±8.2
VR-36§		
Physical component	36.1±9.9	37.6±9.7
Mental component	40.0±13.5	43.3±12.8
Physical functioning	56.3±27.6	63.1.0±25.6
Role limitations due to physical health	36.2±39.3	39.7±38.1
Role limitations due to emotional problems	46.2±49.9	56.7±46.0
Vitality	34.9±19.0	41.4±20.5
Mental health index	57.8±22.8	63.6±22.5
Social functioning	49.6±30.2	55.6±26.2
Pain index	40.1±22.1	43.9±21.7
General health	45.9±21.3	51.8±20.9
PHQ-9 score¶	10.4±6.4	9.2±7.2
GAD-7 score‖	8.2±6.3	7.5±6.9

* Plus–minus values are means ±SD. There were no significant differences between the two populations for any characteristic except age (P=0.003), baseline score on the Gastroesophageal Reflux Disease–Health Related Quality of Life (GERD-HRQL) index (P=0.04), and score for vitality on the Veterans RAND 36-Item Health Survey (VR-36) (P=0.04). PPI denotes proton-pump inhibitor.
† Race was determined by patient report.
‡ Scores on the GERD-HRQL index range from 0 to 50, with higher scores indicating worse symptoms. The baseline score was assessed after a 2-week trial of omeprazole at a dose of 20 mg twice daily, with instructions to take omeprazole 30 minutes before breakfast and dinner.
§ The VR-36 measures health-related quality of life on multiple dimensions. Scores range from 0 to 100 for all dimensions, except role limitations due to physical health (range, −7 to 110) and role limitations due to emotional problems (range, −2 to 114), with higher scores indicating better function. The number of patients with functional heartburn who had missing data was as follows: for physical component, two; for mental component, two; for role limitations due to physical health, one; for role limitations due to emotional problems, two; and for pain index, one.
¶ Scores on Patient Health Questionnaire 9 (PHQ-9) range from 0 to 27, with higher scores indicating worse depression. Data were missing for two patients with functional heartburn.
‖ Scores on the Generalized Anxiety Disorder 7-Item Questionnaire (GAD-7) range from 0 to 21, with higher scores indicating worse anxiety.

patients who received a diagnosis of reflux-related, PPI-refractory heartburn and subsequently underwent randomization and in patients who did not undergo randomization because they had functional heartburn, had a response to PPI treatment, or dropped out before completing evaluation (Table S2). Baseline demographic characteristics and scores on quality of life (GERD-HRQL and VR-36), depression (PHQ-9), and anxiety (GAD-7) were similar in patients

Table 2. Demographic Features, GERD-HRQL Scores, and Multichannel Intraluminal Impedance (MII)–pH Monitoring Results, According to Treatment Group.*

Variable	Surgery (N=27)	Active Medical Treatment (N=25)	Control Medical Treatment (N=26)
Sex — no. (%)			
Male	23 (85)	18 (72)	23 (88)
Female	4 (15)	7 (28)	3 (12)
Race — no. (%)			
White	17 (63)	20 (80)	17 (65)
Black	2 (7)	2 (8)	5 (19)
Other or mixed	8 (30)	3 (12)	4 (15)
Mean age (95% CI) — yr	44.9 (40.2–49.5)	43.9 (38.9–49.0)	47.2 (42.5–52.0)
Mean GERD-HRQL score at baseline (95% CI)	25.8 (22.5–29.1)	21.0 (18.6–23.4)	24.7 (20.9–28.5)
Mean results of MII-pH monitoring (95% CI)†			
Total % of time with pH <4‡	6.1 (3.4–8.8)	7.7 (2.6–12.7)	8.8 (2.2–15.5)
DeMeester score§	20.5 (11.8–29.2)	27.6 (10.5–44.7)	21.5 (11.9–31.2)
Total nonacidic reflux events¶	38.9 (26.4–51.4)	50.6 (27.2–74.0)	47.8 (35.6–60.1)
Heartburn SAP‖	92.2 (84.2–100)	89.2 (79.6–98.7)	86.5 (73.3–99.6)
MII-pH monitoring randomization criteria — no. (%)†			
SAP of >95% alone**	14 (52)	12 (48)	11 (42)
Abnormal acid reflux alone	6 (22)	5 (20)	4 (15)
Abnormal acid reflux and SAP of >95%	7 (26)	7 (28)	11 (42)

* There were no significant differences among the randomized groups for any feature included in this table. Symptom association probability (SAP) values of more than 95% indicate significant (P<0.05) associations between periods of reflux episodes and periods of heartburn symptoms. Percentages may not total 100 because of rounding. CI denotes confidence interval.
† Data for MII-pH monitoring were missing for one patient in the active medical group.
‡ A value of 4.2% or higher for the total percentage of time with an esophageal pH of less than 4 indicates abnormal acid reflux.
§ DeMeester scores of more than 14.7 indicate abnormal acid reflux.
¶ Reflux events with a pH of more than 4 were considered nonacidic. Shown is the total number of nonacidic reflux events recorded during the 24-hour monitoring period, irrespective of whether they were associated with heartburn.
‖ Heartburn SAP values are those for all reflux events (acidic and nonacidic).
** An SAP of 95% alone indicates reflux hypersensitivity.

who received a diagnosis of functional heartburn and those who received a diagnosis of reflux-related, PPI-refractory heartburn and subsequently underwent randomization (Table 1).

RANDOMLY ASSIGNED PATIENTS

A total of 78 patients (64 men; mean age, 45.4±11.8 years; 54 white, 9 black, 14 other race, and 1 unknown race) with reflux-related, PPI-unresponsive heartburn were randomly assigned to receive surgical treatment (27 patients), active medical treatment (25), or control medical treatment (26). MII-pH monitoring showed an SAP of more than 95% alone (reflux hypersensitivity) in 37 patients, abnormal acid reflux alone in 15, and both an SAP of more than 95% and abnormal acid reflux in 25 (data were missing for 1 patient) (Table 2). There were no significant differences among treatment groups in demographic and clinical features or in the distribution of MII-pH monitoring results (i.e., groups had similar percentages of patients with an SAP of >95% alone, abnormal acid reflux alone, and an SAP of >95% and abnormal acid reflux).

After trial completion, review of primary data collection forms revealed discrepancies in eligibility criteria data for 5 randomly assigned patients, which raised uncertainty about whether these patients met all eligibility criteria. Analyses that excluded these 5 patients were similar to our primary analyses (see the Supplementary Appendix).

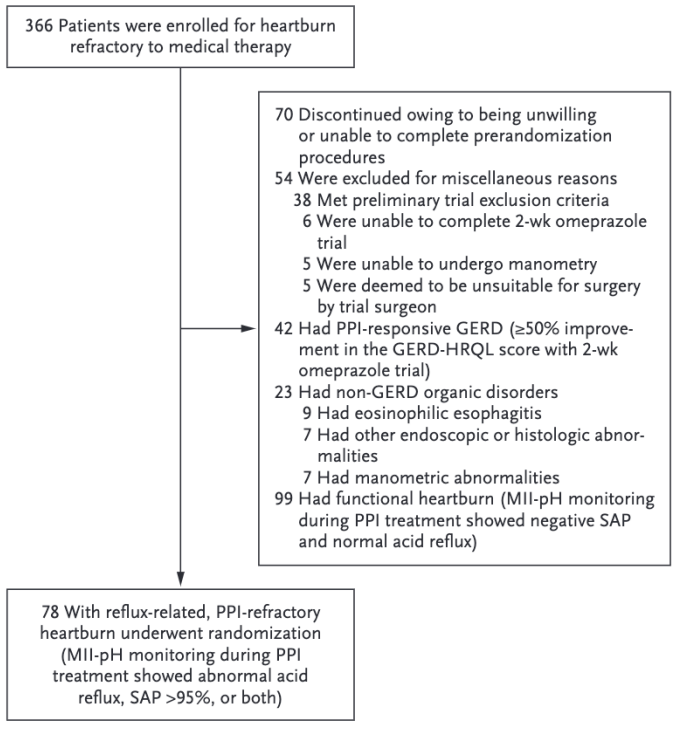


Figure 2. Patient Exclusions.

The medical records of consecutive patients who were referred to gastroenterology clinics at participating VA medical centers for heartburn refractory to PPIs were screened for preliminary exclusion criteria, and patients with no identified exclusions were invited to participate in the trial with the understanding that they would enter the randomized trial of medical or surgical treatment if trial procedures documented reflux-related heartburn. These 366 patients accepted those terms and provided written informed consent. Data on how many patients declined to provide written informed consent are not available. Of the 38 patients who met preliminary trial exclusion criteria after enrollment, 6 had coexisting conditions because of which the patient was considered by the local trial investigator to be unable to participate, 6 had non-GERD disorders that can cause heartburn sensation, 4 had a "heartburn" description did not meet the trial definition, 4 had laboratory exclusions, 3 had previous upper gastrointestinal surgery, 3 were ruled to be not suitable for the trial or were enrolled in another trial, 2 had an initial GERD-HRQL score that conferred ineligibility, 2 had contraindications to trial medications, 2 had morbid obesity, 2 were using forbidden medications, 1 had schizophrenia, 1 had a paraesophageal hernia, 1 was older than 70 years of age, and 1 had a seizure disorder. A total of 5 patients who were randomly assigned to the surgery group did not undergo surgery (2 declined, 1 was lost to follow-up before surgery, 1 moved from the area before surgery, and 1 had a change in medical status that precluded surgery). One patient who was assigned to receive active medical treatment was lost to follow-up before receiving medical treatment. All patients who were assigned to receive control medical treatment were treated as assigned.

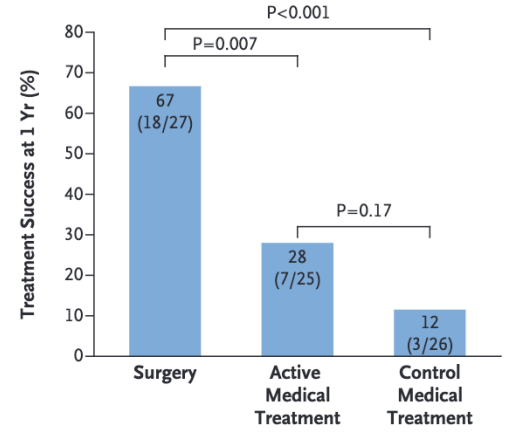


Figure 3. Treatment Success at 1 Year (Intention-to-Treat Analyses).

Treatment success was defined as an improvement (decrease) of 50% or more in the GERD-HRQL score from baseline.

OUTCOMES

At 12 months, treatment success (≥50% improvement in the GERD-HRQL score) occurred in 18 of 27 patients (67%) in the surgery group, 7 of 25 (28%) in the active medical group, and 3 of 26 (12%) in the control medical group (Fig. 3 and Fig. S1A and S1B). The incidence of treatment success with surgery was significantly superior to that with active medical treatment (P=0.007; Hochberg-adjusted significance threshold, 0.025) or control medical treatment (P<0.001; Hochberg-adjusted significance threshold, 0.017). The difference in the incidence of treatment success between the active medical group and the control medical group was 16 percentage points (95% confidence interval [CI], −5 to 38; P=0.17). The relative risk of treatment success was 2.38 (95% CI, 1.20 to 4.71) for surgery versus active medical treatment, 5.78 (95% CI, 1.93 to 17.31) for surgery versus control medical treatment, and 2.43 (95% CI, 0.71 to 8.35) for active medical treatment versus control medical treatment (unadjusted comparisons). A prespecified subgroup analysis assessed the incidence of treatment success with surgery among patients with reflux hypersensitivity (SAP of >95% alone) and patients with abnormal acid reflux (acid reflux alone or with an SAP of >95%). The incidence of success in the surgery group was 71% among the 14 patients with reflux hypersensitivity and 62% among the 13 with abnormal acid reflux.

There were five serious adverse events in 4 patients in the surgery group, four serious adverse events in 4 patients in the active medical group, and five serious adverse events in 3 patients in the control medical group (Table S3).

One surgical patient had a herniated Nissen fundoplication that resulted in repeat surgery complicated by postoperative pneumonia; recovery was complete, and repeat surgery successfully controlled heartburn. There were no deaths.

DISCUSSION

Among 366 patients enrolled in our trial of medical and surgical treatments for PPI-refractory heartburn, systematic evaluation revealed that GERD underlay truly PPI-refractory heartburn in only a minority of patients. In 42 patients referred because of "PPI-refractory" heartburn, heartburn was relieved during a standardized, 2-week trial of omeprazole twice daily. A systematic evaluation showed that GERD was not the likely cause of heartburn for an additional 122 patients — 99 received a diagnosis of functional heartburn and 23 received a diagnosis of a non-GERD organic disorder. Only 78 patients completed the full assessment and were found to have GERD that was truly unresponsive to twice-daily PPIs. In that highly selected group, the incidence of treatment success with laparoscopic Nissen fundoplication at 1 year (67%) was significantly superior to that with active medical treatment (28%) or control medical treatment (12%).

In our trial, a substantial minority of patients who were referred to gastroenterology clinics with "PPI-refractory" heartburn got relief when prescribed omeprazole twice daily with explicit instruction on how to take it properly. There are two likely explanations as to why some previously PPI-refractory patients had a response to this standardized PPI trial. First, trial patients were given explicit instructions to take omeprazole 30 minutes before meals. This is important because PPIs bind only to gastric proton pumps that are actively secreting acid. Fewer than 10% of those pumps are active during fasting, whereas approximately 70% are active when stimulated by meals.[26] Consequently, PPIs are most effective when taken before meals. Second, patients taking PPIs other than omeprazole at trial entry were switched to omeprazole. Relative potencies of different PPIs vary widely,[27] and individual patients can exhibit considerable variability in response to different PPIs.[28]

This trial highlights the critical importance of systematic evaluation, similar to that recommended by Gyawali and Fass,[17] for managing the care of patients with PPI-refractory heartburn. Many patients would not complete this rigorous evaluation, and among those who did, the cause of heartburn in most of them was not GERD. Furthermore, no demographic or clinical characteristics distinguished patients with reflux-related heartburn from those with functional heartburn, those whose heartburn responded to omeprazole taken properly, and those who would not complete diagnostic evaluation. Although coexisting psychological conditions are common in patients with functional gastrointestinal disorders,[29,30] we found no significant differences in PHQ-9 depression and GAD-7 anxiety scores between patients who received a diagnosis of functional heartburn and those who received a diagnosis of reflux-related, PPI-refractory heartburn, with both groups having mean scores in the "moderate" range (i.e., 7 to 10)[31] (Table 1).

Our finding that reflux hypersensitivity can respond to fundoplication is noteworthy because reflux hypersensitivity is considered a functional disorder, which might not be expected to improve with a procedure that eliminates reflux without altering abnormal esophageal pain perception.[10] Furthermore, we identified hypersensitivity to nonacidic reflux by SAP values on esophageal MII monitoring, a newer technology whose usefulness in selecting patients for fundoplication has been questioned,[21,32] as has the validity of the SAP in general.[33,34] For our surgical patients, the incidence of treatment success was 71% among the 14 with reflux hypersensitivity and 62% among the 13 with abnormal acid reflux. In support of our findings, observational studies have noted that patients chosen for fundoplication on the basis of MII results can do well.[35-37] However, the overall 1-year incidence of treatment success among our surgical patients with PPI-refractory heartburn (67% in an intention-to-treat analysis) is considerably lower than the more than 90% incidence of success with fundoplication that is commonly described in observational studies involving patients with typical, PPI-responsive GERD.[38] The reasons for this are not entirely clear, but patients considering surgery for reflux-related, PPI-refractory heartburn should be advised that surgery was successful in only approximately 2 of 3 cases in our trial.

PPIs are inactivated through the hepatic cytochrome P450 isoenzyme CYP2C19, and *CYP2C19* mutations can influence PPI inactivation rates and clinical efficacy.[39] We did not test for these mutations, because this seldom is done in clini-

cal practice and because we documented PPI acid-control efficacy by MII-pH monitoring.

Limitations of our trial include its relatively small sample size and predominance of white men (reflecting the veteran patient population). With no sham-surgery group, we cannot determine the contribution of the placebo effect to the incidence of treatment success with surgery. Furthermore, because MII-pH monitoring was performed only while patients were taking PPIs, we cannot determine how many would have abnormal acid reflux when not taking these drugs.

Another limitation involves the intratrial protocol amendments required to enable trial completion. Overly restrictive entry criteria that limit trial enrollment, generalizability, and completion are a common problem in trials involving patients with functional gastrointestinal disorders.[40,41] We found early trial recruitment inadequate because most patients with PPI-refractory GERD who were referred to our gastroenterology clinics had contraindications to desipramine (e.g., concomitant use of other antidepressants). Consequently, we amended the protocol to allow the entry of patients with contraindications to desipramine. We also amended power calculations to detect only large differences between medical and surgical treatments, reasoning that physicians would not recommend surgery unless it were considerably more effective than medical therapy. Our findings document the considerable superiority of antireflux surgery over feasible medical therapy for patients with reflux-related heartburn that is resistant to PPIs. Although the incidence of treatment success did not differ significantly between active medical treatment and control medical treatment (28% and 12%, respectively), our amended trial was insufficiently powered to rule out an important benefit of medical therapy (the 95% confidence interval around the 16-percentage-point difference was −5 to 38).

In conclusion, for patients referred to our clinics for heartburn unrelieved by PPIs, systematic workup revealed that heartburn was both truly PPI-refractory and reflux-related in a minority of patients. In that highly selected group, laparoscopic Nissen fundoplication was significantly superior to medical therapy. We conclude that systematic workup including esophageal MII-pH monitoring can identify a subgroup of patients with PPI-refractory heartburn, including those with reflux hypersensitivity, who can have a response to antireflux surgery.

The views expressed herein are the opinions of the authors and do not necessarily represent the views of the Department of Veterans Affairs or U.S. government.

A data sharing statement provided by the authors is available with the full text of this article at NEJM.org.

Supported by the Department of Veterans Affairs Cooperative Studies Program.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Leslie Erin Norman for her assistance with project coordination, Ellen Dematt and Rebecca Ann Horney for help in the programming of trial data, and Brandi Read for her service as national trial coordinator.

APPENDIX

The authors' full names and academic degrees are as follows: Stuart J. Spechler, M.D., John G. Hunter, M.D., Karen M. Jones, M.S., Robert Lee, M.D., Brian R. Smith, M.D., Hiroshi Mashimo, M.D., Ph.D., Vivian M. Sanchez, M.D., Kerry B. Dunbar, M.D., Ph.D., Thai H. Pham, M.D., Uma K. Murthy, M.D., Taewan Kim, M.D., Christian S. Jackson, M.D., Jason M. Wallen, M.D., Erik C. von Rosenvinge, M.D., Jonathan P. Pearl, M.D., Loren Laine, M.D., Anthony W. Kim, M.D., Andrew M. Kaz, M.D., Roger P. Tatum, M.D., Ziad F. Gellad, M.D., Sandhya Lagoo-Deenadayalan, M.D., Ph.D., Joel H. Rubenstein, M.D., Amir A. Ghaferi, M.D., Wai-Kit Lo, M.D., Ronald S. Fernando, M.D., Bobby S. Chan, M.D., Shirley C. Paski, M.D., Dawn Provenzale, M.D., Donald O. Castell, M.D., David Lieberman, M.D., Rhonda F. Souza, M.D., William D. Chey, M.D., Stuart R. Warren, Pharm.D., Anne Davis-Karim, Pharm.D., Shelby D. Melton, M.D., Robert M. Genta, M.D., Tracey Serpi, Ph.D., Kousick Biswas, Ph.D., and Grant D. Huang, Ph.D.

The authors' affiliations are as follows: the Department of Medicine and Center for Esophageal Diseases, Baylor University Medical Center at Dallas, Baylor Scott & White Health, and the Department of Medicine, Veterans Affairs (VA) North Texas Health Care System (S.J.S., R.F.S.), and the Departments of Medicine (K.B.D.), Surgery (T.H.P.), and Pathology (S.D.M.), VA North Texas Health Care System and University of Texas Southwestern Medical Center, Dallas, and the Department of Pathology, Baylor College of Medicine, Houston (R.M.G.); the Departments of Surgery (J.G.H.) and Medicine (D.L.), Oregon Health and Science University, and the Department of Medicine, VA Portland Health Care System (D.L.) — both in Portland; the VA Cooperative Studies Program Coordinating Center, Perry Point (K.M.J., T.S., K.B.), and the Departments of Medicine (E.C.R.) and Surgery (J.P.P.), VA Maryland Health Care System and University of Maryland School of Medicine, Baltimore — all in Maryland; the Departments of Medicine (R.L.) and Surgery (B.R.S.), VA Long Beach Healthcare System and University of California, Irvine, Irvine, the Departments of Medicine (C.S.J., R.S.F., B.S.C.) and Surgery (J.M.W.), VA Loma Linda Healthcare System and Loma Linda University Medical Center, Loma Linda, and the Department of Surgery, University of Southern California, Los Angeles (A.W.K.) — all in California; the Department of Medicine, VA Boston Healthcare System, Harvard Medical School (H.M., W.-K.L.), and the Department of Surgery, VA Boston Healthcare System, Boston University School of Medicine (V.M.S.), Boston; the Departments of Medicine (U.K.M.) and Surgery (T.K.), Syracuse VA Medical Center and SUNY Upstate Medical University, Syracuse, NY; the Department of Medicine, VA Connecticut Healthcare System and Yale University, New Haven, CT (L.L.); the Departments of Medicine (A.M.K., S.C.P.) and Surgery (R.P.T.), VA Puget Sound Health Care System and University of Washington School of Medicine, Seattle; the Departments of Medicine (Z.F.G., D.P.) and Surgery (S.L.-D.), Durham VA Medical Center and Duke University Medical Center, Durham, NC; the Departments of Medicine (J.H.R.) and Surgery (A.A.G.), VA Ann Arbor Healthcare System, and the Departments of Medicine (J.H.R., W.D.C.) and Surgery (A.A.G.), University of Michigan — both in Ann Arbor; the Department of

Medicine, Medical University of South Carolina, Charleston (D.O.C.); the VA Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque, NM (S.R.W., A.D.-K.); and the VA Office of Research and Development, Washington, DC (G.D.H.).

REFERENCES

1. Locke GR III, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ III. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. Gastroenterology 1997;112:1448-56.
2. Everhart JE, Ruhl CE. Burden of digestive diseases in the United States part I: overall and upper gastrointestinal diseases. Gastroenterology 2009;136:376-86.
3. Dean BB, Crawley JA, Schmitt CM, Wong J, Ofman JJ. The burden of illness of gastro-oesophageal reflux disease: impact on work productivity. Aliment Pharmacol Ther 2003;17:1309-17.
4. Farup C, Kleinman L, Sloan S, et al. The impact of nocturnal symptoms associated with gastroesophageal reflux disease on health-related quality of life. Arch Intern Med 2001;161:45-52.
5. Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. Am J Gastroenterol 2013;108:308-28.
6. Scarpellini E, Ang D, Pauwels A, De Santis A, Vanuytsel T, Tack J. Management of refractory typical GERD symptoms. Nat Rev Gastroenterol Hepatol 2016;13:281-94.
7. Crawley JA, Schmitt CM. How satisfied are chronic heartburn sufferers with their prescription medications? Results of the Patient Unmet Needs Survey. J Clin Outcomes Manag 2000;7(11):29-34.
8. Fass R, Sifrim D. Management of heartburn not responding to proton pump inhibitors. Gut 2009;58:295-309.
9. Spechler SJ. Surgery for gastroesophageal reflux disease: esophageal impedance to progress? Clin Gastroenterol Hepatol 2009;7:1264-5.
10. Aziz Q, Fass R, Gyawali CP, Miwa H, Pandolfino JE, Zerbib F. Functional esophageal disorders. Gastroenterology 2016;150:1368-79.
11. Gawron AJ, Rothe J, Fought AJ, et al. Many patients continue using proton pump inhibitors after negative results from tests for reflux disease. Clin Gastroenterol Hepatol 2012;10:620-5.
12. Li S, Shi S, Chen F, Lin J. The effects of baclofen for the treatment of gastroesophageal reflux disease: a meta-analysis of randomized controlled trials. Gastroenterol Res Pract 2014;2014:307805.
13. Weijenborg PW, de Schepper HS, Smout AJ, Bredenoord AJ. Effects of antidepressants in patients with functional esophageal disorders or gastroesophageal reflux disease: a systematic review. Clin Gastroenterol Hepatol 2015;13(2):251-259.e1.
14. Drossman DA, Tack J, Ford AC, Szigethy E, Törnblom H, Van Oudenhove L. Neuromodulators for functional gastrointestinal disorders (disorders of gut-brain interaction): a Rome Foundation Working Team report. Gastroenterology 2018;154(4):1140-1171.e1.
15. Sifrim D, Zerbib F. Diagnosis and management of patients with reflux symptoms refractory to proton pump inhibitors. Gut 2012;61:1340-54.
16. Hillman L, Yadlapati R, Thuluvath AJ, Berendsen MA, Pandolfino JE. A review of medical therapy for proton pump inhibitor nonresponsive gastroesophageal reflux disease. Dis Esophagus 2017;30(9):1-15.
17. Gyawali CP, Fass R. Management of gastroesophageal reflux disease. Gastroenterology 2018;154:302-18.
18. Morgenthal CB, Lin E, Shane MD, Hunter JG, Smith CD. Who will fail laparoscopic Nissen fundoplication? Preoperative prediction of long-term outcomes. Surg Endosc 2007;21:1978-84.
19. Patel A, Sayuk GS, Gyawali CP. Prevalence, characteristics, and treatment outcomes of reflux hypersensitivity detected on pH-impedance monitoring. Neurogastroenterol Motil 2016;28:1382-90.
20. Ravi K, Katzka DA. Esophageal impedance monitoring: clinical pearls and pitfalls. Am J Gastroenterol 2016;111:1245-56.
21. Jobe BA, Richter JE, Hoppo T, et al. Preoperative diagnostic workup before antireflux surgery: an evidence and experience-based consensus of the Esophageal Diagnostic Advisory Panel. J Am Coll Surg 2013;217:586-97.
22. Velanovich V. The development of the GERD-HRQL symptom severity instrument. Dis Esophagus 2007;20:130-4.
23. Efron B. Forcing a sequential experiment to be balanced. Biometrika 1971;58:403-17.
24. Ganz RA, Peters JH, Horgan S, et al. Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 2013;368:719-27.
25. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800-2.
26. Shin JM, Sachs G. Pharmacology of proton pump inhibitors. Curr Gastroenterol Rep 2008;10:528-34.
27. Graham DY, Tansel A. Interchangeable use of proton pump inhibitors based on relative potency. Clin Gastroenterol Hepatol 2018;16(6):800-808.e7.
28. Katz PO, Koch FK, Ballard ED, et al. Comparison of the effects of immediate-release omeprazole oral suspension, delayed-release lansoprazole capsules and delayed-release esomeprazole capsules on nocturnal gastric acidity after bedtime dosing in patients with night-time GERD symptoms. Aliment Pharmacol Ther 2007;25:197-205.
29. Fass R, Tougas G. Functional heartburn: the stimulus, the pain, and the brain. Gut 2002;51:885-92.
30. Rubenstein JH, Nojkov B, Korsnes S, et al. Oesophageal hypersensitivity is associated with features of psychiatric disorders and the irritable bowel syndrome. Aliment Pharmacol Ther 2007;26:443-52.
31. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32:345-59.
32. Ward MA, Dunst CM, Teitelbaum EN, et al. Impedance-pH monitoring on medications does not reliably confirm the presence of gastroesophageal reflux disease in patients referred for antireflux surgery. Surg Endosc 2018;32:889-94.
33. Kahrilas PJ. When proton pump inhibitors fail. Clin Gastroenterol Hepatol 2008;6:482-3.
34. Choksi Y, Slaughter JC, Sharda R, Higginbotham T, Lal P, Vaezi MF. Symptom association probability does not reliably distinguish functional heartburn from reflux hypersensitivity. Aliment Pharmacol Ther 2018;47:958-65.
35. Mainie I, Tutuian R, Agrawal A, Adams D, Castell DO. Combined multichannel intraluminal impedance-pH monitoring to select patients with persistent gastro-oesophageal reflux for laparoscopic Nissen fundoplication. Br J Surg 2006;93:1483-7.
36. Frazzoni M, Piccoli M, Conigliaro R, Manta R, Frazzoni L, Melotti G. Refractory gastroesophageal reflux disease as diagnosed by impedance-pH monitoring can be cured by laparoscopic fundoplication. Surg Endosc 2013;27:2940-6.
37. Desjardin M, Luc G, Collet D, Zerbib F. 24-Hour pH-impedance monitoring on therapy to select patients with refractory reflux symptoms for antireflux surgery: a single center retrospective study. Neurogastroenterol Motil 2016;28:146-52.
38. Jackson PG, Gleiber MA, Askari R, Evans SR. Predictors of outcome in 100 consecutive laparoscopic antireflux procedures. Am J Surg 2001;181:231-5.
39. Ichikawa H, Sugimoto M, Sugimoto K, Andoh A, Furuta T. Rapid metabolizer genotype of CYP2C19 is a risk factor of being refractory to proton pump inhibitor therapy for reflux esophagitis. J Gastroenterol Hepatol 2016;31:716-26.
40. Vakil NB, Howden CW, Moayyedi P, Tack J. White Paper AGA: functional dyspepsia. Clin Gastroenterol Hepatol 2017;15:1191-4.
41. Talley NJ, Kellow JE, Boyce P, Tennant C, Huskic S, Jones M. Antidepressant therapy (imipramine and citalopram) for irritable bowel syndrome: a double-blind, randomized, placebo-controlled trial. Dig Dis Sci 2008;53:108-15.

Gastroenterology 2015;148:324–333

Efficacy of Transoral Fundoplication vs Omeprazole for Treatment of Regurgitation in a Randomized Controlled Trial

John G. Hunter,[1,*] **Peter J. Kahrilas**,[2,*] Reginald C. W. Bell,[3] Erik B. Wilson,[4] Karim S. Trad,[5,6] James P. Dolan,[1] Kyle A. Perry,[7] Brant K. Oelschlager,[8] Nathaniel J. Soper,[2] Brad E. Snyder,[4] Miguel A. Burch,[9] William Scott Melvin,[7] Kevin M. Reavis,[1,10] Daniel G. Turgeon,[5,6] Eric S. Hungness,[2] and Brian S. Diggs[1]

[1]Department of Surgery, Oregon Health & Science University, Portland, Oregon; [2]Department of Surgery and Department of Gastroenterology, Northwestern University, Chicago, Illinois; [3]SurgOne Foregut Institute, Englewood, Colorado; [4]Department of Surgery, University of Texas Health Science Center, Houston, Texas; [5]Department of Surgery, The George Washington University School of Medicine and Health Sciences, Washington, District of Columbia; [6]Reston Surgical Associates, Reston, Virginia; [7]Department of Surgery, The Ohio State University, Columbus, Ohio; [8]Department of Surgery, University of Washington, Seattle, Washington; [9]Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, California; and [10]Department of Surgery, Oregon Clinic, Portland, Oregon

See Covering the Cover synopsis on page 265; see editorial on page 280.

BACKGROUND & AIMS: Transoral esophagogastric fundoplication (TF) can decrease or eliminate features of gastroesophageal reflux disease (GERD) in some patients whose symptoms persist despite proton pump inhibitor (PPI) therapy. We performed a prospective, sham-controlled trial to determine if TF reduced troublesome regurgitation to a greater extent than PPIs in patients with GERD. **METHODS:** We screened 696 patients with troublesome regurgitation despite daily PPI use with 3 validated GERD-specific symptom scales, on and off PPIs. Those with at least troublesome regurgitation (based on the Montreal definition) on PPIs underwent barium swallow, esophagogastroduodenoscopy, 48-hour esophageal pH monitoring (off PPIs), and high-resolution esophageal manometry analyses. Patients with GERD and hiatal hernias $\leq$2 cm were randomly assigned to groups that underwent TF and then received 6 months of placebo (n = 87), or sham surgery and 6 months of once- or twice-daily omeprazole (controls, n = 42). Patients were blinded to therapy during follow-up period and reassessed at 2, 12, and 26 weeks. At 6 months, patients underwent 48-hour esophageal pH monitoring and esophagogastroduodenoscopy. **RESULTS:** By intention-to-treat analysis, TF eliminated troublesome regurgitation in a larger proportion of patients (67%) than PPIs (45%) ($P = .023$). A larger proportion of controls had no response at 3 months (36%) than subjects that received TF (11%; $P = .004$). Control of esophageal pH improved after TF (mean 9.3% before and 6.3% after; $P < .001$), but not after sham surgery (mean 8.6% before and 8.9% after). Subjects from both groups who completed the protocol had similar reductions in GERD symptom scores. Severe complications were rare (3 subjects receiving TF and 1 receiving the sham surgery). **CONCLUSIONS:** TF was an effective treatment for patients with GERD symptoms, particularly in those with persistent regurgitation despite PPI therapy, based on evaluation 6 months after the procedure. Clinicaltrials.gov no: NCT01136980.

Keywords: TIF; EsophyX; Stomach; Esophagus.

Gastroesophageal reflux disease (GERD) remains one of the most common condition for which Americans take daily medication, and proton-pump inhibitor (PPI) use has more than doubled in the last decade.[1] Despite this, up to 40% of PPI-dependent GERD patients have troublesome symptoms of GERD, despite PPI therapy.[2,3] Although laparoscopic antireflux surgery has been suggested for this group of patients, fear of surgery, side effects, and recurrent symptoms have kept patient and referring physician interest to <10% of those otherwise qualifying for surgery.[4,5] Transoral endoscopic methods of treating GERD have been available for many years, but only one of these technologies allows the creation of a fundoplication, by folding the stomach anteriorly around the esophagus and securing it with multiple fasteners. Although this device has been in use for 9 years in Europe and 7 years in the United States, and has been proven effective in registry trials and one randomized controlled trial (RCT), comparison of effectiveness in patients with persistent symptoms on PPI has been absent.[6–8] Our aim was to determine whether or not transoral fundoplication (TF) was better than PPI treatment of troublesome GERD symptoms, particularly regurgitation, in a population of chronic PPI-dependent GERD patients.

Methods

Ethics Statement

This study was approved by the Institutional Review Board of each site and was conducted in accordance with the Good Clinical Practices and Declaration of Helsinki. All patients provided written informed consent form. All authors had access to the study data and reviewed and approved the final manuscript.

*Author share co-first authorship.

Abbreviations used in this paper: EGD, esophagogastroduodenoscopy; GERD, gastroesophageal reflux disease; ITT, intention to treat; PPI, proton pump inhibitor; RDQ, Reflux Disease Questionnaire; TF, transoral fundoplication.

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2014.10.009

Study Design and Patients

The Randomized EsophyX vs Sham, Placebo-Controlled Transoral Fundoplication (RESPECT) trial was carried out at 8 academic and community medical centers across the United States. We recruited patients between the ages of 18 and 80 years with more than 6 months of GERD symptoms and troublesome regurgitation, despite a minimum PPI dose of 40 mg daily. Troublesome regurgitation was defined as mild symptoms for 2 or more days per week or moderate to severe symptoms more than 1 day per week, per Montreal consensus criteria.[3,9] Symptom assessment used the following 3 validated tools: the Reflux Disease Questionnaire (RDQ), the Gastroesophageal Reflux Symptom Score, and the GERD-Health Related Quality of Life on PPI and off PPI for at least 7 days. Abnormal amounts of gastroesophageal reflux off PPI for 7 days was confirmed by distal esophageal pH <4 for $>5.3\%$ of at least 1 of the 2 days that pH was measured with a Bravo (radiotelemetry) probe (Given Imaging, Yoqneam, Israel). High-resolution esophageal manometry confirmed the absence of esophageal motor dysfunction. Esophagogastroduodenoscopy (EGD) was performed to grade the appearance of the antireflux barrier (Hill grade), to confirm the absence of long segment Barrett's esophagus, and to grade esophagitis, if present. Cine-esophagography was performed to confirm the absence of hiatal hernia or a hiatal hernia ≤ 2 cm in length. Exclusion criteria included systemic disease not well controlled, obesity determined by body mass index >35, esophageal ulcer, stricture, Barrett's esophagus >2 cm in length, hiatal hernia >2 cm in length, Los Angeles grade C or D esophagitis, esophageal dysmotility, previous esophageal or gastric surgery, peptic ulcer disease, gastric outlet obstruction, gastroparesis, pregnancy or plans for pregnancy in the next 12 months, immunosuppression, portal hypertension, and coagulopathy. Patients were randomized 2:1 to either TF (study group) or sham surgery (control group). A computer-generated block-randomization method was used to assign patients to study or control group. After informed consent and administration of general anesthesia with endotracheal intubation, a sealed envelope, provided by an independent statistician, was opened by the operating team that indicated group allocation.

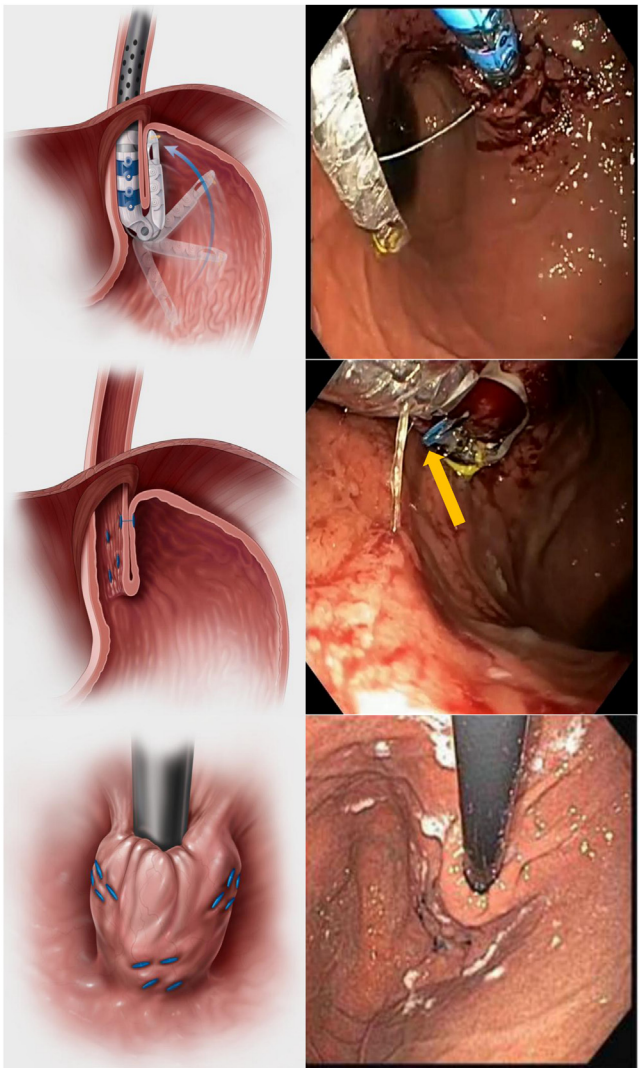


Figure 1. Transoral fundoplication creates a 3 cm flap valve, 180–270 degrees in circumference.

Operative Procedure

Patients allocated to the TF group underwent a standardized technique using the EsophyX-2 device (EndoGastric Solutions, Redmond, WA) as described previously.[10] The valve was created with a minimum of 13 fasteners, and was at least 1 cm long at either corner and 3 cm long in its mid-portion (Figure 1). Each participating surgeon submitted a video of a qualifying TF procedure that was reviewed and approved by Hunter and Bell before enrolling patients into the trial (Video 1). Patients in the control group had a sham procedure performed for 45–60 minutes, which included EGD for 30 minutes, and passage of a 50F Maloney dilator for 15 minutes, to simulate TF procedure and oropharyngeal irritation caused by TF.

Postoperative Care and Follow-Up

Patients were kept in the hospital overnight and were generally discharged the next day on omeprazole 40 mg for 14 days to help promote mucosal healing around fasteners if reflux control was incomplete. Thereafter, TF patients were switched to placebo, and sham patients were continued on omeprazole in an identical-appearing capsule. For the first 2 weeks postoperatively, patients were kept on a liquid diet. Soft foods were given from weeks 3 to 7, and a regular diet was reinstated 2 months after the operative procedure. Neither the patient nor their family was aware of allocation group until the 6-month point, or when they were declared failures and allowed to cross over to the other treatment arm. The perioperative caregivers (other than the operative team) were unaware of treatment allocation.

Follow-up occurred at weeks 2, 12, and 26 after TF or sham procedure. If troublesome symptoms of GERD recurred after 2 weeks, the medication dose was doubled (omeprazole 40 mg bid or placebo bid). If troublesome symptoms persisted at 3 months, despite bid medication use, the patient was declared a failure and the blind was broken. Once the blind was broken, failed TF patients were given PPI and sham patients were offered TF both for ethical reasons and to make study enrollment more attractive to potential participants (Figure 2).

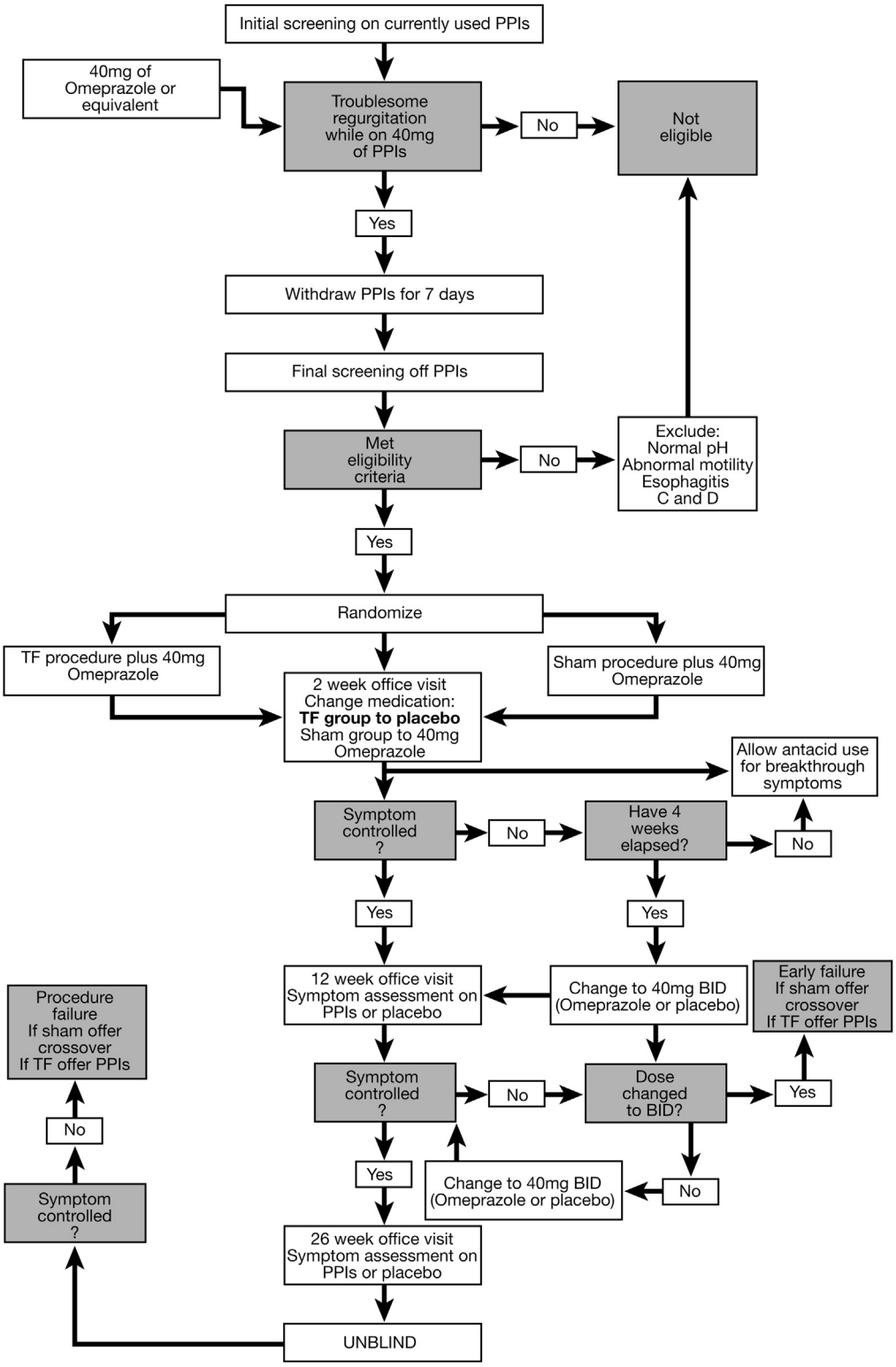


Figure 2. Study flow chart.

Six-month follow-up included repeating the 3 questionnaires on and off medication (PPI or placebo), 48-hour esophageal pH monitoring off medications (7 days), and EGD. After completion of these evaluation steps, the study was considered complete. Symptomatic sham patients were offered the opportunity to cross over to TF and TF patients with troublesome symptoms were offered PPIs.

Primary and Secondary End Points

The primary study end point was the elimination of troublesome regurgitation, per Montreal consensus definition, on placebo (TF group) and on PPI (control group). The Montreal consensus defines troublesome symptoms as mild symptoms occurring 2 or more days a week, or moderate to severe symptoms occurring more than 1 day a week.[2] The elimination

of troublesome regurgitation was evaluated with the RDQ. This instrument asks 12 questions addressing the symptom domains of heartburn, regurgitation, and dyspepsia using a scale from 0 to 5 to rate the severity and frequency of 6 symptoms.[11] A severity score of 2 or more and a frequency score of 3 or more for the regurgitation questions were required to meet the Montreal consensus criteria for troublesome regurgitation, a threshold supported by a recent analysis of the impact of regurgitation on quality of life.[9] Our primary hypothesis was that the proportion of transoral fundoplication/placebo patients who are relieved of troublesome regurgitation will be statistically significantly greater than those randomized to the sham/PPI group.

Secondary end points included early failure (defined as moderate to severe regurgitation at any time >12 weeks after surgery and after a doubling of medication, PPI, or placebo) and control of intraesophageal acid exposure. Other secondary outcomes assessed included improvement in various symptom scores (particularly heartburn), healing of esophagitis, common side effects associated with treatment (bloating and dysphagia), and significant adverse events.

Statistical Analysis

A sample size of 80 TF/placebo and 40 sham/PPI patients was necessary for an 85% power to detect a significant difference between the 2 treatment groups. Sample size was determined assuming a 30% greater elimination of troublesome regurgitation with TF as compared with PPI, based on previous reports.[12] The primary study end point of elimination of troublesome regurgitation was assessed using a χ^2 test. Binary secondary outcomes were also assessed with a χ^2 test, and continuous outcomes used a Wilcoxon matched pairs test for comparisons between screening and end-of-study values or a Mann-Whitney U test for comparisons between groups. Spearman ρ statistics was used to estimate correlation between postoperative pH parameters and symptom control as assessed by the quality of life symptom scores.

The primary end point was analyzed using the intent-to-treat population (ITT) and per protocol population. For the ITT analysis, a patient was declared a treatment failure if the 3-month and 6-month follow-up visits were not completed.[13] If a patient reported insufficient control of regurgitation on increased dose of medication at 3-month visit, and missed the 6-month visit, the patient was considered a treatment failure.

Results

Patient Population

Between June 2011 and September 2013, there were 3388 initial contacts made, mostly through web-based announcement of the trial. Six hundred and ninety-six patients were screened for eligibility and 567 were excluded. The most frequent reasons for exclusion were the presence of hiatal hernia >2 cm, absence of troublesome regurgitation, normal esophageal pH monitoring, and long segment Barrett's esophagus (Supplementary Figure 1). One hundred and twenty-nine patients were randomized, underwent sham surgery or TF, and were analyzed using the ITT population. Upon review of the entry criteria, 10 patients were excluded after treatment (6 in the TF arm and 4 in the sham arm), because they did not meet the entry criteria of troublesome regurgitation, as defined by Montreal criteria (8 patients), or did not have an abnormal pH study (2 patients). Of these 10 patients, 2 of 6 (33%) in the TF/placebo group and 2 of 4 (50%) in the sham/placebo group were declared early failures ($P > .999$). These patients did not receive 6-month follow-up with questionnaires and testing. Therefore, the PP analysis includes 81 TF/placebo and 38 sham/PPI patients. One patient in each group was lost to follow-up. The baseline and disease-related characteristics of the ITT study population are shown in Table 1.

Procedure

The mean operating time for TF was 49 minutes (range, 21–119 minutes). A mean of 23 fasteners was used (range, 13–37). As assessed by immediate post-procedure endoscopy, performance of 270-degree fundoplication (range, 200–340 degrees) resulted in the conversion of Hill grade 2 and 3 valves to Hill grade 1 in 79 of 82 (96%) patients. At discharge, epigastric pain was the only symptom that occurred more commonly in the TF than the sham group (34 of 83 vs 8 of 40; $P = .026$). Significant adverse events occurred in 7 patients in the TF/placebo group, and 1 in the sham/PPI group (Table 2). None of these events led to additional procedures, and all resolved without residual effect. Two patients with prolonged epigastric pain where treated with over-the-counter pain medication and did not report pain 4 weeks after TF.

Follow-Up and Early Failure (Intention to Treat)

At 3 months follow-up, 15 of 42 patients (36%) in the sham group met criteria for early failure, and 12 of 15 patients (80%) underwent crossover to TF. The 3 sham patients who had not crossed over completed the 6-month follow-up testing. In the TF/placebo group 10 of 87 patients (11%) met the criteria for early failure ($P = .002$) and all 10 returned to PPI treatment. Four of these 10 patients completed their 6-month follow-up testing. In total, 28 sham patients and 76 TF patients completed 6-month evaluation (Supplementary Figure 1).

Primary Outcomes

In the ITT analysis at 6-month follow-up, 58 of 87 (67%) TF/placebo patients reported the elimination of troublesome regurgitation vs 19 of 42 (45%) patients in the sham/PPI arm ($P = .023$).

The PP analysis revealed similar outcomes; 54 of 81 (67%) patients in the TF/placebo arm reported the elimination of troublesome regurgitation, and 17 of 38 (45%) patients in the sham/PPI arm reported elimination of troublesome regurgitation ($P = .028$).

Secondary Outcomes

As measured with the RDQ in those patients completing their 6-month follow up, TF provided equivalent improvement in symptom scores to sham/PPI on medication (Figure 3). TF provided greater reduction in heartburn and

Table 1. Demographics and Baseline Characteristics of the Study Patients

Variables	TF/placebo (n = 87)	Sham/PPI (n = 42)	*P* value
Female, n (%)[a]	40 (45.9)	26 (61.9)	.096
Age, *y*, median (range)	52 (22–74)	55 (22–73)	.513
50 y, n (%)[a]	35 (40.2)	13 (30.9)	.337
50–65 y, n (%)[a]	43 (49.4)	25 (59.5)	.348
>65 y, n (%)[a]	9 (10.3)	4 (9.6)	>.999
Body mass index, median (range)	27.1 (20.3–35.5)	27.8 (20.4–38.9)	.326
<25, n (%)	22 (25.3)	10 (24.3)	>.999
25–30, n (%)	45 (51.7)	19 (45.2)	.574
>30, n (%)	20 (23.0)	13 (30.5)	.391
GERD symptom duration, *y*, median (range)	10 (0.6–37)	10 (0.9–38)	.546
PPI therapy duration, *y*, median (range)	9 (1–30)	8 (1–23)	.541
Esophagitis (Los Angeles grade), n (%)[a]	17 (19.5)	6 (14.3)	.625
A[a]	10 (58.8)	3 (50.0)	>.999
B[a]	7 (41.2)	3 (50.0)	>.999
Hill grade, n (%)[a,b]	86 (98.8)	41 (97.6)	.547
I[a]	4 (4.6)	5 (12.2)	.147
II[a]	57 (66.3)	26 (63.4)	.842
III[a]	25 (29.1)	10 (24.4)	.674
Hiatal hernia, n (%)[a]	60 (69.8)	29 (69.0)	>.999
Axial length ≤1 cm[a]	33 (55.0)	18 (62.1)	.649
Axial length >1 cm and ≤2 cm[a]	27 (45.0)	11 (37.9)	.649
GTD, n (%)[a]			
≤1 cm	20 (33.9)[b]	13 (46.4)[b]	.345
>1 cm and ≤2 cm[a]	36 (61.0)[b]	15 (53.6)[b]	.642
>2 cm[a]	3 (5.1)[b]	0 (0)[b]	.548
RDQ score, median (range)			
On PPIs	2.8 (1.1–4.8)	3.3 (0.9–5.0)	.094
Off PPIs (n = 85 TIF; n = 40 sham)	3.3 (1.2–5.0)	3.6 (0.6–5.0)	.085
GERD-HRQL score, median (range)			
On PPIs	25 (0–41)	27 (7–45)	.108
Off PPIs (n = 85 TIF; n = 40 sham)	29 (3–47)	31 (9–50)	.450
GERSS, median (range)			
On PPIs	22 (3–54)	27 (8–56)	.052
Off PPIs (n = 85 TIF; n = 40 sham)	30 (5–60)	34 (9–60)	.185

NOTE. Esophagitis, Hill grade were evaluated with screening endoscopy. Hiatal hernia size was graded with videofluoroscopy. *P* values were calculated using Mann-Whitney *U* test unless indicated otherwise.
GERD-HRQL, Gastroesophageal Reflux Disease Health-Related Quality of Life; GERSS, Gastroesophageal Reflux Symptom Score; GTD, greatest transverse dimension; RDQ, Reflux Disease Questionnaire.
[a]Two-tailed Fisher's exact test.
[b]One patient in the transoral fundoplication (TF)/placebo groupo and one patient in the sham/PPI group have a missing data point.

Table 2. Significant Adverse Events

Randomization group	Significant adverse event	Maximum severity	Onset after procedure	Duration
Sham	Nausea	Severe	PPD 1	2 Days
TF	Temporary epigastric /abdominal pain	Severe	PPD 5	2 Weeks
	Chest pain	Severe	PPD 5	3 Days
	Musculoskeletal pain	Severe	PPD 1	1 Day
	Temporary epigastric /abdominal pain	Moderate	PPD 1	4 Weeks
	Dysphagia	Moderate	PPD 1	8 Days
	Dysphagia	Mild	PPD 1	1 Day
	Nausea	Mild	PPD 1	1 Day

NOTE. Per-protocol definition, the events reported were classified as serious adverse events as they required in-patient hospitalization or prolonged hospitalization. All reported serious adverse events resolved without residual effect.
PPD, post-procedure day; TF, transoral fundoplication.

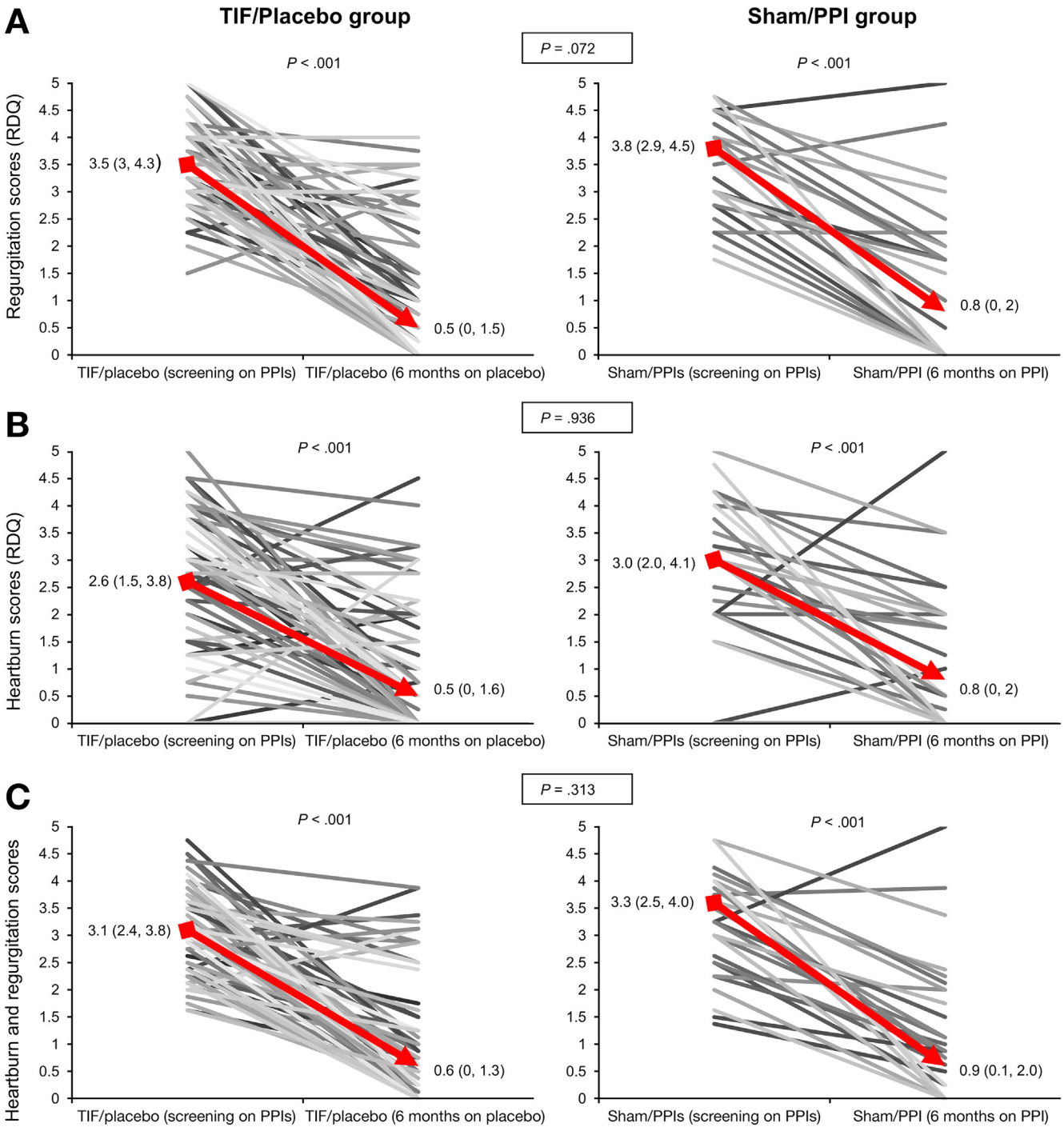


Figure 3. (*A*) Individual total regurgitation scores on placebo (TF group) and on PPI (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. (*B*) Individual total heartburn scores on placebo (TF group) and on PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. (*C*) Individual total composite heartburn and regurgitation scores on placebo (TF group) and on PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. All scores were assessed using RDQ. *Red lines* represent improvement in the median (25%, 75% quartiles) scores. The *P* values in *boxes* represent comparisons between treatment groups.

regurgitation scores than the sham group off medication (Supplementary Figure 2).

TF was associated with significant decrease in intra-esophageal acid exposure in all parameters measured (Figure 4). Mean number of reflux episodes fell from 135 before TF to 94 after TF ($P < .001$). Mean percent total time pH <4 improved from 9.3 before TF to 6.4 after TF ($P < .001$). Mean DeMeester score fell from 33.6 before TF to 23.9 after TF ($P < .001$). Of these 3 measures, only the number of reflux episodes was normalized by the

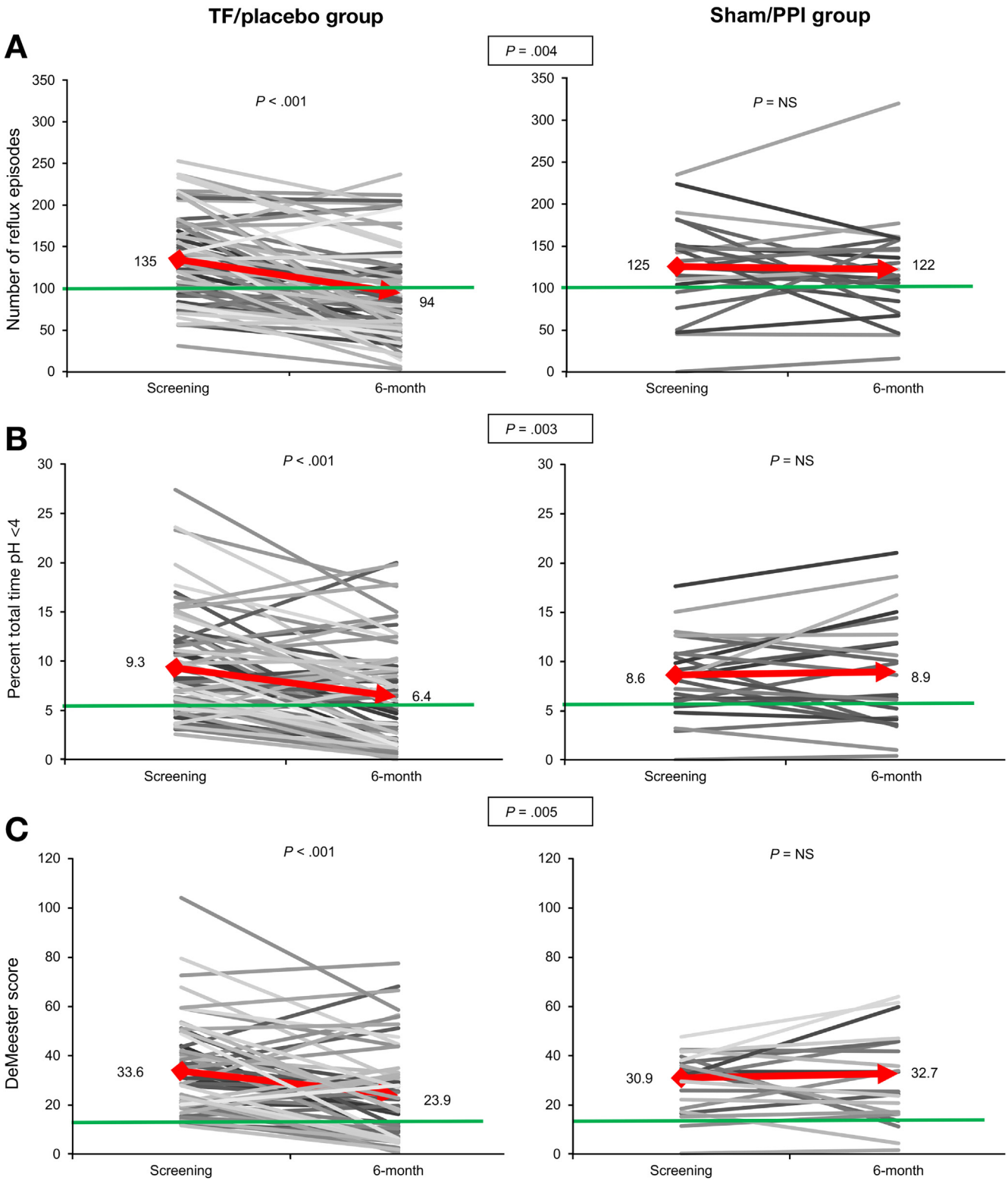


Figure 4. Total number of reflux episodes (*A*), percent time pH <4 (*B*), and DeMeester Score (*C*) were improved in TF/placebo group, but not in sham PPI group. The *red lines* represent change in mean scores. The *green lines* represent the cut-off for the normal values (reflux episodes = 100, percent total time pH <4 = 5.3% and DeMeester Score = 14.72). The *P* values in *boxes* represent comparisons between treatment groups.

performance of TF. After sham surgery, no improvement in pH control was detected, as measured with 48-hour pH testing off PPIs for 7 days. Mean number of reflux episodes were 125 before sham surgery and 122 after sham surgery (P = NS). Mean percent total time pH <4 was 8.6 before sham surgery and 8.9 after sham surgery (P = NS). Mean

DeMeester score was 30.9 before sham surgery and 32.7 after sham surgery (P = NS).

EGD revealed esophagitis in the minority of patients at baseline (17 in the TF group and 6 in the sham group). Of 17 patients in the transoral fundoplication/placebo group who had esophagitis on screening, 13 (76%) underwent endoscopy at 6 months. Reflux esophagitis was healed in 10 of 13 (77%); esophagitis improved from grade B to A in an additional 2 patients; in the last patient grade B esophagitis remained unchanged. In the sham/PPI group, of 6 patients with esophagitis on screening, 2 (50%) underwent endoscopy at 6 months. Esophagitis was healed in 1 patient (50%) improved from grade B to grade A in the other. At 6-month follow-up, de novo esophagitis was present in 4 TF/placebo patients (3 grade B and 1 grade A) and 5 sham/PPI patients (3 grade B and 2 grade A) (P = NS).

With the exception of postoperative epigastric pain, complications, and adverse effects were not different between TF and sham groups. On medication, bloating and dysphagia, as evaluated by Gastroesophageal Reflux Symptom Score, were improved after the procedure in both TF and sham groups (Supplementary Figure 3). One patient in the TF group and 2 patients in the sham group developed de novo dysphagia.

Including the early failures mentioned and follow-up out to 18 months, 30 of 42 patients (71%) in the sham/PPI arm have crossed over to TF. Including the early failures of TF mentioned, 24 of 87 (28%) in the TF/placebo group have resumed PPI ($P < .001$).

Discussion

A variety of endoscopic devices have been introduced to treat GERD over the past 2 decades. Most of these have been removed from the market because they were ineffective or unsafe. The only device available over the past 5 years that is capable of creating an antireflux valve endoscopically is the EsophyX device. Several case series and several registry reports have guided the evolution of the TF technique with this device.[7,12,14–16] As is common, early case series of this new surgical technique had mixed results, but as more experience was gained with the procedure, outcomes improved, and the number of related complications decreased.[6] One relevant trend observed was that procedures using fewer fasteners were associated with less favorable outcomes,[17] an observation that led us to use a mean of 23 fasteners in this series. A recent open-label randomized controlled trial comparing PPI treatment with TF demonstrated benefit for TF over PPI in control of troublesome GERD symptoms, with 54% of patients achieving normalization of intra-esophageal pH off PPI after TF. Similar pH normalization was achieved with high-dose PPI (on high-dose PPI), but GERD symptoms, particularly regurgitation and atypical symptoms, were better controlled with TF than with high-dose PPI.[8]

The Montreal definition of reflux is either mucosal damage or troublesome symptoms attributable to reflux. Consistent with this, we used the elimination of troublesome regurgitation (defined as that of sufficient magnitude to impair quality of life), rather than an improvement in regurgitation score as our primary end point. This approach has been recommended in previous published literature on assessing regurgitation in GERD management.[9,18] The primary end point in this study, elimination of troublesome regurgitation, was achieved in a greater proportion of patients treated with TF than with omeprazole: 67% vs 45%. That the reduction in composite symptom scores associated with treatment show no statistical difference between treatment groups at 6 month (Figure 3) is potentially confusing because these comparisons do not include data from the early failures, a group that was overrepresented in the sham/PPI treatment arm. Additionally, reduction in a symptom score is not measuring the same thing as the elimination of a troublesome symptom, and might yield different results, even if the populations queried were identical.

Secondary end points included response of other symptoms to TF, using well-validated questionnaires, and objective testing (48-hour esophageal pH monitoring and EGD). Evidence that TF was effective at improving GERD symptoms, heartburn, and regurgitation was well demonstrated with the improvement in 6-month RDQ scores as compared with baseline scores (Supplementary Figure 2). Improvement of intra-esophageal acid control was greater after TF than sham (Figure 4). Some studies evaluating TF,[15] PPI therapy,[19] and traditional laparoscopic fundoplication[20,21] demonstrated poor correlation between post-treatment pH parameters and symptom control, as evaluated with various disease-specific symptom scores. This study also found no significant correlation between objective and subjective outcomes in either treatment group (Supplementary Table 1). Although some studies have used pH normalization as a primary end point, the elimination of troublesome symptoms and the healing of reflux esophagitis are more clinically relevant goals of GERD treatment; symptom control might not require pH normalization. With traditional antireflux surgery, there has long been the concern that reflux control comes at the expense of new symptoms and side effects (primarily dysphagia and bloating). This did not appear to be the case in this study, as dysphagia and bloating scores were improved in both treatment groups, and new onset symptoms (dysphagia or bloating) were rare and evenly balanced between groups (Supplementary Figure 3).

Reflux esophagitis was healed in 77% of TF/placebo patients in this study, mirroring results from other recent reports from the United States.[7,15] However, these results must be interpreted from the perspective that this study was not designed to evaluate esophagitis healing and only a limited number of enrolled subjects had esophagitis at entry; 17 patients in the TF/placebo group and 6 patients in the sham/PPI group.

TF can fill the “therapeutic gap” that exists between PPI and laparoscopic fundoplication. Up to 40% of GERD patients have troublesome symptoms, despite adequately dosed PPI.[3] Although this group of patients might be treated with laparoscopic fundoplication or the LINX device,[22] the absence of hiatal hernia or advanced esophageal disease begs the question as to whether or not a less invasive and more calibrated treatment might be available to fill this gap. When comparing this trial with those using the LINX device,

in should be kept in mind that this trial was a prospective randomized trial, that patients in this study had less response to PPI therapy at baseline than in previously published case series of LINX use (Table 1), and that TF cannot effectively close a hiatal hernia, as is a part of the LINX procedure in many patients. Both interventions seem to have particular benefit in improving the symptom of regurgitation. Considering the virtual absence of dysphagia and bloating after TF, which can be problematic with LINX, it would appear that TF is an option for patients with troublesome regurgitation, as well as for patients with troublesome GERD symptoms who wish not to take PPI for a protracted period of time.

This study was not designed to evaluate the cost-effectiveness of TF compared with other treatments for chronic GERD. Currently, it is unclear if the benefit of TF would offset higher upfront cost of TF as compared with long-term PPI therapy. Higher upfront cost of TF can be offset by improvement in patients' quality of life and lower health care utilization in patients who do not fully respond to PPI therapy. Cost-effectiveness models can be developed from these and other data when longer term follow-up becomes available.

There are several limitations to this study. Our ITT analysis included 12 patients with limited follow-up data. Assessment of the primary end point at 6 months can be viewed as premature by some; however, we believed it likely that delaying the primary end point beyond 6 months would risk patients not entering or dropping out of the study prematurely. That 15 of 42 (36%) patients in the control group were early failures and 12 of these decided to cross over to TF is further evidence that they felt incompletely treated on escalating doses of PPI. Although there is a plan to follow both groups of patients beyond 6 months, the proof of efficacy was achieved in a 6-month window. Studies that have followed TF patients for more than 3 years have demonstrated little deterioration in the response measured shortly after operation.[16] Screening of interested patients eliminated about 81% of the patients who had GERD symptoms on PPI. The most frequent reason for exclusion was a hiatal hernia >2 cm, which eliminated 31% of those screened. TF has been shown to be capable of reducing hiatal hernias up to 2 cm in axial height, but patients with hiatal hernias >2 cm in height and troublesome GERD symptoms despite appropriate medical therapy should be considered for laparoscopic hiatal hernia repair with fundoplication.[23]

In this sham-controlled randomized controlled trial, transoral fundoplication was effective in eliminating troublesome GERD symptoms, especially regurgitation, with a low failure rate and good safety profile for 6 months. We believe TF has a role in treating GERD patients with small or absent hiatal hernia who suffer from troublesome regurgitation despite PPI therapy.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2014.10.009.

References

1. Rotman SR, Bishop TF. Proton pump inhibitor use in the US ambulatory setting, 2002–2009. PloS One 2013; 8:e56060.
2. Vakil N, van Zanten SV, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101:1900–1920; quiz 43.
3. Fass R. Proton pump inhibitor failure—what are the therapeutic options? Am J Gastroenterol 2009;104(Suppl 2): S33–S38.
4. Liu JY, Finlayson SR, Laycock WS, et al. Determining an appropriate threshold for referral to surgery for gastroesophageal reflux disease. Surgery 2003;133:5–12.
5. Finks JF, Wei Y, Birkmeyer JD. The rise and fall of antireflux surgery in the United States. Surg Endosc 2006; 20:1698–1701.
6. Wendling MR, Melvin WS, Perry KA. Impact of transoral incisionless fundoplication (TIF) on subjective and objective GERD indices: a systematic review of the published literature. Surg Endosc 2013;27:3754–3761.
7. Wilson EB, Barnes WE, Mavrelis PG, et al. The effects of transoral incisionless fundoplication on chronic GERD patients: 12-month prospective multicenter experience. Surg Laparosc Endosc Percutan Tech 2014;24:36–46.
8. Trad KS, Barnes WE, Simoni G, et al. Transoral incisionless fundoplication effective in eliminating GERD symptoms in partial responders to proton pump inhibitor therapy at 6 months: the TEMPO Randomized clinical trial. Surg Innov 2014 Apr 21 [Epub ahead of print].
9. Kahrilas PJ, Jonsson A, Denison H, et al. Impact of regurgitation on health-related quality of life in gastro-oesophageal reflux disease before and after short-term potent acid suppression therapy. Gut 2014;63: 720–726.
10. Bell RC, Cadiere GB. Transoral rotational esophagogastric fundoplication: technical, anatomical, and safety considerations. Surg Endosc 2011;25:2387–2399.
11. Shaw M, Talley NJ, Beebe T, et al. Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. Am J Gastroenterol 2001;96:52–57.
12. Cadiere GB, Buset M, Muls V, et al. Antireflux transoral incisionless fundoplication using EsophyX: 12-month results of a prospective multicenter study. World J Surg 2008;32:1676–1688.
13. Food and Drug Administration. Statistical guidance for clinical trials of non diagnostic medical devices. Available at: www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm106757.htm. Accesed August 25, 2014.
14. Testoni PA, Vailati C, Testoni S, et al. Transoral incisionless fundoplication (TIF 2.0) with EsophyX for gastroesophageal reflux disease: long-term results and findings affecting outcome. Surg Endosc 2012;26:1425–1435.
15. Bell RC, Mavrelis PG, Barnes WE, et al. A prospective multicenter registry of patients with chronic

gastroesophageal reflux disease receiving transoral incisionless fundoplication. J Am Coll Surg 2012;215:794–809.
16. Muls V, Eckardt AJ, Marchese M, et al. Three-year results of a multicenter prospective study of transoral incisionless fundoplication. Surg Innov 2013;20:321–330.
17. Testoni PA, Corsetti M, Di Pietro S, et al. Effect of transoral incisionless fundoplication on symptoms, PPI use, and ph-impedance refluxes of GERD patients. World J Surg 2010;34:750–757.
18. Kahrilas PJ, Howden CW, Hughes N. Response of regurgitation to proton pump inhibitor therapy in clinical trials of gastroesophageal reflux disease. Am J Gastroenterol 2011;106:1419–1425.
19. Milkes D, Gerson LB, Triadafilopoulus G. Complete elimination of reflux symptoms does not guarantee normalization of intraesophageal and intragastric pH in patients with gastroesophageal reflux disease (GERD). Am J Gastroenterol 2004;99:991–996.
20. Markus PM, Horstmann O, Kley C, et al. Laparoscopic fundoplication. Surg Endosc 2002;16:48–53.
21. Velanovich V, Karmy-Jones R. Measuring gastroesophageal reflux disease: relationship between the Health-Related Quality of Life score and physiologic parameters. Am Surg 1998;64:649–653.
22. Bonavina L, DeMeester TR, Ganz RA. LINX(TM) Reflux Management System: magnetic sphincter augmentation in the treatment of gastroesophageal reflux disease. Expert Rev Gastroenterol Hepatol 2012;6:667–674.
23. Boeckxstaens G, El-Serag HB, Smout AJ, et al. Symptomatic reflux disease: the present, the past and the future. Gut 2014;63:1185–1193.

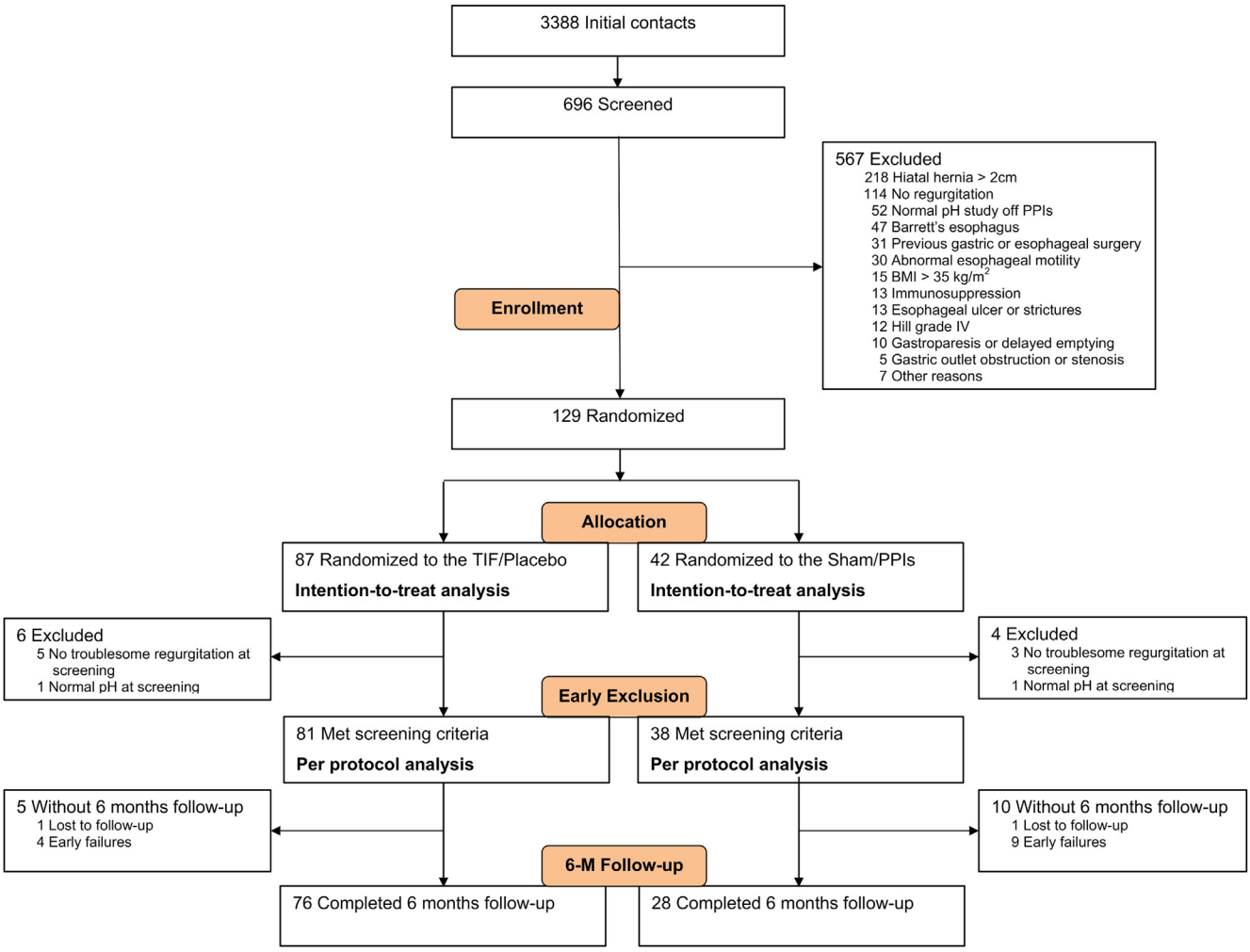
Received July 16, 2014. Accepted October 7, 2014.

Reprint requests
Address requests for reprints to: John G. Hunter, MD, FACS, Department of Surgery, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Mail Code L223A, Portland, Oregon 97239. e-mail: hunterj@ohsu.edu; fax: (503) 494-5615.

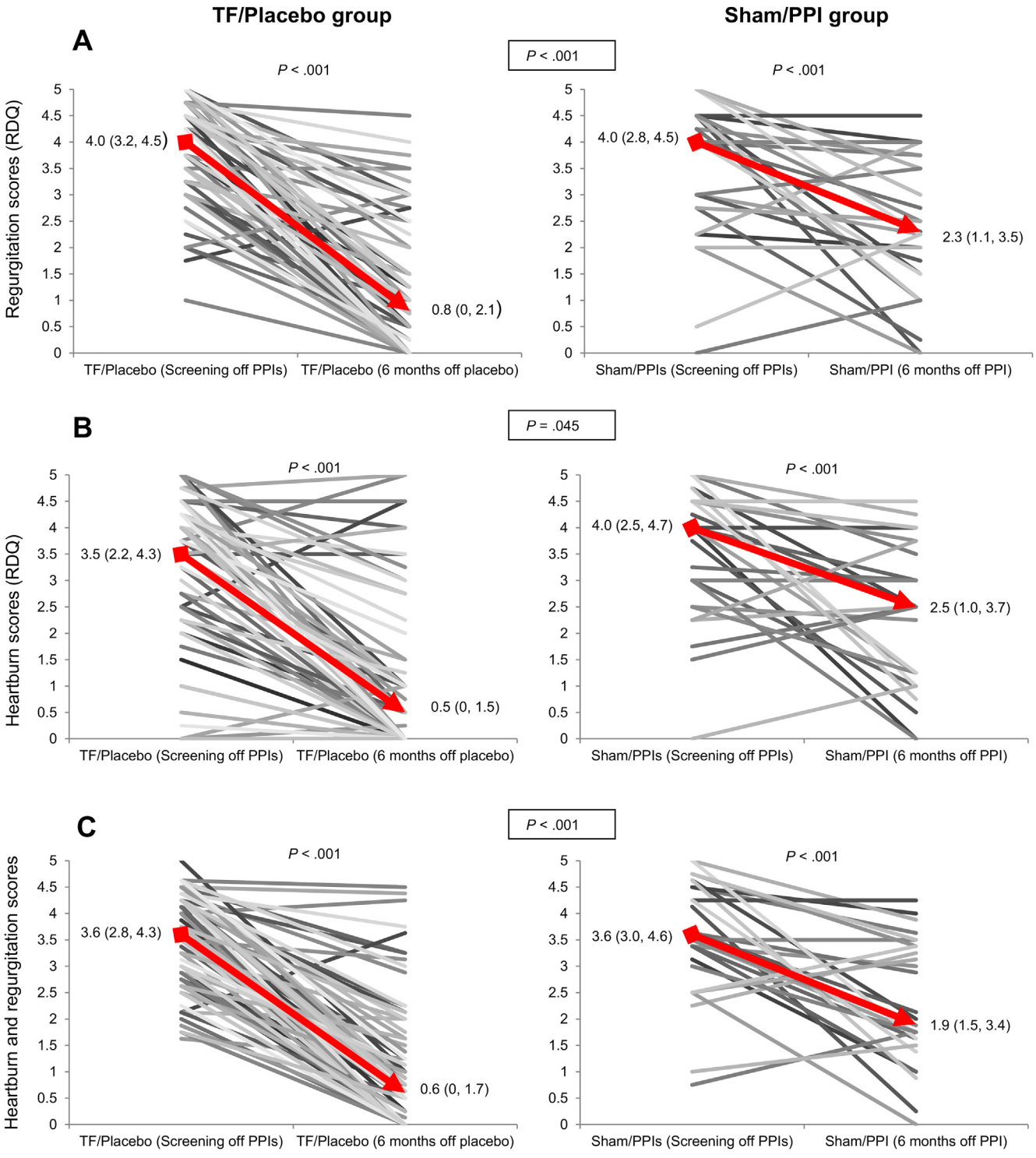
Acknowledgments
The abstract of this work was Presented at the 2014 Annual Scientific Meeting of American College of Gastroenterology, October 21, 2014, Philadelphia, Pennsylvania.

Conflicts of interest
These authors disclose the following: John Hunter is a consultant for EndoGastric Solutions. Peter Kahrilas is a consultant for AstraZeneca, Pfizer, and Trimedyne; has been a consultant for GlaxoSmithKline; and has been on advisory boards for Torax Medical and Reckitt Benckiser. Reginald Bell has received research grant from EndoGastric Solutions. Erik Wilson is a consultant for Apollo, Gore Medical, and Ethicon. Karim Trad has acted as a speaker bureau member and has received speaking honoraria from EndoGastric Solutions. Brant Oelschlager is a consultant and has received a research grant from EndoGastric Solutions. Kevin Reavis is a consultant for EndoGastric Solutions. Eric Hungness received an honorarium as part of being Northwestern University faculty for a surgical training course with Baxter. The remaining authors disclose no conflicts.

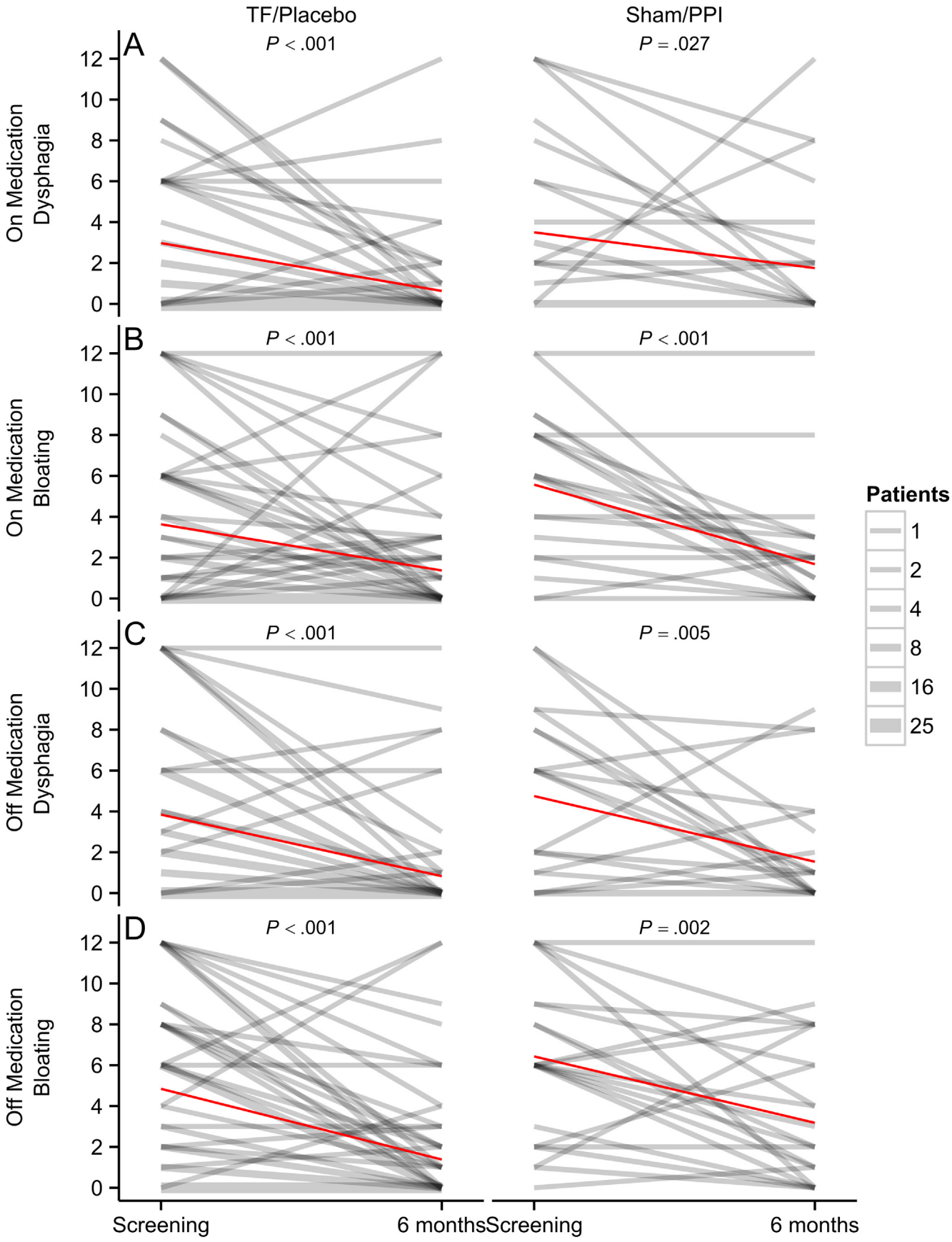
Funding
This study was sponsored by EndoGastric Solutions, Redmond, WA. The study was designed by Kahrilas and Hunter in collaboration with the sponsor. The sponsor monitored the study conduct. The statistical analyses were performed independently by Brian S. Diggs, PhD.



Supplementary Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.



Supplementary Figure 2. (*A*) Individual regurgitation scores off placebo (TF group) and off PPIs (sham group) undergoing symptomatic assessment before treatments and at 6-month follow-up. (*B*) Individual heartburn scores off placebo (TF group) and off PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. (*C*) Individual composite heartburn and regurgitation scores off placebo (TF group) and off PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. All scores were assessed using RDQ. *Red lines* represent improvement in the median (25%, 75% quartiles) scores. The *P* values in *boxes* represent comparisons between treatment groups.


TF/Placebo
Sham/PPI
A
P < .001
P = .027
On Medication Dysphagia
B
P < .001
P < .001
On Medication Bloating
C
P < .001
P = .005
Off Medication Dysphagia
D
P < .001
P = .002
Off Medication Bloating
0
2
4
6
8
10
12
Screening
6 months
Patients
1
2
4
8
16
25

Supplementary Figure 3. (*A*) Individual dysphagia scores on placebo (TF group) and on PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. (*B*) Individual bloating scores on placebo (TF group) and on PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. (*C*) Individual dysphagia scores off placebo (TF group) and off PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. (*D*) Individual bloating scores off placebo (TF group) and off PPIs (sham group) in all patients undergoing symptomatic assessment before treatments and at 6-month follow-up. All scores were assessed using Gastroesophageal Reflux Symptom Score. Thickness of lines represents the number of patients with the same initial and final values. *Red lines* indicate the overall trend connecting the mean at screening and follow-up. *P* values are from Wilcoxon matched pairs tests.

Supplementary Table 1. Correlation Between pH Parameters and Symptom Scores in Both Treatment Groups

Parameters	Regurgitation	Heartburn	R&H	DMS	% Total time	NORE
TF group off placebo						
Regurgitation	1.00					
Heartburn	0.59 (<.001)	1.00				
R&H	0.91 (<.001)	0.85 (<.001)	1.00			
DMS	0.02 (.839)	0.14 (.249)	0.09 (.439)	1.00		
% Total time	0.01 (.889)	0.15 (.213)	0.09 (.433)	0.99 (<.001)	1.00	
NORE	0.03 (.771)	0.08 (.518)	0.07 (.556)	0.82 (<.001)	0.84 (<.001)	1.00
Sham group off Omeprazole						
Regurgitation	1.00					
Heartburn	0.50 (.009)	1.00				
R&H	0.83 (<.001)	0.86 (<.001)	1.00			
DMS	−0.08 (.695)	−0.09 (.639)	−0.12 (.552)	1.00		
% Total time	0.01 (.989)	−0.07 (.718)	−0.06 (.738)	0.98 (<.001)	1.00	
NORE	−0.03 (.869)	−0.04 (.831)	−0.08 (.696)	0.69 (<.001)	0.73 (<.001)	1.00

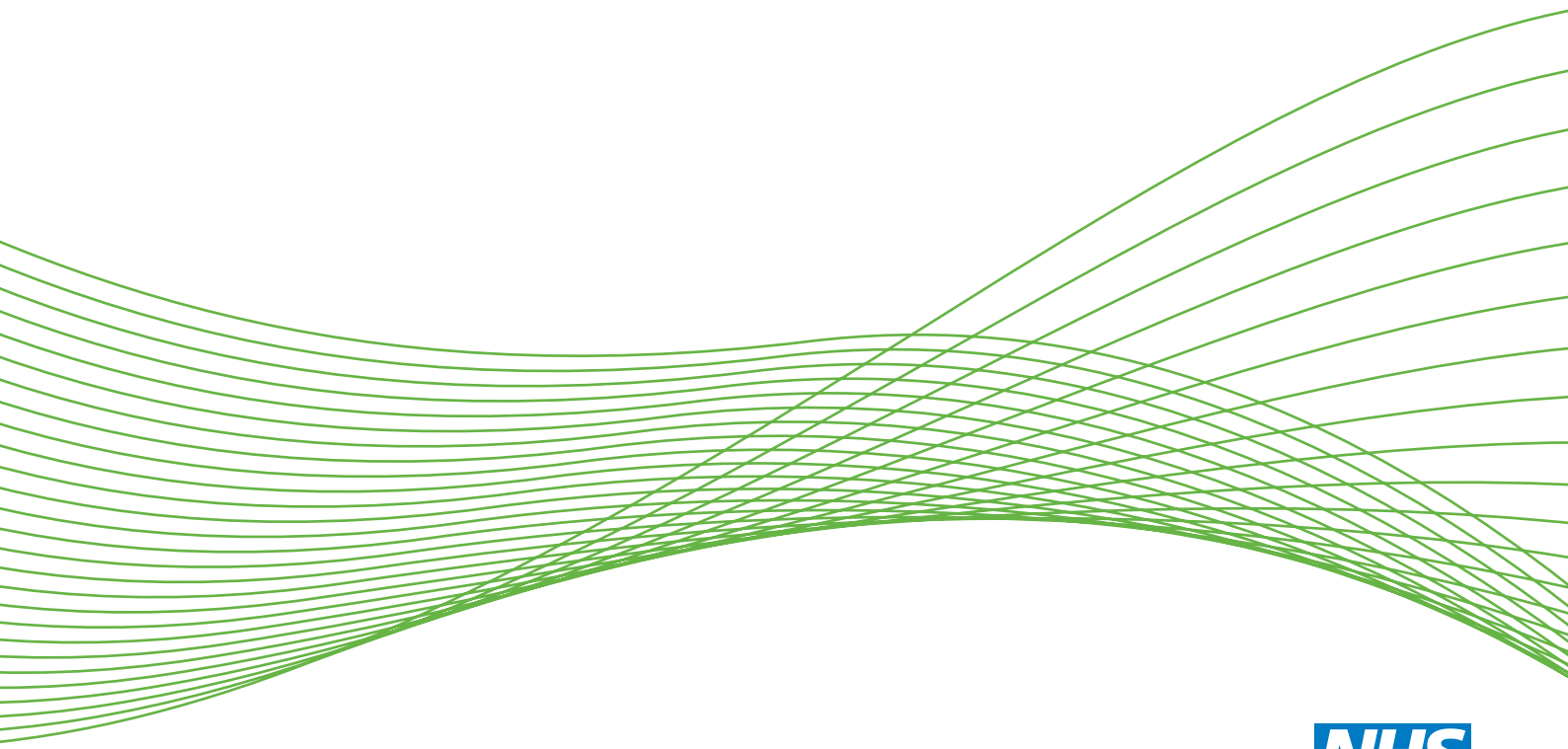
NOTE. Values are Spearman's ρ (P value).
DMS, DeMeester score; NORE, number of reflux episodes; R&H, regurgitation and heartburn composite score.
Symptom scores were assessed using RDQ.

HEALTH TECHNOLOGY ASSESSMENT

VOLUME 17 ISSUE 22 JUNE 2013
ISSN 1366-5278

Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial)

AM Grant, C Boachie, SC Cotton, R Faria, L Bojke, DM Epstein, CR Ramsay, B Corbacho, M Sculpher, ZH Krukowski, RC Heading and MK Campbell on behalf of the REFLUX trial group



National Institute for Health Research

DOI 10.3310/hta17220

Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial)

AM Grant,[1]* C Boachie,[1] SC Cotton,[1] R Faria,[2] L Bojke,[2] DM Epstein,[2] CR Ramsay,[1] B Corbacho,[2] M Sculpher,[2] ZH Krukowski,[3] RC Heading[4] and MK Campbell[1] on behalf of the REFLUX trial group

[1]Health Services Research Unit, Health Sciences Building, University of Aberdeen, Aberdeen, UK
[2]Centre for Health Economics, University of York, York, UK
[3]Aberdeen Royal Infirmary, Aberdeen, UK
[4]School of Medicine and Health, Wolfson Research Institute, Durham University, Stockton-on-Tees, UK

*Corresponding author

Declared competing interests of authors: Adrian Grant has received salary support from the NIHR as director of the NIHR Programme Grants for Applied Research programme.

Published June 2013
DOI: 10.3310/hta17220

This report should be referenced as follows:

Grant AM, Boachie C, Cotton SC, Faria R, Bojke L, Epstein DM, *et al*. Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial). *Health Technol Assess* 2013;**17**(22).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (*SciSearch*®) and *Current Contents*®/Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.596

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 97/10/99. The contractual start date was in May 2007. The draft report began editorial review in October 2011 and was accepted for publication in July 2012. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial)

AM Grant,[1]* C Boachie,[1] SC Cotton,[1] R Faria,[2] L Bojke,[2] DM Epstein,[2] CR Ramsay,[1] B Corbacho,[2] M Sculpher,[2] ZH Krukowski,[3] RC Heading[4] and MK Campbell[1] on behalf of the REFLUX trial group

[1]Health Services Research Unit, Health Sciences Building, University of Aberdeen, Aberdeen, UK
[2]Centre for Health Economics, University of York, York, UK
[3]Aberdeen Royal Infirmary, Aberdeen, UK
[4]School of Medicine and Health, Wolfson Research Institute, Durham University, Stockton-on-Tees, UK

*Corresponding author a.grant@abdn.ac.uk

Background: Despite promising evidence that laparoscopic fundoplication provides better short-term relief of gastro-oesophageal reflux disease (GORD) than continued medical management, uncertainty remains about whether benefits are sustained and outweigh risks.

Objective: To evaluate the long-term clinical effectiveness, cost-effectiveness and safety of laparoscopic surgery among people with GORD requiring long-term medication and suitable for both surgical and medical management.

Design: Five-year follow-up of a randomised trial (with parallel non-randomised preference groups) comparing a laparoscopic surgery-based policy with a continued medical management policy. Cost-effectiveness was assessed alongside the trial using a NHS perspective for costs and expressing health outcomes in terms of quality-adjusted life-years (QALYs).

Setting: Follow-up was by annual postal questionnaire and selective hospital case notes review; initial recruitment in 21 UK hospitals.

Participants: Questionnaire responders among the 810 original participants. At entry, all had documented evidence of GORD and symptoms for >12 months. Questionnaire response rates (years 1–5) were from 89.5% to 68.9%.

Interventions: Three hundred and fifty-seven participants were recruited to the randomised comparison (178 randomised to surgical management and 179 randomised to continued medical management) and 453 to the preference groups (261 surgical management and 192 medical management). The surgeon chose the type of fundoplication.

Main outcome measures: Primary: disease-specific outcome measure (the REFLUX questionnaire); secondary: Short Form questionnaire-36 items (SF-36), European Quality of Life-5 Dimensions (EQ-5D), NHS resource use, reflux medication, complications.

Results: The randomised groups were well balanced. By 5 years, 63% in the randomised surgical group and 13% in the randomised medical management group had received a total or partial wrap fundoplication (85% and 3% in the preference groups), with few perioperative complications and no associated deaths. At 1 year (and 5 years) after surgery, 36% (41%) in the randomised surgical group – 15% (26%) of those who had surgery – were taking proton pump inhibitor medication compared with 87% (82%) in the randomised medical group. At each year, differences in the REFLUX score significantly favoured the randomised surgical group (a third of a SD; $p<0.01$ at 5 years). SF-36 and EQ-5D scores also favoured surgery, but differences attenuated over time and were generally not statistically significant at 5 years. The worse the symptoms at trial entry, the larger the benefit observed after surgery. Those randomised to medical management who subsequently had surgery had low baseline scores that markedly improved after surgery. Following fundoplication, 3% had surgical treatment for a complication and 4% had subsequent reflux-related operations – most often revision of the wrap. Dysphagia, flatulence and inability to vomit were similar in the two randomised groups. The economic analysis indicated that surgery was the more cost-effective option for this patient group. The incremental cost-effectiveness ratio for surgery in the base case was £7028 per additional QALY; these findings were robust to changes in approaches and assumptions. The probability of surgery being cost-effective at a threshold of £20,000 per additional QALY was >0.80 for all analyses.

Conclusions: After 5 years, laparoscopic fundoplication continues to provide better relief of GORD symptoms with associated improved health-related quality of life. Complications of surgery were uncommon. Despite being initially more costly, a surgical policy is highly likely to be cost-effective.

Trial registration: Current Controlled Trials ISRCTN15517081.

Funding: This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 17, No. 22. See the HTA programme website for further project information.

Contents

List of abbreviations

BMI	body mass index
CDSR	Cochrane Database of Systematic Reviews
CI	confidence interval
CONSORT	Consolidated Standards of Reporting Trials
DARE	Database of Abstracts of Reviews of Effects
DMC	Data Monitoring Committee
EQ-5D	European Quality of Life-5 Dimensions
EVPI	expected value of perfect information
GERSS	Gastro-Esophageal Reflux Symptom Score
GORD	gastro-oesophageal reflux disease
GP	general practitioner
GSRS	Gastrointestinal Symptoms Rating Scale
H_2RA	histamine receptor antagonist
HRQoL	health-related quality of life
HTA	Health Technology Assessment
HUI3	Health Utilities Index Mark 3
ICER	incremental cost-effectiveness ratio
ITT	intention to treat
LOTUS	LOng-Term Usage of esomeprazole versus Surgery for treatment of chronic GERD
MAR	missing at random
MCAR	missing completely at random
MCS	mental component score
MICE	multiple imputation using chained equations
MNAR	missing not at random
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMB	net monetary benefit
OLS	ordinary least squares
PCS	physical component score
PGWI	Psychological General Well-Being Index
PP	per protocol
PPI	proton pump inhibitor
QALY	quality-adjusted life-year
QoL	quality of life
QOLRAD	Quality of Life in Reflux and Dyspepsia
RCT	randomised controlled trial
SD	standard deviation
SF-36	Short Form questionnaire-36 items
SF-6D	Short Form questionnaire-6 dimensions

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

In the Health Technology Assessment (HTA)-commissioned REFLUX trial, laparoscopic fundoplication for people with chronic symptoms of gastro-oesophageal reflux disease (GORD) was shown to significantly improve reflux-specific and general health-related quality of life (HRQoL) at least up to 12 months after surgery. However, cost-effectiveness was uncertain without more reliable information about longer-term costs and benefits. Here, we report the findings from longer-term follow-up of the REFLUX trial.

Objective

To evaluate, at 5 years after surgery, the clinical effectiveness, cost-effectiveness and safety of a policy of relatively early laparoscopic surgery compared with continued medical management among people with GORD symptoms that are reasonably controlled by medication and who are judged suitable for both surgical and medical management.

Methods

Design

1. Long-term follow-up of a pragmatic randomised controlled trial (with parallel non-randomised preference groups) comparing a laparoscopic surgery-based policy with a continued medical management policy to assess relative clinical effectiveness.
2. An economic evaluation of laparoscopic surgery for GORD to compare the cost-effectiveness of the two management policies, based on a within-trial (5-year) economic analysis and exploration of the need for a longer-term model.

Setting

Participants had originally been recruited in 21 UK hospitals through local partnership between surgeon(s) and gastroenterologist(s) who shared the secondary care of patients with GORD. After operation (surgical groups) and after optimisation of anti-reflux therapy (medical groups), participants were returned to the care of their general practitioners (GPs). Follow-up was by annual postal questionnaire and selective case notes review when questionnaires indicated reflux-related health-care events.

Participants

Participants in this study were questionnaire responders among the 810 original participants. At trial entry, all had both documented evidence of GORD and symptoms for >12 months. Annual questionnaire response rates (years 1–5) were 89.5%, 77.5%, 76.7%, 69.8% and 68.9%.

Intervention

Of the 810 participants, 357 were recruited to the randomised comparison (178 randomised to surgical management and 179 randomised to continued medical management) and 453 to the parallel non-randomised preference arm (261 surgical management and 192 medical management). The type of fundoplication was left to the discretion of the surgeon.

Main outcome measures

The principal outcome measure was a disease-specific instrument (the REFLUX questionnaire developed specifically for this study). Secondary measures were the Short Form questionnaire-36 items (SF-36), the

European Quality of Life-5 Dimensions (EQ-5D), surgical events including complications, reflux medication use, GP visits, hospital outpatient consultations, day and overnight hospital admissions, and their costs.

Results

At entry to the original trial, participants had been taking GORD medication for a median of 32 months and had a mean age of 46 years, and 66% were men; the randomised groups had been well balanced. Responders at 5 years were older, had been on medication for a shorter time prior to trial entry and had higher baseline quality-of-life scores than non-responders; however, the randomised groups of responders were similar in baseline characteristics. Primary analyses were based on the 'intention-to-treat' (ITT) principle, with secondary per-protocol analyses based on those who, at 1 year, had received their allocated treatment.

By 5 years, 63% ($n = 112$) of the 178 randomised surgery participants and 13% ($n = 24$) of the 179 randomised medical management participants had actually received fundoplication (equivalent figures in the preference groups were 85% and 3%). There had been a mixture of clinical and personal reasons for those allocated surgery not receiving it, sometimes related to long waiting times. A total or partial wrap procedure had been performed depending on surgeon preference; perioperative complications had been uncommon with no deaths associated with surgery.

By the equivalent to 12 months after surgery, 36% in the randomised surgical group (15% among those who had surgery) were taking proton pump inhibitor medication compared with 87% in the randomised medical group. At 5 years, the equivalent figures were 41% (26%) in those randomised to surgery and 82% in those randomised to medical management.

At each year, there were significant differences in the REFLUX score (a third of a SD; $p < 0.01$ at 5 years) favouring the randomised surgical group, reflecting differences in general discomfort (particularly), wind and frequency, nausea and vomiting, and activity limitation subscores. SF-36 and EQ-5D scores also favoured the randomised surgical group, especially SF-36 norm-based general health, but differences attenuated over time and were generally not statistically significant at 5 years [EQ-5D difference (ITT) 0.047, 95% confidence interval (CI) –0.013 to 0.108; $p = 0.13$]. The lower the REFLUX score and hence the worse the symptoms at trial entry, the larger the benefit observed after surgery. Post hoc exploratory analyses showed that those randomly allocated to medical management who subsequently had surgery had worse symptoms (lower baseline scores) than those who continued on medical management as allocated; following surgery, the scores of these patients markedly improved and this explains, at least in part, why differences in outcome between the randomised groups became less marked over time.

The preference surgical group also had low REFLUX scores at baseline. These scores improved substantially after surgery and at 5 years they were slightly better than those in the preference medical group.

Overall, 4% ($n = 16$) of the total 364 in the study who had fundoplication had a subsequent reflux-related operation, of whom two had a further (i.e. third) operation. Reoperation was most often conversion to a different type of wrap or a reconstruction of the same wrap. There were only two cases of reversal of the fundoplication and neither was in the randomised comparison. In total, 3% ($n = 12$) of those who had fundoplication required surgical treatment for a complication directly related to the original surgery, including oesophageal dilatation ($n = 4$) and repair of incisional hernia ($n = 3$). Patterns of 'difficulty swallowing', flatulence and 'wanting to vomit but being physically unable to do so' – all problems that have previously been associated with anti-reflux surgery – were similar in the two randomised groups.

Economic evaluation

Differences in mean costs and mean quality-adjusted life-years (QALYs) at 5 years were used to derive an estimate of the cost-effectiveness of laparoscopic fundoplication and continued medical management from the perspective of the NHS. Conventional decision rules were used to estimate incremental cost-effectiveness ratios (ICERs). Sensitivity analysis (including probabilistic sensitivity analysis) was used to explore and quantify uncertainty in the cost-effectiveness results.

Health-care resource-use data were collected prospectively as part of the clinical report forms and patient questionnaires at each follow-up point. The cost for each individual patient in the trial was calculated by multiplying their use of NHS resources by the associated unit costs (from published sources) and discounting at an annual rate of 3.5%. For the base-case analysis, total costs constituted the costs of surgery, complications due to surgery, reoperations, reflux-related prescribed medication, reflux-related visits to and from the GP and reflux-related hospital inpatient, outpatient and day visits. For the sensitivity analysis, all GP visits and all hospital admissions were included in the calculation of total costs. Health outcomes were expressed in terms of QALYs. HRQoL was assessed at each follow-up point using the EQ-5D. Incremental mean QALYs between randomised treatment groups were estimated with and without adjustment for baseline utility, using ordinary least squares regression.

The extent of missing data throughout the trial follow-up was significant; for this reason, the base case drew on the multiple imputed data set ITT analysis. A separate scenario – the complete-case analysis, in which only participants who returned all questionnaires and completed all EQ-5D profiles are included – was employed for both ITT and per-protocol analyses. Multiple imputation provides unbiased estimates of treatment effect if data are missing at random. Sensitivity analysis was used to test the impact on the cost-effectiveness results if data were missing not at random, that is, if patients with worse outcomes or greater costs were more likely to have missing data.

The results show that, for the base-case analysis (multiple imputed data set), the participants randomised to fundoplication accrued greater costs (incremental mean cost £1518; 95% CI £1006 to £2029) but also reported greater overall HRQoL (incremental mean QALYs 0.2160; 95% CI 0.0205 to 0.4115) than participants randomised to continued medical management. Laparoscopic fundoplication is a cost-effective strategy for GORD patients eligible for the REFLUX trial on the basis of the range of cost-effectiveness thresholds used by the National Institute for Health and Care Excellence (NICE) (£20,000–30,000 per additional QALY). The results for the complete-case analysis concurred with the multiple imputed data set: across analyses adjusted and unadjusted for baseline EQ-5D, ICERs ranged between £5468 and £8410, well below the NICE cost-effectiveness thresholds. For both data sets (multiple imputation and complete case), the probability of surgery being the more cost-effective intervention was >0.82 for incremental analyses unadjusted for baseline EQ-5D and >0.93 once incremental QALYs were adjusted for baseline EQ-5D.

A sensitivity analysis was carried out comparing the groups according to their 'per-protocol' status at 1 year. A per-protocol analysis compares the efficacy of the treatments received, whereas an ITT analysis compares the effectiveness of the strategies as offered to patients. The per-protocol analysis (in complete cases) suggested that surgery was more cost-effective than medical management. Other sensitivity analyses were carried out using a wider set of resource-use data. The results of the first alternative scenario, using the costs of primary care visits for any reason rather than only reflux-related reasons, increased the ICER slightly in relation to the base case. Nevertheless, the ICER remains well below conventional thresholds, and the probability of surgery being cost-effective was >0.85 for both adjusted and unadjusted analyses. In the second alternative scenario, replacing reflux-related hospital costs by all hospital costs, medical management was 'dominated' by the surgical policy; the probability of surgery being cost-effective was >0.90.

The base-case analysis imputes missing data. This assumes that missing data are missing at random, that is, their values can be predicted (with uncertainty) from observed data. This assumption is impossible to confirm or repute but its effect on the results can be tested in sensitivity analysis. The base-case analysis may be biased if the values of a missing variable are different from the observed values (for given values of other covariates). Sensitivity analysis using the multiple imputation data set showed that the cost-effectiveness of surgery was relatively insensitive to any increase in costs: cost-effectiveness changed little when costs were increased for patients with missing data in both treatment groups and when costs were increased just for patients randomly allocated surgery with missing data. A similar result was observed after reducing the total QALYs for all patients with missing data. In contrast, the cost-effectiveness of surgery was highly sensitive to the assumption that patients randomly allocated surgery with missing data experience lower HRQoL than patients with complete data. A 10% decrease in QALYs for patients randomised to surgery with missing data results in the cost-effectiveness increasing above £20,000 per QALY gained. This scenario shows that missing data can have an impact on the results. Nevertheless, although it is impossible to empirically confirm or refute this scenario from the data in the trial, it would seem improbable in practice that surgical patients with poor quality of life are less likely to respond to follow-up questionnaires than similar participants undergoing medical management.

Comparison with similar randomised trials

The findings of the REFLUX trial were considered in the context of the three other randomised trials that have compared laparoscopic surgery with medical management. In respect of benefits, the trials consistently show better relief of GORD symptoms following surgery, with parallel, though less marked, improvements in generic HRQoL. The four trials are also consistent in respect of complications of surgery, with small numbers having associated visceral injuries, postoperative problems and dilatation of the fundoplication wrap. The REFLUX trial suggests that 4.5% have a reoperation and the other trials are broadly consistent with this. Difficulty swallowing (dysphagia), flatulence and bloating have been linked with fundoplication in the other trials. In contrast, although a small number of REFLUX participants had a dilatation of the fundoplication wrap, responses to the questionnaires did not show a difference between those randomised to surgery and those randomised to medical management in these respects.

Conclusions

After 5 years' follow-up, a policy of relatively early laparoscopic fundoplication among patients for whom reasonable control of GORD symptoms requires long-term medication and for whom both surgery and medical management are suitable continues to provide better relief of GORD symptoms with associated better quality of life. Complications of surgery were rare. Despite being initially more costly, a surgical policy is likely to be more cost-effective for such patients suffering from GORD who were eligible for the REFLUX trial.

Implications for health care

Extending the use of laparoscopic fundoplication to people whose GORD symptoms require long-term medication for reasonable control and who would be suitable for surgery would provide health gains that extend over a number of years. The longer-term data reported here indicate that this would also be a cost-effective use of resources. The more troublesome the symptoms, the greater the potential benefit from surgery.

Recommendations for research

Most patients taking anti-reflux medication are managed in general practice. It is uncertain how many of these people might be suitable for surgery and hence what the most efficient provision of future care might be. Further research to explore the feasibility and resource impact of alternative policies for fundoplication within the NHS is therefore recommended.

Trial registration

This study is registered as ISRCTN15517081.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Introduction

This report describes the long-term follow-up of the REFLUX trial assessing the clinical effectiveness and cost-effectiveness of laparoscopic surgery compared with continued medical management for people with gastro-oesophageal reflux disease (GORD). This comparison was identified as a priority for research by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme, which funded the trial in two stages. The first stage, encompassing preliminary economic modelling, outcome development, trial recruitment, initial clinical management, follow-up to a time equivalent to 1 year after surgery and modelling of cost-effectiveness based on results available at that time, was reported in 2008.[1–5] The second stage, reported here, describes analyses based on further follow-up to 5 years after surgery.

Gastro-oesophageal reflux disease

The lower oesophagus, at its junction with the stomach, normally acts as a sphincter to prevent the contents of the stomach flowing back up the oesophagus. When the sphincter does not work adequately, the acid stomach contents leak, or 'reflux', into the oesophagus. The commonest symptom that this causes is heartburn, a burning sensation in the chest or throat. GORD has been defined through an international consensus process as 'a condition which develops when the reflux of stomach contents causes troublesome symptoms and/or complications'; in this consensus, symptoms were considered 'troublesome' 'if they adversely affected a patient's well-being'.[6]

Symptoms caused by gastro-oesophageal reflux are common: between 20% and 30% of a 'Western' adult population experience heartburn and/or reflux intermittently.[7–9]

Treatment of GORD includes both medical and surgical management, the options depending on the severity of symptoms. The majority of people with reflux have only mild symptoms and require little, if any, medication. The simplest is self-administered antacids with advice to alter lifestyle factors such as dietary modification, smoking cessation and weight reduction. A minority have severe symptoms and develop overt complications, despite full medical therapy, and require surgical intervention. Among the remainder, control of symptoms requires regular or continuous acid-suppression therapy using either histamine receptor antagonists (H_2RAs) or proton pump inhibitors (PPIs); initial high-dose therapy may be followed by maintenance treatment using these drugs either intermittently or continuously at a reduced dose sufficient to suppress symptoms. It is from this intermediate group of patients with significant disease requiring maintenance medical treatment that most of the treatment costs for the health service arise.

Laparoscopic fundoplication

Interest in surgery as an alternative to long-term medical therapy for GORD has been considerable since the introduction of the minimal access laparoscopic approach in the early 1990s.[10] Randomised trials conducted comparing laparoscopic with open surgery showed similar improvement in symptoms but with clear benefits of the laparoscopic approach in terms of recovery and fewer postsurgical complications.[11] As a consequence, surgery was suggested as an alternative to long-term maintenance medical treatment with anti-reflux drugs.

The operative method, whether using an open or a laparoscopic approach, involves performing a fundoplication by wrapping the fundus of the stomach around the lower oesophagus to create a high-pressure zone, thus reducing gastro-oesophageal reflux. The wrap created can be either complete (360°)

or partial. Many operative variants have been described. The commonest operation is a 1-cm complete wrap fashioned over a large bougie, the so-called 'short-floppy Nissen'.[12,13] There has been debate about the use of a partial rather than a total fundoplication. The partial approach has a number of potential advantages (such as fewer postoperative complications) but several controlled studies have shown broad equivalence between the two approaches;[14] for the purpose of this study they were therefore regarded as equivalent. Although fundoplication is reported to produce resolution of reflux symptoms in upwards of 90% of patients, like all surgery it carries risks and can have side effects. There is also uncertainty about the durability of benefit and frequency and severity of side effects following surgical therapy. Long-term follow-up to 12 years after open reflux surgery suggested attenuated but continuing better control of reflux symptoms; however, other symptoms such as difficulties swallowing (dysphagia), rectal flatulence and inability to belch or vomit were more common in surgical patients.[15] An important objective of this study was to determine if the long-term pattern of symptoms following laparoscopic surgery was similar.

Medical management

Proton pump inhibitors, sometimes supplemented with prokinetics or alginates, are the most effective medical treatment for moderate to severe GORD. Once started on PPIs, the majority of patients with significant GORD remain on long-term treatment.[16] It is estimated that around 1% or more of the UK adult population are prescribed PPI maintenance therapy.[17–19] The cost to the NHS of medical management of GORD is considerable. In England alone, the cost of PPIs is estimated to be £220M per year.[20] Of this budget, most of this prescribing occurs within the primary care setting.[21,22]

Although PPIs are generally considered safe, there is increasing acknowledgement of their possible adverse effects.[23,24] Gastric acid suppression predisposes to enteric infections and the sustained hypergastrinaemia resulting from PPI use causes rebound acid hypersecretion and the development of acid-related symptoms if the drug is stopped. Acute severe hypomagnesaemia has been recognised relatively recently as a rare adverse reaction to PPIs; the mechanism underlying it is not known. The clinical significance of impaired vitamin B_{12} and iron absorption due to PPIs is uncertain; there is also controversy about the risk of fractures and pneumonia and about the occurrence and significance of gastric mucosal atrophy and intestinal metaplasia, which have been seen in *Helicobacter pylori*-positive patients taking PPIs. Drug–drug interactions have also been a cause for concern,[25] although unequivocal evidence of their occurrence does not in itself establish clinical significance.

For the purpose of this study, medical therapy was taken to mean long-term therapy with PPIs (or H_2RAs if intolerant to PPIs).

Rationale for the study design

The original study design was based on the belief that decisions about the management of GORD should be made using unbiased, statistically precise comparisons of alternative policies. At study entry all patients fulfilled three criteria: they were on long-term acid suppression with PPIs; they had symptoms that were thought to be adequately controlled; and they were suitable in terms of fitness and comorbidity for either surgical or continuing medical treatment for their GORD. At the time that the study was planned, the consensus opinion of clinicians was that these three criteria identified GORD patients for whom surgical and continuing medical treatment could be considered equally acceptable treatment options and that, consequently, the comparison should be undertaken in patients meeting these criteria.

The most likely sources of bias were in the ways in which the groups being compared were selected; how their outcomes were assessed; and how the management was actually delivered. This is the basis for using a pragmatic randomised controlled trial (RCT) design. Random allocation protected against selection bias.

Confining the trial to those with no clear treatment preference limits biased patient-centred assessment of outcome, and pragmatic comparison of alternative policies [with intention-to-treat (ITT) analysis] avoids bias introduced by individual cases of non-compliance. This approach had limitations, however, and for this reason we chose to incorporate two parallel, non-randomised preference groups.

Including those with a clear preference for one policy or the other allows broader extrapolation and generalisability. Study of this group may give insights into the reasons for preference and hence give pointers to patient choices after the study.[26] Furthermore, preference may influence outcome and, if so, this may also help when making treatment decisions.[26,27] A third reason for the parallel, non-randomised preference groups[28] was that the addition of data from the preference groups may reduce imprecision around the estimates from the randomised comparison and this may be particularly useful for rare events, such as complications that can be confidently ascribed to one or other treatment. (The limitation is that the preference groups are not derived by random allocation, and hence the comparisons are exposed to the biases of non-randomised studies.)

Reliable comparisons within and between randomised and preference groups require valid measurement of treatment outcome. Although there were a number of quality-of-life (QoL) tools available, none was sufficiently specific to assess the spectrum of gastrointestinal symptoms associated with the treatment of GORD, particularly those due to surgery. For this reason we developed and validated a new outcome measure (the REFLUX questionnaire). We have continued to use this as the primary outcome measure in the longer-term follow-up reported here. Details of the REFLUX questionnaire and its derivation have been described elsewhere.[1,4]

Gastro-oesophageal reflux disease and its management represent a very significant call on NHS resources. Although clinical effectiveness, acceptability and safety will be important determinants of future policy, the issues of cost and resource use may be over-riding. This is the reason for the economic evaluation component of this study. Policy should be guided by both assessment of the relative cost-effectiveness of alternative policies and assessment of the impact that possible policy changes would have for the NHS and for patients with GORD.

The cost of laparoscopic fundoplication appears to be equivalent to the cost of 2–3 years of maintenance treatment with PPIs, although it is acknowledged that the costs of PPIs are falling.[29] The costs of surgery are related largely to two factors: the incidence of complications/length of hospital stay and the number of patients requiring long-term medical interventions after surgery.

We addressed cost-effectiveness in our report of the first phase of the REFLUX trial.[1] We reported both a within-trial cost-effectiveness analysis based on the results up to 12 months after surgery and an extended cost-effectiveness model that explored a number of scenarios beyond 12 months. The within-trial analysis related the extra mean costs associated with the surgical policy to the estimated increase in mean quality-adjusted life-years (QALYs) associated with surgery up to that time. The incremental cost-effectiveness ratio (ICER) was around £19,000 when the ITT analysis was used. Taking into account the uncertainties around the estimates of both costs and utilities, it was calculated that the chance that the surgical policy would be cost-effective at a threshold of £20,000 per QALY was 46%. This indicated considerable uncertainty at thresholds that are currently commonly applied to costs per QALY. The limitations of the within-trial analysis were discussed in detail in the earlier report, in particular that it ignored costs and benefits that accrued after 1 year.

The economic model was designed to address the limitations of the within-trial analysis. It explored a range of scenarios of varying lifetime benefits and costs, and analyses gave a wide range of incremental costs per QALY of £1000–44,000, again indicative of wide uncertainty. The factors contributing most to this uncertainty were the projected health-related quality-of-life (HRQoL) parameters and the long-term uptake of medication following surgery.

Thus, although data available up to a time equivalent to 1 year after surgery provided promising evidence that surgical management might well be cost-effective, there was too much uncertainty, especially about longer-term costs and benefits, to provide clear guidance for decision-makers. This was the justification for the longer-term follow-up to 5 years reported here.

Chapter 2 Methods

Original study design

The study had two complementary components:

1. a multicentre, pragmatic[30] RCT (with parallel non-randomised preference groups) comparing a laparoscopic surgery-based policy with a continued medical management policy to assess their relative clinical effectiveness
2. an economic evaluation of laparoscopic surgery for GORD to compare the cost-effectiveness of the two management policies, identify the most efficient provision of future care and describe the resource impact that various policies for fundoplication would have on the NHS.

Eligible patients who consented to participate in the RCT were randomly allocated to either laparoscopic surgery or continued medical management. Those patients who had a strong preference for one or other of the two treatment options could be recruited to the preference study. Clinical history was recorded at study entry. Participants completed health status questionnaires at the time of recruitment to the study and then at specified times equivalent to 3 and 12 months and then 2, 3, 4 and 5 years after surgery.

Approval for this study was obtained from the Scottish Multicentre Research Ethics Committee and the appropriate Local Research Ethics Committees.

Clinical centres

Clinical centres were based on local partnerships between surgeons with experience of laparoscopic fundoplication and gastroenterologists, with whom they shared the secondary care of patients with GORD. Centres were eligible if they included:

- a surgeon who had performed at least 50 laparoscopic fundoplication operations
- one or more gastroenterologists who agreed to collaborate with the surgeon(s) in the trial.

Study population

Eligible patients were those for whom care had been provided by a participating clinician who was uncertain which management policy (surgical or medical) was better. In addition, patients had to have documented evidence of GORD (based on endoscopy and/or manometry/24-hour pH monitoring) as well as symptoms for >12 months requiring maintenance PPI therapy for reasonable symptom control. Patients who were intolerant to PPIs and therefore required H_2RA therapy to control their symptoms were also eligible. Patients who were morbidly obese [body mass index (BMI) >40 kg/m^2] or who had Barrett's oesophagus of >3 cm or evidence of dysplasia, a paraoesophageal hernia or an oesophageal stricture were all excluded.

Eligible patients who did not want to take part in the randomised trial because of a strong preference for one type of management or the other were invited to take part in the preference arm of the study. For logistical reasons and to maintain a balance between the sizes of the randomised and the preference groups, we aimed to cap the numbers of participants recruited to the preference arms to 20 per arm in each centre.

All participants gave their informed consent.

Health technology policies being compared

Laparoscopic surgery policy

For those participants allocated to the randomised surgical group or recruited to the preference surgical group of the trial, subsequent deferring or declining of surgery, by either the participant or the surgeon, was always an option (i.e. even after trial entry), particularly among those recruited by a gastroenterologist and referred to a surgeon for consideration of surgery within the trial. Participants who had not had manometry/pH studies underwent these tests before surgery to exclude achalasia.

The surgery was performed either by an experienced surgeon who had undertaken >50 laparoscopic fundoplications or by a less experienced surgeon working under the supervision of an experienced surgeon. It was recommended that crural repair be routine and that non-absorbable synthetic sutures (not silk) be used for the repair. The type of fundoplication used was left to the discretion of the experienced surgeon. For the purposes of the main comparisons, the different surgical techniques for laparoscopic fundoplication were considered as part of a single policy. The study design, however, allowed for indirect comparisons between techniques.

Medical management policy

Those allocated to the medical management policy had their therapy reviewed and adjusted as necessary by the local gastroenterologist to be 'best medical management'. It was recommended that management conformed to the principles of the Genval Workshop Report.[31] These include stepping down antisecretory medication in most patients to the lowest dose that maintained acceptable symptom control. Following the therapy review by the gastroenterologist, trial participants had their medication managed by their general practitioner (GP). Although, in general, trial participants allocated to medical management were managed in this way, the protocol did include the option of surgery if a clear indication for it subsequently developed.

Study registration (and treatment allocation when randomised)

The treatment allocation for participants in the randomised component of the trial was computer generated; it was stratified by centre, with balance in respect of other key prognostic variables – age (18–49 years or 50+ years), sex (male or female) and BMI (≤28 or >29 kg/m^2) – by a process of minimisation. Randomisation was organised centrally at the Health Services Research Unit, University of Aberdeen, and was independent of all clinical collaborators.

Clinical management

Participants who were allocated to surgical management were invited to a consultation with the collaborating surgeon. During this consultation, the surgeon confirmed that there was no contraindication to surgery and discussed the operation in more detail, before arranging an operation date. The surgeon recorded intraoperative details on specially designed study forms. All other in-hospital data collection was the responsibility of the local study nurse. In all respects, other than the trial interventions, clinical management was left to the discretion of the clinician responsible for care. This continued to be the case in the extended follow-up phase, which is the focus of this report, with GPs monitoring subsequent care needs throughout the follow-up period.

Data collection

Follow-up by postal questionnaire was first performed at 3 months after surgery, or at an equivalent time among those who did not have surgery, and then annually. The questionnaire used for the follow-up at 2–5 years was similar to the questionnaire that had been used in the earlier phase of the trial up to 12 months after surgery. Non-responders received up to two reminder telephone calls or letters to encourage return of their postal questionnaires. On occasion, and at the participants' convenience, a shortened version of the questionnaire was completed over the telephone.

From around half-way through the 5-year follow-up, participants were sent a £5 gift voucher with their final postal reminder to compensate for their time in completing the questionnaire. This decision was taken based on the findings of a systematic review of the effects of incentives on postal questionnaire return[32] and specific randomised trials that evaluated the use of vouchers.[33–35]

All data were sent to the trial office in Aberdeen for processing. A random 10% sample of all data was double-entered to check accuracy and no significant errors were identified. Extensive range and consistency checks further enhanced the quality of the data.

The principal study outcome measure

The primary outcomes for measuring the differences in effects between medical and surgical management were:

- a 'disease-specific' measure incorporating assessment of reflux and other gastrointestinal symptoms and the side effects and complications of both therapies (the REFLUX questionnaire was developed specifically for this study[4])
- NHS costs including treatments, investigations, consultations and other contacts with the health service.

The secondary outcome measures were:

- HRQoL – measured using the European Quality of Life-5 Dimensions (EQ-5D)[36] and Short Form questionnaire-36 items (SF-36)[37]
- patient costs, including loss of earnings, reduction in activities and the costs of prescriptions and travel to health care
- other serious morbidity, such as operative complications
- (further) anti-reflux surgery
- mortality.

An example of the annual questionnaire used for collecting this information is provided in *Appendix 1*.

Sample size

The original aim was to recruit 600 participants to the randomised trial to give 80% power to identify a difference between the two groups of 0.25 of a standard deviation (SD) in respect of the disease-specific instrument and other continuous variables such as EQ-5D and SF-36, using a significance level of 5%. Based on the same arguments, it was planned that 300 people would be recruited to each arm of the preference study. The cost savings of a surgical policy largely depend on the number of patients managed surgically who no longer require PPI treatment, and a trial with 300 surgically managed patients would have estimated this proportion to within about 5% with 95% statistical confidence.

However, prompted by a lower rate of recruitment than expected, this target was revised in January 2003 in consultation with the Data Monitoring Committee (DMC) and representatives of the HTA programme. It was agreed that a larger benefit (0.3 of a SD) was clinically plausible based on improvements seen after surgery in the accruing literature among more severely affected people (who were not eligible for the trial). This was calculated to require 196 in each group to give 80% power ($2p = 0.05$).

Statistical considerations

This report describes analyses of annual questionnaire data up to 5 years after surgery (or an equivalent time if managed medically). As a general rule, in the tables and analyses presented in this report, the participants in the randomised groups are separate from those in the preference groups. A sizeable group of patients allocated to surgery did not receive surgery. Therefore, to investigate the potential influence of this non-compliance with allocation, summary statistics in the results tables are given for four main analysis populations (comprising eight groups of participants):

1. Randomised ITT population (groups that were randomised to either surgery or medical management).
2. Per-protocol (PP) population (groups that were either randomised to surgery and received surgery in the first year or randomised to medical management and did not receive surgery in the first year).
3. Preference ITT population (groups that preferred either surgery or medical management at recruitment).
4. Preference PP population (groups that either preferred surgery at recruitment and received surgery in the first year or preferred medical management and did not receive surgery in the first year).

The primary outcome measure (REFLUX QoL score) and secondary outcome measures (SF-36, EQ-5D, REFLUX symptom scores, anti-reflux surgery and use of reflux-related drugs) were analysed using general linear models. The analyses adjusted for the minimisation covariates (age, BMI and sex) and where appropriate (defined by significant at the 5% significance level) also adjusted for baseline measures and baseline measures by treatment interaction. A secondary, pre-stated subgroup analysis explored the differential effects of surgeon's preferred operative procedure on the primary outcome measure. All analyses were reported with 95% confidence intervals (CIs).

The primary analysis of the randomised groups was by ITT. The ITT approach sustains the integrity of the randomisation and gives the least biased estimate of effectiveness of the two forms of management. Given that a sizeable minority of the randomised surgical participants did not receive surgery, we were also interested in estimating the efficacy of the initial treatment received as a secondary comparison (i.e. commonly known as a PP analysis). In an open trial design a PP analysis can have substantial selection bias. To minimise the effects of selection bias we used the method of 'adjusted treatment received' as described by Nagelkerke *et al.*[38] and others.[39,40] The method used a two-stage least-squares approach whereby treatment randomised was regressed onto treatment received and the residuals from that model were used as an independent variable in a second model, together with the treatment received, to estimate the effects on the various primary and secondary outcome measures.

For the preference study, only the primary outcome was analysed statistically. The analysis compared the preference surgical group with the preference medical group and adjusted for the minimisation factors. As described above, for logistical reasons and to maintain balance between the randomised and preference groups, we capped the number of preference participants at 20 per group per centre. The study design was not therefore a true comprehensive cohort. We did consider modelling differences between the randomised and preference groups; however, it is not universally accepted that formal modelling is appropriate in this context. In this case we knew from the randomised arms that there was a strong interaction between treatment effects and baseline REFLUX QoL, and in addition we knew that there was a large difference in QoL between preference arms at baseline (and patient demographics such as age and

sex). We therefore decided that formal modelling of the arms would add little to the comparison given the large confounding between preference groups.

Sensitivity analyses

The sensitivity of the primary outcome analysis result was investigated using two approaches – the effect of excluding a large centre and the effects of missing data. In the first approach the largest recruiting centre, Aberdeen, was excluded and the analysis as described above was rerun. Second, previous work demonstrated that the primary outcome was likely missing at random (MAR) or missing completely at random (MCAR) and that a repeated measures analysis (using all available data) was an appropriate statistical method for analysing data up to 12 months.[41] We therefore used a repeated measures analysis on the primary outcome across all of the follow-up data (12 months to 5 years) to investigate the effect of incorporating a profile of measures for each participant. No further imputation for missing values was necessary.

Data monitoring

During recruitment, an independent DMC met on three occasions and each time saw no reason to recommend any fundamental changes to the protocol. The committee did not meet after recruitment was completed.

Chapter 3 Trial results and clinical effectiveness

Recruitment to the trial

Participants were recruited in 21 clinical centres, all within the UK (their locations are listed on the left-hand side of *Table 1*). Recruitment to the trial was open from March 2001 until the end of June 2004, although not all centres enrolled over the total period because of the staggered introduction of centres and early closure for logistical reasons in a few places.[1]

A total of 357 participants were recruited to the randomised component: 178 allocated to surgery and 179 allocated medical management. 453 participants agreed to join the preference component: 261 choosing surgery and 192 choosing medical management. *Table 1* shows recruitment by centre. Around one-fifth of the randomised participants were enrolled in Aberdeen; no centre contributed >10% of participants in the preference component.

Analysis populations

Throughout the analyses presented later in this chapter, the participants in the randomised component are kept separate from those in the preference component (other than for rare surgical events). The numbers of participants in each of the four main analysis populations are shown in *Table 2*. All 357 who joined the randomised component are in the randomised ITT population; only the 280 within this group who actually received their allocated management over the first year are in the randomised PP population. All 453 participants who joined the preference component are in the preference ITT population; the 407 of these who, by the end of the first year, were managed as originally chosen were in the preference PP population.

Trial conduct

The derivation of the main study groups and their progress through the stages of follow-up in the trial are shown in *Figure 1*. This is in the form of a CONSORT (Consolidated Standards of Reporting Trials) flow diagram. In total, 1078 patients were considered for trial entry and 200 of these were found not to meet one or more of the eligibility criteria. Of the 68 patients eligible for the study but not recruited, 51 declined to participate, six were subsequently deemed inappropriate for the study by the surgeon responsible for care and the remaining 11 were missed.

Details of the clinical management actually received are described later in this chapter.

The mean (SD) time intervals in months between the receipt by the trial office of each subsequent annual postal questionnaire are shown in *Table 3*; all were near 12 months, as would be expected. There was, however, a difference between the randomised groups in the time interval between the 1-year and the 2-year questionnaires (mean 12.2 months surgical group vs 13.9 months medical group). In part, this was due to more late returns in the medical management group – the median intervals were closer: 12.00 and 13.00 months respectively. As described previously,[1] early follow-up was adjusted to be at a time equivalent to 3 and 12 months after surgery. The adjustments in the medical group to match this could be only approximate and this is the explanation for the difference that remained between the randomised groups. An advantage of long-term follow-up to 5 years is that any difference in the timing of follow-up becomes proportionately smaller over time.

TABLE 1 Number of participants by centre

	Randomised participants, *n* (%)		Preference participants, *n* (%)	
	Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)
Aberdeen: Aberdeen Royal Infirmary	38 (21.3)	40 (22.3)	20 (7.7)	21 (10.9)
Belfast: Royal Victoria Hospital	15 (8.4)	14 (7.8)	4 (1.5)	20 (10.4)
Bournemouth: Royal Bournemouth Hospital	4 (2.2)	3 (1.7)	20 (7.7)	3 (1.6)
Bristol: Bristol Royal Infirmary	12 (6.7)	11 (6.1)	18 (6.9)	20 (10.4)
Bromley: Princess Royal Infirmary	3 (1.7)	3 (1.7)	20 (7.7)	17 (8.9)
Edinburgh: Royal Infirmary of Edinburgh	11 (6.2)	11 (6.1)	1 (0.4)	15 (7.8)
Guildford: Royal Surrey County Hospital	10 (5.6)	10 (5.6)	17 (6.5)	10 (5.2)
Hull: Hull Royal Infirmary	7 (3.9)	7 (3.9)	1 (0.4)	2 (1.0)
Inverness: Raigmore Hospital	7 (3.9)	8 (4.5)	2 (0.8)	8 (4.2)
Leeds: Leeds General Infirmary	1 (0.6)	2 (1.1)	10 (3.8)	3 (1.6)
Leicester: Leicester Royal Infirmary	0 (0.0)	0 (0.0)	3 (1.1)	1 (0.5)
London: St Mary's Hospital	8 (4.5)	7 (3.9)	4 (1.5)	10 (5.2)
London: Whipps Cross Hospital	4 (2.2)	3 (1.7)	16 (6.1)	5 (2.6)
Poole: Poole Hospital	10 (5.6)	10 (5.6)	25 (9.6)	13 (6.8)
Portsmouth: Queen Alexandra Hospital	10 (5.6)	10 (5.6)	15 (5.7)	1 (0.5)
Salford: Hope Hospital	0 (0.0)	1 (0.6)	6 (2.3)	3 (1.6)
Stoke-on-Trent: North Staffordshire Hospital	5 (2.8)	6 (3.4)	20 (7.7)	9 (4.7)
Swansea: Morriston Hospital	8 (4.5)	8 (4.5)	14 (5.4)	9 (4.7)
Telford: Princess Royal Hospital	11 (6.2)	12 (6.7)	24 (9.2)	8 (4.2)
Yeovil: Yeovil District Hospital	9 (5.1)	8 (4.5)	18 (6.9)	8 (4.2)
York: York District Hospital	5 (2.8)	5 (2.8)	3 (1.1)	6 (3.1)
Total	178 (100)	179 (100)	261 (100)	192 (100)

TABLE 2 Number of participants in each analysis population

	Surgical, *n* (%)	Medical, *n* (%)	Total, *n*
Randomised ITT	178 (49.9)	179 (50.1)	357
Randomised PP	111 (39.6)	169 (60.4)	280
Preference ITT	261 (57.6)	192 (42.4)	453
Preference PP	218 (53.6)	189 (46.4)	407

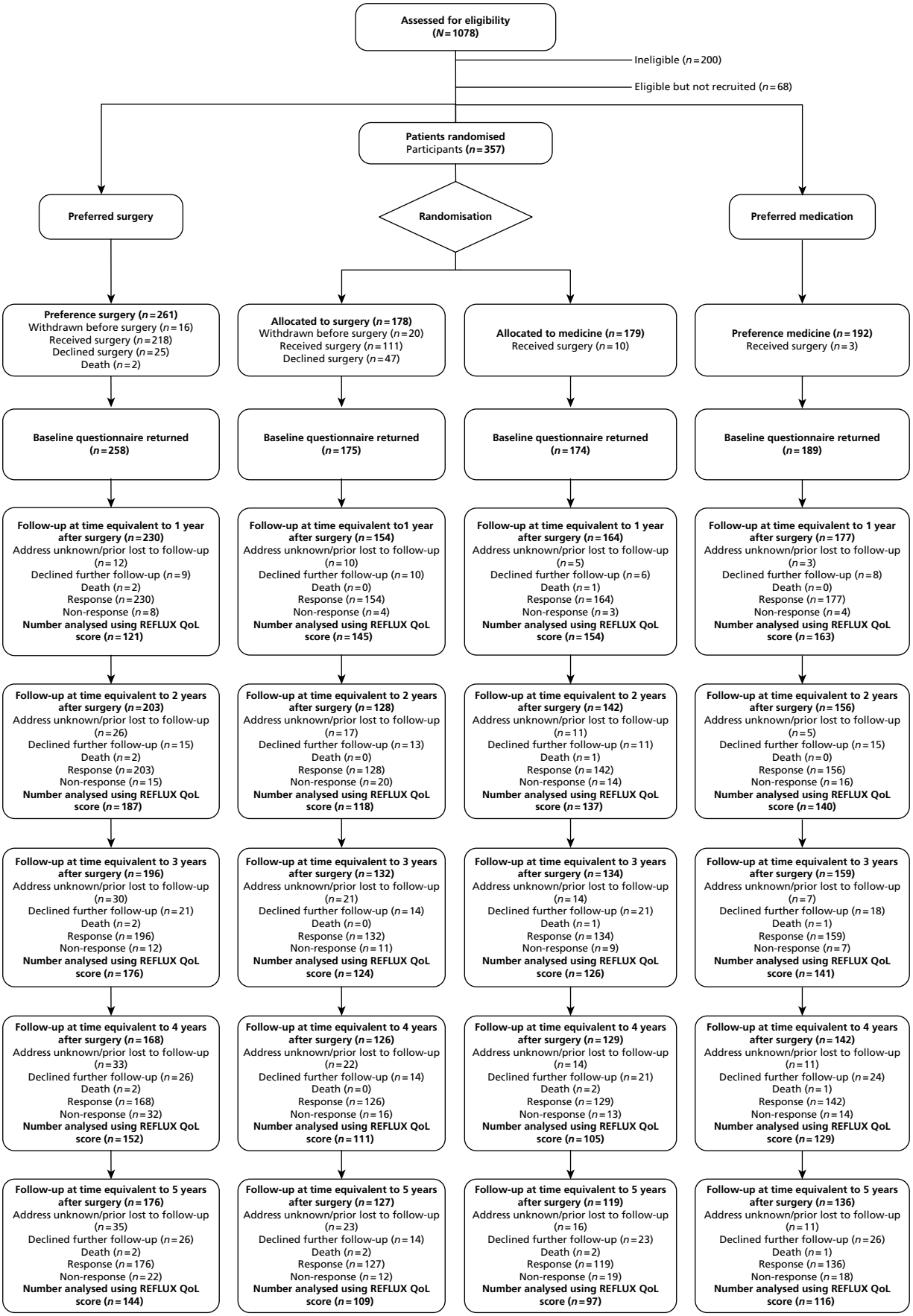


FIGURE 1 The CONSORT diagram.

TABLE 3 Interval between randomisation and follow-up (months), mean (SD)

	Randomised participants		Preference participants	
	Surgical	Medical	Surgical	Medical
	ITT (*n* = 178)	ITT (*n* = 179)	ITT (*n* = 261)	ITT (*n* = 192)
1 year to 2 years	12.2 (1.9)	13.9 (3.1)	12.4 (1.8)	12.9 (4.6)
2 years to 3 years	11.8 (1.2)	11.6 (1.2)	11.6 (1.5)	11.8 (1.2)
3 years to 4 years	12.0 (1.5)	12.0 (1.4)	12.1 (1.2)	12.0 (1.1)
4 years to 5 years	11.8 (1.3)	12.0 (1.3)	12.1 (1.5)	12.0 (1.3)

More details of the response rates to the annual questionnaires are provided in *Table 4*. The overall rates of return of annual follow-up questionnaires (years 1–5) were 89.5%, 77.7%, 76.7%, 69.8% and 68.9% of the study participants. Seven participants are known to have died up to the end of the 5-year follow-up; equivalent response rates among those not known to have died are 89.8%, 77.9%, 77.0%, 70.2% and 69.5%. There were no substantive differences in response rates between the groups.

Three participants died before the 1-year follow-up was reached: two in the preference surgery group and one in the randomised medical group. None of these participants actually had surgery. Four died subsequently; there is no evidence linking these deaths to trial participation.

Description of the groups at trial entry

Sociodemographic and clinical factors

Table 5 provides a description of the groups at trial entry. The main division within the table is between participants in the randomised component and those in the preference component. These two halves of the table are further divided according to the allocation of participants and then subdivided according to ITT or PP.

Randomised arms

Within the randomised groups there were no apparent imbalances between the medical and surgical intervention arms. The patients were, on average, 46 years old, 66% were men, around two-thirds were in full employment and participants had been on GORD medication for a median of 32 months. The baseline characteristics in the randomised PP groups were similar.

Preference arms

The sociodemographic characteristics of the preference participants were broadly similar to those of the randomised participants. However, preference medical participants tended to be older (mean age 50 years) and were more likely to be female, fewer were in full-time employment and participants had been on GORD medication for a shorter period (approximately 6 months less than randomised participants).

Prescribed medications

The prescribed medications at the time of trial entry are shown in *Table 6*. There was a similar profile of prescribed medications across the randomised and preference groups. As would be expected, nearly all participants reported taking a reflux-related drug in the previous 2 weeks. Over 90% had taken a PPI, of which lansoprazole was the most common.

TABLE 4 CONSORT table

Year	Category	Randomised participants, *n* (%)		Preference participants, *n* (%)	
		Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)
1	Responded	154 (87)	164 (92)	230 (88)	177 (92)
	Declined further follow-up	10 (6)	6 (3)	9 (3)	8 (4)
	Deceased	0 (0)	1 (1)	2 (1)	0 (0)
	Address unknown/lost to follow-up	10 (6)	5 (3)	12 (5)	3 (2)
	Non-responder	4 (2)	3 (2)	8 (3)	4 (2)
2	Responded	128 (72)	142 (79)	203 (78)	156 (81)
	Declined further follow-up	13 (7)	11 (6)	15 (6)	15 (8)
	Deceased	0 (0)	1 (1)	2 (1)	0 (0)
	Address unknown/lost to follow-up	17 (10)	11 (6)	26 (10)	5 (3)
	Non-responder	20 (11)	14 (8)	15 (6)	16 (8)
3	Responded	132 (74)	134 (75)	196 (75)	159 (83)
	Declined further follow-up	14 (8)	21 (12)	21 (8)	18 (9)
	Deceased	0 (0)	1 (1)	2 (1)	1 (1)
	Address unknown/lost to follow-up	21 (12)	14 (8)	30 (11)	7 (4)
	Non-responder	11 (6)	9 (5)	12 (5)	7 (4)
4	Responded	126 (71)	129 (72)	168 (64)	142 (74)
	Declined further follow-up	14 (8)	21 (12)	26 (10)	24 (13)
	Deceased	0 (0)	2 (1)	2 (1)	1 (1)
	Address unknown/lost to follow-up	22 (12)	14 (8)	33 (13)	11 (6)
	Non-responder	16 (9)	13 (7)	32 (12)	14 (7)
5	Responded	127 (71)	119 (66)	176 (67)	136 (71)
	Declined further follow-up	14 (8)	23 (13)	26 (10)	26 (14)
	Deceased	2 (1)	2 (1)	2 (1)	1 (1)
	Address unknown/lost to follow-up	23 (13)	16 (9)	35 (13)	11 (6)
	Non-responder	12 (7)	19 (11)	22 (8)	18 (9)

Health status

Randomised arms

The HRQoL scores at study entry are displayed in *Table 7*. The scores were broadly similar in the randomised surgical and randomised medical groups, although they were slightly higher (better health) in the randomised medical group. When the DMC first met after the initial 143 participants had been recruited to the randomised component, the committee did ask us to change the enrolment procedure to ensure that baseline questionnaires were completed *before* formal entry and randomisation. We understand that this was because they were concerned about an apparent imbalance between the randomised groups in baseline health status at that time. After satisfying themselves that this was not due to a breakdown in the randomisation procedure, the DMC surmised that this might be due to prior knowledge of the treatment allocation affecting questionnaire responses (with those allocated surgery tending to project worse health status than those allocated medical management). Certainly, the groups

TABLE 5 Description of groups at trial entry

Characteristic	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
Baseline questionnaire returned, *n* (%)	175 (98.3)	111 (100.0)	174 (97.2)	165 (97.6)	256 (98.1)	216 (99.1)	189 (98.4)	186 (98.4)
Age (years), mean (SD)	46.7 (10.3)	46.3 (10.2)	45.9 (11.9)	45.9 (11.9)	44.4 (12.0)	44.5 (12.2)	49.9 (11.8)	50 (11.7)
Male, *n* (%)	116 (65.2)	68 (61.3)	120 (67.0)	115 (68.0)	170 (65.1)	139 (63.8)	111 (57.8)	110 (58.2)
BMI (kg/m^2), mean (SD)	28.5 (4.3)	28.7 (4.1)	28.4 (4.0)	28.3 (4.0)	27.7 (4.0)	27.5 (3.7)	27.4 (4.1)	27.4 (4.1)
Duration of prescribed medication for GORD (months), median (IQR)	33 (15–83)	30 (16–76)	31 (16–71)	30 (15–71)	35 (14–71)	36 (14–65)	27 (13–60)	26.5 (13–60)
Employment status, *n* (%)								
Employed full-time	116 (66.3)	72 (65.5)	110 (61.8)	104 (61.9)	168 (65.1)	138 (64.2)	100 (52.4)	97 (51.6)
Employed part-time	13 (7.4)	12 (10.9)	16 (9.0)	15 (8.9)	35 (13.6)	29 (13.5)	20 (10.5)	20 (10.6)
Student	5 (2.9)	3 (2.7)	3 (1.7)	3 (1.8)	2 (0.8)	2 (0.9)	3 (1.6)	3 (1.6)
Retired	12 (6.9)	9 (8.2)	22 (12.4)	20 (11.9)	18 (7.0)	16 (7.4)	35 (18.3)	35 (18.6)
Housework	11 (6.3)	6 (5.5)	10 (5.6)	10 (6.0)	17 (6.6)	15 (7.0)	15 (7.9)	15 (8.0)
Seeking work	6 (3.4)	1 (0.9)	3 (1.7)	2 (1.2)	5 (1.9)	5 (2.3)	2 (1.0)	2 (1.1)
Other	12 (6.9)	7 (6.4)	14 (7.9)	14 (8.3)	13 (5.0)	10 (4.7)	16 (8.4)	16 (8.5)

Characteristic	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
Age (years) left full-time education, *n* (%)								
≤16	110 (62.5)	68 (62.4)	108 (60.7)	102 (60.7)	151 (58.5)	128 (59.3)	105 (55.3)	104 (55.6)
17–19	38 (21.6)	24 (22.0)	40 (22.5)	40 (23.8)	63 (24.4)	51 (23.6)	45 (23.7)	43 (23.0)
20+	28 (15.9)	17 (15.6)	30 (16.9)	26 (15.5)	44 (17.1)	37 (17.1)	40 (21.1)	40 (21.4)
Current smoker, *n* (%)	46 (25.8)	29 (26.1)	40 (22.3)	36 (21.3)	71 (27.2)	61 (28.0)	39 (20.3)	39 (20.6)
Erosive oesophagitis, *n* (%)	85 (54.8)	48 (50.0)	97 (62.2)	91 (62.3)	104 (46.4)	80 (43.2)	87 (50.9)	86 (51.2)
Comorbidity: *Helicobacter pylori* status, *n* (%)								
Positive (subsequently treated)	12 (9.0)	5 (6.1)	14 (10.4)	13 (10.3)	18 (8.4)	14 (7.9)	15 (10.5)	15 (10.7)
Positive (subsequently untreated)	1 (0.8)	0 (0.0)	3 (2.2)	3 (2.4)	8 (3.7)	8 (4.5)	2 (1.4)	2 (1.4)
Negative	75 (56.4)	48 (58.5)	73 (54.1)	67 (53.2)	118 (54.9)	101 (56.7)	74 (51.7)	72 (51.4)
Uncertain	45 (33.8)	29 (35.4)	45 (33.3)	43 (34.1)	71 (33.0)	55 (30.9)	52 (36.4)	51 (36.4)
Hiatus hernia present, *n* (%)	94 (57.3)	64 (61.0)	102 (60.4)	94 (59.1)	168 (68.9)	146 (71.2)	101 (59.8)	99 (59.6)
Asthma, *n* (%)	21 (11.9)	14 (12.7)	21 (11.8)	19 (11.3)	30 (11.5)	23 (10.6)	36 (18.8)	36 (19.0)

IQR, interquartile range.

TABLE 6 Description of groups at trial entry: prescribed medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
PPIs, *n* (%)								
Any PPI	161 (92.0)	105 (94.6)	162 (93.1)	153 (92.7)	225 (87.9)	191 (88.4)	173 (91.5)	170 (91.4)
Omeprazole	46 (26.3)	32 (28.8)	46 (26.4)	43 (26.1)	49 (19.1)	36 (16.7)	61 (32.3)	61 (32.8)
Lansoprazole	77 (44.0)	47 (42.3)	72 (41.4)	69 (41.8)	100 (39.1)	92 (42.6)	69 (36.5)	66 (35.5)
Pantoprazole	6 (3.4)	6 (5.4)	11 (6.3)	11 (6.7)	21 (8.2)	17 (7.9)	11 (5.8)	11 (5.9)
Rabeprazole	12 (6.9)	6 (5.4)	13 (7.5)	13 (7.9)	21 (8.2)	16 (7.4)	14 (7.4)	14 (7.5)
Esomeprazole	20 (11.4)	14 (12.6)	20 (11.5)	17 (10.3)	37 (14.5)	33 (15.3)	18 (9.5)	18 (9.7)
H_2RAs, *n* (%)								
Any H_2RA	14 (8.0)	6 (5.4)	12 (6.9)	9 (5.5)	22 (8.6)	16 (7.4)	13 (6.9)	13 (7.0)
Ranitidine	13 (7.4)	6 (5.4)	8 (4.6)	6 (3.6)	11 (4.3)	7 (3.2)	11 (5.8)	11 (5.9)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.4)	1 (0.5)	1 (0.5)	1 (0.5)
Cimetidine	1 (0.6)	0 (0.0)	1 (0.6)	0 (0.0)	1 (0.4)	1 (0.5)	0 (0.0)	0 (0.0)
Nizatidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.2)	3 (1.4)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	0 (0.0)	0 (0.0)	4 (2.3)	3 (1.8)	7 (2.7)	4 (1.9)	2 (1.1)	2 (1.1)

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
Prokinetics, *n* (%)								
Any prokinetics	12 (6.9)	7 (6.3)	8 (4.6)	6 (3.6)	11 (4.3)	10 (4.6)	5 (2.6)	4 (2.2)
Domperidone	8 (4.6)	5 (4.5)	4 (2.3)	3 (1.8)	7 (2.7)	6 (2.8)	4 (2.1)	3 (1.6)
Metoclopramide	4 (2.3)	2 (1.8)	4 (2.3)	3 (1.8)	4 (1.6)	4 (1.9)	1 (0.5)	1 (0.5)
Any reflux-related drug, *n* (%)	170 (97.1)	108 (97.3)	169 (97.1)	160 (97.0)	235 (91.8)	198 (91.7)	184 (97.4)	181 (97.3)
Other prescribed drugs, *n* (%)[a]								
Alginates	22	12	21	18	37	33	14	13
Antispasmodics (e.g. dicycloverine)	0	0	2	2	3	3	0	0
Chelates (e.g. sucralfate)	1	1	0	0	0	0	0	0
Other ulcer-healing drugs	0	0	0	0	1	1	0	0
Mucogel® (Chemidex)	0	0	1	1	1	1	1	1
Asilone® (Thornton & Ross)	0	0	1	1	0	0	0	0
Non-gastrointestinal	7	2	4	4	5	4	6	6
Anti-nausea	0	0	1	1	1	1	1	1

a More than one prescription per person possible.

based on the first 143 participants were well balanced in other respects, and there was subsequently good balance in health status as well. The apparent small imbalance between the randomised groups in health status measures is therefore likely to be a reflection of the imbalance in the first 143 participants.

The most prevalent reflux symptoms (those with lowest scores) were general discomfort and wind. The participants had lower SF-36 and EQ-5D scores than a normal UK population with the same average age and sex characteristics (SF-36 population norm approximately 50 for all domains; EQ-5D norm 0.88).

Preference arms

The preference for surgery participants reported worse REFLUX QoL scores and worse health in general than the preference for medicine participants. It can be seen from *Table 7* that the randomised participants reported QoL measures in between these two extremes.

Baseline characteristics of groups compared at 5 years

There were differences in baseline characteristics between those who had completed a questionnaire at 5 years and those who had not (*Table 8*). For example, responders had a higher mean age (47.9 years vs 43.6 years), had been on prescribed medication for a shorter period at recruitment to the REFLUX trial (50.5 months vs 60.2 months) and had higher QoL scores at baseline (measured on the disease-specific REFLUX instrument, EQ-5D and SF-36).

However, the baseline characteristics of those in the randomised surgical and randomised medical groups who completed a questionnaire at 5 years were very similar, with the only notable difference being in BMI (*Table 9*). The mean baseline BMI among responders in the randomised surgical group was higher (29.0 kg/m^2) than that for responders in the randomised medical management group (27.7 kg/m^2). As described in *Chapter 2*, these results confirmed that a repeated measures analysis assuming no differential loss to follow-up could be considered.

Surgical management

Table 10 summarises the use of surgery in the four study groups over the full 5-year follow-up period. At the end of the first year, 111 participants (62.4%) randomised to surgery had actually undergone fundoplication. Over the next 4 years, one more member of this group had fundoplication, bringing the total to 112 (62.9%). In the randomised medical group, 10 participants (5.6%) had fundoplication in the first year, with a further 14 participants having fundoplication in subsequent years, bringing the total at 5 years to 24 (13.4%). In the preference surgical group, 218 participants (83.5%) had fundoplication in the first year, with four more in the period up to 5 years, taking the percentage to 85.1%. Surgical management applied to only three participants (1.6%) in the preference medical group in the first year, with a further three being operated on in the subsequent 4 years (total 3.1%).

Information about the reasons why participants allocated surgery did not receive it in the first year is available for 47. For 25 of these 47, this was a clinical decision, most commonly the surgeon deciding that surgery was not appropriate; most of the other 22 changed their minds about surgery for a variety of work- or home-related reasons. A further 20 withdrew for unknown reasons. There is no doubt, however, that a number of these participants suffered long delays before being formally offered surgery, and this was an important factor in their eventual decision to choose not to have surgery after all. The trial was conducted at a time when there was great pressure on surgical services in the NHS, with long delays for elective surgery for non-life-threatening benign conditions being common. Indeed, the average time between trial entry and surgery in the trial was 8–9 months.[1]

TABLE 7 Description of groups at trial entry: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
HRQoL instrument	ITT (*n* = 178)	PP (*n* = 111)	ITT (*n* = 179)	PP (*n* = 169)	ITT (*n* = 261)	PP (*n* = 218)	ITT (*n* = 192)	PP (*n* = 189)
REFLUX QoL, mean (SD)	63.6 (24.1)	61.9 (24.5)	66.8 (24.5)	68.2 (24.2)	55.8 (23.2)	55.9 (23.2)	77.5 (19.7)	78.0 (19.1)
REFLUX symptom score, mean (SD)								
General discomfort symptom score	58.5 (24.5)	57.1 (25.1)	61.3 (25.8)	62.4 (25.7)	49.1 (24.4)	48.7 (25.2)	73.1 (21.3)	73.6 (20.9)
Wind and frequency symptom score	48.1 (20.9)	46.2 (20.9)	49.3 (21.4)	49.5 (21.7)	47.1 (21.4)	47.5 (21.2)	59.6 (22.7)	59.8 (22.7)
Nausea and vomiting symptom score	81.5 (19.5)	81.6 (18.8)	80.7 (21.9)	81.6 (21.7)	76.9 (19.9)	77.5 (19.5)	89.7 (13.6)	90.1 (12.9)
Activity limitation symptom score	78.5 (16.9)	77.6 (16.3)	78.9 (17.3)	79.5 (17.1)	74.4 (16.1)	73.9 (16.2)	86.8 (13.0)	87.0 (13.0)
Constipation and swallowing symptom score	77.5 (19.9)	77.3 (20.3)	74.8 (21.0)	75.6 (20.4)	75.8 (22.0)	74.8 (22.6)	83.0 (17.7)	83.3 (17.6)
SF-36 score, mean (SD)								
Norm-based physical functioning	46.8 (10.2)	46.1 (10.3)	47.5 (10.5)	47.7 (10.5)	46.3 (9.4)	46.1 (9.3)	47.1 (10.8)	47.0 (10.9)
Norm-based role physical	46.9 (10.7)	46.6 (10.8)	46.8 (10.6)	47.0 (10.4)	44.7 (10.9)	44.6 (10.7)	46.7 (10.9)	46.6 (10.9)
Norm-based bodily pain	44.4 (10.1)	44.1 (9.9)	44.6 (10.4)	44.9 (10.3)	41.8 (9.5)	41.9 (9.6)	47.1 (9.8)	47.2 (9.8)
Norm-based general health	40.9 (9.9)	40.2 (9.6)	41.1 (10.6)	41.4 (10.6)	40.6 (10.2)	40.8 (10.0)	42.4 (10.0)	42.4 (9.9)
Norm-based vitality	43.5 (10.5)	43.9 (10.3)	44.0 (11.7)	44.4 (11.4)	42.8 (11.1)	42.8 (11.3)	45.5 (10.7)	45.6 (10.7)
Norm-based social functioning	44.4 (11.1)	44.1 (10.6)	44.7 (11.7)	45.2 (11.5)	42.2 (11.6)	42.1 (11.5)	46.8 (10.2)	46.7 (10.2)
Norm-based role emotional	46.6 (11.5)	47.2 (11.5)	45.8 (12.9)	46.3 (12.6)	45.9 (12.2)	46.1 (12.1)	46.9 (11.8)	46.8 (11.8)
Norm-based mental health	46.0 (11.6)	46.9 (11.0)	46.7 (11.6)	47.1 (11.3)	44.6 (11.4)	44.6 (11.6)	46.4 (10.7)	46.3 (10.8)
EQ-5D, mean (SD)	0.711 (0.258)	0.718 (0.239)	0.720 (0.255)	0.732 (0.246)	0.682 (0.259)	0.679 (0.259)	0.750 (0.223)	0.752 (0.222)
EQ-5D$_{VAS}$, mean (SD)	68.6 (17.1)	69.2 (15.9)	70.5 (18.1)	71.2 (17.6)	67.2 (18.5)	67.0 (18.5)	71.3 (16.7)	71.5 (16.6)

VAS, visual analogue scale.

TABLE 8 Baseline characteristics of responders and non-responders at 5 years

Characteristic	Responder (max. *n* = 558)	Non-responder (max. *n* = 252)	*p*-value (two- sided)
BMI (kg/m^2), mean (SD), *n*	27.9 (4.0), 557	28.2 (4.3), 252	0.37
Age (years), mean (SD), *n*	47.9 (11.2), 558	43.6 (12.2), 252	<0.01
Sex, *n/N* (%)			
Male	343/558 (61)	174/252 (69)	0.04
Female	215/558 (39)	78/252 (31)	–
Duration of prescribed medication (months), mean (SD), *n*	50.5 (62.9), 544	60.2 (65.2), 250	0.05
Erosive oesophagitis, *n/N* (%)			
Yes	262/493 (53)	111/213 (52)	0.78
No	231/493 (47)	102/213 (48)	
Helicobacter pylori status, *n/N* (%)			
Positive (subsequently treated)	39/440 (9)	20/186 (11)	0.81
Positive (subsequently untreated)	9/440 (2)	5/186 (3)	–
Negative	238/440 (54)	102/186 (55)	–
Uncertain	154/440 (35)	59/186 (32)	–
Hiatus hernia, *n/N* (%)			
Yes	330/524 (63)	135/222 (61)	0.58
No	194/524 (37)	87/222 (39)	–
Age (years) left full-time education, *n/N* (%)			
≤16	304/552 (55)	170/250 (68)	<0.01
17–19	143/552 (26)	43/250 (17)	–
20+	105/552 (19)	37/250 (15)	–
Employment status, *n/N* (%)			
Full-time	348/551 (63)	146/251 (58)	0.01
Part-time	65/551 (12)	19/251 (8)	–
Student	6/551 (1)	7/251 (3)	–
Retired	62/551 (11)	25/251 (10)	–
Housework	32/551 (6)	21/251 (8)	–
Seeking work	10/551 (2)	6/251 (2)	–
Other	28/551 (5)	27/251 (11)	–
REFLUX QoL, mean (SD), *n*	66.6 (24.2), 533	61.3 (24.1), 226	<0.01
REFLUX symptom score, mean (SD), *n*			
General discomfort symptom score	61.1 (25.5), 544	55.4 (25.4), 231	<0.01
Wind and frequency symptom score	51.5 (21.7), 546	48.9 (23.0), 235	0.13
Nausea and vomiting symptom score	83.8 (18.3), 549	77.1 (21.5), 239	<0.01
Activity limitation symptom score	79.9 (16.1), 547	77.5 (17.4), 232	0.06
Constipation and swallowing symptom score	78.8 (20.0), 550	75.2 (21.7), 236	0.03

TABLE 8 Baseline characteristics of responders and non-responders at 5 years (*continued*)

Characteristic	Responder (max. *n* = 558)	Non-responder (max. *n* = 252)	*p*-value (two- sided)
EQ-5D, mean (SD), *n*	0.735 (0.234), 544	0.662 (0.279), 239	<0.01
SF-36 score, mean (SD), *n*			
SF-36 physical	45.2 (9.5), 530	44.0 (9.7), 232	0.10
SF-36 mental	46.3 (11.2), 530	42.7 (12.9), 232	<0.01
Norm-based physical functioning	47.2 (9.9), 545	46.1 (10.7), 239	0.15
Norm-based role physical	46.6 (10.7), 546	45.0 (11.0), 238	0.06
Norm-based bodily pain	45.1 (10.1), 546	42.3 (9.9), 236	<0.01
Norm-based general health	42.0 (9.8), 544	39.3 (10.7), 236	<0.01
Norm-based vitality	44.3 (10.8), 549	42.8 (11.4), 237	0.07
Norm-based social functioning	45.5 (10.8), 542	41.8 (12.0), 237	<0.01
Norm-based role emotional	47.0 (11.5), 543	44.5 (13.2), 239	0.01
Norm-based mental health	47.0 (10.6), 549	42.9 (12.4), 237	<0.01
Any PPI, *n/N* (%)	508/552 (92)	213/242 (88)	0.07
Any reflux drug, *n/N* (%)	530/552 (96)	225/242 (93)	0.07

max., maximum.

TABLE 9 Baseline characteristics of responders at 5 years by randomised allocation

Characteristic	Surgical (max. *n* = 127)	Medical (max. *n* = 119)	*p*-value (two-sided)
BMI (kg/m^2), mean (SD), *n*	29.0 (4.3), 127	27.7 (3.8), 119	0.01
Age (years), mean (SD), *n*	48.5 (9.3), 127	46.4 (11.6), 119	0.12
Sex, *n/N* (%)			
Male	79/127 (62)	76/119 (64)	0.79
Female	48/127 (38)	43/119 (36)	–
Duration of prescribed medication (months), mean (SD), *n*	57.2 (63.4), 124	46.3 (60.1), 117	0.17
Erosive oesophagitis, *n/N* (%)			
Yes	63/111 (57)	68/107 (64)	0.35
No	48/111 (43)	39/107 (36)	–
Helicobacter pylori status, n/*N* (%)			
Positive (subsequently treated)	6/96 (6)	10/91 (11)	0.52
Positive (subsequently untreated)	1/96 (1)	2/91 (2)	–
Negative	55/96 (57)	45/91 (49)	–
Uncertain	34/96 (35)	34/91 (37)	–
Hiatus hernia, *n/N* (%)			
Yes	73/117 (62)	71/114 (62)	0.99
No	44/117 (38)	43/114 (38)	–

continued

TABLE 9 Baseline characteristics of responders at 5 years by randomised allocation (*continued*)

Characteristic	Surgical (max. *n* = 127)	Medical (max. *n* = 119)	*p*-value (two-sided)
Age (years) left full-time education, *n/N* (%)			
≤16	77/125 (62)	70/119 (59)	0.46
17–19	27/125 (22)	31/119 (26)	–
20+	21/125 (17)	18/119 (15)	–
Employment status, *n/N* (%)			
Full-time	86/124 (69)	76/118 (64)	0.77
Part time	13/124 (10)	10/118 (8)	–
Student	2/124 (2)	1/118 (1)	–
Retired	9/124 (7)	13/118 (11)	–
Housework	4/124 (3)	7/118 (6)	–
Seeking work	4/124 (3)	3/118 (3)	–
Other	6/124 (5)	8/118 (7)	–
REFLUX QoL, mean (SD), *n*	65.9 (23.7), 121	68.6 (24.0), 110	0.38
REFLUX symptom score, mean (SD), *n*			
General discomfort symptom score	60.1 (24.1), 123	63.9 (25.2), 115	0.23
Wind and frequency symptom score	48.0 (19.7), 125	48.7 (20.9), 117	0.78
Nausea and vomiting symptom score	82.9 (18.9), 125	84.7 (18.9), 117	0.46
Activity limitation symptom score	79.9 (15.2), 124	79.9 (16.8), 117	0.99
Constipation and swallowing symptom score	78.2 (19.2), 124	75.9 (20.0), 118	0.35
EQ-5D, mean (SD), *n*	0.736 (0.223), 122	0.755 (0.228), 118	0.51
SF-36 score, mean (SD), *n*			
SF-36 physical	44.8 (10.0), 121	46.1 (9.1), 114	0.30
SF-36 mental	46.6 (11.0), 121	46.5 (11.1), 114	0.98
Norm-based physical functioning	46.8 (10.0), 123	48.4 (10.2), 117	0.22
Norm-based role physical	46.9 (10.8), 124	47.0 (10.8), 116	0.96
Norm-based bodily pain	44.6 (10.1), 123	45.7 (10.1), 117	0.39
Norm-based general health	41.4 (9.4), 124	42.4 (10.2), 116	0.41
Norm-based vitality	43.9 (10.4), 125	44.9 (11.2), 117	0.47
Norm-based social functioning	45.4 (10.5), 124	46.4 (10.8), 115	0.45
Norm-based role emotional	47.2 (11.4), 124	46.7 (12.1), 116	0.74
Norm-based mental health	47.3 (10.9), 125	48.0 (10.6), 117	0.60
Any PPI, *n/N* (%)	120/125 (96)	109/118 (92)	0.23
Any reflux drug, *n/N* (%)	124/125 (99)	113/118 (96)	0.08

max., maximum.

TABLE 10 Initial fundoplication operations

Surgery	Randomised participants		Preference participants	
	Surgical ($n = 178$)	Medical ($n = 179$)	Surgical ($n = 261$)	Medical ($n = 192$)
First fundoplication in first year, *n* (%)	111 (62.4)	10 (5.6)	218 (83.5)	3 (1.6)
First fundoplication after first year, *n*	1	14	4	3
In second year	0	1	2	0
In third year	0	7	1	2
In fourth year	1	4	1	1
In fifth year	0	2	0	0
Fundoplication at any time during 5-year follow-up, *n* (%)	112 (62.9)	24 (13.4)	222 (85.1)	6 (3.1)

Details of the surgery received by the 111 participants (62.4%) randomised to surgery and the 218 preference participants (83.5%) who actually received surgery in the first year, the perioperative complications that they experienced and their hospital stay have been reported previously but are summarised in *Appendix 2* for completeness. There were no perioperative deaths.

Table 11 shows the numbers of those who had fundoplication who subsequently had a second reflux-related operation during the 5 years of follow-up. Overall, this applied to 16 participants (4.4%) among the 364 who had a first operation: five (4.5%) in the randomised surgery group; one (4.2%) in the randomised medical group; eight (3.6%) in the preference surgery group; and two (33.3%) in the preference medical group. In total, five of the 16 operations were reconstructions of the same wrap, three were repairs of hiatus hernia only, six were conversions to a different type of wrap and two were reversals of the fundoplication. Two of these 16 participants had a third reflux-related operation; both were in the preference surgery group – one a reconstruction of the same wrap and one a repair of hiatus hernia only.

Late postoperative complications

Table 12 describes late postoperative complications among those participants who had surgery, in each of the study groups and overall. Of the total 364 who had fundoplication, 12 (3.3%) had a late complication: four (1.1%) were oesophageal dilatations/stricture dilatations; three (0.8%) were repairs of incisional hernias; and five (1.4%) were a heterogeneous group of other complications as detailed in the table.

Medication

Figure 2 summarises reported use of any PPI medication in the previous 2 weeks across the follow-up time points of the trial. Full details are provided in the tables in *Appendix 3*. From the time of the first annual follow-up onwards, rates in both medical groups were consistently around 80%. The rates in the randomised surgical ITT group at the first, second and third annual follow-ups were approximately 36–38%, rising to 43% in the fifth year. The extent to which these rates reflected medication taking among those allocated to surgery and who had fundoplication (rather than those who did not have surgery) can be gauged from the randomised surgery PP group: 7.3% (3 months), 12.5% (1 year), 15.1% (2 years), 19.6% (3 years), 23.9% (4 years) and 25.6% (5 years).

Table 13 allows further exploration of the reasons for the rise in medication use in the randomised surgery group. It distinguishes those reporting taking medication at the end of the first year of follow-up from

TABLE 11 Subsequent reflux-related operations among participants who had fundoplication

Surgery	Randomised participants		Preference participants		
	Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)	Total cohort
First fundoplication operation at any time, *n*	112	24	222	6	364
Second reflux-related reoperation, *n* (%)	5 (4.5)	1 (4.2)	8 (3.6)	2 (33.3)	16 (4.4)
Reconstruction of same wrap	2	1	1	1	5
Repair of hiatus hernia only	1	0	2	0	3
Conversion of type of wrap	2	0	4	0	6
Reversal of fundoplication	0	0	1	1	2
Third reflux-related reoperation, *n*					
Reconstruction of same wrap	0	0	1	0	1
Repair of hiatus hernia only	0	0	1	0	1
Conversion of type of wrap	0	0	0	0	0
Reversal of fundoplication	0	0	0	0	0

TABLE 12 Late postoperative complications (>1 month after surgery)

Complication	Randomised participants		Preference participants		
	Surgical (*n* = 178)	Medical (*n* = 179)	Surgical (*n* = 261)	Medical (*n* = 192)	Total cohort
First fundoplication operation at any time	112	24	222	6	364
Late postoperative complications (within first year of original operation), *n*					
Oesophageal dilatation/stricture dilatation	0	0	3	0	3
Repair of incisional hernia	0	0	1	0	1
Other (admission for deep-vein thrombosis/ pulmonary embolism)	0	0	1	0	1
Late postoperative complications (within second year following operation), *n*					
Oesophageal dilatation/stricture dilatation	1	0	0	0	1
Repair of incisional hernia	0	0	0	0	0
Other (pain from operation; hole between stomach and liver)	0	0	1	1	2
Late postoperative complications (beyond second year), *n*					
Oesophageal dilatation/stricture dilatation	0	0	0	0	0
Repair of incisional hernia	0	0	2	0	2
Other (pain due to original wrap shifting; bleed in stomach/bowel)	1	0	0	1	2
Total late postoperative complications, *n* (%)	2 (1.8)	0 (0.0)	8 (3.6)	2 (33.3)	12 (3.3)

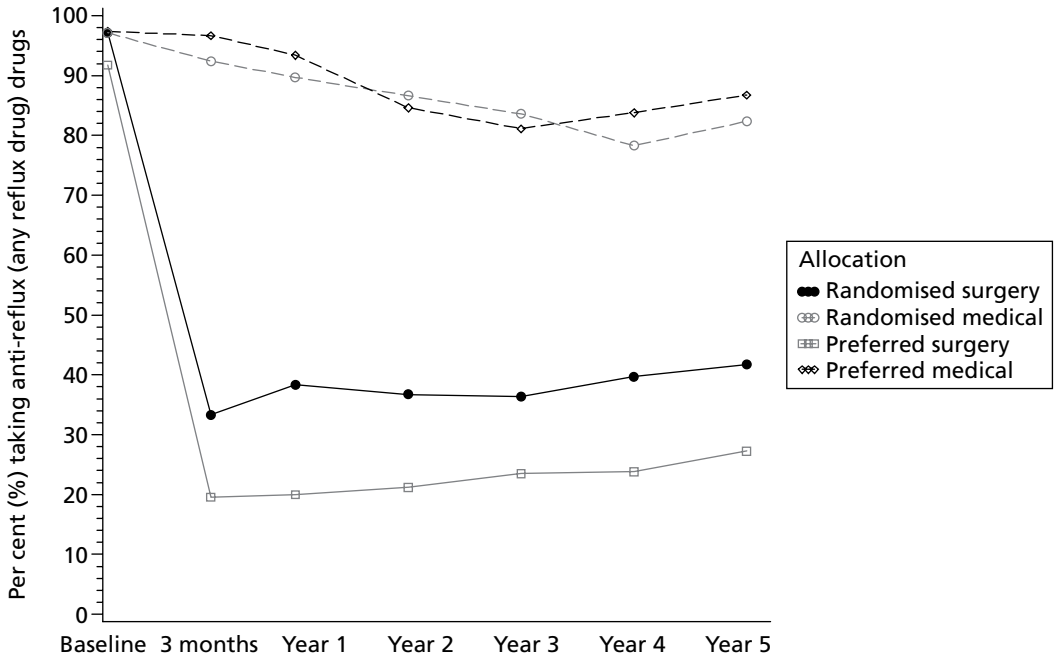


FIGURE 2 Use of PPI medication at baseline and at follow-up points up to 5 years.

those who indicated that they were not taking medication at that time. It shows that around 10–20% of those taking medication at the end of the first year did not report medication use at subsequent annual follow-up. Among those not taking medication at the first annual follow-up in the surgical groups, around 10% rising to around 20% reported medication use at subsequent annual follow-up. This contrasts with the rates in the medical groups, with around 50–60% of those not taking medication at the end of the first year reporting anti-reflux drug use in subsequent annual follow-up. The pattern of type of PPI used changed over the course of the study. Although lansoprazole had been the most commonly used PPI at trial entry, omeprazole use increased over time to become the predominant PPI.

Outcome

Health status

Full details of the health status and QoL measures at each time point of follow-up are in the tables in *Appendix 4*. Details of the statistical testing of the health status and QoL scores can be found in the next section of this chapter.

REFLUX score

Figure 3 summarises changes in the disease-specific REFLUX score over the follow-up period. From this it can be seen that the scores at all time points are highest (indicating fewest symptoms) in the randomised surgical and preference surgical groups. However, the differences between the surgical and medical groups narrow over time. This is due principally to the scores in the randomised medical group improving over the first 3 years and, to a lesser extent, those in the preference medical group improving over the latter end of the follow-up period. The scores for the five components of the measure are summarised graphically in *Figures 4–8*. These show that the overall difference between the groups is principally due to the 'general discomfort' component and, to a lesser extent, the 'nausea and vomiting' and 'activity limitations' components.

TABLE 13 Anti-reflux medication use after the first year

	Randomised participants				Preference participants			
	Surgery		Medical		Surgery		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Known whether or not taking medication at end of first year, *n*	154	104	165	156	232	205	181	178
Group taking anti-reflux drugs at end of the first year								
Taking anti-reflux drugs at end of the first year, *n*	51	10	140	137	42	20	154	152
Taking any anti-reflux drug at end of, *n*/*N* (%)								
Second year	37/42 (88)	7/8 (88)	111/119 (93)	110/116 (95)	27/34 (79)	14/17 (82)	117/130 (90)	116/128 (91)
Third year	34/41 (83)	8/9 (89)	101/115 (88)	101/114 (89)	26/33 (79)	12/14 (86)	117/134 (87)	116/132 (88)
Fourth year	34/41 (83)	7/9 (78)	94/112 (84)	93/110 (85)	20/28 (71)	9/13 (69)	106/119 (89)	105/117 (90)
Fifth year	33/39 (85)	8/9 (89)	89/101 (88)	88/99 (89)	20/29 (69)	10/12 (83)	105/117 (90)	103/115 (90)
Group not taking anti-reflux drugs at end of the first year								
Not taking anti-reflux drugs at end of first year, *n*	103	94	25	19	190	185	27	26
Taking any anti-reflux drug at end of, *n*/*N* (%)								
Second year	10/86 (12)	7/78 (9)	12/23 (52)	11/19 (58)	16/169 (9)	14/165 (8)	15/26 (58)	15/25 (60)
Third year	14/91 (15)	10/83 (12)	11/19 (58)	11/19 (58)	20/163 (12)	20/161 (12)	12/25 (48)	12/24 (50)
Fourth year	16/85 (19)	14/79 (18)	7/17 (41)	7/17 (41)	20/140 (14)	20/139 (14)	13/23 (57)	13/22 (59)
Fifth year	20/88 (23)	16/81 (20)	9/18 (50)	9/17 (53)	28/147 (19)	27/146 (18)	13/19 (68)	12/18 (67)

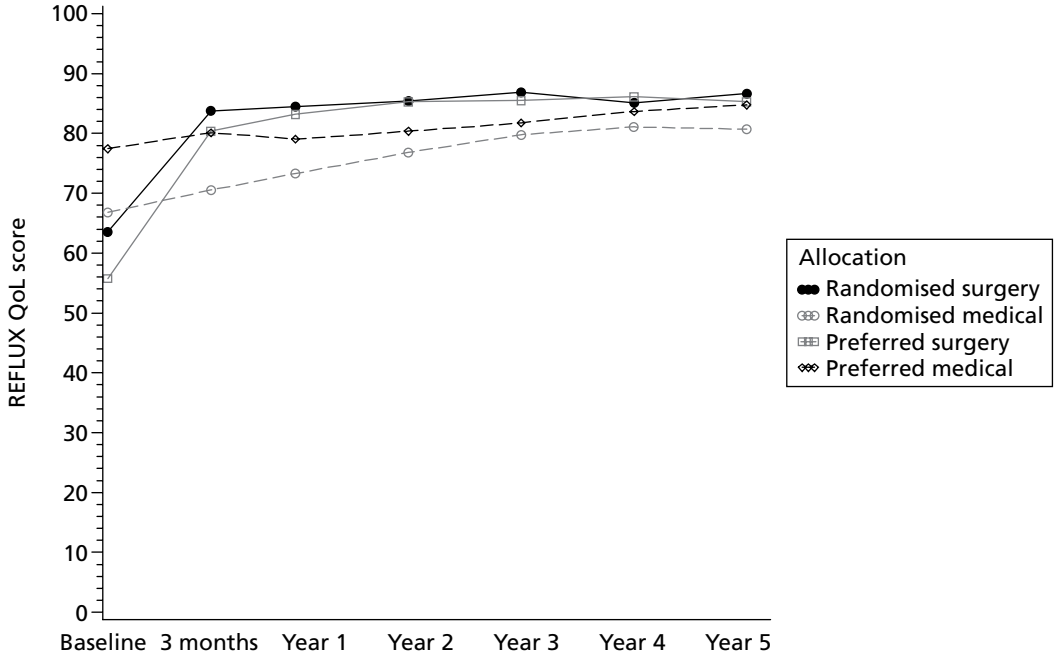


FIGURE 3 Mean REFLUX QoL score at baseline and at follow-up points up to 5 years (score range 0–100; the higher the score, the better the patient felt).

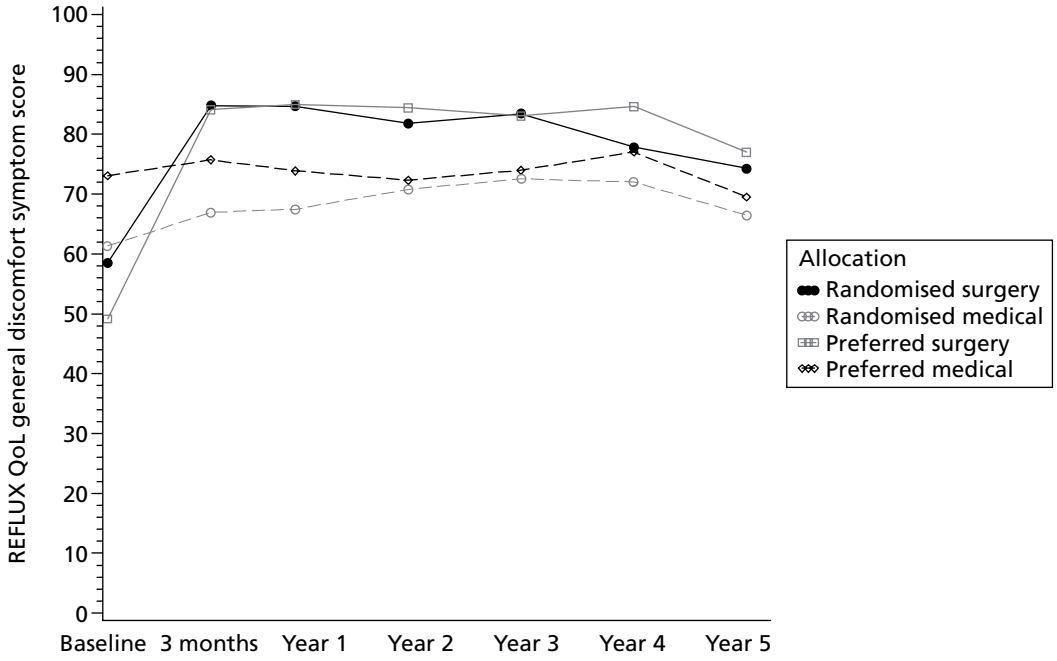


FIGURE 4 Mean REFLUX QoL general discomfort symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

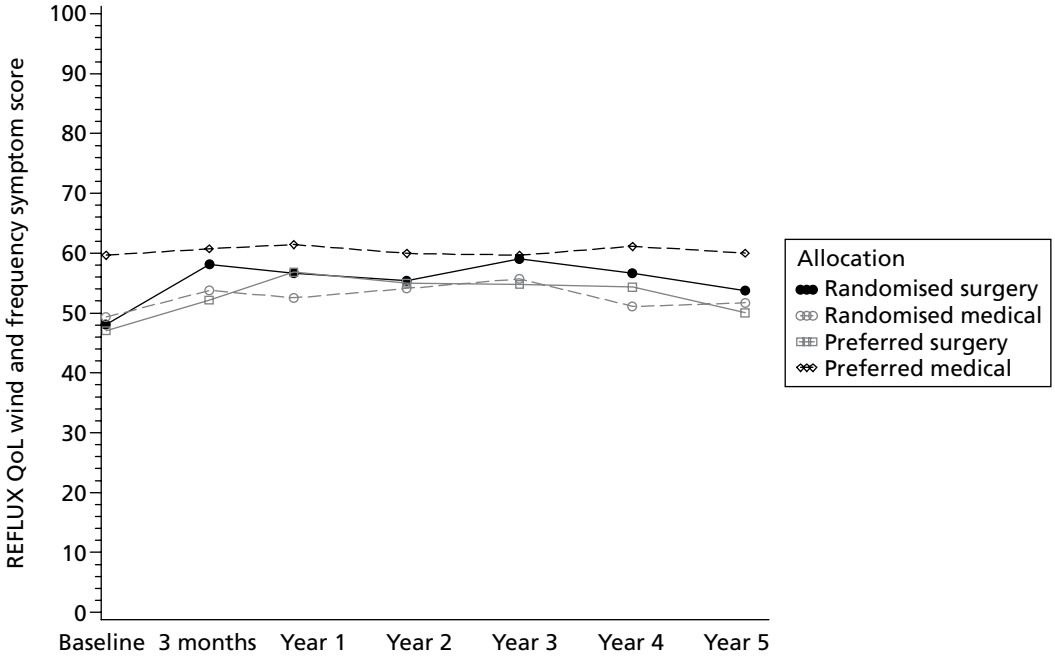


FIGURE 5 Mean REFLUX QoL wind and frequency symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

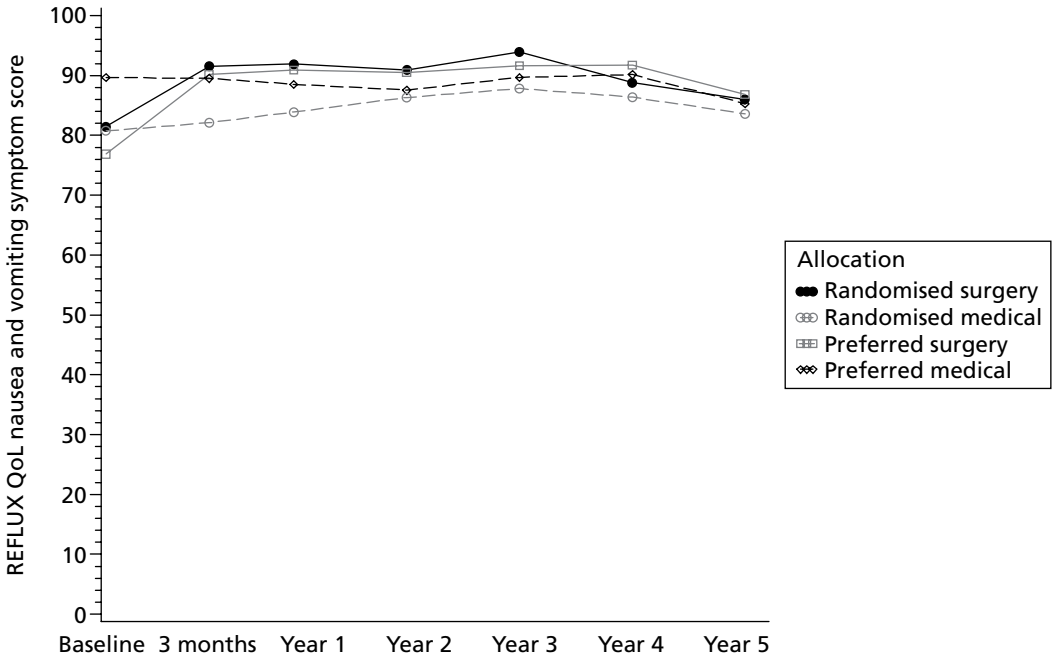


FIGURE 6 Mean REFLUX QoL nausea and vomiting symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

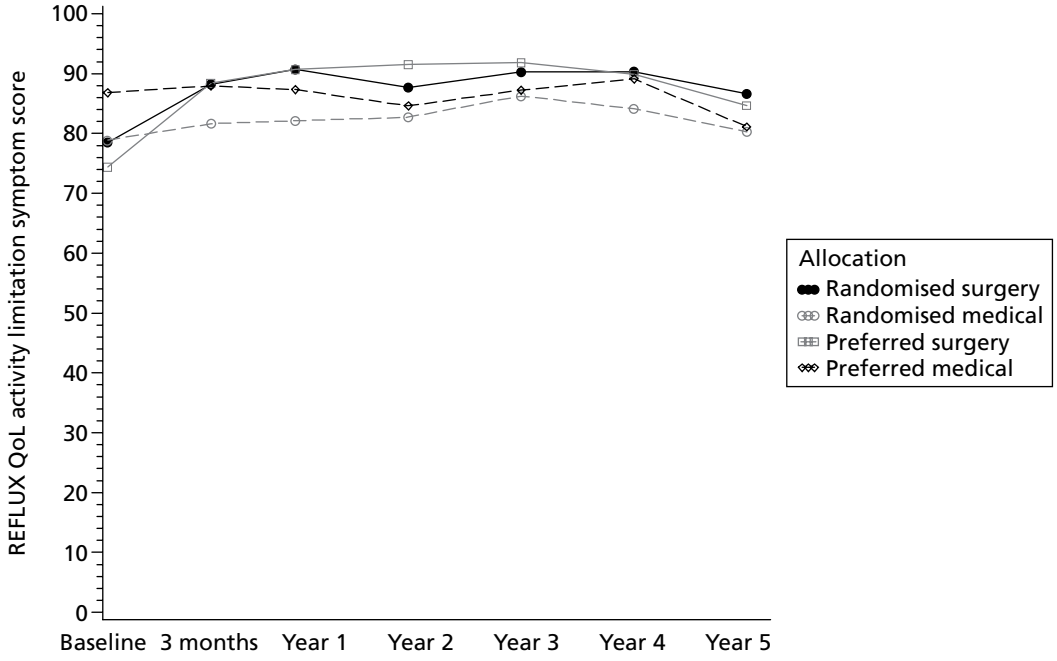


FIGURE 7 Mean REFLUX QoL activity limitation symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

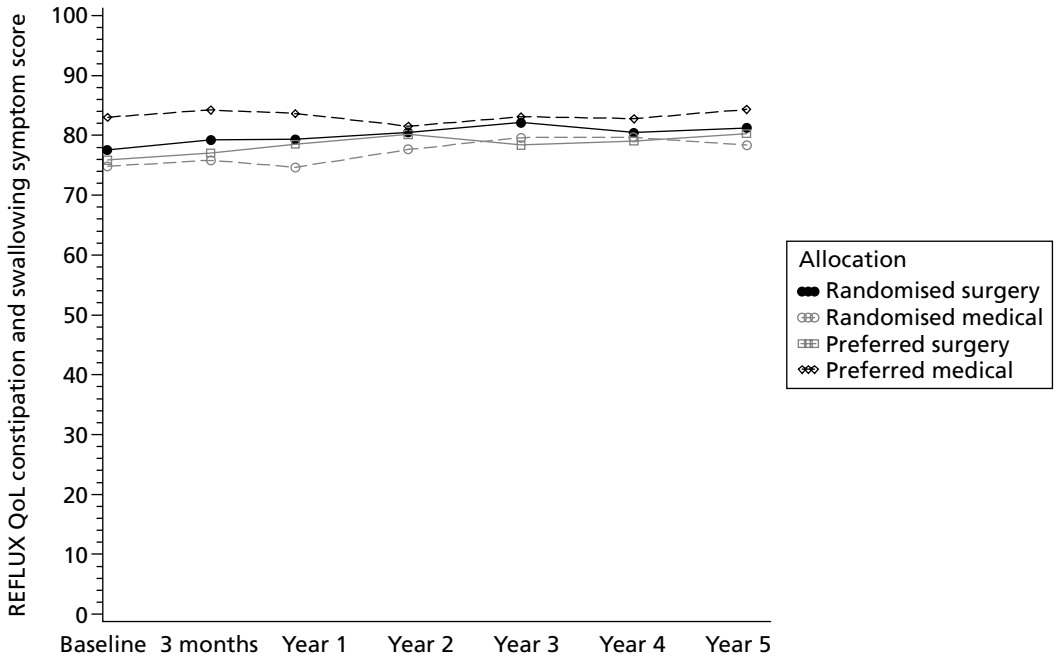


FIGURE 8 Mean REFLUX QoL constipation and swallowing symptom score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

Short Form questionnaire-36 items

The pattern of SF-36 scores, both for the composite physical and mental scores and for the individual dimensions (*Figures 9–16*), was similar to that seen for the REFLUX score, although more compact. Differences narrowed over the 5 years of follow-up, with the 'general health' dimension showing the clearest differences between the surgery and the medical management groups.

European Quality of Life-5 Dimensions

Figure 17 graphically displays the EQ-5D scores over the course of the follow-up period. The pattern is similar to that seen for the REFLUX score although differences are less marked and only clearly seen over the first 3 years.

Use of health services

Table 14 shows use of health services for the randomised groups. The larger number of overnight hospital admissions in the medical group largely reflected admissions for surgery; as described above, 14 participants allocated to medical management had fundoplication after the first year. However, seven participants in the medical group compared with one in the surgical group had admissions for a non-surgery-related reason (data not shown).

Numbers of day-case hospital admissions were similar in the two groups. The larger number of visits to or from a GP for a reflux-related reason in the randomised medical group reflected both more individuals attending their GPs and a higher frequency of visits for those who sought GP care.

Individual symptoms of gastro-oesophageal reflux disease or its treatment

Table 15 shows the frequency with which participants reported symptoms of GORD or its treatment at 3 and 5 years of follow-up for the randomised groups. At both 3 and 5 years, heartburn was reported by a higher proportion of participants in the randomised medical group than in the randomised surgical group. In addition, a higher proportion of participants in the randomised medical group reported more frequent heartburn than in the randomised surgical group. At both time points, a higher proportion of participants

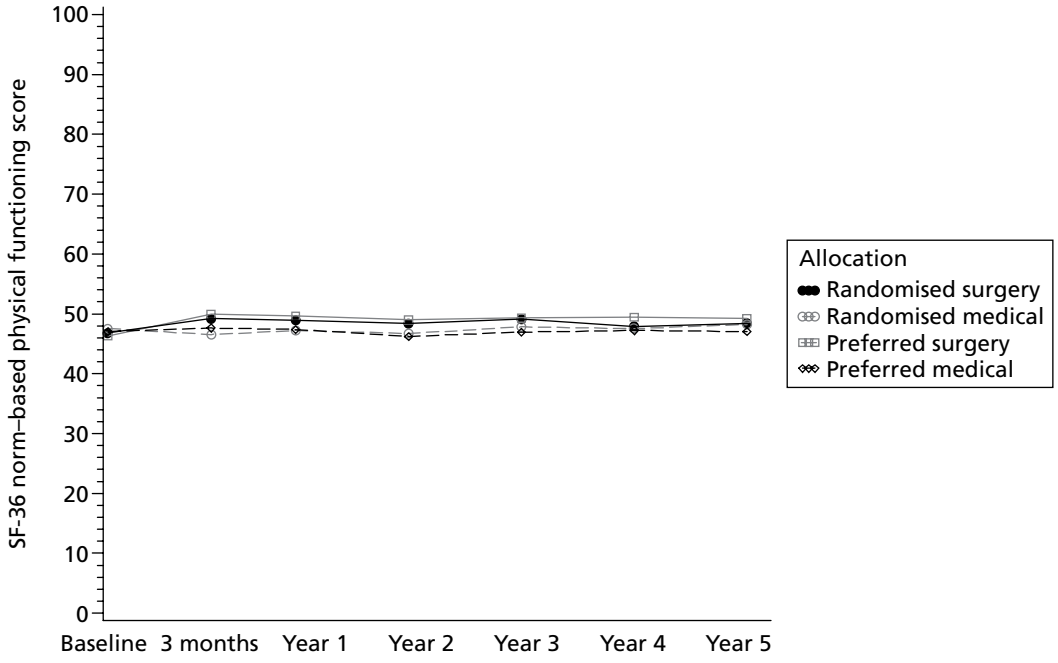


FIGURE 9 Mean SF-36 norm-based physical functioning score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

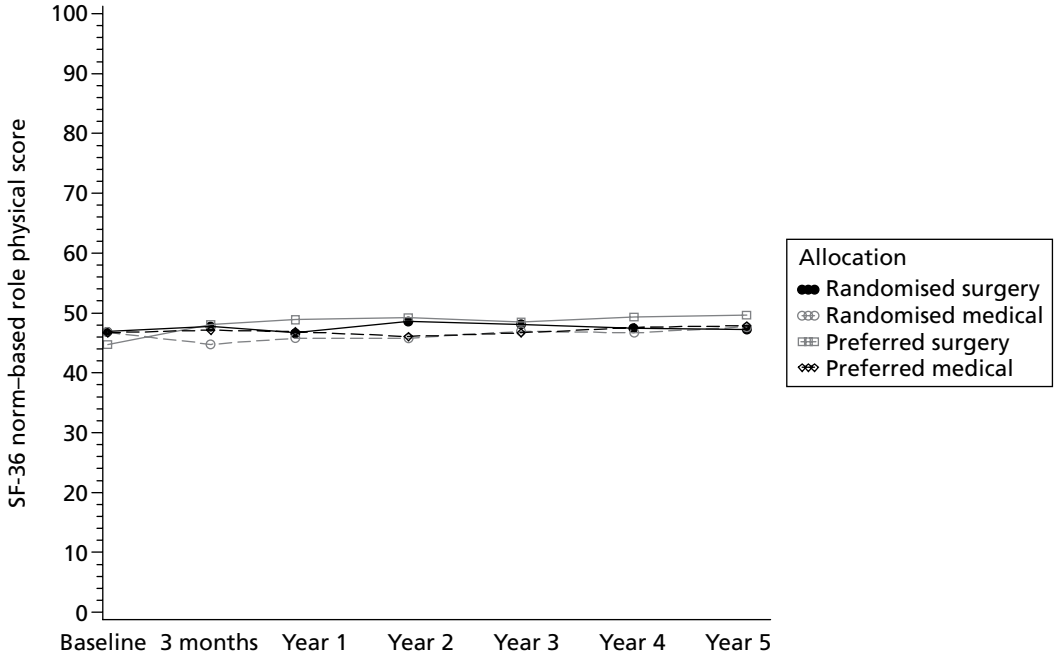


FIGURE 10 Mean SF-36 norm-based role physical score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

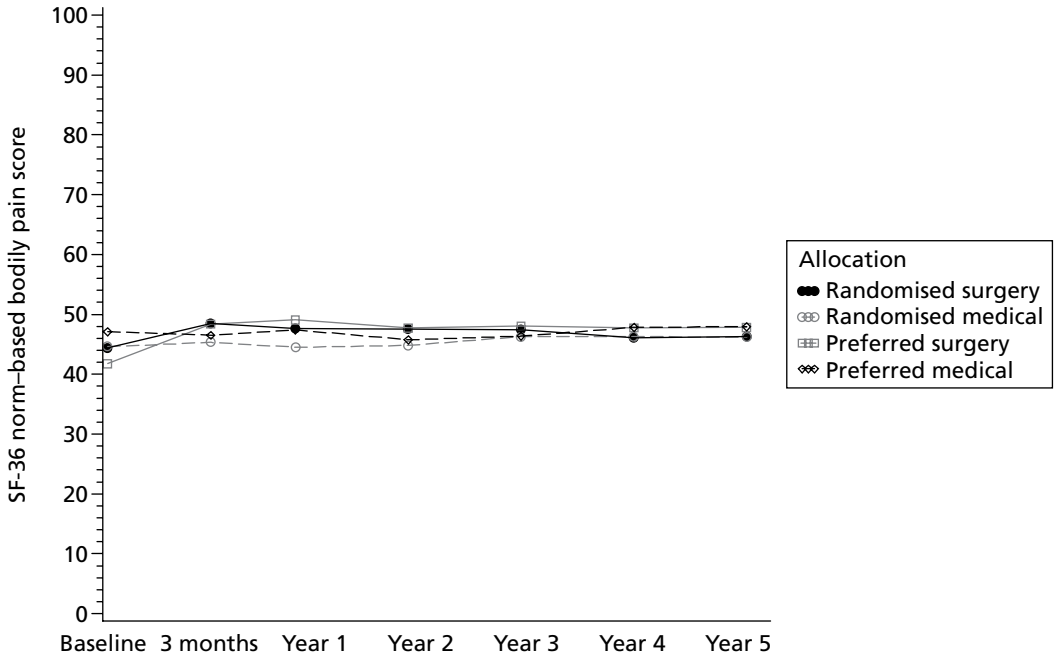


FIGURE 11 Mean SF-36 norm-based bodily pain score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

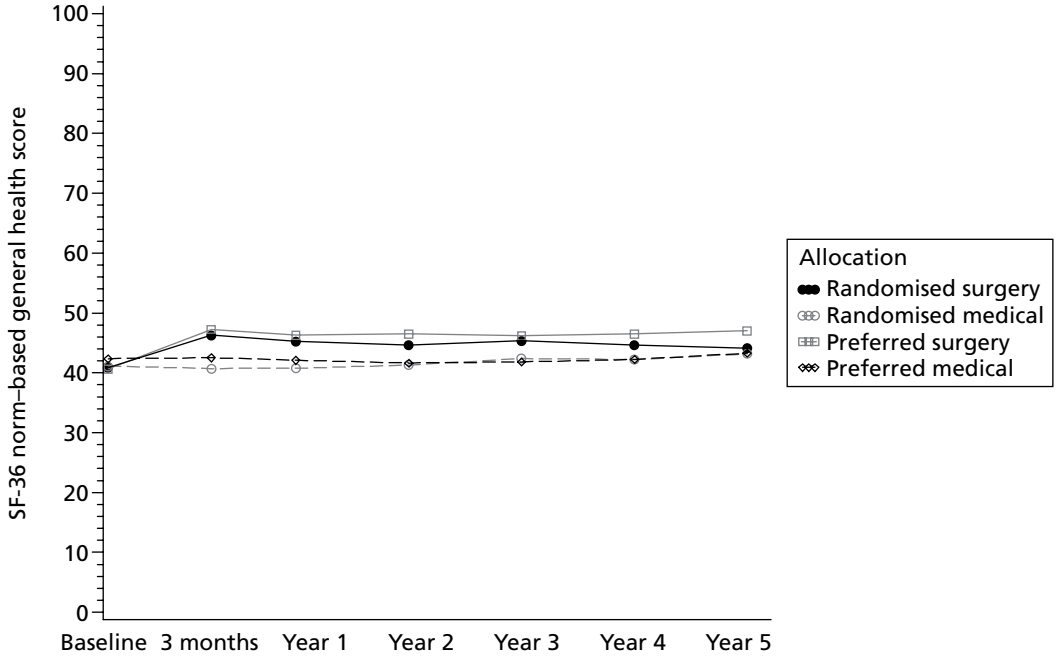


FIGURE 12 Mean SF-36 norm-based general health score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

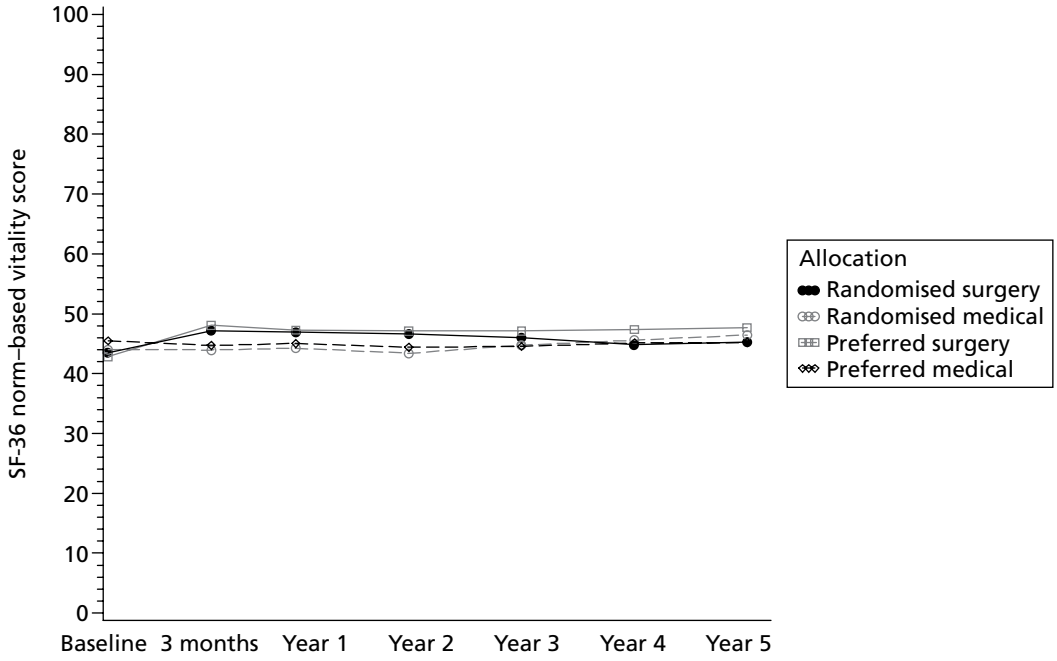


FIGURE 13 Mean SF-36 norm-based vitality score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

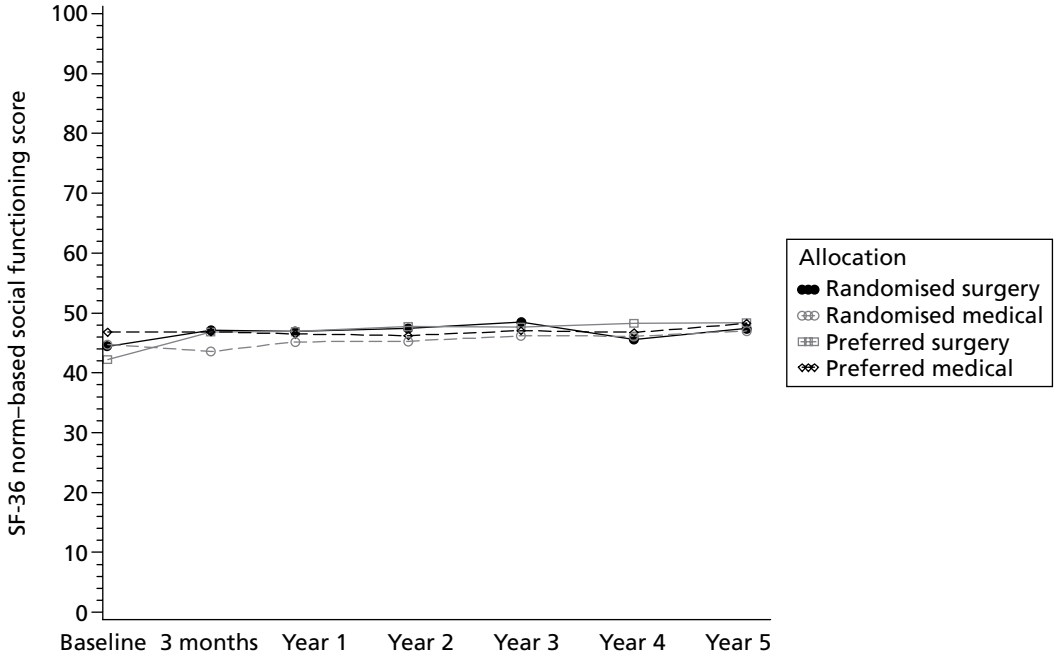


FIGURE 14 Mean SF-36 norm-based social functioning score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

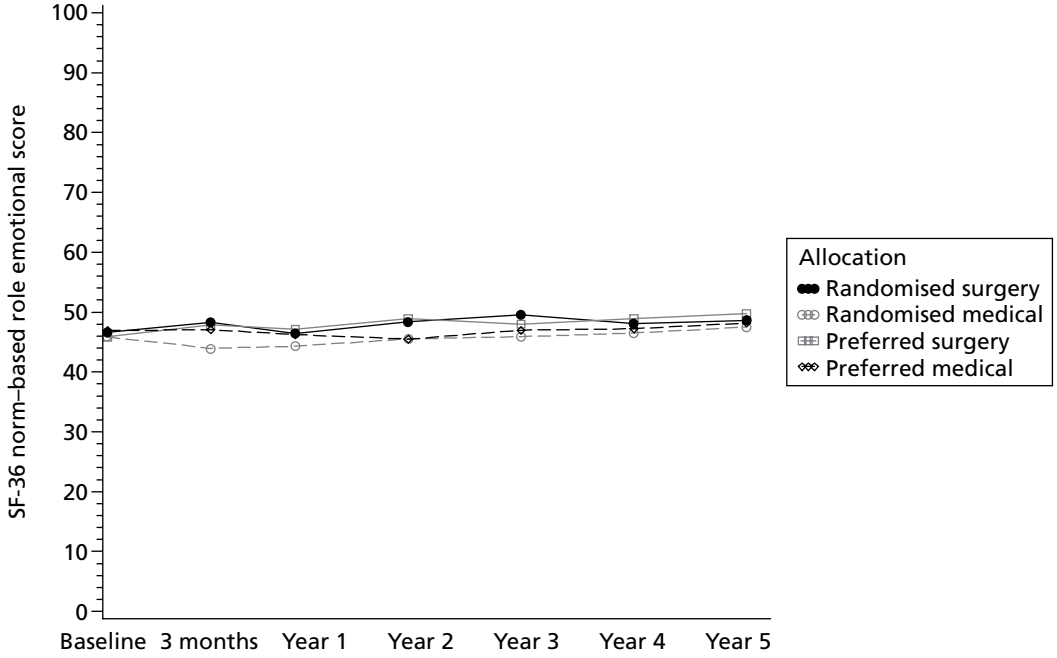


FIGURE 15 Mean SF-36 norm-based role emotional score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

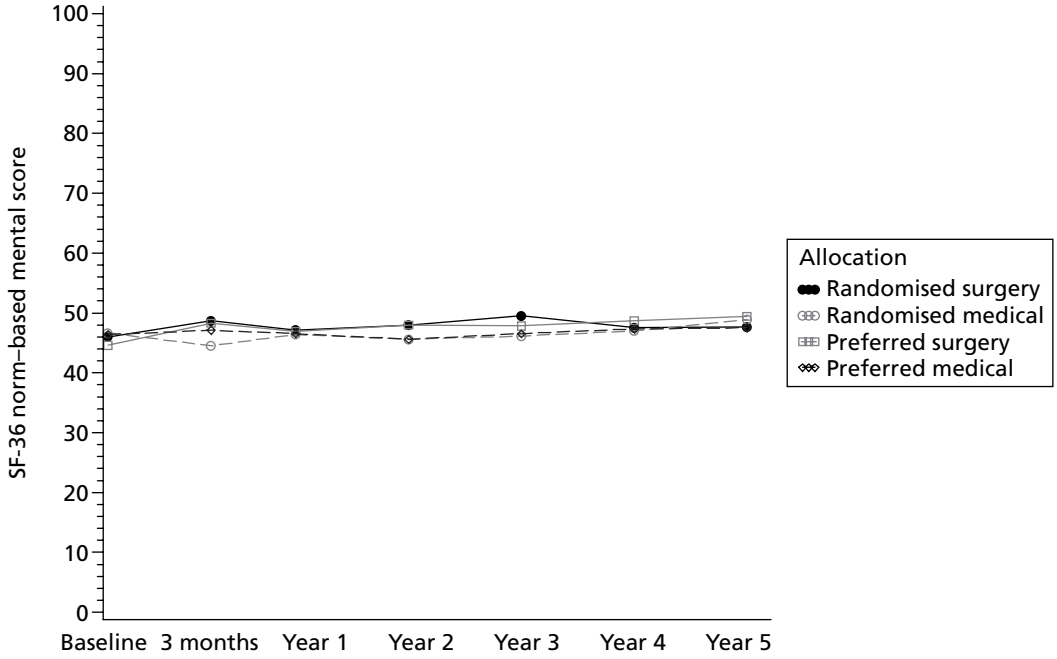


FIGURE 16 SF-36 norm-based mental score at baseline and follow-up points to 5 years (score range 0–100; the higher the score, the better the patient felt).

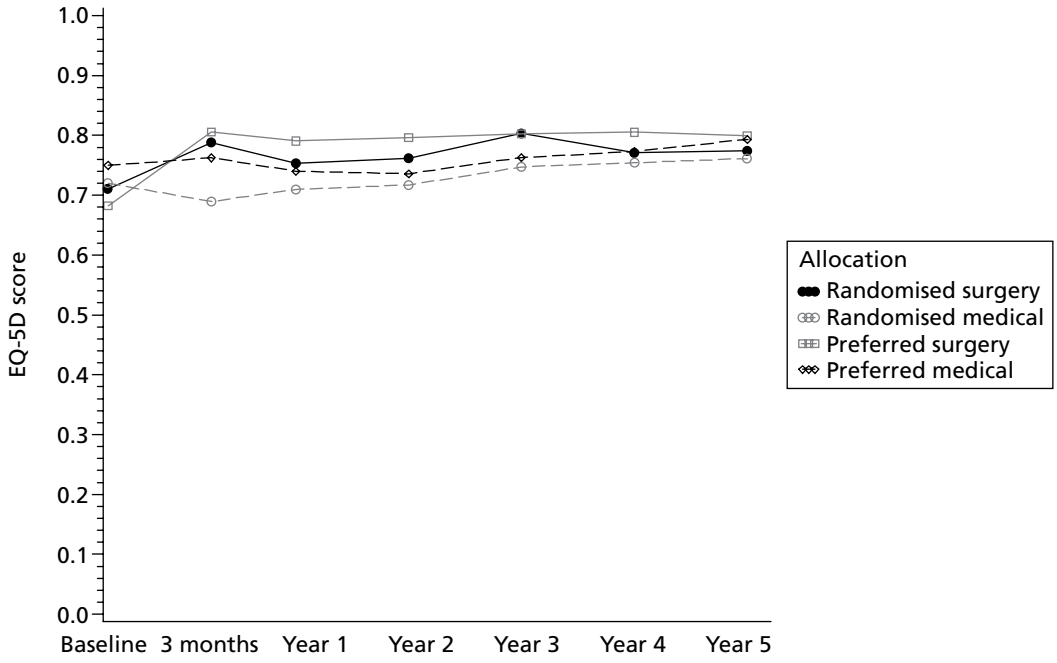


FIGURE 17 EQ-5D at baseline and follow-up points to 5 years.

TABLE 14 Use of health services

Use of health service	Year	Randomised surgical	Randomised medical
Overnight hospital admissions: reflux-related (and all reasons), *n*	1	4 (8)	2 (8)
	2	1 (8)	2 (10)
	3	2 (6)	9 (10)
	4	2 (2)	9 (10)
	5	0 (1)	8 (11)
Day hospital admissions: reflux related (and all reasons), *n*	1	22 (40)	24 (53)
	2	5 (23)	4 (24)
	3	4 (4)	6 (10)
	4	12 (13)	9 (11)
	5	4 (7)	11 (14)
Visits to and from the GP: reflux related (and all reasons), *n*	1	110 (394)	103 (376)
	2	34 (269)	115 (373)
	3	38 (381)	99 (386)
	4	55 (422)	126 (469)
	5	36 (404)	119 (370)

in the randomised medical management group also reported regurgitation symptoms and burping/belching than in the randomised surgical group. At both 3 and 5 years, the proportions who reported no difficulty swallowing and no wind from the lower bowel were similar between the randomised surgical and the randomised medical groups. There was also little difference between the groups at each time point in the proportion of participants who reported a feeling of wanting to be sick but being physically unable to do so.

Statistical analyses

Primary outcome

The pre-chosen primary outcome was the REFLUX QoL score after 5 years of follow-up. The differences between groups with corresponding 95% CIs are shown in *Table 16*. Two types of analysis are presented for the randomised participants – ITT and adjusted treatment received. *Table 16* also displays the impact of including adjustment for baseline score and randomised group*baseline score interaction terms.

Intention to treat

For the ITT analysis there was a mean difference of 6.4 between the groups in favour of surgery when only the minimisation variables were adjusted for (95% CI 1.6 to 11.2; $p = 0.009$). A repeated measures analysis across the 5 years gave a difference of 8.1 (95% CI 4.4 to 11.7). This was not the most parsimonious model – there was strong evidence of an interaction effect between randomised group and baseline REFLUX QoL score (interaction term was –0.23, 95% CI –0.43 to –0.03; $p = 0.023$). This implied that as baseline REFLUX QoL score increased the treatment effect decreased. Estimating the treatment difference at the trial baseline mean REFLUX QoL score of 65.2 resulted in a trial effect size of 8.5 (95% CI 3.9 to 13.1; $p < 0.001$). If the average patient had a lower mean REFLUX QoL score at baseline of 56.0, the effect size increased to 10.6 (95% CI 5.3 to 15.8). If the patient had a higher baseline score of 78.0, the treatment effect size decreased to 5.5 (95% CI 0.6 to 10.4). All results, however, showed strong evidence of increases in REFLUX QoL scores favouring surgery.

TABLE 15 Frequency of GORD symptoms at 3 and 5 years

GORD symptom	3 years		5 years	
	Randomised surgery	Randomised medical	Randomised surgery	Randomised medical
Frequency of heartburn, *n* (%)				
None at all	77 (58.8)	46 (34.8)	65 (58.6)	28 (26.4)
One to three times per week	44 (33.6)	64 (48.5)	38 (34.2)	64 (60.4)
More than three times per week	10 (7.6)	22 (16.7)	8 (7.2)	14 (13.2)
Frequency of regurgitation, *n* (%)				
None at all	102 (77.3)	83 (61.9)	89 (75.4)	71 (63.4)
One to three times per week	27 (20.5)	47 (35.1)	26 (22.0)	37 (33.0)
More than three times per week	3 (2.3)	4 (3.0)	3 (2.5)	4 (3.6)
Frequency of difficulty swallowing, *n* (%)				
None at all	100 (75.8)	102 (76.1)	91 (77.1)	82 (74.5)
One to three times per week	30 (22.7)	27 (20.1)	25 (21.2)	25 (22.7)
More than three times per week	2 (1.5)	5 (3.7)	2 (1.7)	3 (2.7)
Frequency of wind from the bowel, *n* (%)				
None at all	19 (14.4)	20 (15.0)	14 (11.9)	14 (12.7)
One to three times per week	37 (28.0)	35 (26.3)	27 (22.9)	30 (27.3)
More than three times per week	76 (57.6)	78 (58.6)	77 (65.3)	66 (60.0)
Frequency of burping/belching, *n* (%)				
None at all	53 (40.2)	33 (24.8)	46 (39.3)	27 (24.5)
One to three times per week	39 (29.5)	48 (36.1)	40 (34.2)	37 (33.6)
More than three times per week	40 (30.3)	52 (39.1)	31 (26.5)	46 (41.8)
Frequency of wanting to be sick but being physically unable to, *n* (%)				
None at all	116 (87.9)	110 (83.3)	101 (85.6)	92 (82.9)
One to three times per week	15 (11.4)	17 (12.9)	15 (12.7)	16 (14.4)
More than three times per week	1 (0.8)	5 (3.8)	2 (1.7)	3 (2.7)

TABLE 16 Primary outcome: REFLUX QoL scores after 5 years of follow-up

REFLUX QoL score	Randomised participants					
	ITT			Adjusted treatment received		
	Mean difference[a]	95% CI	*p*-value	Mean difference[a]	95% CI	*p*-value
Adjusted for minimisation variables	6.4	1.6 to 11.2	0.009	9.4	1.7 to 17.0	0.017
Adjusted for minimisation variables and baseline REFLUX QoL score	7.6	3.0 to 12.2	0.001	10.6	3.3 to 17.9	0.004
Adjusted for minimisation variables, baseline score and treatment*baseline REFLUX QoL score interaction	8.5	3.9 to 13.1	<0.001	11.5	4.2 to 18.7	0.002

a Difference is surgery group minus medical group.

Adjusted treatment received

The adjusted treatment received analyses attempted to mitigate the effect of non-compliance with the allocated treatment and hence provide an estimate of 'efficacy'.[40] As expected, this approach gave a larger difference, but with wider CIs (9.4, 95% CI 1.7 to 17.0; $p = 0.017$).

Preference groups

The preference for surgery participants reported considerably worse mean REFLUX QoL scores at baseline than the preference for medicine participants (55.8 vs 77.5) (see *Table 7*). Despite starting from a much lower baseline score, at follow-up, the REFLUX QoL score slightly favoured the surgical group using an ITT analysis (difference = 0.61; 95% CI –3.44 to 4.66; $p = 0.767$) and an adjusted treatment received analysis (difference = 0.10; 95% CI –4.77 to 4.97; $p = 0.967$). The differences were not, however, statistically significant.

Secondary outcomes

The secondary outcomes were the health status measures (EQ-5D, SF-36) and REFLUX symptom score at times equivalent to 3 months and then annual follow-up after surgery, and REFLUX QoL (at time points other than 5 years, when it was the primary end point). Analyses of these outcomes are shown in *Tables 17–22*.

REFLUX symptom score

There were statistically significantly higher REFLUX QoL scores at all time points, albeit with some diminution over time in the surgical group (see *Figure 3*). Although symptom category scores favoured surgery across all domains at all time points, the most marked and sustained difference was in 'general discomfort'.

Short Form questionnaire-36 items

The SF-36 scores in all domains also favoured the surgical group at all time points. Differences decreased over time and this was reflected in most *p*-values being <0.05 up to 3 years, whereas at year 5 this applied to only 'norm-based general health' and 'norm-based role emotional'.

European Quality of Life-5 Dimensions

Differences in EQ-5D had a similar pattern to differences in REFLUX QoL and SF-36 scores – differences all favoured the surgical group but tended to narrow such that scores at years 2 and 3 were statistically significantly different, but at later time points they were not. Variability tended to increase over time. Despite the general narrowing of the EQ-5D difference over time, at year 5 it was actually the same as that at 12 months after surgery but with wider CIs.

Adjusted treatment received

As would be expected, all (with a small number of exceptions) the adjusted treatment received analyses had larger differences than the corresponding ITT analyses (around 25–50% higher), but with wider CIs.

Subgroup analyses

Removal of data from the single largest clinical centre (Aberdeen)

No formal exploration of centre effects was undertaken because of the small numbers of participants recruited in many of the clinical centres. However, a sensitivity analysis removing the data from the Aberdeen centre, the centre where the largest number of participants were recruited, did not significantly change the conclusions (adjusted difference in REFLUX score at 60 months = 5.43, 95% CI 0.96 to 9.90).

Partial compared with total wrap procedure

In an observational analysis, there was no evidence of a difference between a total wrap procedure and a partial wrap procedure. The difference in the REFLUX QoL score between these procedures at time equivalent to 5 years post surgery was –1.0 (95% CI –5.4 to 3.7; $p = 0.649$).

TABLE 17 Secondary outcomes at a time equivalent to 3 months after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	15.0	10.5 to 19.4	<0.001	20.7	13.9 to 27.5	<0.001
REFLUX symptom score						
General discomfort symptom score	19.2	14.9 to 23.6	<0.001	26.0	19.6 to 32.4	<0.001
Wind and frequency symptom score	4.6	0.5 to 8.6	0.027	5.1	–1.0 to 11.3	0.101
Nausea and vomiting symptom score	8.8	5.8 to 11.9	<0.001	12.4	7.7 to 17.1	<0.001
Activity limitation symptom score	7.1	3.2 to 11.0	<0.001	9.1	3.2 to 15.1	0.003
Constipation and swallowing symptom score	2.0	–1.9 to 6.0	0.318	2.1	–3.9 to 8.2	0.486
SF-36 score						
Norm-based physical functioning	3.1[b]	1.3 to 4.9	0.001	4.4[b]	1.5 to 7.2	0.003
Norm-based role physical	2.7	0.5 to 4.9	0.018	3.4	–0.04 to 6.8	0.053
Norm-based bodily pain	3.2[b]	1.1 to 5.3	0.003	4.1[b]	0.9 to 7.2	0.012
Norm-based general health	5.8[b]	3.8 to 7.8	<0.001	7.8[b]	4.8 to 10.7	<0.001
Norm-based vitality	3.0	0.9 to 5.1	0.006	3.9	0.7 to 7.1	0.018
Norm-based social functioning	3.6	1.3 to 5.8	0.002	4.6	1.1 to 8.1	0.010
Norm-based role emotional	3.3	0.7 to 5.8	0.012	4.1	0.2 to 8.0	0.042
Norm-based mental health	4.2[b]	2.1 to 6.2	<0.001	5.5[b]	2.4 to 8.6	0.001
EQ-5D, mean (SD)	0.099[b]	0.048 to 0.150	<0.001	0.129[b]	0.051 to 0.207	0.001

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 18 Secondary outcomes at a time equivalent to 12 months after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	14.0	9.6 to 18.4	<0.001	19.4	13.0 to 25.8	<0.001
REFLUX symptom score						
General discomfort symptom score	18.3	13.8 to 22.9	<0.001	26.1	19.6 to 32.5	<0.001
Wind and frequency symptom score	4.9	0.8 to 9.1	0.019	6.7	0.6 to 12.8	0.033
Nausea and vomiting symptom score	7.8	4.6 to 10.9	<0.001	11.5	7.0 to 16.0	<0.001
Activity limitation symptom score	8.4	5.2 to 11.7	<0.001	12.0	7.3 to 16.7	<0.001
Constipation and swallowing symptom score	3.5	–0.5 to 7.5	0.085	5.0	–0.9 to 10.9	0.099
SF-36 score						
Norm-based physical functioning	2.3[b]	0.6 to 4.0	0.007	3.4[b]	0.9 to 5.9	0.008
Norm-based role physical	0.9	–1.1 to 3.0	0.383	1.2	–1.8 to 4.3	0.434
Norm-based bodily pain	3.4[b]	1.4 to 5.5	0.001	5.1[b]	2.1 to 8.0	0.001
Norm-based general health	4.8[b]	2.7 to 6.8	<0.001	7.0[b]	4.0 to 10.0	<0.001
Norm-based vitality	2.5	0.4 to 4.6	0.018	3.7	0.6 to 6.8	0.019
Norm-based social functioning	2.3	0.1 to 4.5	0.040	3.3	0.04 to 6.6	0.047
Norm-based role emotional	1.8	–0.8 to 4.4	0.177	2.7	–1.1 to 6.5	0.168
Norm-based mental health	1.0[b]	–1.0 to 3.1	0.312	1.5[b]	–1.5 to 4.5	0.324
EQ-5D	0.047[b]	–0.004 to 0.097	0.07	0.068[b]	–0.006 to 0.142	0.072

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 19 Secondary outcomes at a time equivalent to 2 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	11.4	6.8 to 16.0	<0.001	15.7	8.5 to 22.9	<0.001
REFLUX symptom score						
General discomfort symptom score	13.08	7.99 to 18.17	<0.001	17.66	9.82 to 25.50	<0.001
Wind and frequency symptom score	3.74[b]	–1.06 to 8.53	0.126	5.67[b]	–2.05 to 13.38	0.149
Nausea and vomiting symptom score	6.34	2.85 to 9.83	<0.001	9.48	4.04 to 14.92	0.001
Activity limitation symptom score	7.02	3.38 to 10.65	<0.001	10.03	4.25 to 15.80	0.001
Constipation and swallowing symptom score	3.29[b]	–1.11 to 7.68	0.142	4.98[b]	–2.09 to 12.05	0.167
SF-36 score						
Norm-based physical functioning	2.73	0.83 to 4.63	0.005	4.27	1.21 to 7.34	0.007
Norm-based role physical	3.11	0.99 to 5.22	0.004	4.69	1.27 to 8.10	0.007
Norm-based bodily pain	3.64	1.51 to 5.77	0.001	5.46	2.04 to 8.88	0.002
Norm-based general health	4.13	1.91 to 6.35	<0.001	5.96	2.39 to 9.54	0.001
Norm-based vitality	3.48	1.20 to 5.76	0.003	5.38	1.66 to 9.09	0.005
Norm-based social functioning	2.74	0.30 to 5.19	0.028	3.79[b]	–0.14 to 7.72	0.059
Norm-based role emotional	2.03[b]	–0.80 to 4.85	0.159	3.06[b]	–1.49 to 7.61	0.187
Norm-based mental health	2.33	0.08 to 4.59	0.043	3.86	0.22 to 7.49	0.038
EQ-5D	0.068[b]	0.005 to 0.131	0.036	0.098[b]	–0.003 to 0.199	0.057

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 20 Secondary outcomes at a time equivalent to 3 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	9.0	4.9 to 13.1	<0.001	12.9	6.3 to 19.5	<0.001
REFLUX symptom score						
General discomfort symptom score	11.86	6.84 to 16.88	<0.001	16.25	8.37 to 24.14	<0.001
Wind and frequency symptom score	4.98[b]	–0.26 to 10.22	0.063	15.95[b]	8.03 to 23.87	<0.001
Nausea and vomiting symptom score	6.69	3.65 to 9.73	<0.001	9.71	4.98 to 14.44	<0.001
Activity limitation symptom score	4.61	0.99 to 8.22	0.013	6.37	0.58 to 12.15	0.031
Constipation and swallowing symptom score	2.62[b]	–1.51 to 6.76	0.212	6.51[b]	0.73 to 12.29	0.027
SF-36 score						
Norm-based physical functioning	2.61	0.56 to 4.67	0.013	3.83	0.52 to 7.14	0.023
Norm-based role physical	1.82[b]	–0.43 to 4.07	0.113	3.82[b]	0.52 to 7.12	0.024
Norm-based bodily pain	2.33	0.24 to 4.42	0.029	3.74	0.36 to 7.12	0.030
Norm-based general health	3.69	1.50 to 5.87	0.001	5.21	1.70 to 8.73	0.004
Norm-based vitality	2.29[b]	–0.23 to 4.81	0.075	5.29[b]	1.77 to 8.81	0.003
Norm-based social functioning	3.27	0.87 to 5.68	0.008	4.81	0.93 to 8.69	0.015
Norm-based role emotional	4.03	1.50 to 6.57	0.002	6.89	2.77 to 11.01	0.001
Norm-based mental health	4.60	2.29 to 6.91	<0.001	7.39	3.65 to 11.14	<0.001
EQ-5D, mean (SD)	0.070[b]	0.015 to 0.126	0.013	0.108	0.016 to 0.201	0.022

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 21 Secondary outcomes at a time equivalent to 4 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX QoL	8.3	3.2 to 13.4	0.001	11.6	3.5 to 19.8	0.005
REFLUX symptom score						
General discomfort symptom score	8.81	3.49 to 14.13	0.001	11.48	3.11 to 19.84	0.007
Wind and frequency symptom score	5.98	0.70 to 11.26	0.027	9.55	0.95 to 18.14	0.030
Nausea and vomiting symptom score	2.93[b]	–1.00 to 6.86	0.143	3.25[b]	–3.01 to 9.51	0.307
Activity limitation symptom score	4.38	0.64 to 8.12	0.022	5.95[b]	–0.03 to 11.93	0.051
Constipation and swallowing symptom score	0.26[b]	–4.21 to 4.74	0.908	0.54[b]	–6.72 to 7.80	0.884
SF-36 score						
Norm-based physical functioning	2.14	0.00 to 4.28	0.050	3.10[b]	–0.36 to 6.55	0.079
Norm-based role physical	1.36[b]	–1.23 to 3.96	0.302	2.42[b]	–1.79 to 6.62	0.259
Norm-based bodily pain	1.72[b]	–0.57 to 4.02	0.140	2.59[b]	–1.13 to 6.31	0.172
Norm-based general health	4.02	1.61 to 6.44	0.001	5.74	1.84 to 9.63	0.004
Norm-based vitality	0.17[b]	–2.25 to 2.60	0.888	0.28[b]	–3.66 to 4.22	0.890
Norm-based social functioning	1.26[b]	–1.60 to 4.12	0.387	1.92[b]	–2.72 to 6.56	0.416
Norm-based role emotional	1.79[b]	–1.28 to 4.85	0.253	2.77[b]	–2.21 to 7.75	0.274
Norm-based mental health	1.55[b]	–1.03 to 4.12	0.238	1.85[b]	–2.31 to 6.00	0.382
EQ-5D	0.036[b]	–0.020 to 0.091	0.212	0.052[b]	–0.039 to 0.142	0.265

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

TABLE 22 Secondary outcomes at a time equivalent to 5 years after surgery: health status

	Randomised participants					
	ITT			Adjusted treatment received		
Secondary outcomes	Difference[a]	95% CI	*p*-value	Difference[a]	95% CI	*p*-value
REFLUX symptom score						
General discomfort symptom score	11.82	6.50 to 17.14	<0.001	15.59	7.52 to 23.66	<0.001
Wind and frequency symptom score	3.34[b]	–1.98 to 8.66	0.218	5.12[b]	–3.50 to 13.73	0.243
Nausea and vomiting symptom score	4.97	1.53 to 8.41	0.005	7.32	2.04 to 12.60	0.007
Activity limitation symptom score	5.97	2.03 to 9.91	0.003	8.27	2.03 to 14.52	0.010
Constipation and swallowing symptom score	2.54[b]	–2.09 to 7.18	0.281	4.11[b]	–3.40 to 11.62	0.282
SF-36 score						
Norm-based physical functioning	2.01[b]	–0.26 to 4.28	0.082	3.35[b]	–0.33 to 7.03	0.074
Norm-based role physical	0.57[b]	–2.10 to 3.24	0.674	1.14[b]	–3.20 to 5.47	0.606
Norm-based bodily pain	1.52[b]	–0.90 to 3.94	0.218	1.65[b]	–2.25 to 5.54	0.406
Norm-based general health	2.76	0.21 to 5.31	0.034	3.79[b]	–0.29 to 7.88	0.068
Norm-based vitality	0.37[b]	–2.23 to 2.98	0.777	0.19[b]	–4.03 to 4.41	0.928
Norm-based social functioning	1.72[b]	–1.05 to 4.49	0.221	2.36[b]	–2.13 to 6.84	0.301
Norm-based role emotional	2.67	0.07 to 5.27	0.044	4.56	0.34 to 8.79	0.034
Norm-based mental health	0.59[b]	–1.96 to 3.14	0.650	0.40[b]	–3.72 to 4.51	0.849
EQ-5D	0.047[b]	–0.013 to 0.108	0.126	0.069[b]	–0.029 to 0.167	0.168

a Difference is the mean of the surgery group minus the mean of the medical group. All analyses adjusted for BMI, age, sex, baseline score and baseline*group interaction.

b Adjusted for BMI, age, sex and baseline score. Baseline*group interaction term not fitted.

Discussion

Follow-up to 5 years after laparoscopic surgery described here provides clear evidence of sustained improvement in GORD symptoms, as judged by the REFLUX QoL scores. Differences between the groups as randomised did tend to diminish over the course of the study; nevertheless, the analyses at 5 years (the primary end point) showed highly statistically significant results with effect sizes of the order of 0.6 of a SD.

This report concentrates on the data collected annually at a time equivalent to between 2 and 5 years post surgery. Data were collected through self-complete postal questionnaires, backed up by postal and telephone reminders and occasional completion of the questionnaire over the telephone. The response rate did drop over time, from 90% at 1 year to around 70% at 5 years. The principal reason for not obtaining a follow-up questionnaire was a loss of contact, such as following a home move; the second most common reason was a decision by a participant to decline further follow-up. The category of 'non-responder' accounted for only around 8% of those without a follow-up questionnaire. Response analysis showed that responders at 5-year follow-up had a higher mean age, had been prescribed anti-reflux medication for a shorter period of time at recruitment and had higher QoL at baseline. However, the characteristics of responders and non-responders at 5 years were similar across the two randomised groups.

Randomised trials, such as the REFLUX trial, that compare surgery with medical management are challenging to mount because of the stark contrast between the treatments compared. As described in the previous report of this study, recruitment was not easy and it is to the credit of the many staff in the 21 centres involved in the trial that this was accomplished successfully. A second challenge was that, after randomisation, a sizable proportion of participants did not receive the treatment to which they had been allocated – again, reflecting the contrasts in the treatments. We explored the impact of this in a number of ways.

Figure 18 shows the results of a supplementary analysis of the group randomly allocated surgery stratified by whether or not they actually had surgery. It shows that those who had surgery started from a lower REFLUX QoL baseline score (had worse symptoms) than those who did not undergo surgery, and then had a sharp rise in score following the operation such that their scores were consistently higher than those who did not actually have fundoplication. To put this another way, the improvement seen among those who had surgery was greater than that in the randomised group overall.

Figure 19 shows a similar supplementary analysis of the group allocated medical management stratified by whether or not they in fact had surgery in the first year. This shows that those who had fundoplication (the lowest line) had more severe symptoms of GORD (low REFLUX QoL scores) at the time of trial entry, worse even than the preference surgical group. In contrast, those solely managed medically had relatively high baseline scores. Scores among those randomised to medical management who had surgery improved markedly over the course of the follow-up, such that by years 4 and 5 the scores in the two strata were similar. This indicates that much of the narrowing of the scores in the ITT groups over the 5 years can be explained by surgery in the randomised medical group.

We assessed more formally the extent to which surgery in the randomised medical management group might have affected the results by undertaking adjusted treatment received analyses. We decided to base these on treatment status at the first year follow-up point. We chose this partly to be consistent with our previous report of the results up to 1 year and partly because we considered that those who had surgery after that time point were likely to be highly selected. To put this another way, we were concerned that a PP analysis up to 5 years would be particularly prone to bias. The adjusted treatment received analyses, as expected, indicated larger effects of surgery – with differences in score around 25–50% higher. As illustrated by the preference groups in this study, the proportion of those recommended surgery and willing to have it who subsequently go on to have fundoplication is likely to be higher in everyday practice. Hence, we would argue that the results of the adjusted treatment received analyses are likely to

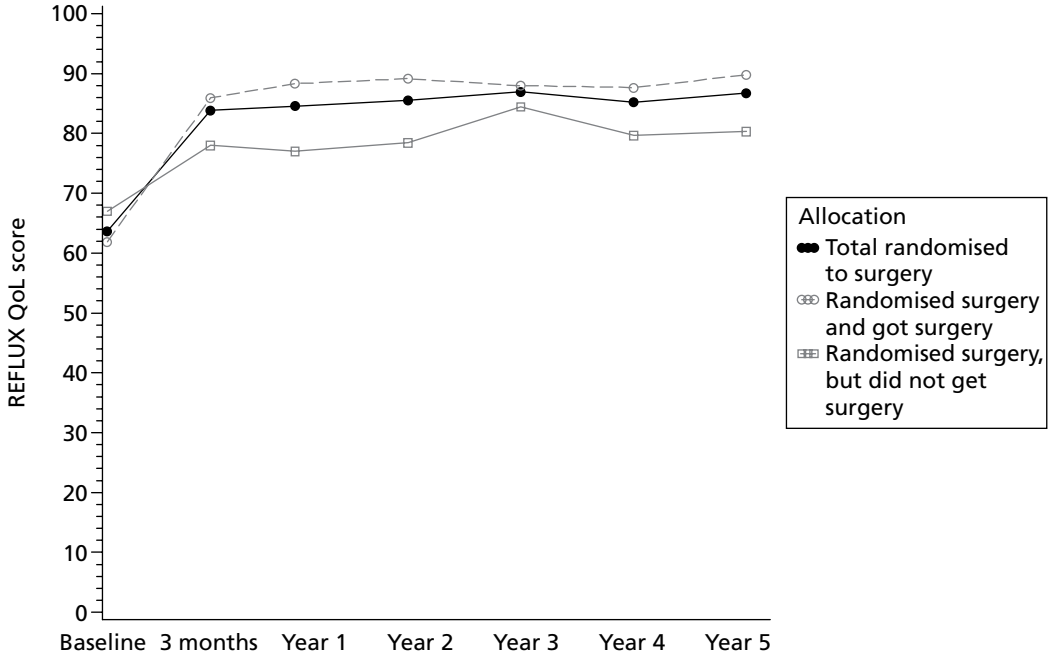


FIGURE 18 Mean REFLUX QoL scores for (a) all randomised to surgery, (b) those randomised to surgery who had fundoplication and (c) those randomised to surgery who did not have surgery.

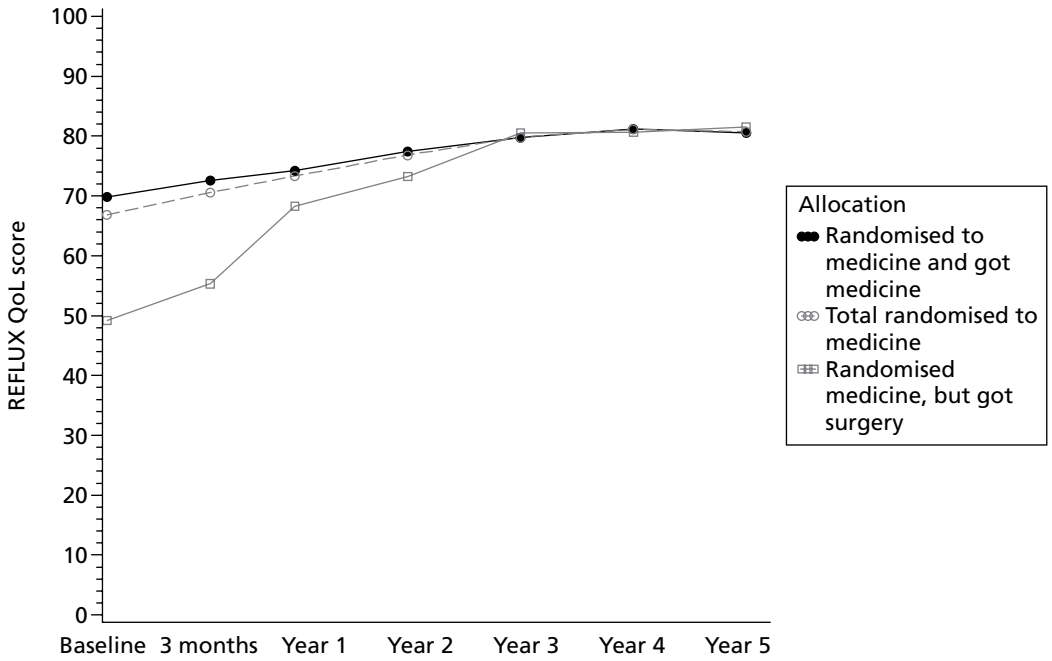


FIGURE 19 Mean REFLUX QoL scores for (a) all randomised to medical management, (b) those randomised to medical management who did not have surgery and (c) those randomised to medical management who had fundoplication.

provide a better estimate of the benefits of a policy of laparoscopic fundoplication as would apply in the health service.

The principal concern about laparoscopic fundoplication is possible risks associated with the surgery. We described intra- and postoperative surgical outcomes in our previous report.[1] Among the 329 patients in the randomised surgical and preference surgical groups who had fundoplication in the first year,

there were no major surgical complications. Two patients (0.6%; 95% CI 0.1% to 2.2%) required conversion to an open procedure; eight (2.4%; 95% CI 1.2% to 4.7%) had a visceral injury; and one (0.3%; 95% CI <0.1% to 1.7%) had a blood transfusion. Three were admitted to a high-dependency unit, but none to an intensive care unit. The 5-year follow-up provides information about longer-term risks. We are aware of seven deaths among trial participants; however, none has an apparent link to the trial. Twelve (3.3%) of the total of 364 participants who had a fundoplication had a late complication: four were oesophageal dilatations/stricture dilatations, three had repairs of incisional hernias and five were a heterogeneous group of other complications (see *Table 12*). Sixteen (4.4%) of those who had fundoplication required further surgery (see *Table 11*): five reconstruction of the same wrap, six conversion to another type of wrap, three repair of hiatus hernia only and two reversal of fundoplication. These, albeit uncommon, complications need to be taken into account when surgery is being considered.

Proton pump inhibitor use in the randomised medical group was consistently around 80%, although these participants were not always the same people at each follow-up. In our questionnaire, we chose to ask about anti-reflux drug use over the preceding 2 weeks as we thought that a recollection over a longer period would be unreliable. Nevertheless, taking of PPIs seems to be dynamic (patients stopping and restarting) and rates of use at any time over a longer period would likely have been higher. We did observe more visits to GPs in the medical groups for reflux-related reasons during the 5 years of follow-up but are not able to say whether this was due to routine reassessments or because symptom control was less stable or inadequately controlled in the medical group.

The pattern of PPIs used did change over the course of the study. At baseline, the commonest PPI was lansoprazole, but omeprazole superseded this over the course of the trial. Much of this change occurred in the first year and hence could be a consequence of the review of medical management that was part of the trial management for those randomised to medical management.

The larger number of overnight hospital admissions in the randomised medical management group was largely, but not totally, explained by the minority who went on to have surgery; as discussed in *Chapter 5* describing the economic evaluation, this was the principal driver of extra resource use by the medical group during the longer-term follow-up.

Despite the methodological challenges alluded to above, the study, through the data presented here, has successfully addressed the first of the objectives of this longer-term follow-up: to assess whether or not short-term clinical benefits, principally in terms of symptom control, are sustained – they are, albeit attenuated. In the next chapter we consider the REFLUX trial in the context of the three other randomised trials that have been conducted worldwide comparing laparoscopic fundoplication with medical management, and assess whether or not the results of the REFLUX trial are consistent with those of the other trials.

Chapter 4 Comparison of the REFLUX trial with other randomised trials of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease

Introduction

The REFLUX trial is one of four randomised trials that have compared laparoscopic surgery with medical management of GORD. Although the REFLUX trial has similarities to the other trials, its design is the most pragmatic[42] and this is reflected in significant differences in comparison with the other trials. The characteristics of the four trials are summarised in some detail in *Appendix 5*; key similarities and differences in characteristics between the REFLUX trial and the other trials will be highlighted here. This overview draws heavily on the relevant Cochrane review,[43] two of whose authors are authors of this report, but incorporates reports published since the Cochrane review, identified primarily through an updated search using a similar strategy to the one described in the Cochrane review.

The three comparable trials

The Anvari *et al.* trial[44–46] is a publicly funded single-centre trial conducted in Canada, led by upper gastrointestinal surgeons. It is the smallest of the four trials (104 randomised). The two intervention policies were standardised and the surgery was undertaken by only four surgeons (*Table 23*). Reflecting this, nearly all participants – unlike in the REFLUX trial – were managed in the way allocated. Like the REFLUX trial, its primary outcome was a GORD-related QoL instrument (the GERSS or Gastro-Esophageal Reflux Symptom Score), and HRQoL was measured with the same instruments as in REFLUX (SF-36 and EQ-5D). The first report described the trial up to 12 months after surgery,[44] and recent papers have reported 3-year results[45] and an economic evaluation.[46] At 3 years, participants in the medical group were offered surgery and a large proportion (42%) accepted; hence, although further follow-up is reported to be ongoing, it will be of limited usefulness in comparing laparoscopic surgery with medical management.

The LOng-Term Usage of esomeprazole versus Surgery for treatment of chronic GERD (LOTUS) trial[47–50] is the largest of the four trials (554 randomised). The study was funded by a pharmaceutical company, AstraZeneca, and the reports all include authors based in the company. The trial involved 39 centres in 11 European countries and was led by an upper gastrointestinal surgeon. The trial is described as 'not designed as a superiority or equivalence trial but, rather, was an exploratory study to estimate the efficacy of laparoscopic anti-reflux surgery and PPI treatment in PPI responders'. Unlike in the REFLUX trial, all participants had shown response to PPI treatment in a run-in phase, and both clinical management policies were strictly standardised (see *Table 23*).

The method by which the total fundoplication approach was standardised has been described in detail.[50] In the medically managed group, the only PPI used was esomeprazole, initially at the standard dose of 20 mg. Both the surgical and medically treated patients were followed up by the investigators at 6-monthly intervals and symptoms were assessed using the Gastrointestinal Symptoms Rating Scale (GSRS) questionnaire. In the medically treated group, esomeprazole could be increased to 40 mg once a day and then to 20 mg twice a day if symptom control was insufficient. Another key difference from the REFLUX trial was that the primary outcome measure was 'treatment failure'. A single definition of treatment failure could not be used for both trial groups; rather, this was specifically defined for each group (including

TABLE 23 Surgical procedure/experience in the four trials[a]

Trial	Surgeon experience	Crural repair	Gastric division	No. of surgeons participating
Anvari[44–46]	>50 procedures performed	Not reported	Short vessels divided	4
LOTUS[47–50]	>40 procedures performed and current workload ≥20 per annum	Protocol specified posterior repair	Protocol specified division	40 trained
Mahon[51–53]	'Experienced'	Yes, all patients	Short vessels divided	2
REFLUX[1–3]	>50 procedures performed	Surgeon discretion	Surgeon discretion	Not reported

a Adapted from Wileman *et al.*[43]

in the medical group need for escalation of medication and in the surgical group, need for regular medication). The concern is that the thresholds for these may not reflect similar levels of GORD. A GORD-specific QoL instrument (Quality of Life in Reflux and Dyspepsia or QOLRAD[54]) was among the secondary outcomes but was given relatively little emphasis in the reporting of the trial. No HRQoL instruments were used and there was no economic evaluation. Although the main analysis was said to be carried out on an ITT basis, it seems that the 40 people allocated surgery who did not receive it were excluded from analyses. Results were first reported after 3 years' follow-up[47] and recently 5-year data have been published.[48]

The Mahon *et al.* trial[51–53] was a two-centre UK trial led by and involving two upper gastrointestinal surgeons. It is not clear how the main trial was funded but supplementary funds were provided by Jansen Pharmaceuticals 'for physiological studies' and by Ethicon Endo-Surgery for the economic analysis.[52] In total, 217 people were randomised; the sequence was 'computerised' but the randomisation process and extent of concealment were not described. The two surgeons used a similar Nissen fundoplication method (see *Table 23*) and there was the option of four different PPI regimens depending on what PPI a participant had been taking prior to the trial. A range of outcome measures were reported and these included a gastrointestinal symptom score (GSRS) and a HRQoL measure [Psychological General Well-Being Index (PGWI)].[55] All those allocated to medical management were offered surgery after 1 year (and apparently this was made clear to potential participants before trial entry) and the majority [54/94 (57%)] then had surgery. The 1-year follow-up was thus essentially the end of this randomised trial, even though a further follow-up has been reported.[53]

Gastro-oesophageal reflux disease-related quality-of-life and symptom scores

Data available for each of the trials that describe GORD QoL or symptom scores at 1, 3 and 5 years' follow-up are summarised in *Tables 24–26*. Although it is not possible to combine data because different instruments (or subscales of instruments) were used in the trials, the results are consistent.

At 1 year there are eligible data from all four trials (see *Table 24*). In each case there are highly statistically significant differences all favouring the surgically managed groups. As mentioned above, the randomised element of the Mahon *et al.* trial[51–53] ended at 1 year but data at 3 years are available for the other three trials (see *Table 25*). Again, all favour the surgical group and this was statistically significant in both the LOTUS[47–50] and the REFLUX[1–3] trials.

Only the LOTUS and (now) the REFLUX trial have reported 5-year follow-up. GORD-related QoL scores significantly favour the surgical groups in both trials (see *Table 26*).

TABLE 24 Gastro-oesophageal reflux disease-related QoL and symptom scores at 1 year

Trial	Surgical		Medical		Mean difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
GERSS	52	8.3 (8.4)	52	13.6 (9.5)	–5.3 (–8.7 to –2.0)	0.002
***LOTUS*[47–50]**						
QOLRAD						
Vitality	203	6.84 (0.52)	220	6.42 (0.92)	0.42 (0.28 to 0.56)	<0.001
Food and drink	203	6.78 (0.60)	220	6.34 (0.98)	0.44 (0.28 to 0.60)	<0.001
Sleep	203	6.87 (0.49)	220	6.53 (0.76)	0.34 (0.22 to 0.46)	<0.001
Physical/social	203	6.93 (0.36)	220	6.72 (0.52)	0.21 (0.12 to 0.30)	<0.001
GSRS						
REFLUX dimension	248	1.18 (0.44)	266	1.66 (0.88)	–0.48 (–0.60 to –0.36)	<0.001
***Mahon*[51–53]**						
GSRS	80	37.0 (5.4)	86	35.0 (7.3)	2.00 (0.003 to 3.94)	0.003
***REFLUX*[1–3]**						
REFLUX QoL	178	84.6 (17.9)	179	73.4 (23.3)	14.0 (9.6 to 18.4)	<0.001

TABLE 25 Gastro-oesophageal reflux disease-related QoL and symptom scores at 3 years

Trial	Surgical		Medical		Mean difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
GERSS	49	6.21 (8.66)	44	9.05 (10.40)	–2.84 (–6.77 to 1.09)	0.166
***LOTUS*[47–50]**						
QOLRAD						
Vitality	181	6.90 (0.31)	189	6.53 (0.85)	0.37 (0.24 to 0.50)	<0.001
Food and drink	181	6.85 (0.40)	189	6.38 (0.91)	0.47 (0.33 to 0.61)	<0.001
Sleep	181	6.92 (0.33)	189	6.53 (0.82)	0.39 (0.26 to 0.52)	<0.001
Physical/social	181	6.94 (0.25)	189	6.74 (0.58)	0.20 (0.11 to 0.29)	<0.001
Mahon*[51–53] – *trial terminated at 1 year						
***REFLUX*[1–3]**						
REFLUX QoL	132	87.0 (15.0)	134	79.7 (20.1)	9.0 (4.9 to 13.1)	<0.001

Health-related quality of life

No general HRQoL measure has been reported for the LOTUS trial.[47–50] Data for the other three trials are shown in *Tables 27–29*. The SF-36 was used in the Anvari *et al.* trial[44–46] as it was in the REFLUX trial.[1–3] Unfortunately, it is reported only as the two summary component scores, physical (PCS) and mental (MCS), plus the 'general health' domain score. For comparability, in *Tables 27* and *28* the same score formats

TABLE 26 Gastro-oesophageal reflux disease-related QoL and symptom scores at 5 years

Trial	Surgical		Medical		Difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46] – no data available**						
***LOTUS*[47–50]**						
QOLRAD						
Vitality	160	6.86 (0.44)	179	6.49 (0.99)	0.37 (0.20 to 0.54)	<0.001
Food and drink	160	6.80 (0.51)	179	6.47 (0.80)	0.33 (0.18 to 0.48)	<0.001
Sleep	160	6.89 (0.47)	179	6.61 (0.72)	0.28 (0.15 to 0.41)	<0.001
Physical/social	160	6.94 (0.23)	179	6.75 (0.51)	0.19 (0.10 to 0.28)	<0.001
***Mahon*[51–53] – trial terminated at 1 year**						
***REFLUX*[1–3]**						
REFLUX QoL	127	86.7 (13.8)	119	80.7 (20.3)	6.42 (1.61 to 11.23)	0.009

are shown for the REFLUX trial but it should be borne in mind that the eight domain scores shown in *Chapter 3* for the REFLUX trial are more informative.

At 1 year, in both trials, the PCS and MCS favour the surgical group, although only the difference in the PCS in the REFLUX trial[1–3] is statistically significant. Both trials showed marked differences in the 'general health' domain score. There was also a statistically significant difference favouring surgery in the Mahon *et al.* trial[51–53] (based on the PGWI).

Although EQ-5D data were collected in the Anvari *et al.* trial,[44–46] they were not reported in a way that allows interpretation. At baseline, scores were markedly lower in the surgery group [mean 0.68 (SD 0.28) vs 0.76 (SD 0.21)] and the reason for this imbalance is not clear. At 1 year the equivalent results were 0.79 (SD 0.23) compared with 0.81 (SD 0.19), that is, still lower in the surgery group. As shown in *Table 27*, in the REFLUX trial,[1–3] the mean 1-year EQ-5D score was higher in the surgery group ($p = 0.07$).

At 3 years, the report of the Anvari *et al.* trial mentions collection of the SF-36 'every 3 months' but the only data reported are for the 'general health' domain score. This, as in the REFLUX trial, significantly favours the surgical group (see *Table 28*). There is no mention of collection of EQ-5D data in the 3-year follow-up of the Anvari *et al.* trial. At 5 years, the only data describing generic HRQoL are from the REFLUX trial (as the LOTUS trial has not included a measure) (see *Table 29*).

Individual symptoms of gastro-oesophageal reflux disease or its management

Data describing individual symptoms are available for all trials, although only dysphagia was reported in the Mahon *et al.* trial.[51–53]

Heartburn

As would be expected from the overall GORD-related QoL and symptom scores, all three trials providing data reported less heartburn in their surgical groups. At 1 year in the Anvari *et al.* trial,[44–46] the GERSS heartburn subscore is lower in the surgical group ($p < 0.001$); in the LOTUS trial[47–50] there is clearly less heartburn in the surgical group but data are presented only graphically; and in the REFLUX trial[1–3] heartburn rates in the surgical group are around half those in the medical group. At 3 years, Anvari

TABLE 27 Health-related quality of life at 1 year

Trial	Surgery		Medical		Difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
SF-36						
PCS	52	46.4 (10.9)	52	43.9 (10.3)	3.15 (–0.94 to 7.23)	0.13
MCS	52	52.7 (10.9)	52	51.5 (9.1)	0.98 (–2.8 to 4.76)	0.61
General health domain score	52	75.4 (23.2)	52	66.4 (23.6)	12.3 (3.7 to 20.8)	0.005
***LOTUS*[47–50] – not reported**						
***Mahon*[51–53]**						
PGWB	79	106.2 (16.3)	86	100.4 (18.9)	5.8 (0.43 to 11.17), adjusted 7.1 (2.5 to 11.7)	
***REFLUX*[1–3]**						
SF-36						
PCS	150	48.0 (10.2)	161	45.1 (9.7)	3.51 (1.77 to 5.25)	<0.001
MCS	150	46.6 (12.8)	161	45.1 (13.1)	1.63 (–0.79 to 3.85)	0.195
General health domain score	178	45.2 (11.1)	179	40.7 (11.2)	4.8 (2.7 to 6.8)	<0.001
EQ-5D	178	0.75 (0.25)	179	0.71 (0.27)	0.047 (–0.004 to 0.097)	0.07

TABLE 28 Health-related quality of life at 3 years

Trial	Surgery		Medical		Difference (95% CI)	*p*-value
	n	Mean (SD)	*n*	Mean (SD)		
***Anvari*[44–46]**						
SF-36						
PCS – not reported						
MCS – not reported						
General health domain score	49	78.50 (19.76)	44	71.41 (21.73)	12.19[a] (2.65 to 21.72)	0.0124
***LOTUS*[47–50] – not reported**						
***Mahon*[51–53] – trial terminated at 1 year**						
***REFLUX*[1–3]**						
SF-36						
PCS	128	47.2 (9.9)	127	46.6 (10.0)	1.43 (–0.45 to 3.32)	0.136
MCS	128	48.9 (10.6)	127	45.6 (12.6)	4.05 (1.57 to 6.52)	0.001
General health score	132	45.3 (10.0)	134	42.4 (11.8)	3.69 (1.50 to 5.87)	0.001
EQ-5D	132	0.803 (0.231)	134	0.747 (0.262)	0.070 (0.0015 to 0.126)	0.013

a Presumably adjusted.

TABLE 29 Health-related quality of life at 5 years

	Surgery		Medical			
Trial	*n*	Mean (SD)	*n*	Mean (SD)	Difference (95% CI)	*p*-value
Anvari[44–46] – no data available						
LOTUS[47–50] – not reported						
Mahon[51–53] – trial terminated at 1 year						
REFLUX[1–3]						
SF-36						
PCS	113	46.1 (9.9)	109	46.1 (10.5)	1.47 (–0.84 to 3.79)	0.211
MCS	113	47.8 (11.7)	109	47.9 (11.7)	1.27 (–1.36 to 3.90)	0.343
General health domain score	117	44.1 (10.3)	111	43.2 (11.5)	2.76 (0.21 to 5.31)	0.034
EQ-5D	127	0.774 (0.259)	119	0.761 (0.282)	0.047 (–0.013 to 0.108)	0.126

et al.[44–46] report significantly more heartburn-free days in the surgical group ($p = 0.008$); in the LOTUS trial,[47–50] less heartburn in the surgical group is shown graphically and the *p*-value is reported as <0.001; and in the REFLUX trial[1–3] 51% of the randomised surgical group compared with 75% of the randomised medical management group report any heartburn (see *Table 15*). At 5 years, data are available only from the LOTUS and REFLUX trials. In LOTUS,[47–50] 8% in the surgery group compared with 16% in the medical group are reported to have heartburn, 'although there was no significant difference in the severity of heartburn ($p = 0.14$)'. In the REFLUX trial,[1–3] 41% in the surgery group compared with 74% in the medical group reported any heartburn (see *Table 15*).

Regurgitation

Again, as would be expected from the overall GORD-related QoL and symptom scores, all three trials providing data reported less regurgitation in the surgical groups. At 1 year in the Anvari *et al.*[44–46] trial, the GERSS regurgitation subscore is significantly lower in the surgical group ($p = 0.002$); in the LOTUS trial,[47–50] graphical presentation clearly indicates less regurgitation in the surgical group, although no figures are reported; and in the REFLUX trial,[1–3] regurgitation rates in the surgical group are half those in the medical group. At 3 years, information is available only for the LOTUS and REFLUX trials and both report lower rates in the surgical groups. At 5 years in the LOTUS trial, 2% in the surgical group compared with 13% in the medical group ($p<0.001$) have regurgitation, and in the REFLUX trial 25% in the surgical group compared with 37% in the medical group report any regurgitation.

Dysphagia

As mentioned in *Chapter 1*, dysphagia following both open fundoplication and laparoscopic fundoplication has been reported. At 1 year, Anvari *et al.*[44–46] report a higher GERSS dysphagia subscore in the surgical group but this was not statistically significant ($p = 0.8$); in the LOTUS trial[47–50] there were more reports of dysphagia in the surgical group but data were presented only graphically; in the Mahon *et al.* trial,[51–53] dysphagia persisting beyond 3 months was reported in 5 out of 104 (4.8%) having surgery; and in the REFLUX trial,[1–3] rates of 'difficulty swallowing' were the same in the two randomised groups. At 3 and 5 years, information is available only from the LOTUS and REFLUX trials. In the LOTUS trial there is more dysphagia in the surgical group ($p<0.001$) at both time points: at 5 years 11% in the surgical group report dysphagia compared with 5% in the medical group. In the REFLUX trial, one further participant had undergone oesophageal dilatation (see *Table 12*), but the numbers reporting difficulty swallowing were the same in the two randomised groups (see *Table 15*, e.g. any difficulty swallowing 24.2% vs 23.9%).

Flatulence

Flatulence has also been reported as more common after both open and laparoscopic fundoplication. Information is available only from the LOTUS[47–50] and REFLUX[1–3] trials. In the LOTUS trial, flatulence was more common in the surgery group than in the medical management group at 1, 3 and 5 years. At 5 years, the rates are 57% in the surgical group and 40% in the medical group ($p < 0.001$). In the REFLUX trial, rates of 'wind from the lower bowel' are not statistically significantly different between the groups [more than three times per week: 65.0% in the randomised surgical group vs 59.4% in the randomised medical group at 1 year; 57.6% vs 58.5% at 3 years; and 65.3% vs 60.0% at 5 years (see *Table 15* for more detail)].

Other symptoms

In the LOTUS trial,[47–50] 'bloating' was reported more commonly in the surgical group (40% vs 28% at 5 years). In contrast, 'bloating/trapped wind' was reported less commonly in the surgical group in the REFLUX trial[1–3] (at 1 year: 72.1% vs 82.4%). A particular concern following fundoplication is an inability to vomit despite wanting to. In the REFLUX trial we attempted to address this through a question on 'frequency of wanting to be sick but being physically unable to' and found no difference between the groups (see *Table 15*).

Surgical complications

Like all procedures involving surgery under general anaesthesia, laparoscopic fundoplication carries risks. *Table 30* summarises intra and early postoperative complications reported in the four trials.

Conversion to an open procedure

The decision to convert from a laparoscopic to an open approach is usually indicative of difficulties experienced during the procedure. Rates varied from 0% in the Anvari *et al.* trial[44–46] to 2.4% in the LOTUS trial[47–50] (see *Table 30*).

Intraoperative complications

In the Mahon *et al.*[51–53] and REFLUX[1–3] trials combined, the 10 intraoperative complications reported (overall rate 2.3%) were injuries to the spleen ($n = 3$), liver ($n = 3$), pleura ($n = 3$) and oesophagus ($n = 1$). In the LOTUS trial[47–50] it was unclear whether intraoperative complications occurred or whether they were incorporated within all postoperative complications; however, the report noted that 29 participants encountered a variety of operative difficulties that were described as 'trivial'.

TABLE 30 Intra- and early postoperative events in the four trials[a]

Trial	*n* having operation	Conversion, *n* (%)	Intraoperative complications, *n* (%)	Postoperative complications, *n* (%)
Anvari[44–46]	51	0 (0.0)	0 (0.0)	7 (13.7)
LOTUS[47–50]	248	6 (2.4)	Unclear	7 (2.8)
Mahon[51–53]	109	1 (0.9)	4 (3.7)	6 (5.5)
REFLUX[1–3]				
Randomised	111	2 (1.8)	2 (1.8)	1 (0.9)
Preference	218	0 (0.0)	4 (1.8)	2 (0.9)

a Adapted from Wileman *et al.*[43]

Early postoperative complications

In the Anvari *et al.* trial,[44–46] seven (14%) participants had postprandial bloating, two of whom were treated with a single dilatation of the wrap. No details are given of the postoperative complications in the LOTUS trial. In the Mahon *et al.* trial[51–53] there were three wrap migrations, two respiratory tract infections and one case of a sutured nasogastric tube. In the REFLUX trial,[1–3] one participant in the randomised group and two in the preference group were admitted to a high-dependency unit immediately after the surgical procedure.

Reoperations

By the time of the 3-year follow-up in the Anvari *et al.* trial,[44–46] 4 of 51 (7.8%) participants had undergone a second fundoplication operation. Four (3.7%) in the Mahon *et al.* trial[51–53] required reoperation within 3 months of their first fundoplication, one of whom had a gastric resection because of necrosis. It is not clear if anyone in the LOTUS trial[47–50] had a reoperation. As shown in *Table 11*, in the REFLUX trial,[1–3] 5 of the 112 (4.5%) randomised to surgery who actually had a fundoplication had a second reflux-related operation, and this applied to 16 (4.4%) of the total 364 participants in the study who had a laparoscopic fundoplication.

Other late postoperative complications

Dilatation of the wrap was reported for two (3.9%) people in the Anvari *et al.* trial[44–46] and four (3.7%) in the Mahon *et al.* trial.[51–53] It is not stated whether or not dilatation occurred in the LOTUS trial.[47–50] In the REFLUX trial,[1–3] two (1.8%) participants in the randomised surgical group (plus two in the preference surgical group – giving an overall rate of 1.1%) had stricture dilatation or food disimpaction (see *Table 12*). There were three cases (0.8%) of repair of incisional hernia in the REFLUX trial – all in the preference group – but this complication was not mentioned in the other trials' reports. There were no deaths in any of the trials associated with surgical or medical management.

Surgery-related mortality

No perioperative deaths were reported among the 771 people in the four trials who had fundoplication surgery.

Discussion

Of the four trials, the REFLUX trial is the most pragmatic in design. It involved a large proportion of UK centres where laparoscopic anti-reflux surgery is undertaken and the surgery was undertaken by NHS upper gastrointestinal surgeons within these centres, all of whom had experience of carrying out the procedure. The exact method of fundoplication was left to the discretion of the surgeon, so he or she was comfortable with the approach. After surgery and, in the medically treated patients, after optimisation of their PPI medication, care of the participants was the responsibility of GPs. The principal measure of outcome was a patient-reported disease-specific QoL measure. Unlike the other trials, the REFLUX trial was coordinated from an accredited trials unit, local recruitment was led by gastroenterologist/gastrointestinal surgeon partnerships rather than by gastrointestinal surgeons alone, and the trial was publicly funded through the HTA programme rather than by industry.

In respect of potential benefits of surgery, the four trials appear to be consistent. All show significantly better relief of GORD symptoms for as long as the length of their current follow-up. (Surprisingly, the LOTUS trial report[48] does not draw attention to this but, judged on data describing the QOLRAD reported in an e-table, there are significant differences between the groups in all dimensions of this instrument, favouring surgery.) Data available describing the principal symptoms of GORD (heartburn and regurgitation) show large differences, again favouring surgery. Only limited data are available from generic QoL measures, and much of this is from the REFLUX trial; although differences are less marked than for the GORD-related QoL instruments, they are consistent with benefit from surgery.

The four trials are broadly consistent in respect of intraoperative and early postoperative complications: a small number of operations are converted to an open procedure, a small number of laparoscopic procedures have associated visceral injuries, a small number of people have problems postoperatively and a small number require dilatation of the wrap. The REFLUX trial suggests that 4.5% have reoperations and the other trials are broadly consistent with this. None of the trials had a reported perioperative death. Data from the Finnish Registry[56] suggest a mortality of 0.1%, but this is based on a single case among 1162 people who had laparoscopic fundoplication; furthermore, the registry included all cases of fundoplication and hence went beyond the sorts of patients recruited to the REFLUX trial.

The other trials, particularly the LOTUS trial, show higher rates of dysphagia and flatulence following laparoscopic fundoplication than in the medically managed group. As mentioned above, a small number of participants in the REFLUX trial did have a dilatation procedure, presumably because of difficulty swallowing, but this was not reflected in responses to the REFLUX questionnaire, suggesting that there were only a few isolated cases of dysphagia following surgery in this trial. Similarly, there were no significant differences in flatulence in the REFLUX trial.

Hence, taking all four trials together, it is now possible to give a clear picture of most of the potential benefits and risks of laparoscopic fundoplication, at least up to 5 years. There are, however, differing resource implications of surgery and medical management. In the next chapter we explore whether or not the benefits of surgery in patients with established GORD requiring long-term PPI therapy for reasonable control and suitable for either clinical policy (average age around 45 years) are sufficient to outweigh any differences in costs.

Chapter 5 Economic analysis

The economic evaluation aimed to determine the cost-effectiveness of laparoscopic fundoplication compared with continued medical management in patients with GORD symptoms that are reasonably controlled by medication and who are judged suitable for both surgical and medical management. The analysis entailed three components:

1. systematic review of existing cost-effectiveness evidence
2. within-trial (5-year) economic analysis
3. validation of within-trial analysis and exploration of the need for a longer-term model.

Systematic review of existing cost-effectiveness evidence

The aim of this systematic review is to identify any existing cost-effectiveness studies that compare laparoscopic fundoplication with medical management for GORD. A previous HTA report included a review of the evidence available from 1995 to December 2005 and identified three relevant studies (described below).[1] The updated search focuses on the period from December 2005 to April 2011. The methods used to identify studies and the results of the systematic search are discussed in the sections below.

Methods

The following data sets were searched to identify published evidence: MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (1948 to present), EMBASE (1996 to week 15, 2011), Cochrane Database of Systematic Reviews (CDSR) and the NHS Centre for Reviews and Dissemination databases [Database of Abstracts of Reviews of Effects (DARE), NHS Economic Evaluation Database (NHS EED), HTA]. The search strategy incorporated broad reflux-related search terms as used in a recent Cochrane Review.[57] The search also focused on identifying health-related and GORD-specific QoL evidence.

Studies were considered relevant for inclusion in the review if they were published in English and were full health economic evaluations (cost-effectiveness, cost-utility or cost–benefit analysis) comparing costs and outcomes associated with laparoscopic fundoplication and medical management. For the purpose of this study laparoscopic fundoplication includes both complete and partial wrap procedures. Publications outside the above criteria were excluded from this review. Details of the updated search strategy are presented in *Appendix 6*.

Results

A total of 3662 references were identified from the searches (MEDLINE: 1640, EMBASE: 1825, CDSR: 44, DARE: 56, NHS EED: 85, HTA: 12). Titles and/or abstracts were reviewed and studies that satisfied all inclusion criteria were included in the review. Papers describing five additional studies were obtained for inclusion. These were published between 2007 and 2011 and were related to the UK and Canadian settings. Of the total of eight studies, five are linked to three of the randomised trials described in *Chapter 4*: Anvari *et al.*,[44–46] Mahon *et al.*[51–53] and the REFLUX trial,[1] the long-term follow-up of which is the topic of this report. There is no economic evaluation in the LOTUS trial.[48] Three of the studies were based on the REFLUX trial. These were published as part of the earlier HTA report[1] and in two journal articles.[3,5] Summaries of the two within-trial economic evaluations are presented in *Appendix 7*. Below is a brief description of the eight reports – the five linked to the three randomised trials are considered first, followed by the three studies based on observational data.

Economic analyses based on clinical trials

Economic evaluation based on the Anvari et al. *trial*[46]

This was an economic evaluation conducted alongside the Anvari *et al.* trial described in *Chapter 4*. Laparoscopic fundoplication was compared with PPI for patients with chronic GORD. The follow-up period was 3 years and the analysis was conducted from a societal perspective. Cost-effectiveness was reported in terms of cost per QALY gained.

Three generic preference-based questionnaires were administered during the trial: Health Utilities Index Mark 3 (HUI3), EQ-5D and Short Form questionnaire-6 dimensions (SF-6D). Although these instruments have been valued by large general public samples, they differ in the attributes used for their descriptive system and the method of valuation applied. The EQ-5D has been valued using time trade-off whereas the SF-6D and HUI3 use the standard gamble. Utility scores showed an improvement in patients' HRQoL in both groups across the three utility instruments; however, the degree of improvement varied according to the utility instrument used. The base-case analysis (using the HUI3 instrument), after adjustment for baseline differences, indicated that, over the 3 years, laparoscopic fundoplication patients experienced a 0.109 gain in QALYs compared with PPI patients. The ICER for laparoscopic fundoplication patients was around C$29,400 (£19,000) per QALY gained. An increased ICER of C$76,300 (£49,300) was obtained using the EQ-5D as the HRQoL measure.

Economic evaluation based on the Mahon et al. *trial*[52]

This study looked at the cost-effectiveness of laparoscopic fundoplication compared with maintenance PPI medication for severe GORD based on the Mahon *et al.* randomised trial described in *Chapter 4*. Results based on the 12-month follow-up were extrapolated using other published data sets. Costs and outcomes for up to 12 months were obtained from a sample of patients in the trial (the first 100) and resource use was quantified using data from hospital records and GPs' notes. The incremental cost of laparoscopic fundoplication compared with PPI therapy per additional patient returned to a physiologically normal acid score (<13.9) at 3 months was £5515 (95% CI £3655 to £13,400) and the incremental cost per point improvement in combined gastrointestinal and psychological well-being score at 12 months was £293 (90% CI £149 to £5250). The authors concluded that laparoscopic surgery would break even compared with medical management after 8 years and would be cost saving thereafter.

Economic evaluation based on the REFLUX trial[1,3,5]

Bojke *et al.*[5] present a preliminary cost-effectiveness analysis conducted before the availability of the 1-year REFLUX trial results. The analysis compared the cost-effectiveness of surgery (laparoscopic fundoplication) with long-term medical management (PPIs) for GORD disease in an average 45-year-old man. A lifetime (30 years) Markov model that adopted the perspective of the NHS was developed. Effectiveness data were obtained from a fixed-effect meta-analysis that synthesised data from multiple sources. QALYs were estimated using utility scores (measured by the EQ-5D instrument) derived from a subset of UK patients included in the REFLUX trial. Over a lifetime, expected costs associated with surgery (£5014) were higher than expected costs associated with PPI (£4890). Expected QALYs associated with surgery (13.04) were greater than QALYs associated with PPIs (12.36). The incremental cost per QALY gained (ICER) for surgery compared with medical care was £180. The estimated probability that surgery was cost-effective at the threshold of £30,000 per QALY was 0.639. The authors highlighted important areas for further research, such as the HRQoL of patients on PPIs or post surgery.

The within-trial cost-effectiveness analysis, comparing laparoscopic fundoplication with medical management 1 year post surgery, was described in full in the 2008 report of the REFLUX trial.[1] The analysis was conducted on an ITT basis from a NHS perspective. HRQoL was assessed at baseline and at 3 and 12 months' follow-up using the EQ-5D. Cost-effectiveness was reported in terms of the difference in mean QALYs between the treatment groups. This difference was estimated using ordinary least squares (OLS) regression, adjusting for baseline differences in EQ-5D between individuals. The estimated difference in mean costs between the groups was £1280 (95% CI £1054 to £1468). The HRQoL of patients randomised

to surgery tended to improve on average by 0.066 more QALYs (95% CI 0.023 to 0.107) than in the medical management group. The estimated mean ICER was around £19,000. At a threshold of £30,000 per QALY, the probability of surgery being cost-effective was 0.86.

Epstein *et al.*[3] developed a Markov model using 12-month data from the REFLUX trial and other sources in order to extrapolate the cost-effectiveness of laparoscopic fundoplication compared with medical management over the longer term (lifetime). Cost-effectiveness was reported in terms of the cost per QALY gained from surgery. The analysis was conducted from a NHS perspective. Under base-case assumptions, surgery had an additional mean cost of £847 and additional mean QALYs of 0.37 over the lifetime of the patients. The incremental cost per additional QALY gained was around £3000. At a threshold of £20,000 per QALY, the probability that surgery was cost-effective was around 0.74.

Economic analyses based on observational data

Economic evaluation based on Romagnuolo et al.[58]

This study is based on observational data and compares the cost-effectiveness of maintenance regimens of omeprazole and laparoscopic fundoplication within the Canadian medical system. The effectiveness, HRQoL and resource-use data were derived from studies published between 1985 and 2000. Outcomes were expressed as QALYs and costs were estimated from the perspective of a provincial health ministry. A two-stage Markov model (healing and maintenance phases) was used to estimate costs and utilities using a time horizon of 5 years. Laparoscopic fundoplication was the most cost-effective option at 3.3 years of follow-up and was cost saving at 5 years. These results were sensitive to the price of omeprazole. QALYs did not differ significantly between treatment groups.

Economic evaluation based on Arguedas et al.[59]

This study, also based on observational data, compared the cost-effectiveness of laparoscopic fundoplication and medical management in patients with severe reflux oesophagitis. Outcomes were quantified using QALYs with model inputs derived from the published literature. A Markov simulation model was used to extend a previous analysis to a 10-year time horizon. Procedure and hospitalisation costs were estimated using Medicare reimbursement rates from the authors' institution. Medical therapy was associated with a total cost of $8798 and 4.59 QALYs, whereas the surgery was more expensive ($10,475) and less effective (4.55 QALYs). The authors concluded that medical therapy dominated surgery.

Economic evaluation based on Comay et al.[60]

This is a cost-effectiveness analysis, based on observational data, principally concerned with assessing an endoscopic therapy (Stretta procedure) compared with PPIs and laparoscopic fundoplication in the management of GORD. The Strettra procedure is out of the scope of our analysis; however, the data on costs and QALYs provided by the authors allow us to better understand QoL related to these technologies and make comparisons with other authors' estimates. The authors constructed a Markov model that tracked patients over a period of 5 years. Analysis was undertaken from the Canadian Ministry of Health perspective. A literature review for published studies before 2004 was carried out to derive effectiveness and utility data. Symptom-free months and QALYs were used to measure benefit. PPI was the dominant strategy, producing more symptom-free months at lower costs than the other strategies. Laparoscopic fundoplication was associated with higher costs and generated more QALYs. The discounted mean QALYs over 5 years were 4.6487 for laparoscopic fundoplication and 4.6357 for PPI. The ICER for laparoscopic fundoplication compared with PPI was C$384,692 (£240,470). This is unlikely to be considered cost-effective.

Conclusions

The different outcomes used make it difficult to compare the results of the various studies analysed here. For those studies quantifying the benefits associated with the two treatments using QALYs, the results differ depending on the type of analysis conducted. Although the trial-based results suggest that there

is good short- and medium-term evidence indicating that surgery may well represent a cost-effective alternative intervention, the model-based studies are not so optimistic.

The ICER for surgery ranged from £180 to £49,000 per QALY gained. However, the limitations of the studies included in this review suggest that we should be cautious when interpreting these results. The decision model developed as part of the REFLUX trial extrapolated from data at 12 months and was based on the assumption that the treatment effect of surgery (in terms of impact on HRQoL) remains constant over the lifetime of patients. However, as would be expected, the results of the sensitivity analysis suggested that surgery was less cost-effective when the beneficial effect of surgery was limited to 5 years (increasing the ICER to £11,300) and when HRQoL was worse in those for whom surgery failed (increasing the ICER to £11,310 when considering very high rates of surgical failure).

The value of conducting additional research to reduce any uncertainty in the REFLUX model was demonstrated. The expected value of perfect information (EVPI) is the maximum amount that a decision-maker should be willing to pay to eliminate all uncertainty that arises because of imprecision in the parameters of the model. The value of information analysis suggested that further research could be worthwhile. At a threshold of £30,000, the per-patient EVPI was £15,106.

Within-trial economic evaluation

Follow-up data from the REFLUX trial up to 5 years after surgery are now available. These economic data represent the longest follow-up of randomised patients currently available. These data can help to inform the question regarding the sustainability of initial improvement in HRQoL following surgery. This section describes the updating of the cost-effectiveness analysis using these data to reduce the level of uncertainty about the cost-effectiveness of surgery and thus its role in the NHS.

Overview

Differences in mean costs and QALYs at 5 years (based on data collected within the REFLUX trial) were used to derive an estimate of the cost-effectiveness of laparoscopic surgery (laparoscopic fundoplication) and continued medical management. The extent of missing data throughout the trial follow-up is significant; therefore, the base case consists of the multiple imputed data set following ITT analysis. A separate scenario – complete-case analysis, in which patients with any missing data are excluded – was employed for ITT and PP for 1-year analyses. Costs and QALYs were evaluated on the basis of costs falling on the NHS and Personal Social Services expressed in UK pounds sterling at a 2010 price base. All analysis and modelling were undertaken in Stata/SE 11.1 (StataCorp LP, College Station, TX, USA).

Methods

Patient population

As described in earlier chapters, the patient population in the REFLUX trial was patients with GORD whose symptoms required medication for reasonable control and for whom either surgery or continued medical management appeared to be an acceptable treatment option. A policy of offering relatively early laparoscopic fundoplication was compared with the alternative policy of continued medical management. The analysis used data only from the randomised trial component of the REFLUX trial (i.e. not from the preference groups). As described in *Chapter 3*, 357 patients were randomised to either surgical treatment ($n = 178$) or medical management ($n = 179$) and patients were followed for up to 5 years.

Health-care resource use

Health-care resource-use data were collected prospectively as part of the clinical report forms and patient questionnaires at 3 and 12 months and 2, 3, 4 and 5 years. Patient questionnaires at 3 and 12 months collected information for the previous 3 and 9 months respectively. In addition, a questionnaire at 12 months recorded resource use for the whole of the first year (see following section on costs). Patient

questionnaires from the second year onwards collected information for the previous 12 months on hospital admissions (day and overnight admissions) and GP visits, and data on medication for the previous 2 weeks. Clinical report forms collected data on surgery and perioperative complications of surgery.

Costs

The cost for each individual patient in the trial was calculated by multiplying his or her use of health-care resources by the associated unit costs (*Table 31*). Discount was applied from year 2. Unit costs were all sourced from published data (see *Table 31*). Total costs include the costs of surgery, GP visits, hospital admissions and medication. Incremental costs (laparoscopic fundoplication vs medical management) for each year and per category of resource use, according to ITT allocation, were calculated using OLS regression.

The questionnaires asked for details of anti-reflux medication taken in the previous 2 weeks: name, dose and number of tablets/capsules. The cost of anti-reflux medication during these 2 weeks was calculated by multiplying the prices published in the Drug Tariff for December 2010[61] for each medicine by the number of tablets taken. Yearly medication costs are calculated using the area under the curve method,[62] which assumes linear interpolation between follow-up points. The costs of reflux-related inpatient, outpatient and day-case visits were derived from the *NHS Reference Costs 2009–10*,[63] in which the relevant codes were weighted by activity level.

For the base-case analysis, total costs included the costs of surgery, complications due to surgery, reoperations, reflux-related prescribed medication, reflux-related visits to and from the GP and reflux-related hospital inpatient, outpatient and day visits. For the sensitivity analysis, all GP visits and all hospital admissions are included in the calculation of total costs (see *Incremental analysis* for more details on sensitivity analysis). Costs of hospital admissions and GP visits were obtained by multiplying the relevant unit costs by the numbers of admissions and visits reported by the patients respectively. Patients themselves classified how many visits and admissions were reflux related in relation to the total number of visits. There is a possibility that patients may not have fully understood the clinical consequences of GORD; hence, they may misclassify the reason for a consultation. If such misclassification is different across treatment groups, estimates of incremental costs may be biased.

For the first year of the trial, data on resource use were collected at 3 months and 12 months, and for the whole year using an additional questionnaire. To make the most efficient use of the data available for the first year of the trial, resource use at 1 year was estimated as the greater of the area under the curve between the first and second questionnaire and the 12-month health-care survey. This is in line with the procedure employed for the earlier publication evaluating the REFLUX trial.[1]

The cost of surgery included the costs of (1) presurgical procedures (endoscopy, pH monitoring and manometry), (2) the surgery team, (3) operative complications, (4) hospital stay, (5) capital costs and overheads and (6) consumables. The cost of reoperations was assumed to be equivalent to the mean cost of the first surgery. The cost of reflux-related visits to and from the GP was assumed to be equivalent to the average cost of visits to and from the GP.[64]

Quality-adjusted life-years

Health outcomes were expressed in terms of QALYs. HRQoL was assessed in the REFLUX trial at baseline and 3 months and then yearly until 5 years using the EQ-5D.[65,66] The EQ-5D is a standardised and validated generic instrument for the measurement of HRQoL. It has five dimensions: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort, and anxiety and depression. Each dimension has three possible responses (no problems, moderate problems or severe problems), creating 245 mutually exclusive health states. Each of these health states has been valued in a large UK population study using the time trade-off method, in which 1 corresponds to perfect health (thus the maximum value possible) and 0 corresponds to death.[65,66]

TABLE 31 Unit costs employed to calculate the costs of reflux-related health-care use

Health-care activity	Resource	Cost (£)	Source
Laparoscopic fundoplication surgery	Endoscopy	218.52	Grant *et al.*[1] (inflated to 2009–10 prices using Curtis[64])
	pH test	81.85	
	Manometry	76.94	
	Operation cost per minute	6.36	
	Capital cost per surgery	11.71	
	Consumables	1080.96	
	High-dependency unit per night	797.86	
	General ward per night[a]	282.78	*NHS Reference Costs 2009–10* (excess bed stay)[63]
Hospital admissions[b]	Overnight admission due to surgery	2108.22	Mean surgery cost
	Overnight admission due to complications[a]	1534.76	*NHS Reference Costs 2009–10* (elective inpatient)[63]
	Day case	559.00	*NHS Reference Costs 2009–10* (day case)[63]
	Outpatient	221.98	*NHS Reference Costs 2009–10* (outpatient)[63]
GP use	Visit from GP	120.00	Curtis[64]
	Visit to GP	36.00	

Health-care activity	Resource	Cost (£)	Source
Medication			
PPI	Omeprazole 10 mg, 28 capsules	1.92	Drug Tariff December 2010[61]
	Omeprazole 20 mg, 28 capsules	1.81	
	Omeprazole 40 mg, 7 capsules	1.95	
	Lansoprazole 15 mg, 28 capsules	1.44	
	Lansoprazole 30 mg, 28 capsules	2.23	
	Pantoprazole 20 mg, 28 tablets	1.79	
	Pantoprazole 40 mg, 28 tablets	2.82	
	Rabeprazole 10 mg, 28 tablets	11.56	
	Rabeprazole 20 mg, 28 tablets	19.55	
	Esomeprazole 20 mg, 28 tablets	18.50	
	Esomeprazole 40 mg, 28 tablets	25.19	
H_2RA	Ranitidine 150 mg, 60 tablets	1.97	
	Ranitidine 300 mg, 30 tablets	2.17	
	Famotidine 20 mg, 28 tablets	4.40	
	Famotidine 40 mg, 28 tablets	5.55	
	Nizatidine 150 mg, 30 capsules	12.04	
	Nizatidine 300 mg, 30 capsules	15.34	
	Cimetidine 400 mg, 60 tablets	7.61	
	Cimetidine 800 mg, 30 tablets	21.63	
Prokinetic	Domperidone 10-mg tablets[c]	1.53	
	Metoclopramide 10 mg, 28 tablets	1.01	

a Average of the relevant cost for diagnoses codes FZ24B, FZ24C, FZ24D, FZ25A, FZ25B, FZ27A, FZ27B, FZ27C, FZ27D, FZ28A, FZ28B, FZ28C, FZ29Z, FZ30Z, FZ31D, FZ31E and FZ31F, weighted by activity levels.

b Costing hospital admissions included the cost of the individual procedures (endoscopy, pH test, manometry, barium meal) using the unit cost data used for costing surgery.

c Average cost of 30-tablet and 100-tablet pack.

DOI: 10.3310/hta17220

QALYs for each patient were calculated as the area under the curve following the trapezium rule,[67] which assumes linear interpolation between follow-up points. Incremental mean QALYs between treatment groups were estimated with and without adjustment for baseline utility, using OLS regression.

Discounting

Costs and outcomes from year 2 were discounted using a 3.5% annual discount rate, in line with current guidelines.[65,68]

Missing data and multiple imputation

Given the extent of missing data, the multiple imputed data set is presented as the base case. This was created using all available data and multiple imputation with chained equations.[69] Mean imputation was used to predict missing data at baseline,[70] as randomisation should ensure equal distribution of potentially confounding variables. Complete-case analysis refers to only those patients who returned all questionnaires and completed all EQ-5D profiles.

Missing or inconsistent answers to questions on resource use were dealt with as follows. For medication use, patients were asked at each follow-up questionnaire whether or not they were using prescribed medication for reflux and, if so, to indicate the name, strength and the number of tablets taken in the past 2 weeks. It was evident from preliminary analyses that the answers to the first question were not necessarily consistent with the answers to the second question. Therefore, the following rule was applied for the costing of drugs: (1) if the patient provided the name, strength and number of tablets taken, he/she was assumed to be taking medication; (2) if the patient did not specify either a drug or the number of tablets taken, he/she was considered not to be taking medication; (3) if the patient specified a particular drug but no dosage, the missing data were imputed as the median of all other patients on that medication. Similarly, missing answers to the questions regarding GP visits and hospital admissions were assumed to indicate that no visits or admissions occurred. Because of the nature of the questionnaire, it is reasonable to assume that absence of an answer indicates no use of services.

Multiple imputation[71] was the statistical technique chosen to deal with missing cost and HRQoL data because of non-returned questionnaires and incomplete EQ-5D profiles, using the user-defined programme 'ice' in Stata 11.1. Multiple imputation presents three major advantages over standard ad hoc methods for dealing with missing data (such as mean imputation and last value carried forward): (1) it makes full use of all of the available data, (2) it incorporates uncertainty associated with the missing data and (3) it ensures unbiased estimates and standard errors as long as data are MAR.[69] [Little and Rubin[72] defined three missing data mechanisms: (1) MCAR if the probability of data being unobserved is independent of both observed and unobserved values; (2) MAR if the probability of data being unobserved is dependent on the observed values but independent of unobserved ones and (3) missing not at random (MNAR) if the probability of data being unobserved is dependent on unobserved values.]

Multiple imputation follows three steps. First, regression models are used to predict plausible values for the missing observations from the observed values. A random component is included to reflect the uncertainty around the predictions. These values are then used to fill in the gaps in the data set. This process is repeated *m* number of times (*m* being the number of imputations), creating *m* number of imputed data sets. Second, each data set is analysed independently using complete-case methods. Third, the estimates obtained from each imputed data set are combined to generate mean estimates of costs and QALYs, variances and CIs using Rubin's rules,[73] in such a way that the uncertainty around the predicted values is fully taken into account.[69,74] Because the REFLUX trial has missing data for both costs and EQ-5D scores, multiple imputation using chained equations (MICE) was employed. For MICE, each variable is predicted with its own regression model. Each imputed data set is created by running the regression models over several cycles, in which each variable informs the prediction of the other variables.[69,74] To obtain overall estimates of mean and incremental costs and QALYs across all of the imputed data sets, the 'mim' command was used.[75] Semi-parametric bootstrapping in Stata 11.1 was employed to estimate the

probability that surgery is cost-effective, while maintaining the correlation between costs and QALYs (see *Incremental analysis* for more details).[76]

Plausible prediction of the missing data depends on the appropriate specification of the regression models used in MICE.[74] If a model is misspecified, the distribution of imputed values may not resemble that of the observed values, and thus the estimates of treatment effect may be biased.[69] The regression model specified will depend on the type and distribution of the variable to be predicted.[70] The variables required for the economic evaluation are costs for each year and EQ-5D scores at each time point. Both are continuous variables and neither is normally distributed; EQ-5D scores in the REFLUX trial are bounded between –0.594 and 1,[66] and costs are bounded at zero and tend to present a positive skew. Two approaches to deal with non-normality with MICE have been suggested in the literature:[69] (1) transformation towards normality and (2) predictive mean matching. [In predictive mean matching the missing observation is imputed with an observed value from an individual with a similar linear predictor.[70] Consequently, the distribution of imputed values tends to closely match the distribution of the observed values.[69]] Using the REFLUX data set none of the transformation approaches (Box–Cox,[77] log-transformation and log-transformation of non-zero values with generation of an indicator variable[78]) were successful in transforming the data distribution to normality. As a result, predictive mean matching was the strategy employed to ensure that the distribution of imputed values closely resembled the distribution of observed values. All known covariates thought to be associated with the missingness mechanism, costs and EQ-5D scores were included in the prediction equations: EQ-5D scores at each follow-up point, costs at each year, allocation, BMI, age and sex. A total of 100 imputations ($m = 100$) was used to ensure efficient and reproducible estimates.[69]

Multiple imputation provides unbiased estimates of treatment effect if data are MAR. Whether or not data are MAR is an untestable assumption by definition, as unobserved values are unknown. Departure from the MAR assumption may have implications for decision-making if the results from the cost-effectiveness analysis differ from those of the base case. Sensitivity analysis was used to test the impact on the cost-effectiveness results if data were MNAR, that is, if patients with worst outcomes or greater costs were more likely to have missing data.[70,79] Four scenarios were tested. In scenario (1), all patients with missing data had their total QALYs reduced by 10%, 20%, 30%, 40% and 50%. Conversely, in scenario (2), for all patients with missing data costs were increased by the same proportions (10%, 20%, 30%, 40% and 50%). In scenario (3), only surgery patients with missing data had their QALYs reduced. In scenario (4), costs were increased only for patients undergoing surgery.

Incremental analysis

The cost-effectiveness of surgery was evaluated by comparing the costs and QALYs incurred in the surgery arm with the costs and QALYs in the medical management arm at 5 years of follow-up, using conventional decision rules and estimating ICERs as appropriate.[80] If one intervention is associated with greater mean QALYs and lower mean costs it is deemed cost-effective by dominance. The ICER is calculated if either treatment arm does not dominate. The ICER summarises the additional costs associated with one intervention over another and relates this to the additional benefits. This ICER is then compared with a threshold for the cost per QALY. The National Institute for Health and Care Excellence (NICE) uses a threshold cost per QALY of around £20,000–30,000 to determine whether or not an intervention represents good value for money in the NHS.[65] Consequently, if the ICER is < £20,000, laparoscopic fundoplication could be considered potentially cost-effective. ICERs between £20,000 and £30,000 per QALY are considered borderline and an ICER > £30,000 is not typically considered cost-effective.

The ICER can be re-expressed using the net monetary benefit (NMB). The NMB of an intervention is the value of the health benefits gained from a particular intervention compared with standard care in monetary terms, minus the incremental costs of the intervention. The translation of health benefits into the monetary scale was made using a cost-effectiveness threshold of £20,000. This is the threshold commonly use by NICE (this corresponds to 1 QALY being valued at £20,000). Therefore, the NMB provides a measure of the gain (or loss) in resources of investing in a particular intervention when those resources could have

been used elsewhere.[81] The NMB of laparoscopic fundoplication and medical management were calculated and used to demonstrate the influence of trial duration on the estimates of cost-effectiveness of surgery.

As discussed previously, the multiple imputed data set was used as the base case for the cost-effectiveness analysis because of the large proportion of data lost for the complete-case analysis. Because total costs and total QALYs are cumulative quantities, any missing data at any of the follow-up points will result in that patient being dropped from a complete-case analysis. The cost-effectiveness results using the complete case are presented for comparison. Complete-case analysis will provide unbiased estimates only if the data are MCAR, that is, the probability of data being unobserved is independent of both observed and unobserved values. Multiple imputation ensures unbiased estimates if the data are MAR (the probability of data being unobserved is dependent on the observed values but independent of unobserved ones). Because unobserved values are unknown, the missing data mechanism and hence the validity of either assumption is untestable. Nevertheless, multiple imputation presents two advantages. First, it requires a less stringent assumption for ensuring unbiased estimates. Second, if data are MCAR, both complete-case and multiple imputation estimates will be unbiased whereas, if data are MAR, complete-case analysis will be biased.

Analysis of uncertainty for incremental analysis

Sensitivity analysis is used to explore and quantify any uncertainty in the cost-effectiveness results. Three types of sensitivity analysis were undertaken: structural, scenario and probabilistic sensitivity analysis. Structural and scenario sensitivity analyses were carried out on the complete-case data set. Probabilistic sensitivity analysis was carried out in both the complete case and the multiple imputation data set.

Structural sensitivity analysis consisted of a PP analysis that classified patients according to treatment compliance at 1 year of follow-up, that is, whose management at 1 year was consistent with their original random allocation. Consequently, the PP data set consisted of the patients randomised to surgery who actually had surgery, and of the patients randomised to medical management who did not undergo surgery at 1 year. Patients randomised to medical management who had surgery might differ from those randomised to medical management who were managed medically without surgery, for several reasons. A patient's condition might have worsened, prompting surgery, or patients might have changed their preferences and wish to be taken off medication. The latter implies that, had they been screened for the study at the point in time when they had surgery, they would not have been eligible for the study. These patients would have had a preference and would not have accepted randomisation. The condition itself is complex because of its recurrent and cyclical nature (patients suffering from reflux have punctual exacerbations, which can lead them to change their preferences and request surgery). Therefore, the reasons for not complying with randomisation are likely to be a combination of the two motives (worsening of condition and change in preference). PP was chosen because it was thought to be more similar to clinical practice, where patients can experience a wait for surgery and change their preferences during this period. Any switching of treatment after 1 year is assumed to be because of a change in clinical status, which would preclude inclusion in the clinical trial.

The base-case analysis included only the costs of reflux-related GP visits and hospitalisations. Two alternative costing scenarios were tested in sensitivity analysis: including either all GP visits or all hospital use, regardless of whether they had been classified as reflux or non-reflux related.

Probabilistic sensitivity analysis attempts to quantify the joint effect of uncertainty around the costs and QALYs. Semiparametric bootstrapping was used to estimate the probability that each intervention is cost-effective for a range of cost-effectiveness threshold values. In bootstrapping, the original data are sampled with replacement to create a new data set, in order to calculate estimates of treatment effect. Repeating this process a large number of times results in a vector of replicated statistics, which ultimately provide an empirical estimate of the CIs around mean incremental costs and QALYs. The probability of an intervention being the most cost-effective is the conventional method of presenting the uncertainty around the cost-effectiveness results. The CIs around the ICER are not presented because they are difficult to interpret and

are not easy to use: a negative ICER can indicate that an intervention dominates (because it is associated with more benefits and lower costs than its comparator) or it is dominated (because it is associated with fewer benefits and higher costs).[76]

Validation

Several procedures were used to ensure the validity of the analysis. First, two statistical analysis codes (written in Stata) were developed in parallel and their results compared. Second, the code was developed by one analyst and checked independently by another. Third, the results were cross-checked in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) for a sample of the data set. Lastly, selected results were represented graphically and examined for face validity. The validity of the imputation strategy was explored by (1) analysing the data for predictors of missingness,[70] (2) comparing the distributions of the observed and imputed values graphically[70] and (3) estimation of Monte Carlo errors.[69] *Appendix 8* describes the validation process in more detail.

Results

Patient population

Complete-case analysis consisted of the patients who returned all questionnaires and completed all EQ-5D profiles. Overall, there are 172 patients in the complete-case analysis (88 randomised to medical management and 84 randomised to surgery). *Table 32* shows the numbers of questionnaires returned (includes those with some missing data) and the numbers of completed questionnaires returned for each year. As expected, the number of questionnaires returned in each year of follow-up decreases with time. The return of questionnaires does not follow a monotonic pattern, that is, patients who did not return the questionnaire for one particular year may have returned a questionnaire in subsequent years. Therefore, the number of patients in the complete-case analysis is lower than the number of completed questionnaires in year 5. The large number of patients not included in the complete-case analysis because of missing data strengthens the rationale for using the multiple imputation data sets in the base case.

Health-care resource use

Table 33 summarises yearly health-care resource use in the two trial arms according to ITT analysis. During the first year of the trial, 111 patients randomised to surgery and 10 patients randomised to medical management underwent laparoscopic fundoplication. The 111 patients who were randomised to and received surgery constituted the surgery group in the PP analysis. The 169 patients who were randomised to medical management and did not undergo surgery during the first year of follow-up constituted the medical management group in the PP analysis. In the subsequent years of follow-up there were 15 patients who underwent surgery (one patient who had been randomised to surgery and 14 patients

TABLE 32 Numbers of questionnaires returned and completed questionnaires returned and corresponding proportions per trial arm, according to ITT analysis

	Questionnaires returned, *n* (%)		Completed questionnaires,[a] *n* (%)	
Year	Surgery	Medical management	Surgery	Medical management
1	154 (87)	164 (92)	134 (75)	147 (82)
2	128 (72)	142 (79)	121 (68)	134 (75)
3	132 (74)	134 (75)	112 (63)	119 (66)
4	126 (71)	129 (72)	114 (64)	118 (66)
5	127 (71)	119 (66)	115 (65)	113 (63)
Number of patients in complete-case analysis			88 (49)	84 (47)

a Completed questionnaires means that all of the questions on health-care resource use and EQ-5D were filled in.

TABLE 33 Health-care resource use per year per trial arm, according to ITT analysis

Health-care resource	Year	Reflux-related reasons, *n*		All reasons, *n*	
		Surgery (*n* = 178)	Medical management (*n* = 179)	Surgery (*n* = 178)	Medical management (*n* = 179)
Laparoscopic fundoplication (first year)	1	111	10	N/A	N/A
Hospital overnight admissions (excluding surgery in the first year)	1	4	2	8	8
	2	1	2	8	10
	3	2	9	6	10
	4	2	9	2	10
	5	0	8	1	11
	Total	9	30	25	49
Hospital day admissions	1	22	24	40	53
	2	5	4	23	24
	3	4	6	4	10
	4	12	9	13	11
	5	4	11	7	14
	Total	47	54	87	112
Visits to and from GP	1	110	103	394	376
	2	34	115	269	373
	3	38	99	381	386
	4	55	126	422	469
	5	36	119	404	370
	Total	273	562	1870	1974
Number of patients on reflux-related medication	1	58	148	N/A	N/A
	2	48	124	N/A	N/A
	3	51	113	N/A	N/A
	4	51	106	N/A	N/A
	5	56	98	N/A	N/A

N/A, not applicable.

who had been randomised to medical management). These patients are included in the overnight hospital admissions category. Patients randomised to medical management reported more hospital and GP visits than the surgery patients over the 5 years of follow-up.

Table 34 shows the costs associated with health-care use according to ITT analysis for all available cases (see *Appendix 9* for corresponding table for PP). All available cases uses data from all questionnaires returned at each time point. Per annum costs and costs per category refer to all available data, that is, to all participants who returned the questionnaire for that particular year or for that particular category. Therefore, the sum of the costs per category is different from the sum of the costs per annum. Similarly, total costs for complete-case analysis do not correspond to the sum of the costs per category or to the sum of the costs per annum because complete case is a subset of all available data because of

the non-monotone missing data pattern. Total costs for complete-case analysis refer to the patients who returned all questionnaires and completed all EQ-5D profiles (84 surgery patients and 88 medical management patients).

Patients randomised to medical management accumulate lower costs than patients randomised to surgery. *Table 34* indicates that surgery patients accrued a large proportion of the total costs in the first year, and accumulated lower costs during the remaining 4-year follow-up than the medical management group. In contrast, the costs accrued by medical management patients are evenly distributed across the duration of the trial. These results suggest that the cost trend in medical management patients is steeper than in surgery patients; hence, that cumulative costs in medical management patients tend to increase at a greater rate than in surgery patients. Costs associated with surgery were the major cost driver for the surgery group. Costs associated with reflux-related medication were significantly greater for the medical management group than for the surgery group. Costs associated with admissions to hospital and GP visits were not statistically significantly different between the two groups. Surgery during years 2–5 is accounted for in the overnight hospital admissions. There were a few crossovers from medical management to surgery from year 2; hence, the difference in costs associated with overnight hospital admissions between the two treatment groups is small. These results suggest that patients undergoing surgery in subsequent years are not a major cost driver in determining the cost-effectiveness of surgery.

Quality-adjusted life-years

Table 35 summarises the EQ-5D scores reported at each follow-up point for all available cases (see *Appendix 9* for the corresponding table for PP). All available cases uses data from all questionnaires returned at each time point. The surgery group appears to have better HRQoL than the medical

TABLE 34 Costs associated with resource use for all available cases, discounted from year 2 at 3.5%, according to ITT analysis

Returned questionnaires in each year			Mean (SD) resource-use cost (£)		Incremental mean cost (cost surgery – cost medical management) (95% CI[a]) (£)
Surgery	Medical management	Year	Surgery	Medical management	
154	164	1	2500.75 (1697.99)	559.62 (1006.81)	1941.13 (1621.43 to 2260.83)
128	142	2	94.15 (317.63)	150.96 (356.57)	–56.81 (–138.08 to 24.46)
132	134	3	94.35 (340.33)	276.41 (894.16)	–182.05 (–345.87 to –18.24)
126	129	4	111.41 (394.00)	303.50 (1337.26)	–192.09 (–436.56 to 52.28)
127	119	5	58.38 (178.58)	234.03 (629.33)	–175.65 (–290.26 to –61.03)
Cost category					
Surgery in year 1[b]			1734.05 (1407.58)	164.31 (644.63)	1569.74 (1342.05 to 1797.42)
Reflux-related hospital night admissions			343.82 (1176.05)	302.34 (818.41)	41.47 (–247.48 to 330.42)
Reflux-related hospital day admissions			221.67 (633.61)	250.35 (631.37)	–28.68 (–209.24 to 151.87)
Reflux-related GP visits			127.18 (178.96)	200.13 (462.53)	–72.95 (–173.26 to 27.35)
Medication			121.34 (265.05)	365.70 (517.05)	–244.35 (–361.82 to –126.89)

a CIs estimated using OLS regression.

b Only surgery occurring during the first year of the trial was included here. Laparoscopic fundoplication occurring in the subsequent years of the trial has been included in reflux-related hospital admissions. For the 10 medical management patients who had surgery in the first year of the trial, the average (SD) cost was £2679 (£126). For the 111 surgery patients who had surgery in the first year of the trial, the average (SD) cost was £2798 (£501).

TABLE 35 Health-related quality of life (EQ-5D) for all available cases according to ITT analysis

Completed questionnaires returned at each time point			Mean (SD) EQ-5D		
Surgery ($n = 178$[a])	Medical management ($n = 179$[a])	Follow-up	Surgery	Medical management	Difference in EQ-5D (surgery – medical management) (95% CI)[b,c]
171	173	Baseline	0.7107 (0.2581)	0.7201 (0.2545)	–0.0094 (–0.0638 to 0.0445)
149	153	3 months	0.7881 (0.2328)	0.6894 (0.3012)	0.0987 (0.0376 to 0.1597)
152	164	Year 1	0.7537 (0.2468)	0.7097 (0.2715)	0.0440 (–0.0136 to 0.1016)
122	138	Year 2	0.7619 (0.2718)	0.7172 (0.3127)	0.0447 (–0.0273 to 0.1167)
129	132	Year 3	0.8034 (0.2312)	0.7474 (0.2621)	0.0560 (–0.0043 to 0.1163)
125	127	Year 4	0.7713 (0.2438)	0.7544 (0.2719)	0.0169 (–0.0472 to 0.0810)
124	117	Year 5	0.7743 (0.2590)	0.7612 (0.2815)	0.0131 (–0.0555 to 0.0817)

a *n* refers to the number of patients originally randomised to each trial arm.
b CIs estimated using OLS regression.
c Unadjusted for baseline EQ-5D.

management group, despite starting from a lower baseline EQ-5D on average (0.7201 in the medical management group and 0.7107 in the surgery group). The difference in HRQoL between the two treatment groups decreased with time. This may be due to patients randomised to medical management undergoing surgery throughout the follow-up period and/or to diminishing treatment effect over time.

Comparison of costs and quality-adjusted life-years between multiple imputation and complete case

Table 36 shows the comparison of the total costs per year between the complete-case data set and the multiple imputation results. Complete case includes only those participants who returned all questionnaires and fully completed the EQ-5D questionnaires. The similarity of both the means and the CIs provides some reassurance of the validity of the multiple imputation model. The distribution of costs and EQ-5D scores in the imputed data sets matches reasonably well the distribution of the original data (see *Appendix 8* for details). Furthermore, the Monte Carlo errors are <15% of the coefficient and CI estimates, suggesting that 100 imputations are sufficient to ensure reproducibility and statistical efficiency.[69]

For both the complete-case and multiple imputation data sets, the participants randomised to laparoscopic fundoplication accrued greater costs but also reported greater HRQoL than participants randomised to continued medical management. The 95% CI for mean incremental QALYs crosses zero for the unadjusted for baseline estimates, whereas it remains above zero for the adjusted values. This result reflects the baseline imbalance in mean utility between treatment groups. Therefore, these results strongly indicated that surgery is associated with a greater QALY improvement than medical management. The sum of the differences in EQ-5D for the ITT groups does not correspond to the incremental mean QALYs because of the effect of discounting.

Cost-effectiveness

The results of the incremental analysis suggest that laparoscopic fundoplication is a cost-effective strategy for GORD patients eligible for the REFLUX trial (*Table 37*). The results for the complete-case analysis concur with those for the multiple imputation data set; across adjusted and unadjusted ICER for baseline EQ-5D, ICERs range between £5468 and £8410, well below conventional cost-effectiveness thresholds of £20,000 and £30,000 per additional QALY. For both data sets (complete case and multiple imputation), the probability of surgery being the more cost-effective intervention is >0.82 for incremental analysis

TABLE 36 Comparison between the complete-case and multiple imputation data sets for costs and HRQoL, according to ITT allocation

	Incremental mean cost (cost surgery–cost medical management) (95% CI[a]) (£)		Difference in mean EQ-5D (surgery–medical management) (95% CI[a])		
Year	Complete case	Multiple imputation	Follow-up	Complete case	Multiple imputation
N/A	N/A	N/A	Baseline	–0.0388 (–0.1083 to 0.0308)	–0.0091 (–0.0615 to 0.0433)
N/A	N/A	N/A	3 months	0.0848 (0.0122 to 0.1573)	0.0825 (0.0232 to 0.0142)
1	2197.14 (1779.67 to 2614.61)	1958.46 (1617.31 to 2299.62)	Year 1	0.0519 (–0.0198 to 0.1237)	0.0407 (–0.0150 to 0.0963)
2	–139.14 (–237.02 to –41.26)	–44.58 (–129.68 to 40.52)	Year 2	0.0467 (–0.0356 to 0.1289)	0.0445 (–0.0218 to 0.1108)
3	–193.44 (–361.94 to –24.93)	–127.42 (–306.36 to 51.52)	Year 3	0.0508 (–0.0195 to 0.1211)	0.0454 (–0.0150 to 0.1057)
4	–37.66 (–173.02 to 97.70)	–144.61 (–374.78 to 85.56)	Year 4	0.0324 (–0.0395 to 0.1044)	0.0260 (–0.0356 to 0.0875)
5	–165.12 (–304.74 to –25.50)	–123.90 (–236.56 to –11.24)	Year 5	–0.0095 (–0.0871 to 0.0680)	0.0294 (–0.0358 to 0.0945)
Total cost	1661.78 (1130.00 to 2193.55)	1517.95 (1006.49 to 2029.41)	Total QALYs (unadjusted)	0.1976 (–0.0857 to 0.4810)	0.1948 (–0.0356 to 0.4251)
Monte Carlo error[b]		10.16	Monte Carlo error[b]		0.0034
			Total QALYs (adjusted)[c]	0.3039 (0.0928 to 0.5150)	0.2160 (0.0205 to 0.4115)
			Monte Carlo error[b]		0.0035

N/A, not applicable (due to costs being calculated using the area under the curve method).

a 95% CI estimated with OLS regression and the 'mim' command.

b Monte Carlo error – SD across repeated runs of the same imputation procedure with the same data.[69]

c Total QALYs adjusted for baseline EQ-5D.

TABLE 37 Incremental analysis for the ITT analysis at 5 years of follow-up for the complete-case and multiple imputation data sets

Data set	Adjustment for baseline EQ-5D?	Incremental mean costs (£) (95% CI)	Incremental mean QALYs (95% CI)	ICER (£)	Probability cost-effective at £20,000 per QALY[a]	Probability cost-effective at £30,000 per QALY[a]
Complete case	No – unadjusted QALYs	1661.78 (1130.00 to 2193.55)	0.1976 (–0.0857 to 0.4810)	8409.82	0.828	0.866
	Yes – adjusted QALYs	1661.78 (1130.00 to 2193.55)	0.3039 (0.0928 to 0.5150)	5468.36	0.989	0.996
Multiple Imputation	No – unadjusted QALYs	1517.95 (1006.49 to 2029.41)	0.1948 (–0.0356 to 0.4251)	7792.35	0.861	0.906
	Yes – adjusted QALYs	1517.95 (1006.49 to 2029.41)	0.2160 (0.0205 to 0.4115)	7027.55	0.932	0.962

a Probability of intervention being cost-effective calculated with semiparametric bootstrapping.

unadjusted for baseline EQ-5D and >0.93 once incremental QALYs are adjusted for baseline EQ-5D. In the ITT analysis the ICER is higher for the multiple imputed data than for the complete case if QALYs are adjusted for baseline EQ-5D, but lower if QALYs are unadjusted. This might reflect the effect of having baseline EQ-5D in the prediction model, which would preclude the need for adjustment.

Figure 20 shows how the NMB associated with laparoscopic fundoplication increases with the duration of the trial. This reflects the increase in costs associated with the medical group, which offsets the initial investment made in laparoscopic fundoplication in the surgery group.

Structural sensitivity analysis: per-protocol status for the complete case

Structural sensitivity analysis consisted of PP status at 1 year for the complete case. In the PP analysis patients are classified according to the treatment actually received at 1 year of follow-up. The PP group consists of 111 patients who were randomised to surgery and who actually had surgery during the first year of the trial and 169 patients who were randomised to medical management and who did not undergo surgery during this time period. However, complete-case data exist only for 84 medical management patients and 66 laparoscopic fundoplication patients. *Appendix 9* presents detailed results for costs and HRQoL according to PP analysis. As expected, patients who actually had surgery have higher costs than patients who did not undergo surgery, regardless of their randomisation. *Table 38* summarises the incremental results of the PP analysis. Similar to the ITT analysis, the surgical policy is likely to be cost-effective at conventional (NICE) thresholds for cost-effectiveness. The incremental costs are higher and the incremental QALYs lower for the PP analysis (for surgery compared with medical management) than for the ITT analysis if no adjustment is made for baseline imbalances in EQ-5D. Therefore, the ICER is also greater (surgery is less cost-effective than suggested by the ITT analysis). Once total QALYs are adjusted for baseline EQ-5D, however, the incremental mean QALYs increase substantially and the ICER is reduced. Nevertheless, the adjusted ICER in the ITT analysis is lower than that in the PP analysis by around £2000.

Scenario sensitivity analysis: all general practitioner and all hospital costs for complete case

The results of the scenario analyses strengthen the case for the surgical policy (*Table 39*). For scenario 1, replacing reflux-related GP costs by all GP costs, the ICER increased slightly in relation to the base case. Nevertheless, the ICER remains well below conventional thresholds and the probability of surgery being cost-effective is >0.83, for both adjusted and unadjusted analyses. In scenario 2, replacing reflux-related

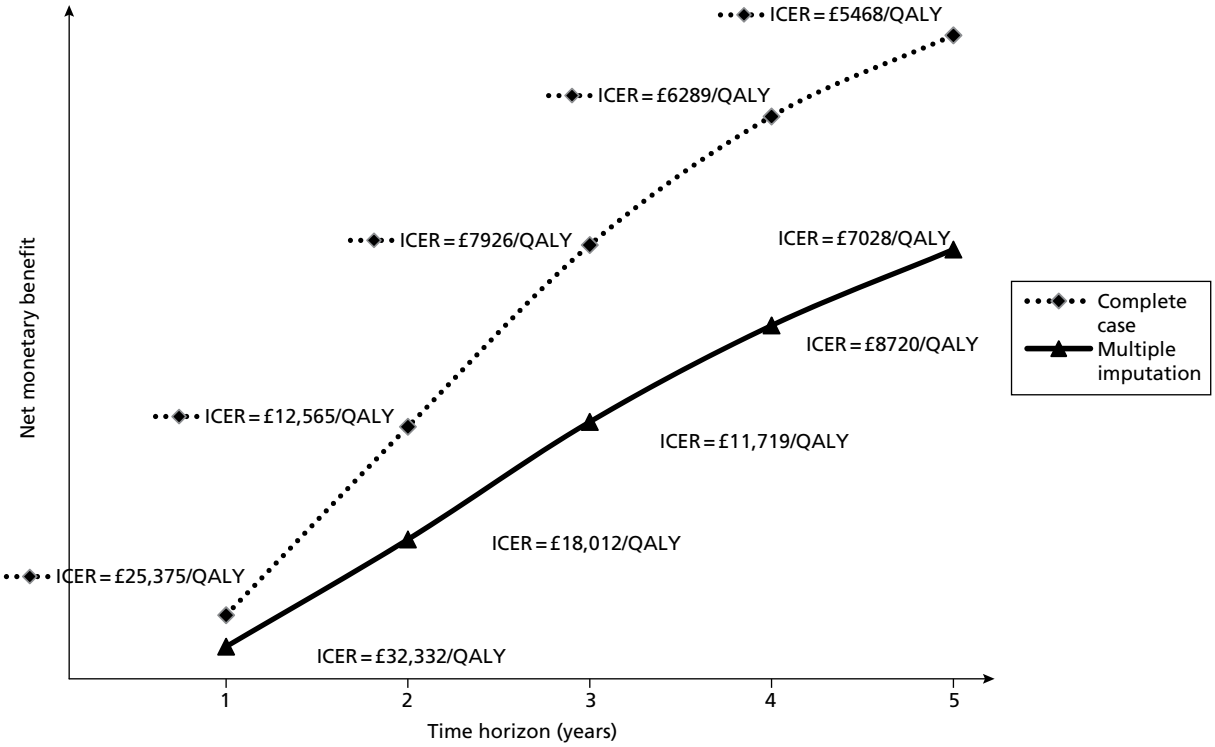


FIGURE 20 Net monetary benefit (incremental QALYs × £20,000 per QALY − incremental costs) over the duration of the REFLUX trial for the multiple imputation and complete-case data sets (QALYs adjusted by baseline EQ-5D).

hospital costs by all hospital costs, medical management was 'dominated' by the surgical policy because of this intervention being associated with greater benefits in terms of QALYs and lower costs. For this scenario the probability of surgery being cost-effective was > 0.93.

Sensitivity analysis for the multiple imputation model: departure from missing at random assumption

The multiple imputation procedure assumes that the individuals who completed and returned all questionnaires are similar to the individuals who did not, conditional on their observed characteristics (MAR assumption).[69,79] However, this may not be the case: patients who did not return a questionnaire may have experienced worse HRQoL and accrued higher health service costs, or vice versa. Sensitivity analysis on the multiple imputation model tested how sensitive the cost-effectiveness results are to the MAR assumption. *Figure 21* represents the change in NMB adjusted for baseline EQ-5D as costs and QALYs are varied in patients with missing data. The origin, marked as 'base case', refers to the incremental results from the multiple imputed data set (ICER = £7028 per additional QALY). The right quadrant plots NMB after increasing the total costs in steps of 10% for patients for whom there was missing data, for both treatment groups and for surgery-allocated patients. The left quadrant plots NMB after decreasing total QALYs in similar fashion. Positive values for NMB indicate that surgery is cost-effective; negative values indicate that surgery is not cost-effective for a threshold of £20,000 per additional QALY.

The cost-effectiveness of surgery is relatively insensitive to any increase in costs; the NMB changes little if costs are increased for patients with missing data in both treatment groups and if costs are increased just for surgery-allocated patients with missing data. A similar result is observed for the reduction in total QALYs for all patients with missing data. In contrast, the cost-effectiveness of surgery is highly sensitive if it is assumed that surgery-allocated patients with missing data experience lower HRQoL than patients with complete data. A 10% decrease in QALYs for patients randomised to surgery with missing data results in NMB decreasing to negative values. This scenario shows that missing data can have an impact on the results under certain conditions. It is impossible to empirically confirm or refute the scenario from

TABLE 38 Incremental analysis for the PP analysis at 5 years of follow-up for the complete-case data set

Adjustment for baseline EQ-5D	Incremental mean costs (95% CI) (£)	Incremental mean QALYs (95% CI)	ICER (£)	Probability cost-effective at £20,000 per QALY[a]	Probability cost-effective at £30,000 per QALY[a]
Unadjusted QALYs	2323.77 (1799.90 to 2847.65)	0.1782 (–0.1316 to 0.4879)	13,043.90	0.672	0.747
Adjusted QALYs	2323.77 (1799.90 to 2847.65)	0.3200 (0.0837 to 0.5562)	7262.85	0.957	0.983

a Probability of intervention being cost-effective was calculated with semiparametric bootstrapping.

TABLE 39 Incremental analysis for the scenario sensitivity analysis at 5 years of follow-up for the complete-case data set

Sensitivity analysis	Adjustment for baseline EQ-5D?	Incremental mean costs (95% CI) (£)	Incremental mean QALYs (95% CI)	ICER (£)	Probability cost-effective at £20,000 per QALY[a]	Probability cost-effective at £30,000 per QALY[a]
Scenario 1: all GP costs	No – unadjusted QALYs	1685.60 (1103.97 to 2267.23)	0.2125 (–0.0748 to 0.4998)	7932.23	0.826	0.863
	Yes – adjusted QALYs	1685.60 (1103.97 to 2267.23)	0.3191 (0.1061 to 0.5321)	5282.36	0.987	0.994
Scenario 2: all hospital costs	No – unadjusted QALYs	–262.72 (–860.08 to 334.65)	0.2125 (–0.0748 to 0.4998)	Medical management dominated	0.930	0.928
	Yes – adjusted QALYs	–£262.72 (–860.08 to 334.65)	0.3191 (0.1061 to 0.5321)	Medical management dominated	0.999	0.999

a Probability of intervention being cost-effective calculated with semiparametric bootstrapping.

the data in the trial, but it could be considered an extreme case. It seems improbable in practice that surgical patients with poor quality of life are less likely to respond to follow-up questionnaires than similar participants undergoing medical management.

Conclusion

The results of the within-trial economic analysis suggest that laparoscopic fundoplication is the more cost-effective option for the management of the sorts of patients suffering from GORD who were eligible for the REFLUX trial. The ICER for the ITT approach in the complete case was between £5468 and £8410 per additional QALY, and for the multiple imputed data set was between £7028 and £7792 per additional QALY, depending on whether QALYs are unadjusted or adjusted for baseline. Adjusted results are likely to be more reflective of the improvement in HRQoL associated with surgery. The probability of surgery being cost-effective was >0.80 for all analyses. The results are robust to the scenario analyses testing assumptions regarding resource-use and missing data mechanism apart from when surgery-allocated patients with missing data were assumed to experience lower HRQoL than other patients. In all scenarios the ICERs were similar to the base case ICERs and well below NICE cost-effectiveness thresholds.

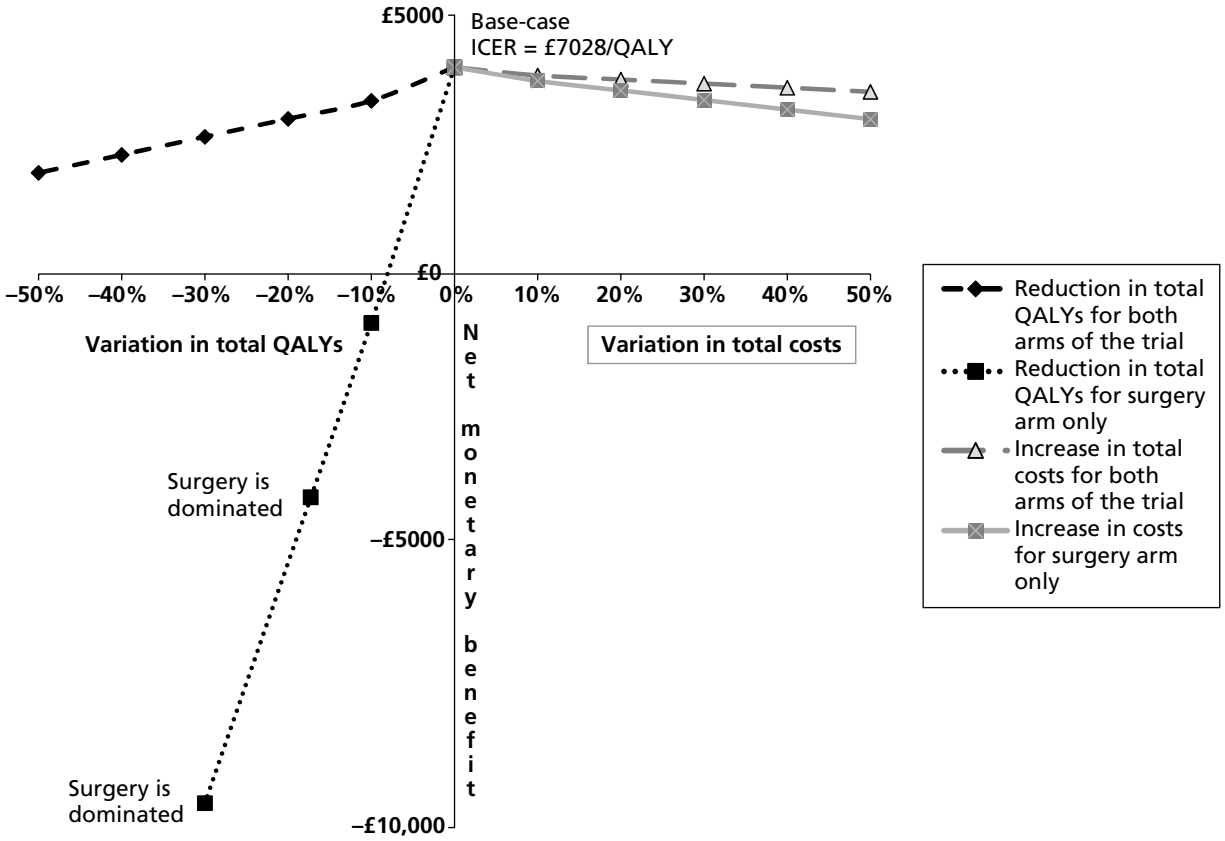


FIGURE 21 Net monetary benefit (incremental QALYs adjusted for baseline EQ-5D × £20,000 per QALY – incremental costs) over variation in total costs and total QALYs in the multiple imputed data set.

Validation of within-trial (5-year) analysis and exploration of the need for a long-term model

Introduction

The within-trial analysis found that surgery was cost-effective over a 5-year time horizon. A sufficient condition for surgery to be unambiguously cost-effective over a longer term is that, in each year after 5 years, HRQoL is lower and costs are the same or increasing faster in the medical group than in the surgical group. The results from both the multiple imputation and the complete-case analysis suggest that surgery is likely to be a cost-effective alternative over the longer term. Based on the ITT analyses undertaken so far, it is unlikely that mean HRQoL in patients who had surgery will become lower than that in patients on medical management after 5 years, and it is also very unlikely that mean annual costs incurred by surgery patients will exceed those incurred by medical management patients. If these results are robust, then there is no need to develop an economic model to extrapolate the 5-year results over a longer time horizon. Surgery would simply become more cost-effective over time.

This section develops a statistical model to investigate whether or not the results obtained in the within-trial economic analysis are robust to alternative assumptions and methods, and uses the results to consider whether or not the evidence supports this sufficient condition over the longer term.

Methods

Overview

The aim of this analysis was to estimate the difference in costs and the difference in HRQoL (measured with the EQ-5D) between the surgical and medical management randomised groups and describe how this difference evolves over time. A simple way of doing this would be to estimate the difference in costs and outcomes at each time point independently. The results of this analysis were shown in *Table 34* (for costs) and *Table 35* (for EQ-5D). These showed that costs were greater in the surgical group in the first year but greater in the medical group thereafter. EQ-5D tended to be higher in the surgical group in years 4 and 5 but the CIs crossed zero. There are two main limitations of this simple analysis:

1. The outcomes at each time point are unlikely to be independent. If the outcomes at one time point are correlated with those at other time points this analysis may lead to biased estimates of standard errors.
2. The analysis does not take account of missing data. If missing data are not MCAR then this analysis may lead to biased estimates of the mean of the coefficients.

The multiple imputation accounts for the correlation of responses from the same individuals and for the missing data (see *Table 36*). However, the validity of this analysis depends on the correct specification of the equations used to impute the missing data. Moreover, other regression-based methods are available for handling missing data in longitudinal studies, principally mixed models, and results may be sensitive to the methods used. This section uses a mixed model to handle the missing data and compares predicted outcomes with those using multiple imputation.

Mixed models

A mixed model is a regression-based method for handling continuous data that is measured at more than one time point during follow-up. It allows estimation of treatment effects under the assumption that the data are MAR, that is, dropout may depend on intermediate values. Analysing each time period separately assumes that dropouts are MCAR, a stronger assumption. A mixed model uses all of the observed data. Individuals who dropped out after providing intermediate data contribute to the estimation of the final outcomes. This analysis has the same aims as multiple imputation but uses a different method and with different assumptions. Therefore, it can also be viewed as a sensitivity analysis to test the robustness of the multiple imputation.

The mixed model can be written as:

Where for an individual i,

$Y_i = \alpha + \beta R_i + X_i + e_i$, $e_i \sim \text{MVN}(0, \Sigma)$
R_i = randomised group
Y_i = vector of all outcomes (at times 1…T)
X_i = vector of covariates

The variance–covariance matrix Σ is unstructured, that is, no prior assumptions are made about the values of the correlations. Separate models are fitted for costs and for EQ-5D. Baseline values of the EQ-5D are included as an 'outcome' (i.e. at $t = 1$). Dummies representing time points 1 to T were included as covariates X_i. Treatment effects are included as time*randomised group interactions although no treatment effect at baseline is allowed. No other covariates are included in the model.

Results

Costs

Figure 22 shows the difference in costs (excluding initial surgery) in years 1–5. Mean costs are greater in the medical management arm of the trial after the first year and the CIs only just cross zero. These results are very similar to those of *Table 34*.

European Quality of Life-5 Dimensions

Figure 23 shows the difference in EQ-5D at 3 months and in years 1–5. Mean HRQoL tends to be greater in the surgical group during the trial, although the CIs cross zero in some periods. These results are very similar to those of *Table 35*.

Conclusion

The results of the mixed model (taking account of correlations and missing data) are very similar to those of the complete-case analyses (which assumed that data at different time points were independent) and the multiple imputation (see *Table 36*). All of these analyses show that follow-up costs are significantly greater in the medical management arm of the trial (because of greater reflux-related hospital admissions, GP visits and use of medication). The analyses also show that surgery tends to be more effective, in terms of HRQoL, than medical management over the 5-year follow-up. Although this treatment difference appears to weaken over time, there is no reason to expect that surgery will become less effective with a longer follow-up. Consequently, the evidence suggests that the cost-effectiveness of laparoscopic fundoplication will not diminish if measured over a longer follow-up time. Nevertheless, there is uncertainty surrounding these conclusions because of the large proportion of missing data.

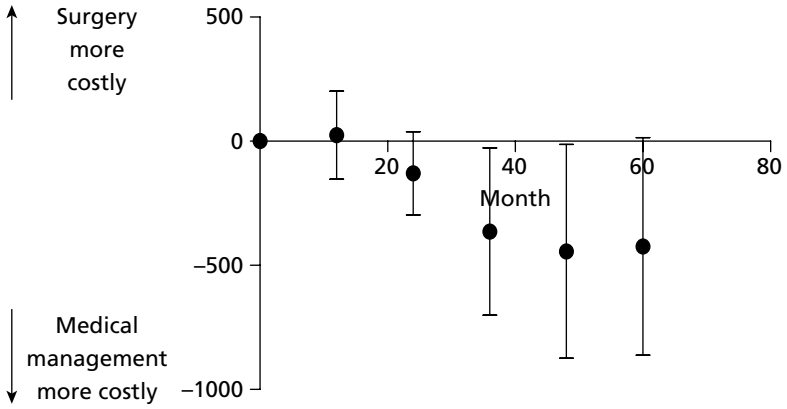


FIGURE 22 Difference in costs (£) excluding initial surgery (mean, 95% CI).

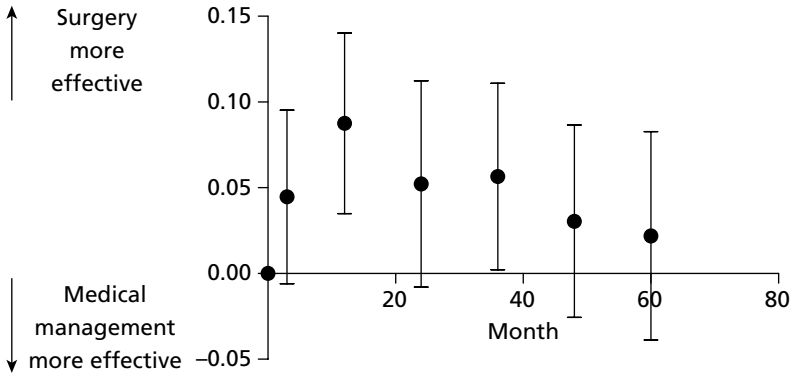


FIGURE 23 Difference in EQ-5D, adjusted for baseline (mean, 95% CI).

Discussion

The results of the cost-effectiveness analysis strongly suggest that a policy of offering laparoscopic fundoplication to people with GORD who require long-term PPI treatment for symptom control is more cost-effective than continuing to manage them with PPIs (with selective use of surgery if symptoms are poorly controlled), assuming that the cost-effectiveness thresholds used by NICE (£20,000–30,000 per QALY) are appropriate for the NHS. Surgery represents a greater initial investment and lower medium-term costs, whereas costs associated with medical management remain relatively constant or slightly increase over time. The difference in HRQoL achieved with surgery is sustained over 5 years, although the results indicate that mean EQ-5D scores for surgery and medical management tended to converge (as discussed in *Chapter 3*, in part this reflects later surgery in patients randomised to medical management). The ICER favours surgery when incremental QALYs are both adjusted and unadjusted for baseline EQ-5D. Nevertheless, adjusted incremental QALYs are likely to be a more reliable estimate of treatment effect as they account for differences in baseline utility. Patients randomised to medical management reported higher baseline utility than patients randomised to surgery. Failure to adjust for these baseline differences could result in a biased ICER, as discussed elsewhere.[62] The results from the multiple imputed data set are likely to be more accurate than the results from the complete-case analysis because of the large number of patients with incomplete data (>50%). Therefore, multiple imputation was chosen for the base case. Nevertheless, the results are similar across the data sets and laparoscopic surgery is the more cost-effective intervention for both.

There is little uncertainty regarding the cost-effectiveness results once adjustment for EQ-5D at baseline is performed. The probability of surgery being cost-effective ranged between 0.932 and 0.999 for the base case and across the scenarios tested. Furthermore, it is clear from the results of the scenario analysis that the base-case results are robust to alternative costing assumptions. The PP analysis is used to test whether or not the ITT analysis is potentially misleading because of the dilution of treatment effect (some patients randomised to surgery did not have surgery and some patients randomised to medical management actually had surgery). The PP analysis has the advantage of mimicking clinical practice and could be thought to be more relevant to decision-makers. However, the PP analysis is not without its limitations. First, and as with any PP analysis, it is sensitive to selection bias because of breaking randomisation. Second, the PP analysis may still underestimate the effect of surgery because patients having surgery between 2 and 5 years are counted as medical management. Third, the PP analysis is actually a subset of the ITT groups, which further reduces the data set. For these reasons, the ITT results are likely to be more reliable. It is important to characterise any uncertainty in the analysis as failure to do so can result in inaccurate estimates of cost-effectiveness, particularly when costs and benefits are highly skewed.[82] In addition, any analyses of uncertainty can help to illustrate where caution should be exercised when interpreting the results of a cost-effectiveness analysis. The results of the sensitivity analyses suggest that the uncertainty is likely to be driven by HRQoL. If QALYs for randomised surgery patients with missing data are reduced, surgery may no longer be cost-effective.

For the within-trial analysis no assumptions are needed about the longer-term effectiveness and costs associated with surgery and medical management. However, the within-trial analysis has some disadvantages. First, it does not account for any differences in costs and QALYs that may be expected over the longer term (>5 years), which could be due to differences in recurrence/relapse, medication use, NHS service utilisation or HRQoL. Second, it uses data only from the REFLUX trial and does not consider other sources of evidence. Third, only a limited range of sensitivity analyses was possible. Finally, the large proportion of missing observations required an assumption regarding the mechanism of missing data, which may have some impact on the cost-effectiveness estimates. The exploration of the need for a longer-term model aimed to address the first limitation of the within-trial analysis. A mixed model was used to examine the trend in the difference in costs and the difference in QALYs between treatment groups over time.

No evidence was found to suggest that the cost-effectiveness of surgery diminishes over a longer follow-up time. Both multiple imputation and mixed models are commonly used methods to handle missing data. Multiple imputation was used in the previous section because, by imputing missing data, it naturally allows the estimation of the total cost and total QALYs for each patient in the trial. Furthermore, it can handle correlation between several outcomes (in this case costs and QALYs) as well as correlation between outcomes over time. Mixed models do not explicitly impute missing data but adjust the estimates of the differences between treatment groups at each discrete follow-up time to take account of the missing data. The approach therefore offers an alternative method to multiple imputation to examine trends in the difference in costs and the difference in QALYs between treatment groups over time. Because the analyses using multiple imputation and mixed models agree, we can have more confidence that the results are valid and that surgery is the most cost-effective intervention.

A number of other studies have quantified the cost-effectiveness of laparoscopic fundoplication and medical management. Not all of these, however, use a common metric (such as QALYs) to measure benefits. Of those studies quantifying the benefits associated with the two treatments using QALYs, the ICER for surgery ranged from £180 to £49,000. There are a number of key differences between the methodologies used in the studies, which limit the extent to which comparisons can be made between results. Importantly, not all of the studies are based on within-trial analysis; in fact, only two are: those by Grant *et al.*[1] and Goeree *et al.*[46] The remainder use modelling techniques to either extrapolate short-term trial results over the longer term or pool available evidence to generate estimates of costs and outcomes. Comparing the results from Grant *et al.*[1] with those from Goeree *et al.*[46] we can see that there are quite significant differences in the estimates of cost per QALY, from £19,000 in Grant *et al.*[1] to £49,000 in Goeree *et al*.[46] This difference is primarily driven by the difference in QALYs. In Goeree *et al.*[46] the EQ-5D score is actually lower in the surgical group than in the medical management group (this is unadjusted for baseline imbalances) whereas the HUI3 score is higher for the surgical group than the medical management group. The reason for the difference between the EQ-5D and the HUI3 scores is not discussed in the paper. The cost differences in the two studies were similar. Comparing the results from Goeree *et al.*[46] with those from the updated trial analysis we see even starker contrast between the ICERs produced (£7028 vs £49,000). Again, this is driven by the differences in EQ-5D scores observed throughout the trial period. The EQ-5D scores in the REFLUX 5-year analysis are consistently higher in the surgery group than in the medical management group, although there is a tendency for convergence towards the end of the follow-up period. Further research is required to look at why the trials produce such different results using the EQ-5D.

Other considerations

The generalisability of these findings to the GORD population in the UK is difficult to ascertain because the proportion of GORD patients meeting the entry criteria for this trial is uncertain. The surgeons participating in the trial may be more proficient in the procedure than those in actual practice. Furthermore, capacity constraints may limit the offer of the surgery policy to all potentially eligible patients.

Chapter 6 Conclusions

In the report of the first phase of the REFLUX trial[1] we concluded that, among the sorts of patients recruited to the trial, laparoscopic fundoplication 'significantly increases general and reflux related QoL measures, at least up to 12 months after surgery'. There was, however, considerable uncertainty about cost-effectiveness, largely because the follow-up period was so short. Varying plausible assumptions about the longer-term effects of surgery, particularly in terms of QALYs gained and costs of medication, led to markedly differing results. This was the basis for this second phase of the trial, in which follow-up has been extended out to a time equivalent to 5 years after surgery.

The trial has a pragmatic design and compared two policies for managing GORD, rather than directly comparing surgery with PPI therapy. This is the basis for the primary analyses being based on the ITT principle as this directly compares the policies. The first policy can be characterised as relatively early surgery for most eligible patients but with the option to take medication if considered helpful, irrespective of whether or not surgery had been performed. The second policy can be described as medical management as appropriate with 'delayed' surgery in selected cases. Hence, we have not made an assumption that those taking medication after surgery are 'failures'. In our view, although surgery may have improved symptoms, the addition of PPIs may give further improvement and hence should be considered to be a component of both policies.

In contrast to the other large randomised trial (the LOTUS trial,[48] discussed in *Chapter 4*), whose primary outcome was 'treatment failure', we chose patient-reported outcome measures as our primary and main secondary outcome measures. The advantage is that they provide a 'common currency' across the two trial policies and do not depend on clinical judgements (as 'treatment failures' do). There is, however, a concern that completion of the patient-reported outcome questionnaires may be influenced by the nature of the management received. We had a reminder of this in the early stages of our trial. The DMC noticed an imbalance in baseline scores of the first few patients randomised, but not in other descriptive characteristics. It seemed that this might have been due to completion of the form after the allocation was known (although it could still have been due to chance); once it was made a requirement that the form had to be filled in before the allocation was known, however, this discrepancy disappeared. We believe that a strength of the long-tem follow-up as reported here is that, as the time from the differentiating event (surgery or no surgery) gets increasingly long, the possibility of such reporting bias becomes remote. Protection was also provided by the partially randomised patient preference design: the randomised component was limited to patients who were uncertain which treatment to choose while those who had strong views were enrolled into the preference groups.

We designed the trial with the aim of making the management policies as similar as possible to normal NHS care. So, for example, a large number of centres were involved (both teaching and non-teaching hospitals); recruitment was based on gastroenterologist–upper gastrointestinal surgeon partnerships; surgeons chose the type of fundoplication and other aspects of the procedure; after optimisation of medical management in secondary care, all subsequent medical care was in general practice; there was no requirement for extra tests or hospital visits; and simple entry criteria identified people with chronic troublesome GORD symptoms that required anti-reflux medication for reasonable control suitable for either policy (average age 46 years). The results should, therefore, be easily generalisable to standard NHS care.

The one area in which we think the trial did not 'mimic' usual care is in the relatively low proportion of those allocated surgery who actually had surgery (62%; see *Table 10*). There are reasons for thinking that the unusual circumstances of a randomised trial comparing medical management with surgery were partly responsible for the large proportion who did not have surgery. We think the rate (84%) in the preference group is likely to be more indicative of 'normal' acceptance rates. For this reason we undertook secondary

adjusted treatment received analyses aimed to compensate for this. These analyses are likely to give a better estimate of differential effects in usual care, but because they depart from the randomised groups and hence may be prone to bias they should be treated with appropriate caution. We also explored this issue through post hoc analyses stratified by whether or not those allocated surgery actually had fundoplication. This showed (see *Figure 18*) that those who had surgery had lower baseline REFLUX scores (worse symptoms) than those who did not have an operation, but that, following surgery, their scores were consistently higher that those who did not have surgery.

Despite our best attempts to retain the cohort of participants there has been some attrition over the course of the follow-up period. The response rate of 69% at 5 years can be considered satisfactory in a study of this type and is similar to the rate in the LOTUS trial (67%).[48] The rate in the REFLUX trial reflects the decision among some participants to withdraw, but with high levels of return among those remaining. Responders did differ from non-responders but we used analysis techniques to make the most of the available data (repeated measures and imputation), and the responders in the two randomised groups were generally reassuringly similar in respect of baseline characteristics.

The new results provide clear evidence of a sustained greater improvement in GORD-related QoL in the group randomised to surgery. The results also suggest sustained benefit in respect of generic health-related measures of QoL, although the differences attenuate over time and are not statistically significant at 5 years. In these respects the REFLUX trial is in line with the results of the other three randomised trials that have compared laparoscopic surgery with medical management. The worse the symptoms at entry (the lower the score at baseline), the greater are the benefits of surgery.

By 5 years, 24 (13%) of the participants randomly allocated to medical management had undergone anti-reflux surgery. Exploratory analyses (see *Figure 19*) showed that, as a group, these 24 had low REFLUX questionnaire scores (worse symptoms) at trial entry, which subsequently improved markedly after surgery. Hence, this group is at least a contributory factor to the narrowing of differences between the randomised groups over time (see, in particular, *Figures 3* and *17*) and a reason for thinking that the ITT-based analyses comparing the two management policies are likely to underestimate the effects of surgery.

The follow-up has clarified the rates of longer-term use of PPI medication in both policies. In the randomised medical group, 87% were taking medication at 1 year, falling gradually to 82% at 5 years (see *Figure 2*). The equivalent figures in the randomised surgery group were 36% at 1 year (15% among those who had surgery) and 41% (26%) at 5 years. This was in response to a question that, to avoid problems with recall, asked just about the preceding 2 weeks (rather than the full year), and we have assumed that the 2 weeks are typical of the previous year. We know, however, that medication use is sometimes dynamic – that patients stop and start. This is apparent in *Table 13*, for example: among those in the medical group who were not taking medication at the end of the first year, 13 (68%) of the 19 respondents reported that they were taking PPIs at 5 years.

Short-term complications of surgery were described in more detail in the first report of this trial. However, the REFLUX trial is consistent with the other three trials in this respect, with small numbers having associated visceral injuries, postoperative problems and dilatation of the wrap. The longer-term follow-up has now clarified the likelihood of further surgery following a fundoplication. Overall, 4% ($n = 16$) of the total 364 in the study who had fundoplication had a subsequent reflux-related operation, of whom two had a further (i.e. third) operation. Reoperation was most often conversion to a different type of wrap or a reconstruction of the same wrap. There were only two cases of reversal of the fundoplication and neither was in the randomised comparison. In total, 3% ($n = 12$) of those who had fundoplication required surgical treatment for a complication directly related to the original surgery, including oesophageal dilatation ($n = 4$) and repair of incisional hernia ($n = 3$). As described in *Chapter 4*, although it is not possible to extract exactly comparable data, these results are broadly in line with those of the other trials.

Where the REFLUX trial results do differ from the results of the other trials, especially the LOTUS trial, is in the likelihood and extent of adverse symptoms associated with fundoplication. Dysphagia, flatulence and bloating, and inability to vomit despite wanting to have all been reported to be problematical after fundoplication. However, in the REFLUX trial, the patterns of difficulty swallowing, flatulence and wanting to vomit but being physically unable to do so were similar in the two randomised groups (see *Table 15*), with no statistically significant differences.

The economic analysis of the 5-year data from the REFLUX trial had two phases. First, a within-trial 5-year cost-effectiveness analysis was undertaken; this was followed by an exploration of the need to develop a longer-term model. Differences in mean costs and mean QALYs at 5 years were used to derive an estimate of relative cost-effectiveness. The base-case approach used multiple imputation (principally because of the extent of missing data), an ITT analysis and adjustment for baseline QALYs. As described in *Chapter 5*, complete-case and PP analyses were also undertaken, as were a range of structural, scenario and probabilistic sensitivity analyses. Costs were estimated from a health-care perspective and consideration was limited to randomised trial participants. Costs for each participant were calculated by multiplying their use of health-care resources by associated unit costs and were discounted at an annual rate of 3.5%. HRQoL was calculated from serial EQ-5D measurements. The mean (SD) costs in the first year were £2501 (£1698) in the surgical group compared with £560 (£1007) in the medical group; in each subsequent year the mean costs were around £175 higher in the medical group. The estimated incremental mean cost of the surgical policy was £1518 (95% CI £1006 to £2029) with incremental mean QALYs of 0.2160 (95% CI 0.0205 to 0.4115), giving an ICER of £7028. The probability of the surgical policy being the more cost-effective was 0.93 at a threshold of £20,000 per QALY and 0.96 at a threshold of £30,000 per QALY. The complete-case analysis gave similar results and the conclusions were robust to plausible changes in assumptions, the only exception being when surgery-allocated patients with missing data were assumed to experience lower HRQoL than other patients. A regression-based mixed-model approach was then used to explore the robustness of the findings and to gauge the likelihood that the current strong evidence for cost-effectiveness might be reversed over subsequent years. The regression-based model gave very similar results to the base-case imputation approach. Given the trends in both costs and benefits, it was concluded that it was highly unlikely that the cost-effectiveness of surgery would be reversed when extrapolated beyond 5 years.

Thus, this second phase of the REFLUX trial has accomplished what it set out to do. After 5 years' follow-up, a policy of relatively early laparoscopic fundoplication among patients for whom reasonable control of GORD symptoms requires long-term medication and for whom both surgery and medical management are suitable continues to provide better relief of GORD symptoms with associated better QoL. Although surgery carries risks, complications were rare. And despite being initially more costly, a surgical policy was found to be highly likely to be cost-effective for such patients at conventional threshold costs per QALY.

Implications for health care

Extending the use of laparoscopic fundoplication to people whose GORD symptoms require long-term medication for reasonable control and who would be suitable for surgery would provide health gain that extends over a number of years. The longer-term data reported here indicate that this is highly likely to be a cost-effective use of resources. The more troublesome the symptoms, the greater the potential benefit from surgery.

Recommendations for research

The practical implications for health services of any extension of the use of laparoscopic fundoplication depend on how many patients might seek such surgery as a consequence. Most patients taking anti-reflux medication are managed in general practice. Currently, it is uncertain how many people require long-term medication for reasonable control of their GORD symptoms, how many of these would be suitable for surgery and how many would seek it; hence, it is not clear what the most efficient provision of future care might be. We therefore recommend further research to address these issues and explore the practical and resource implications of alternative policies for laparoscopic fundoplication, which include extending its use within the NHS to the sorts of patients enrolled in the REFLUX trial.

Acknowledgements

The authors wish to thank Janice Cruden and Pauline Garden for their secretarial support and data management; Samantha Wileman and Julie Bruce for invaluable help with the overview of trials; the following researchers for their assistance in nurse co-ordination, patient recruitment and follow-up: Maureen GC Gillan, Marie Cameron, Christiane Pflanz-Sinclair, Lynne Swan; Sharon McCann, who was involved in the piloting of the practical arrangements of this trial; and Allan Walker, Daniel Barnett and Gladys McPherson for database and programming support.

Members of the (1) Trial Steering Group and (2) Data Monitoring Committee who oversaw the first phase of this study were:

1. Wendy Atkin* (Independent chairperson), John Bancewicz, Garry Barton (1999–2002), Ara Darzi, Robert Heading, Janusz Jankowski,* Zygmunt Krukowski, Richard Lilford,* Iain Martin (1997–2000), Ashley Mowat, Ian Russell, Mark Sculpher and Mark Thursz.
2. Jon Nicholl,* Chris Hawkey* and Iain MacIntyre.*

*Independent of trial.

Members of the REFLUX trial group responsible for recruitment in the clinical centres were as follows:

Aberdeen: Aberdeen Royal Infirmary	A Mowat, Z Krukowski, E El-Omar, P Phull and T Sinclair
Belfast: Royal Victoria Hospital	B Clements, J Collins, A Kennedy and H Lawther
Bournemouth: Royal Bournemouth Hospital	D Bennett, N Davies, S Toop and P Winwood
Bristol: Bristol Royal Infirmary	D Alderson, P Barham, K Green and R Mittal
Bromley: Princess Royal University Hospital	M Asante and S El Hasani
Edinburgh: Royal Infirmary of Edinburgh	A De Beaux, R Heading, L Meekison, S Paterson-Brown and H Barkell
Guildford: Royal Surrey County Hospital	G Ferns, M Bailey, N Karanjia, TA Rockall and L Skelly
Hull: Hull Royal Infirmary	M Dakkak, J King, C Royston and P Sedman
Inverness: Raigmore Hospital	K Gordon, LF Potts, C Smith, PL Zentler-Munro and A Munro
Leeds: Leeds General Infirmary	S Dexter and P Maoyeddi
Leicester: Leicester Royal Infirmary	DM Lloyd
London: St Mary's Hospital	V Loh, M Thursz and A Darzi
London: Whipps Cross Hospital	A Ahmed, R Greaves, A Sawyerr, J Wellwood and T Taylor
Poole: Poole Hospital	S Hosking, S Lowrey and J Snook
Portsmouth: Queen Alexandra Hospital	P Goggin, T Johns, A Quine, S Somers and S Toh
Salford: Hope Hospital	J Bancewicz, M Greenhalgh and W Rees
Stoke-on-Trent: North Staffordshire Hospital	CVN Cheruvu, M Deakin, S Evans, J Green and F Leslie
Swansea: Morriston Hospital	JN Baxter, P Duane, MM Rahman, M Thomas and J Williams
Telford: Princess Royal Hospital	D Maxton, A Sigurdsson, MSH Smith and G Townson
Yeovil: Yeovil District Hospital	S Gore, RH Kennedy, ZH Khan and J Knight
York: York District Hospital	D Alexander, G Miller, D Parker, A Turnbull and J Turvill

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health Directorate.

Contribution of authors

Adrian Grant (Professor, Health Services Research Trialist) was the principal grant applicant and contributed to the development of the trial protocol and the preparation of the report and had overall responsibility for the conduct of the study.

Charles Boachie (Statistician, Health Statistics) conducted the statistical analysis.

Seonaidh Cotton (Trial Co-ordinator, Health Services Research Trialist) was responsible for the day-to-day management of the trial, monitored data collection and assisted in the preparation of the report.

Rita Faria (Research Fellow, Health Economics) was involved in the cost-effectiveness section, namely conducting the economic evaluation and writing the final report.

Laura Bojke (Senior Research Fellow, Health Economics) was responsible for the cost-effectiveness section – supervising the economic evaluation and the systematic review of existing economic evidence and writing the final report.

David Epstein (Honorary Research Fellow, Health Economics) was involved in the cost-effectiveness section – conducting and supervising the economic evaluation and writing the final report.

Craig Ramsay (Professor, Health Services Research Statistician/Trialist) contributed to the grant application and the trial design and conducted the statistical analysis.

Belen Corbacho (Research Fellow, Health Economics) was responsible for the systematic review of existing health economic evidence – study selection, data extraction, validity assessment and writing the final report.

Mark Sculpher (Professor of Health Economics, Health Economics) was responsible for the economic evaluation section of the grant application and overseeing the economic evaluation.

Zygmunt Krukowski (Surgeon, Gastroenterology) advised on clinical aspects of the trial and commented on the draft report.

Robert C Heading (Honorary Professor) advised on clinical aspects of the trial design and the conduct of the trial and commented on the draft report.

Marion Campbell (Director, Health Services Research Statistician/Trialist) contributed to the development of the trial design and all aspects of the conduct of the trial and commented on the draft report.

Publications

1. Grant AM, Cotton SC, Boachie C, Ramsay CR, Krukowski ZH, Heading RC, *et al.* Minimal access surgery compared with medical management for gastro-oesophageal reflux disease: five year follow-up of a randomised controlled trial (REFLUX) *BMJ* 2013;**346**:f1908. DOI: http://dx.doi.org/10.1136/bmj.f1908 (published online on 18 April 2013).
2. Faria R, Bojke L, Epstein D, Corbacho B, Sculpher M, on behalf of the REFLUX trial group. Cost-effectiveness of laparoscopic fundoplication versus continued medical management for the treatment of gastro-oesophageal reflux disease based on long-term follow-up of the REFLUX trial. *Br J Surg* 2013; in press. DOI: http://dx.doi.org/10.1002/bjs.9190.

References

1. Grant A, Wileman S, Ramsay C, Bojke L, Epstein D, Sculpher M, *et al.* The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease – a UK collaborative study. The REFLUX trial. *Health Technol Assess* 2008;**12**(31).
2. Grant AM, Wileman SM, Ramsay CR, Mowat NA, Krukowski ZH, Heading RC, *et al.* Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ* 2008;**337**:a2664. http://dx.doi.org/10.1136/bmj.a2664
3. Epstein D, Bojke L, Sculpher MJ, REFLUX trial group. Laparoscopic fundoplication compared with medical management for gastro-oesophageal reflux disease: cost effectiveness study. *BMJ* 2009;**339**:b2576. http://dx.doi.org/10.1136/bmj.b2576
4. Macran S, Wileman S, Barton G, Russell I. The development of a new measure of quality of life in the management of gastro-oesophageal reflux disease: the Reflux questionnaire. *Qual Life Res* 2007;**16**:331–43. http://dx.doi.org/10.1007/s11136-006-9005-3
5. Bojke L, Hornby E, Sculpher M. A comparison of the cost effectiveness of pharmacotherapy or surgery (laparoscopic fundoplication) in the treatment of GORD. *Pharmacoeconomics* 2007;**25**:829–41. http://dx.doi.org/10.2165/00019053-200725100-00003
6. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R, Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;**101**:1900–20. http://dx.doi.org/10.1111/j.1572-0241.2006.00630.x
7. Locke GR, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. *Gastroenterology* 1997;**112**:1448–56. http://dx.doi.org/10.1016/S0016-5085(97)70025-8
8. Kennedy T, Jones R. The prevalence of gastro-oesophageal reflux symptoms in a UK population and the consultation behaviour of patients with these symptoms. *Aliment Pharmacol Ther* 2000;**14**:1589–94. http://dx.doi.org/10.1046/j.1365-2036.2000.00884.x
9. Heading RC. Prevalence of upper gastrointestinal symptoms in the general population: a systematic review. *Scand J Gastroenterol* 1999;**34**:3–8.
10. Davis CS, Baldea A, Johns JR, Joehl RJ, Fisichella PM. The evolution and long-term results of laparoscopic antireflux surgery for the treatments of gastroesophageal reflux disease. *JSLS* 2010;**13**:332–41. http://dx.doi.org/10.4293/108680810X12924466007007
11. Peters MJ, Mukhtar A, Yunus RM, Khan S, Pappalardo J, Memon B, *et al.* Meta-analysis of randomized clinical trials comparing open and laparoscopic anti-reflux surgery. *Am J Gastroenterol* 2009;**104**:1548–61. http://dx.doi.org/10.1038/ajg.2009.176
12. Donahue PE, Larson GM, Stewardson RH, Bombeck CT. Floppy Nissen fundoplication. *Rev Surg* 1977;**34**:223–4.
13. DeMeester TR, Bonavina L, Albertucci M. Nissen fundoplication for gastroesophageal reflux disease. Evaluation of primary repair in 100 consecutive patients. *Ann Surg* 1986;**204**:9–19. http://dx.doi.org/10.1097/00000658-198607000-00002
14. Broeders JAJL, Mauritz FA, Ahmed Ali U, Draaisma WA, Ruurda JP, Gooszen HG, *et al.* Systematic review and meta-analysis of laparoscopic Nissen (posterior total) versus Toupet (posterior partial) fundoplication for gastro-oesophageal reflux disease. *Br J Surg* 2010;**97**:1318–30. http://dx.doi.org/10.1002/bjs.7174

15. Lundell L, Miettinen P, Myrvold HE, Hatlebakk JG, Wallin L, Engstrom C, *et al.* Comparison of outcomes twelve years after antireflux surgery or omeprazole maintenance therapy for reflux esophagitis. *Clin Gastroenterol Hepatol* 2009;**7**:1292–8. http://dx.doi.org/10.1016/j.cgh.2009.05.021

16. Chiba N, De Gara CJ, Wilkinson JM, Hunt RH. Speed of healing and symptom relief in grade II to IV gastroesophageal reflux disease: a meta analysis. *Gastroenterology* 1997;**112**:1798–810. http://dx.doi.org/10.1053/gast.1997.v112.pm9178669

17. Boutet R, Wilcock M, Mackenzie I. Survey on repeat prescribing for acid suppression drugs in primary care in Cornwall and the Isles of Scilly. *Aliment Pharmacol Ther* 1999;**13**:813–17. http://dx.doi.org/10.1046/j.1365-2036.1999.00524.x

18. McCullagh M, Brown C, Bell D, Powell K. Long term acid suppressing treatment survey shows variation in practice. *BMJ* 1994;**308**:1238. http://dx.doi.org/10.1136/bmj.308.6938.1238

19. Ryder SD, O'Reilly S, Miller RJ, Ross J, Jacyna MR, Levi AJ. Long term acid suppressing treatment in general practice. *BMJ* 1994;**308**:827–30. http://dx.doi.org/10.1136/bmj.308.6932.827

20. Dibley LB, Norton C, Jones R. Non-pharmacological intervention for gastro-oesophageal reflux disease in primary care. *Br J Gen Pract* 2010;**60**:e459–65. http://dx.doi.org/10.3399/bjgp10X544050

21. Roberts SJ, Bateman DN. Prescribing of antacids and ulcer-healing drugs in primary care in the north of England. *Aliment Pharmacol Ther* 1995;**9**:137–43. http://dx.doi.org/10.1111/j.1365-2036.1995.tb00362.x

22. Bashford JNR, Norwood J, Chapman SR. Why are patients prescribed proton pump inhibitors? Retrospective analysis of the link between morbidity and prescribing in the General Practice Research Database. *BMJ* 1998;**317**:452–6. http://dx.doi.org/10.1136/bmj.317.7156.452

23. Yang Y, Metz DC. Reviews in basic and clinical gastroenterology and hepatology safety of proton pump inhibitor exposure. *Gastroenterology* 2010;**139**:1115–27. http://dx.doi.org/10.1053/j.gastro.2010.08.023

24. Sheen E, Triadafilopoulos G. Adverse effects of long-term proton pump inhibitor therapy. *Dig Dis Sci* 2011;**56**:931–50. http://dx.doi.org/10.1007/s10620-010-1560-3

25. Ogawa R, Echizen H. Drug–drug interaction profiles of proton pump inhibitors. *Clin Pharmacokinet* 2010;**49**:509–33. http://dx.doi.org/10.2165/11531320-000000000-00000

26. Henshaw RC, Naji SA, Russell IT, Templeton AA. Comparison of medical abortion with surgical vacuum aspiration: women's preferences and acceptability of treatment. *BMJ* 1993;**307**:714–17. http://dx.doi.org/10.1136/bmj.307.6906.714

27. Cooper KG, Grant AM, Garratt AM. The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. *Br J Obstet Gynaecol* 1997;**104**:1367–73. http://dx.doi.org/10.1111/j.1471-0528.1997.tb11005.x

28. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ* 1989;**299**:313–15. http://dx.doi.org/10.1136/bmj.299.6694.313

29. British Medical Association and Royal Pharmaceutical Society of Great Britain. *British national formulary*. No. 52, September 2006. London: BMA and RPS; 2006.

30. Torgerson D, Sibbald B. Understanding controlled trials: what is a patient preference trial? *BMJ* 1998;**316**:360. http://dx.doi.org/10.1136/bmj.316.7128.360

31. Dent J, Brun J, Fendrick AM, Fennerty MB, Janssens J, Kahrilas PJ, *et al.* An evidence-based appraisal of reflux disease management – the Genval Workshop Report. *Gut* 1999;**44**:S1–16.

32. Edwards PJ, Roberts IG, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, *et al.* Methods to increase response rates to postal questionnaires. *Cochrane Database Syst Rev* 2007;**2**:MR000008.

33. Brealey S, Atwell C, Bryan S. Improving response rates using monetary incentives for patient completion of questionnaires: an observational study. *BMC Med Res Methodol* 2007;**7**:12. http://dx.doi.org/10.1186/1471-2288-7-12

34. Kenyon S, Pike K, Jones D, Taylor D, Salt A, Marlow N, *et al.* The effect of a monetary incentive on return of a postal health and development questionnaire: a randomised trial. *BMC Health Serv Res* 2007;**5**:55. http://dx.doi.org/10.1186/1472-6963-5-55

35. Dirmaier J, Harfst T, Koch U, Schulz H. Incentives increased return rates but did not influence partial nonresponse or treatment outcome in a randomized trial. *J Clin Epidemiol* 2007;**60**:1263–70. http://dx.doi.org/10.1016/j.jclinepi.2007.04.006

36. Brooks R, with the EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. http://dx.doi.org/10.1016/0168-8510(90)90421-9

37. Jenkinson C, Layte R, Wright L, Coulter A. *The UK SF-36: an analysis and interpretation manual*. Oxford: Health Services Research Unit; 1996.

38. Nagelkerke N, Fidler V, Bersen R, Borgdorff M. Estimating treatment effects in randomised clinical trials in the presence of non-compliance. *Stat Med* 2000;**19**:1849–64. http://dx.doi.org/10.1002/1097-0258(20000730)19:14<1849::AID-SIM506>3.0.CO;2-1

39. White IR. Uses and limitations of randomization-based efficacy estimators. *Stat Methods Med Res* 2005;**14**:327–47. http://dx.doi.org/10.1191/0962280205sm406oa

40. Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al.* Contamination in trials of educational interventions. *Health Technol Assess* 2007;**11**(43).

41. Fielding S, Fayers P, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health Qual Life Outcomes* 2009;**7**:57. http://dx.doi.org/10.1186/1477-7525-7-57

42. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998;**316**:285. http://dx.doi.org/10.1136/bmj.316.7127.285

43. Wileman SM, McCann S, Grant AM, Krukowski ZH, Bruce J. Medical versus surgical management for gastro-oesophageal reflux disease (GORD) in adults. *Cochrane Database Syst Rev* 2010;**3**:CD003243.

44. Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, *et al.* A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for treatment of patients with chronic gastroesophageal reflux disease: one-year follow-up. *Surg Innov* 2006;**13**:238–49. http://dx.doi.org/10.1177/1553350606296389

45. Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, *et al.* A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for the treatment of patients with chronic gastroesophageal reflux disease (GERD): 3-year outcomes. *Surg Endosc* 2011;**25**:2547–54. http://dx.doi.org/10.1007/s00464-011-1585-5

46. Goeree R, Hopkins R, Marshall JK, Armstrong D, Ungar WJ, Goldsmith C, *et al.* Cost-utility of laparoscopic Nissen fundoplication versus proton pump inhibitors for chronic and controlled gastroesophageal reflux disease: a 3-year prospective randomized controlled trial and economic evaluation. *Value Health* 2011;**14**:263–73. http://dx.doi.org/10.1016/j.jval.2010.09.004

47. Lundell L, Attwood S, Ell C, Fiocca R, Galmiche J, Hatlebakk J, *et al.* Comparing laparoscopic antireflux surgery with esomeprazole in the management of patients with chronic gastro-oesophageal reflux disease: a 3-year interim analysis of the LOTUS trial. *Gut* 2008;**57**:1207–13. http://dx.doi.org/10.1136/gut.2008.148833

48. Galmiche J, Hatlebakk J, Attwood S, Ell C, Fiocca R, Eklund S, *et al.* Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD. The LOTUS randomized clinical trial. *JAMA* 2011;**305**:1969–77. http://dx.doi.org/10.1001/jama.2011.626

49. Attwood SE, Lundell L, Hatlebakk JG, Eklund S, Junghard O, Galmiche J, *et al.* Medical or surgical management of GERD patients with Barrett's esophagus: the LOTUS trial 3-year experience. *J Gastrointest Surg* 2008;**12**:1646–55. http://dx.doi.org/10.1007/s11605-008-0645-1

50. Attwood SEA, Lundell L, Ell C, Galmiche J, Hatlebakk J, Fiocca R, *et al.* Standardization of surgical technique in antireflux surgery: the LOTUS trial experience. *World J Surg* 2008;**32**:995–8. http://dx.doi.org/10.1007/s00268-007-9409-4

51. Mahon D, Rhodes M, Decadt B, Hindmarsh A, Lowndes R, Beckingham I, *et al.* Randomized clinical trial of laparoscopic Nissen fundoplication compared with proton-pump inhibitors for treatment of chronic gastro-oesophageal reflux. *Br J Surg* 2005;**92**:695–9. http://dx.doi.org/10.1002/bjs.4934

52. Cookson R, Flood C, Koo B, Mahon D, Rhodes M. Short-term cost effectiveness and long-term cost analysis comparing laparoscopic Nissen fundoplication with proton-pump inhibitor maintenance for gastro-oesophageal reflux disease. *Br J Surg* 2005;**92**:700–6. http://dx.doi.org/10.1002/bjs.4933

53. Mehta S, Bennett J, Mahon D, Rhodes M. Prospective trial of laparoscopic Nissen fundoplication versus proton pump inhibitor therapy for gastroesophageal reflux disease: seven-year follow-up. *J Gastrointest Surg* 2006;**10**:1312–17. http://dx.doi.org/10.1016/j.gassur.2006.07.010

54. Wiklund I, Bigard MA, Grace E, Talley NJ, Kamm M, Veldhuyzen van Zanten S, *et al.* Quality of life in reflux and dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD) 11. *Eur J Surg* 1998;**164**(Suppl. 583):41–9. http://dx.doi.org/10.1080/11024159850191238

55. Dimenas E. Methodological aspects of evaluation of quality of life in upper gastrointestinal diseases. *Scand J Gastroenterol* 1993;**28**(Suppl. 199):18–21. http://dx.doi.org/10.3109/00365529309098350

56. Rantanen TK, Salo JA, Sipponen JT. Fatal and life-threatening complications in antireflux surgery: analysis of 5502 operations. *Br J Surg* 1999;**86**:1573–7. http://dx.doi.org/10.1046/j.1365-2168.1999.01297.x

57. van Pinxteren B, Sigterman KE, Bonis P, Lau J, Numans ME. Short-term treatment with proton pump inhibitors, H2-receptor antagonists and prokinetics for gastro-oesophageal reflux disease-like symptoms and endoscopy negative reflux disease. *Cochrane Database Syst Rev* 2010;**11**:CD002095.

58. Romagnuolo J, Meier MA, Sadowski DC. Medical or surgical therapy for erosive reflux oesophagitis: cost–utility analysis using a Markov model. *Ann Surg* 2002;**236**:191–202. http://dx.doi.org/10.1097/00000658-200208000-00007

59. Arguedas MR, Heudebert GR, Klapow JC, Centor RM, Eloubeidi M, Wilcox CM. Re-examination of the cost-effectiveness of surgical versus medical therapy in patients with gastroesophageal reflux disease: the value of long-term data collection. *Am J Gastroenterol* 2004;**99**:1023–8. http://dx.doi.org/10.1111/j.1572-0241.2004.30891.x

60. Comay D, Adam V, da Silveira EB, Kennedy W, Mayrand S, Barkun AN. The Stretta procedure versus proton pump inhibitors and laparoscopic Nissen fundoplication in the management of gastroesophageal reflux disease: a cost-effectiveness analysis. *Can J Gastroenterol* 2008;**22**:552–8.

61. *HMSO Electronic Drug Tariff for the National Health Service of England and Wales*. London: HMSO; 2010.

62. Manca A, Hawkins N. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. http://dx.doi.org/10.1002/hec.944

63. Department of Health (DoH). *NHS reference costs 2009–2010*. London: DoH; 2010.

64. Curtis L. *Unit costs of health and social care*. Canterbury: PSSRU, University of Kent; 2010.

65. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal*. London: NICE; 2008.

66. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. Discussion paper 138. York: Centre for Health Economics, University of York; 1995.

67. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technol Assess* 1999;**3**(10).

68. HM Treasury. *Green book: appraisal and evaluation in central government*. London: The Stationery Office; 2003.

69. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2010;**30**:377–99. http://dx.doi.org/10.1002/sim.4067

70. White IR, Mander A, Wason J, Bond S. Handling missing outcome data in randomised trials: a workshop for the Network of Hubs for Trials Methodology Research. Medical Research Council, Cambridge, 20 June 2011.

71. Royston P. Multiple imputation by the MICe system of chained equations. Implementation in Stata. *J Stat Softw* 2011;**45**(4). URL: www.jstatsoft.org (accessed 5 December 2012).

72. Little RJA, Rubin DB. *Statistical analysis with missing data*. New York, NY: Wiley; 1987.

73. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley; 1987. http://dx.doi.org/10.1002/9780470316696

74. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* 2007;**61**:79–90. http://dx.doi.org/10.1198/000313007X172556

75. Galati JC, Carlin JB, Royston P. MIM: Stata module to analyse and manipulate multiply imputed datasets. EconPapers. URL: http://EconPapers.repec.org/RePEc:boc:bocode:s456825 (accessed 1 July 2011).

76. Briggs AH, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999;**3**(2).

77. Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc Series B* 1964;**2**:211–43.

78. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin Trials* 2007;**4**:154–61. http://dx.doi.org/10.1177/1740774507076914

79. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011;**342**:910–12. http://dx.doi.org/10.1136/bmj.d40

80. Johannesson M, Weinstein S. On the decision rules of cost-effectiveness analysis. *J Health Econ* 1993;**12**:459–67. http://dx.doi.org/10.1016/0167-6296(93)90005-Y

81. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;**18**(Suppl.):S68–80. http://dx.doi.org/10.1177/0272989X9801800209

82. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;**26**:781–98. http://dx.doi.org/10.2165/00019053-200826090-00008

Appendix 1 Annual questionnaire

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

The REFLUX Trial

ANNUAL FOLLOW-UP QUESTIONNAIRE

A questionnaire for people participating in the REFLUX trial,
which aims to find out whether taking medication or having an operation
is the best form of treatment for gastro-oesophageal reflux disease

CONFIDENTIAL

This study is funded by the NHS Research and Development Health Technology Assessment Programme

PLEASE READ ALL THE INSTRUCTIONS BEFORE COMPLETING THE QUESTIONNAIRE

Thank you for agreeing to take part in the study. The responses you give in this questionnaire will help us find out if the treatments you get are helpful for your condition.

The information you provide will be completely confidential.

HOW TO FILL IN THE QUESTIONNAIRE

For each section please put a cross in the appropriate box like this:

Do you drive a car? **Yes** ☐

No ☒

If you make any errors while completing this questionnaire, shade out the incorrect box completely and put a cross in the correct box like this:

Do you drive a car? **Yes** ☒

No ☒

The intended answer above is No.

PLEASE USE A BLUE OR BLACK PEN TO FILL IN YOUR ANSWERS

REFLUX QUESTIONNAIRE

For the questions in section A - F, please put a cross in the box which best describes how often your symptoms have occurred and the effect they have had on your quality of life.

SECTION A - HEARTBURN

A1. In the last two weeks, how often have you experienced heartburn (a burning sensation which moves up from your chest to your throat)?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

A2. In the last two weeks, how often have you experienced any discomfort or pain in your chest?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

A3. In the last two weeks, how much has the heartburn or discomfort/pain in your chest affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

SECTION B - ACID REFLUX

B1. In the last two weeks, how often have you experienced acid reflux and/or had an acid taste in your mouth?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B2. In the last two weeks, how often have you been sick (vomited)?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B3. In the last two weeks, how often have you regurgitated (brought up) quantities of liquid or solids into your mouth?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B4. In the last two weeks, how often have you experienced a feeling of nausea (without actually being sick or regurgitating)?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B5. In the last two weeks, how often have you wanted to be sick but physically been unable to?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

B6. In the last two weeks, how much have these acid reflux symptoms affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION C – WIND

C1. In the last two weeks, how often have you experienced a lot of wind from the lower bowel?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

C2. In the last two weeks, how often have you experienced a lot of burping/belching?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

Participant Study No

☐☐☐☐☐

(for completion by co-ordinating centre in Aberdeen)

C3. In the last two weeks, how often have you experienced bloatedness and/or a feeling of trapped wind, in your stomach?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

C4. In the last two weeks, how often have you experienced loud gurgling noises from your stomach?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

C5. In the last two weeks, how much have these wind problems affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION D - EATING AND SWALLOWING

D1. In the last two weeks, how often have you experienced difficulty swallowing food or have you actually choked on food?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

Participant Study No

☐☐☐☐☐

(for completion by co-ordinating centre in Aberdeen)

D2. In the last two weeks, how often have your eating habits been restricted because of your condition? Examples might be eating more slowly, having smaller portions or eating different foods.

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

D3. In the last two weeks, how much have these problems with eating affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION E – BOWEL MOVEMENTS

E1. In the last two weeks, how often have you experienced diarrhoea and/or loose stools?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

E2. In the last two weeks, how often have you experienced constipation and/or hard stools?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

Participant Study No

☐☐☐☐☐

(for completion by co-ordinating centre in Aberdeen)

E3. In the last two weeks, how often have you had a feeling of an urgent need to have a bowel movement?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

E4. In the last two weeks, how often have you had a feeling of not emptying your bowels?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most days ☐
- Everyday ☐

E5. In the last two weeks, how much have these bowel problems affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION F – SLEEP

F1. In the last two weeks, how often have you experienced difficulty in lying down to sleep?

- Not at all ☐
- Once a week ☐
- Two or three times a week ☐
- Most nights ☐
- Every night ☐

F2. In the last two weeks, how often have you experienced difficulty getting to sleep because of your reflux symptoms?

- ☐ Not at all
- ☐ Once a week
- ☐ Two or three times a week
- ☐ Most nights
- ☐ Every night

F3. In the last two weeks, how often have you been woken up because of your reflux symptoms?

- ☐ Not at all
- ☐ Once a week
- ☐ Two or three times a week
- ☐ Most nights
- ☐ Every night

F4. In the last two weeks, how much have these sleep related problems affected your quality of life?

- ☐ Not at all
- ☐ A little
- ☐ Moderately
- ☐ A lot
- ☐ Extremely

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

SECTION G – WORK, PHYSICAL AND SOCIAL ACTIVITIES

For the following section, please put a cross in the box which best applies to you.

G1. In the last two weeks, have your reflux symptoms affected you at work (paid or voluntary)?

- Not applicable (I do not do paid or voluntary work) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still work ☐
- Yes, I have worked less often because of my symptoms ☐
- Yes, I have not worked in the last two weeks because of my symptoms ☐
- I no longer work because of my symptoms ☐

G2. In the last two weeks, have your reflux symptoms affected your ability to perform less strenuous activities (such as going for a gentle walk, shopping or housework)?

- Not applicable (I do not perform these activities, though this is not due to my reflux symptoms) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still perform these activities as often as ever ☐
- Yes, I perform these activities less often because of my symptoms ☐
- Yes, I have not performed these activities in the last two weeks ☐
- I no longer perform these activities at all because of my symptoms ☐

G3. In the last two weeks, have your reflux symptoms affected your ability to perform strenuous activities (such as brisk walking or swimming)?

- Not applicable (I do not perform these activities, though this is not due to my reflux symptoms) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still perform these activities as often as ever ☐
- Yes, I perform these activities less often because of my symptoms ☐
- Yes, I have not performed these activities in the last two weeks ☐
- I no longer perform these activities at all because of my symptoms ☐

Participant Study No

(for completion by co-ordinating centre in Aberdeen)

G4. In the last two weeks, have you found that your reflux symptoms have affected any of your social activities (such as going out for meals, going out for drinks or socializing with other people)?

- Not applicable (I do not perform these activities, though this is not due to my reflux symptoms) ☐
- No, my symptoms do not affect me ☐
- Yes, my symptoms have affected me but I still perform these activities as often as ever ☐
- Yes, I perform these activities less often because of my symptoms ☐
- Yes, I have not performed these activities in the last two weeks ☐
- I no longer perform these activities at all because of my symptoms ☐

G5. In the last two weeks, how much has the effect of your reflux symptoms on your work, physical or social activities affected your quality of life?

- Not at all ☐
- A little ☐
- Moderately ☐
- A lot ☐
- Extremely ☐

SECTION H – DESCRIBING YOUR OWN HEALTH TODAY

By placing a cross in one box in each group below, please indicate which statements best describe your own health state today

Mobility	I have no problems in walking about	☐
	I have some problems in walking about	☐
	I am confined to bed	☐
Self-care	I have no problems with self-care	☐
	I have some problems washing or dressing myself	☐
	I am unable to wash or dress myself	☐
Usual Activities *(e.g. work, study, housework, family or leisure activities)*	I have no problems with performing my usual activities	☐
	I have some problems with performing my usual activities	☐
	I am unable to perform my usual activities	☐
Pain/Discomfort	I have no pain or discomfort	☐
	I have moderate pain or discomfort	☐
	I have extreme pain or discomfort	☐
Anxiety/Depression	I am not anxious or depressed	☐
	I am moderately anxious or depressed	☐
	I am extremely anxious or depressed	☐

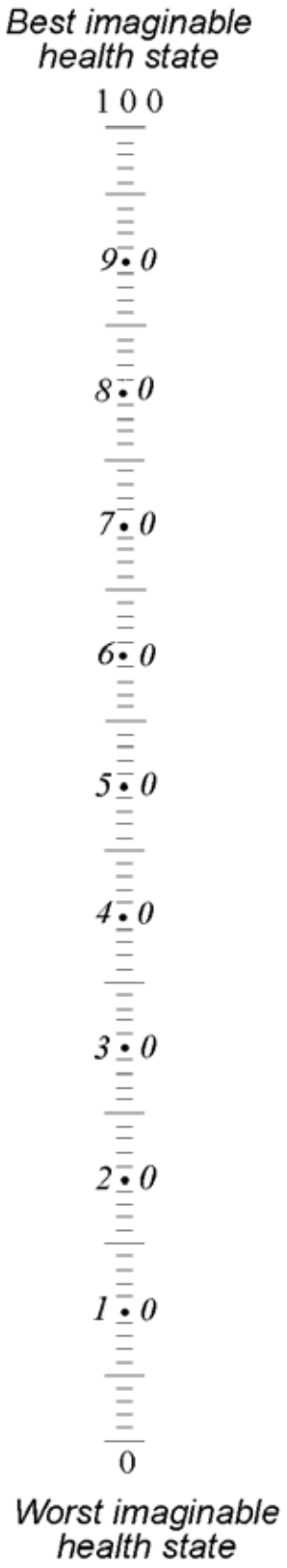
SECTION H - DESCRIBING YOUR OWN HEALTH TODAY

Please indicate on this scale how good or bad your own health state is today.

The best health state you can imagine is marked 100 and the worst health state you can imagine is marked 0.

Please draw a line from the box below to the point on the scale that best indicates how good or bad your health state is today.

Your own health state today



SECTION I – GENERAL HEALTH

Please fill in all the questions again by putting a cross in the relevant box of the answer that applies to you.

These questions ask for your views about your health and how you feel about life in general. Do not spend too much time in answering as your immediate response is likely to be the most accurate.

1. In general, would you say your health is:

Excellent	Very good	Good	Fair	Poor
☐	☐	☐	☐	☐

2. Compared to one year ago, how would you rate your health in general now?

Much better now than one year ago	Somewhat better now than one year ago	About the same as one year ago	Somewhat worse now than one year ago	Much worse now than one year ago
☐	☐	☐	☐	☐

3. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes limited a lot	Yes limited a little	No, not limited at all
a) **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sport	☐	☐	☐
b) **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling or playing golf	☐	☐	☐
c) Lifting or carrying groceries	☐	☐	☐
d) Climbing **several** flights of stairs	☐	☐	☐
e) Climbing **one** flight of stairs	☐	☐	☐
f) Bending, kneeling or stooping	☐	☐	☐
g) Walking **more than one mile**	☐	☐	☐
h) Walking **several hundred yards**	☐	☐	☐
i) Walking **one hundred yards**	☐	☐	☐
j) Bathing or dressing yourself	☐	☐	☐

4. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a) Cut down on the **amount of time** you spent on work or other activities	☐	☐	☐	☐	☐
b) **Accomplished less** than you would like	☐	☐	☐	☐	☐
c) Were limited in the **kind** of work or other activities	☐	☐	☐	☐	☐
d) Had **difficulty** performing the work or other activities (for example, it took extra effort)	☐	☐	☐	☐	☐

5. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a) Cut down on the **amount of time** you spent on work or other activities	☐	☐	☐	☐	☐
b) **Accomplished less** than you would like	☐	☐	☐	☐	☐
c) Did work or other **activities less carefully than usual**	☐	☐	☐	☐	☐

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with the family, friends, neighbours, or groups?

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

7. How much bodily pain have you had during the past 4 weeks?

None	Very mild	Mild	Moderate	Severe	Very severe
☐	☐	☐	☐	☐	☐

8. During the past 4 weeks, how much did pain interfere with your normal work (including both outside the home and housework)?

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks…

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a) Did you feel full of life?	☐	☐	☐	☐	☐
b) Have you been very nervous?	☐	☐	☐	☐	☐
c) Have you felt so down in the dumps that nothing could cheer you up?	☐	☐	☐	☐	☐
d) Have you felt calm and peaceful?	☐	☐	☐	☐	☐
e) Did you have a lot of energy?	☐	☐	☐	☐	☐
f) Have you felt downhearted and depressed?	☐	☐	☐	☐	☐
g) Did you feel worn out?	☐	☐	☐	☐	☐
h) Have you been happy?	☐	☐	☐	☐	☐
i) Did you feel tired?	☐	☐	☐	☐	☐

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
☐	☐	☐	☐	☐

11. How TRUE or FALSE is each of the following statements for you?

		Definitely true	Mostly true	Don't know	Mostly false	Definitely false
a)	I seem to get sick a little easier than other people	☐	☐	☐	☐	☐
b)	I am as healthy as anybody I know	☐	☐	☐	☐	☐
c)	I expect my health to get worse	☐	☐	☐	☐	☐
d)	My health is excellent	☐	☐	☐	☐	☐

SECTION J - HEALTH CARE RELATED QUESTIONS

In the following questions, we are trying to find out about some of the costs you incurred over the last 12 MONTHS as a result of your health problems.

If you are not sure or cannot remember exact details, please give the best answer you can.

1. CURRENT EMPLOYMENT

Please tick the box, which best describes your current employment status.

- ☐ Full time employment
- ☐ Part time employment
- ☐ Student
- ☐ Retired
- ☐ Housework
- ☐ Seeking work
- ☐ Other

2. TIME AWAY FROM WORK, DUE TO ILLNESS

If you are in paid employment, how many days off work have you had in the past 12 MONTHS because of health problems?

☐☐☐ Days in total

☐☐☐ Days because of reflux symptoms

3. VISITS TO NHS HEALTH CARE FACILITIES

a) How many times in the past 12 MONTHS have you personally visited your GP? Do not include visits made on behalf of others, or if you are a woman attending routine visits because of your pregnancy.

☐ Total number of visits

☐ Visits because of your reflux symptoms

b) How many times in the past 12 MONTHS have you personally had a visit from your GP?

☐ Total number of visits

☐ Visits because of your reflux symptoms

Please give details of the visits that you have had TO or FROM your GP in the spaces below
(continue on a separate sheet if necessary).

Visit 1

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

Visit 2

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

Visit 3

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

Visit 4

Date of visit Month ☐☐ Year 2 0 ☐☐

Reason for visit

c) How many times in the past 12 MONTHS have you personally had to attend the outpatients or casualty department of a hospital?

☐ Total number of visits

☐ Visits because of your **reflux** symptoms

d) How many times in the past 12 MONTHS have you personally been admitted to a hospital as a day case (do not stay overnight)?

☐ Total number of day case admissions

☐ Admissions because of your **reflux** symptoms

Please give details of the day case admissions you have had and approximate date, in the spaces below *(continue on a separate sheet if necessary)*.

Admission 1

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

Admission 2

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

Admission 3

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

Admission 4

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Reason for day case admission

e) How many times in the past 12 MONTHS have you personally been admitted to a hospital for treatment as an inpatient (overnight or longer)?

☐ Total number of **inpatient** admissions

☐ Admissions because of your **reflux** symptoms

Please give details of the inpatient stays you have had, in the spaces below. (continue on a separate sheet if necessary)

Admission 1

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

Admission 2

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

Admission 3

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

Admission 4

Date of admission Day ☐☐ Month ☐☐ Year 2 0 ☐☐

Number of nights ☐☐

Reason for admission and details of any procedures

4. PRESCRIBED MEDICATION FOR REFLUX

Are you currently being PRESCRIBED medication for reflux symptoms?

YES ☐ **NO** ☐ → *If NO, please go to question 5 on page 29*

If YES, please put a cross in the box against the current dose you are being prescribed and write in the number of tablets you have taken in the last two weeks.

(Please note the dose can be found on the side of your tablet bottle or packet)

	Dose (mg)						Number of tablets taken in the last 2 weeks
Omeprazole (Losec)	10mg ☐		20mg ☐		40mg ☐		☐
Lansoprazole (Zoton)	15mg ☐		30mg ☐				☐
Pantoprazole (Protium)	20mg ☐		40mg ☐				☐
Rabeprazole (Pariet)	10mg ☐		20mg ☐				☐
Esomeprazole (Nexium)	20mg ☐		40mg ☐				☐
Ranitidine (Zantac)	150mg ☐		300mg ☐				☐
Famotidine (Pepcid)	20mg ☐		40mg ☐				☐
Nizatidine (Axid)	150mg ☐		300mg ☐				☐
Cimetidine (Tagamet)	400mg ☐		800mg ☐				☐
Domperidone (Motilium)	10mg ☐		20mg ☐				☐
Metoclopramide (Maxolon)	10mg ☐		20mg ☐				☐

If you are prescribed any other medication (tablets or liquid) for your reflux symptoms that are not listed above, please list below the name(s) of the medicine(s) and include the number of times you have taken it in the last two weeks.

Names of medication	Number of times taken in last 2 weeks
e.g. Gaviscon	

5. NON PRESCRIBED MEDICATION FOR REFLUX

Please list below the names of any NON PRESCRIBED (over the counter) medication (tablets/liquid) you take for your REFLUX symptoms and include the number of times you have taken it in the last two weeks.

Names of medication	Number of times taken in last 2 weeks
e.g. Rennies	

IF YOU HAVE ANY OTHER COMMENTS about your gastro-oesophageal reflux symptoms, your reflux treatment or this study, please write them below.

THANK YOU FOR YOUR HELP IN COMPLETING THIS QUESTIONNAIRE

Once you have completed the form, please return it in the pre-paid envelope provided or to the following address:

REFLUX Trial Office
Health Services Research Unit
Polwarth Building
Foresterhill
Aberdeen AB25 2ZD
Tel: 01224 XXXXXX
Fax: 01224 XXXXXX
E-mail: reflux@hsru.abdn.ac.uk

Appendix 2 Intra- and postoperative surgical outcomes

TABLE 40 Intra- and post-operative surgical outcomes

Surgical outcome	Surgical participants, *n* (%)	
	Randomised (*n* = 111)	Preference (*n* = 218)
Conversion	2 (1.8)	0 (0.0)
Liver injury	1 (0.9)	1 (0.5)
Splenic injury	0 (0.0)	1 (0.5)
Pleural injury	1 (0.9)	2 (0.9)
Oesophageal injury	0 (0.0)	0 (0.0)
Other visceral injury	0 (0.0)	0 (0.0)
Haemorrhage	1 (0.9)	1 (0.5)
Pneumothorax	0 (0.0)	2 (0.9)
Blood transfusion	0 (0.0)	1 (0.5)[a]
Other postoperative event	3 (2.7)	5 (2.3)
ICU admission	0 (0.0)	0 (0.0)
HDU admission	1 (0.9)	2 (0.9)
Reoperation within 12 months	0 (0.0)	3 (1.4)
Stricture dilatation or food disimpaction required within 12 months	1 (0.9)	2 (0.9)
Discharged status		
Home	107 (96.4)	213 (97.7)
Other	4 (3.6)	5 (2.3)
Length of stay (days), median (IQR)	2 (2–3)	2 (2–3)

HDU, high-dependency unit; ICU, intensive-care unit; IQR, interquartile range.

a Participant was transfused with three units.

Appendix 3 Tables showing medication use in preceding fortnight at each time point of follow-up

TABLE 41 Follow-up at the time point equivalent to 3 months after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	150	109	158	150	230	203	182	178
PPIs, *n* (%)[a]								
Any PPI	47 (31.3)	8 (7.3)	140 (88.6)	133 (88.7)	41 (17.8)	13 (6.4)	167 (91.8)	152 (84.9)
Omeprazole	16 (10.7)	5 (4.6)	45 (28.5)	45 (30.0)	15 (6.5)	3 (1.5)	57 (31.3)	57 (31.8)
Lansoprazole	19 (12.7)	3 (2.8)	55 (34.8)	54 (36.0)	13 (5.7)	7 (3.5)	67 (36.8)	64 (35.8)
Pantoprazole	1 (0.7)	0 (0.0)	9 (5.7)	8 (5.3)	3 (1.3)	2 (1.0)	14 (7.7)	14 (7.8)
Rabeprazole	4 (2.7)	1 (0.9)	9 (5.7)	9 (6.0)	3 (1.3)	0 (0.0)	13 (7.1)	13 (7.3)
Esomeprazole	7 (4.7)	1 (0.9)	22 (13.9)	21 (14.0)	7 (3.0)	3 (1.5)	21 (11.5)	21 (11.7)
H_2RAs, *n* (%)[a]								
Any H_2RA	1 (0.7)	0 (0.0)	12 (7.6)	10 (6.7)	4 (1.7)	2 (1.0)	14 (7.7)	13 (7.3)
Ranitidine	0 (0.0)	0 (0.0)	8 (5.1)	8 (5.3)	2 (0.9)	1 (0.5)	10 (5.5)	9 (5.0)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.6)
Cimetidine	1 (0.7)	0 (0.0)	1 (0.6)	1 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Nizatidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.6)
Over-the-counter H_2RA	0 (0.0)	0 (0.0)	4 (2.5)	4 (2.7)	2 (0.9)	1 (0.5)	3 (1.7)	3 (1.7)
Prokinetics, *n* (%)[a]								
Any prokinetic	7 (4.7)	3 (2.8)	6 (3.8)	5 (3.3)	7 (3.0)	6 (3.0)	5 (2.7)	4 (2.2)
Domperidone	3 (2.0)	1 (0.9)	6 (3.8)	5 (3.3)	3 (1.3)	2 (1.0)	4 (2.2)	3 (1.7)
Metoclopramide	4 (2.7)	2 (1.8)	0 (0.0)	0 (0.0)	4 (1.7)	4 (2.0)	1 (0.5)	1 (0.6)
Any reflux-related drug, *n* (%)[a]	50 (33.3)	10 (9.2)	146 (92.4)	139 (92.7)	45 (19.6)	17 (8.4)	176 (96.7)	161 (89.9)
Other prescribed drugs, *n*[b]								
Alginates	0	0	4	4	0	0	2	2
Antispasmodics (e.g. dicycloverine)	0	0	1	1	2	2	0	0
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0

continued

TABLE 41 Follow-up at the time point equivalent to 3 months after surgery: medications (*continued*)

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Other ulcer-healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	0	0	0	0	1	1
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	1	1	0	0	2	1	1	1
Anti-motility	0	0	0	0	1	1	0	0

a Percentage is for responders completing the relevant section of the questionnaire.
b More than one prescription per person possible.

TABLE 42 Follow-up at the time point equivalent to 12 months after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	154	104	164	155	230	202	177	174
PPIs, *n* (%)[a]								
Any PPI	56 (36.4)	13 (12.5)	142 (86.6)	139 (89.7)	42 (18.3)	19 (9.4)	156 (88.1)	154 (88.5)
Omeprazole	19 (12.3)	6 (5.8)	47 (28.7)	45 (29.0)	14 (6.1)	4 (2.0)	61 (34.5)	60 (34.5)
Lansoprazole	21 (13.6)	2 (1.9)	51 (31.1)	50 (32.3)	17 (7.4)	12 (5.9)	56 (31.6)	5 (31.6)
Pantoprazole	2 (1.3)	1 (1.0)	9 (5.5)	9 (5.8)	3 (1.3)	1 (0.5)	16 (9.0)	16 (9.2)
Rabeprazole	3 (1.9)	1 (1.0)	12 (7.3)	12 (7.7)	2 (0.9)	0 (0.0)	9 (5.1)	9 (5.2)
Esomeprazole	11 (7.1)	3 (2.9)	25 (15.2)	25 (16.1)	8 (3.5)	3 (1.5)	15 (8.5)	15 (8.6)
H_2RAs, *n* (%)[a]								
Any H_2RA	4 (2.6)	3 (2.9)	9 (5.5)	9 (5.8)	5 (2.2)	2 (1.0)	13 (7.3)	13 (7.5)
Ranitidine	3 (1.9)	2 (1.9)	7 (4.3)	7 (4.5)	2 (0.9)	0 (0.0)	8 (4.5)	8 (4.6)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.6)
Cimetidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Nizatidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	1 (0.6)	1 (1.0)	2 (1.2)	2 (1.3)	3 (1.3)	2 (1.0)	5 (2.8)	5 (2.9)
Prokinetics, *n* (%)[a]								
Any prokinetic	6 (3.9)	2 (1.9)	4 (2.4)	4 (2.6)	5 (2.2)	4 (2.0)	6 (3.4)	5 (2.9)
Domperidone	4 (2.6)	1 (1.0)	4 (2.4)	4 (2.6)	1 (0.4)	0 (0.0)	5 (2.8)	4 (2.3)
Metoclopramide	2 (1.3)	1 (1.0)	0 (0.0)	0 (0.0)	4 (1.7)	4 (2.0)	1 (0.6)	1 (0.6)
Any reflux-related drug, *n* (%)[a]	58 (37.7)	15 (14.4)	148 (90.2)	144 (92.9)	46 (20.0)	22 (10.8)	165 (93.2)	163 (93.7)
Other prescribed drugs, *n*[b]								
Alginates	3	0	4	4	1	0	5	5
Antispasmodics (e.g. dicycloverine)	1	1	4	4	1	1	1	1
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0
Other ulcer- healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	1	1	0	0	0	0
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	2	2	6	5	4	4	3	3
Anti-motility	0	0	0	0	0	0	0	0

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 43 Follow-up at the time point equivalent to 2 years after surgery: medications

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	128	86	142	136	203	182	156	153
PPIs, n (%)[a]								
Any PPI	47 (36.7)	13 (15.1)	121 (85.2)	119 (87.5)	43 (21.2)	29 (15.9)	129 (82.7)	128 (83.7)
Omeprazole	17 (13.3)	4 (4.7)	43 (30.3)	42 (30.9)	11 (5.4)	8 (4.4)	48 (30.8)	47 (30.7)
Lansoprazole	17 (13.3)	3 (3.5)	39 (27.5)	39 (28.7)	15 (7.4)	13 (7.1)	42 (26.9)	42 (27.5)
Pantoprazole	1 (0.8)	1 (1.2)	6 (4.2)	5 (3.7)	2 (1.0)	1 (0.5)	12 (7.7)	12 (7.8)
Rabeprazole	4 (3.1)	1 (1.2)	12 (8.5)	12 (8.8)	3 (1.5)	0 (0.0)	6 (3.8)	6 (3.9)
Esomeprazole	8 (6.3)	4 (4.7)	20 (14.1)	20 (14.7)	9 (4.4)	5 (2.7)	14 (9.0)	14 (9.2)
H_2RAs, *n* (%)[a]								
Any H_2RA	2 (1.6)	1 (1.2)	5 (3.5)	5 (3.7)	1 (0.5)	0 (0.0)	13 (8.3)	13 (8.5)
Ranitidine	2 (1.6)	1 (1.2)	4 (2.8)	4 (2.9)	1 (0.5)	0 (0.0)	12 (7.7)	12 (7.8)
Famotidine	0 (0.0)	0 (0.0)	1 (0.7)	1 (0.7)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.7)
Cimetidine	0 (0.0)	0 (0.0)	1 (0.7)	1 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	1 (0.8)	0 (0.0)	2 (1.4)	2 (1.5)	2 (1.0)	2 (1.1)	2 (1.3)	2 (1.3)
Prokinetics, *n* (%)[a]								
Any prokinetic	5 (3.9)	4 (4.7)	7 (4.9)	6 (4.4)	5 (2.5)	3 (1.6)	4 (2.6)	4 (2.6)
Domperidone	5 (3.9)	4 (4.7)	3 (2.1)	3 (2.2)	2 (1.0)	0 (0.0)	3 (1.9)	3 (2.0)
Metoclopramide	0 (0.0)	0 (0.0)	4 (2.8)	3 (2.2)	3 (1.5)	3 (1.6)	1 (0.6)	1 (0.7)
Any reflux-related drug, *n* (%)[a]	48 (37.5)	14 (16.3)	124 (87.3)	122 (89.7)	46 (22.7)	30 (16.5)	140 (89.7)	139 (90.8)
Other prescribed drugs, *n*[b]								
Alginates	5	2	4	4	5	3	6	6
Antispasmodics (e.g. dicycloverine)	1	1	0	0	0	0	1	1
Chelates (e.g. sucralfate)	0	0	0	0	1	1	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	1	1
Mucogel	1	1	0	0	0	0	0	0
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	1	0	0	0	3	3	5	5
Anti-nausea	0	0	1	1	1	1	0	0
Anti-motility	0	0	0	0	0	0	0	0
Over the counter – not prescribed	6	0	8	7	2	1	6	6

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 44 Follow-up at the time point equivalent to 3 years after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	132	92	134	133	196	175	159	156
PPIs, *n* (%)[a]								
Any PPI	50 (37.9)	18 (19.6)	112 (83.6)	112 (84.2)	47 (24.0)	32 (18.3)	129 (81.1)	128 (82.1)
Omeprazole	19 (14.4)	8 (8.7)	43 (32.1)	43 (32.3)	16 (8.2)	9 (5.1)	51 (32.1)	50 (32.1)
Lansoprazole	13 (9.8)	1 (1.1)	37 (27.6)	37 (27.8)	20 (10.2)	16 (9.1)	45 (28.3)	45 (28.8)
Pantoprazole	3 (2.3)	3 (3.3)	4 (3.0)	4 (3.0)	3 (1.5)	3 (1.7)	10 (6.3)	10 (6.4)
Rabeprazole	5 (3.8)	2 (2.2)	8 (6.0)	8 (6.0)	1 (0.5)	0 (0.0)	6 (3.8)	6 (3.8)
Esomeprazole	8 (6.1)	5 (5.4)	20 (14.9)	20 (15.0)	6 (3.1)	4 (2.3)	12 (7.5)	12 (7.7)
H_2RAs, *n* (%)[a]								
Any H_2RA	2 (1.5)	1 (1.1)	2 (1.5)	2 (1.5)	2 (1.0)	0 (0.0)	10 (6.3)	10 (6.4)
Ranitidine	2 (1.5)	1 (1.1)	2 (1.5)	2 (1.5)	2 (1.0)	0 (0.0)	9 (5.7)	9 (5.8)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.6)
Over-the-counter H_2RA	1 (0.8)	0 (0.0)	2 (1.5)	2 (1.5)	6 (3.1)	2 (1.1)	4 (2.5)	2 (1.3)
Prokinetics, *n* (%)[a]								
Any prokinetic	3 (2.3)	2 (2.2)	6 (4.5)	6 (4.5)	5 (2.6)	2 (1.1)	4 (2.5)	4 (2.6)
Domperidone	2 (1.5)	2 (2.2)	4 (3.0)	4 (3.0)	2 (1.0)	0 (0.0)	3 (1.9)	3 (1.9)
Metoclopramide	1 (0.8)	0 (0.0)	2 (1.5)	2 (1.5)	3 (1.5)	2 (1.1)	1 (0.6)	1 (0.6)
Any reflux-related drug, *n* (%)[a]	51 (38.6)	18 (19.6)	113 (84.3)	113 (85.0)	47 (24.0)	32 (18.3)	135 (84.9)	134 (85.9)
Other prescribed drugs, *n*[b]								
Alginates	9	4	12	12	10	3	12	6
Antispasmodics (e.g. dicycloverine)	1	1	1	1	0	0	1	1
Chelates (e.g. sucralfate)	0	0	0	0	1	1	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	1	1
Mucogel	0	0	0	0	1	0	0	0
Asilone	0	0	0	0	0	0	0	0
Non-gastrointestinal	2	2	6	6	1	3	1	5
Anti-nausea	0	0	0	0	0	0	0	0
Anti-motility	0	0	0	0	1	1	0	0
Over the counter – not prescribed	0	0	1	1	2	1	2	6

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 45 Follow-up at the time point equivalent to 4 years after surgery: medications

Medication	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	126	88	129	127	168	152	142	139
PPIs, *n* (%)[a]								
Any PPI	52 (41.3)	21 (23.9)	104 (80.6)	104 (81.9)	42 (25.0)	30 (19.7)	118 (83.1)	117 (84.2)
Omeprazole	21 (16.7)	6 (6.8)	40 (31.0)	40 (31.5)	17 (10.1)	13 (8.6)	44 (31.0)	43 (30.9)
Lansoprazole	15 (11.9)	6 (6.8)	34 (26.4)	34 (26.8)	16 (9.5)	13 (8.6)	48 (33.8)	48 (34.5)
Pantoprazole	1 (0.8)	1 (1.1)	4 (3.1)	4 (3.1)	1 (0.6)	0 (0.0)	6 (4.2)	6 (4.3)
Rabeprazole	5 (4.0)	2 (2.3)	6 (4.7)	6 (4.7)	2 (1.2)	0 (0.0)	4 (2.8)	4 (2.9)
Esomeprazole	8 (6.3)	6 (6.8)	16 (12.4)	16 (12.6)	3 (1.8)	2 (1.3)	12 (8.5)	12 (8.6)
H_2RAs, *n* (%)[a]								
Any H_2Ra	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.8)	4 (2.4)	1 (0.7)	6 (4.2)	6 (4.3)
Ranitidine	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.8)	4 (2.4)	1 (0.7)	5 (3.5)	5 (3.6)
Famotidine	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.7)	1 (0.7)
Over-the-counter H_2RA	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.8)	2 (1.3)	2 (1.4)	2 (1.4)
Prokinetics, *n* (%)[a]								
Any prokinetic	6 (4.8)	4 (4.5)	8 (6.2)	7 (5.5)	5 (3.0)	3 (2.0)	7 (4.9)	7 (5.0)
Domperidone	5 (4.0)	3 (3.4)	5 (3.9)	4 (3.1)	3 (1.8)	1 (0.7)	6 (4.2)	6 (4.3)
Metoclopramide	1 (0.8)	1 (1.1)	3 (2.3)	2 (1.6)	2 (1.2)	2 (1.3)	1 (0.7)	1 (0.7)
Any reflux-related drug, *n* (%)[a]	51 (40.5)	21 (23.9)	106 (82.2)	105 (82.7)	43 (25.6)	31 (20.4)	125 (88.0)	124 (89.2)
Other prescribed drugs, *n*[b]								
Alginates	12	4	11	11	12	9	13	13
Antispasmodics (e.g. dicycloverine)	2	2	2	1	0	0	0	0
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	0	0	0	0	0	0
Asilone	1	1	1	1	0	0	1	1
Non-gastrointestinal	4	4	3	1	1	0	0	0
Anti-nausea	0	0	1	1	1	1	0	0
Anti-motility	0	0	0	1	1	1	0	0
Over the counter – not prescribed	0	0	0	0	3	2	0	0

a Percentage is for responders completing the relevant section of the questionnaire.

b More than one prescription per person possible.

TABLE 46 Follow-up at the time point equivalent to 5 years after surgery: medications

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Medication	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	127	90	119	116	176	158	136	133
PPIs, *n* (%)[a]								
Any PPI	55 (43.3)	23 (25.6)	98 (82.4)	97 (83.6)	48 (27.3)	36 (22.8)	116 (85.3)	113 (85.0)
Omeprazole	24 (18.9)	10 (11.1)	44 (37.0)	43 (37.1)	22 (12.5)	15 (9.5)	48 (35.3)	46 (34.6)
Lansoprazole	16 (12.6)	5 (5.6)	31 (26.1)	31 (26.7)	18 (10.2)	15 (9.5)	45 (33.1)	45 (33.8)
Pantoprazole	1 (0.8)	1 (1.1)	2 (1.7)	2 (1.7)	0 (0.0)	0 (0.0)	7 (5.1)	7 (5.3)
Rabeprazole	3 (2.4)	2 (2.2)	4 (3.4)	4 (3.4)	0 (0.0)	0 (0.0)	2 (1.5)	2 (1.5)
Esomeprazole	9 (7.1)	6 (6.7)	17 (14.3)	17 (14.7)	7 (4.0)	6 (3.8)	12 (8.8)	11 (8.3)
H_2RAs, *n* (%)[a]								
Any H_2RA	0 (0.0)	0 (0.0)	2 (1.7)	2 (1.7)	3 (1.7)	2 (1.3)	6 (4.4)	6 (4.5)
Ranitidine	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.9)	3 (1.7)	2 (1.3)	5 (3.7)	5 (3.8)
Famotidine	0 (0.0)	0 (0.0)	1 (0.8)	1 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Over-the-counter H_2RA	1 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	4 (2.3)	1 (0.6)	4 (2.9)	4 (3.0)
Prokinetics, *n* (%)[a]								
Any prokinetic	6 (4.7)	5 (5.6)	5 (4.2)	4 (3.4)	8 (4.5)	6 (3.8)	5 (3.7)	4 (3.0)
Domperidone	5 (3.9)	4 (4.4)	4 (3.4)	3 (2.6)	4 (2.3)	2 (1.3)	4 (2.9)	3 (2.3)
Metoclopramide	1 (0.8)	1 (1.1)	2 (1.7)	1 (0.9)	3 (1.7)	3 (1.9)	1 (0.7)	1 (0.8)
Any reflux-related drug, *n* (%)[a]	56 (44.1)	24 (26.7)	98 (82.4)	97 (83.6)	49 (27.8)	37 (23.4)	121 (89.0)	118 (88.7)
Other prescribed drugs, *n*[b]								
Alginates	11	3	11	11	12	9	15	15
Antispasmodics (e.g. dicycloverine)	1	1	0	0	1	0	1	1
Chelates (e.g. sucralfate)	0	0	0	0	0	0	0	0
Other ulcer-healing drugs	0	0	0	0	0	0	0	0
Mucogel	0	0	0	0	0	0	0	0
Asilone	1	1	0	0	0	0	0	0
Non-gastrointestinal	1	1	3	1	1	0	1	1
Anti-nausea	0	0	0	0	1	0	0	0
Anti-motility	0	0	0	0	1	1	0	0
Over the counter – not prescribed	0	0	0	0	2	0	0	0

a Percentage is for responders completing the relevant section of the questionnaire.
b More than one prescription per person possible.

Appendix 4 Tables showing health status measures at each time point of follow-up

TABLE 47 Follow-up at the time point equivalent to 3 months after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/ allocated	178	111	179	169	261	218	192	189
Number of responders	149	97	157	141	229	186	182	168
REFLUX QoL, mean (SD)[a]	83.9 (19.4)	85.9 (19.0)	70.6 (24.6)	70.8 (24.4)	80.4 (21.6)	82.5 (20.3)	80.2 (18.2)	80.6 (17.7)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	84.8 (17.3)	89.4 (14.0)	66.9 (26.2)	66.5 (26.0)	84.1 (19.6)	87.2 (16.6)	75.7 (19.6)	76.0 (19.5)
Wind and frequency symptom score	58.1 (19.7)	55.9 (19.7)	53.7 (22.6)	54.4 (22.5)	52.2 (21.1)	52.6 (20.7)	60.7 (22.2)	60.9 (22.3)
Nausea and vomiting symptom score	91.5 (15.7)	93.1 (15.7)	82.1 (20.7)	82.3 (20.2)	90.2 (15.2)	91.6 (13.7)	89.5 (12.9)	90.0 (11.9)
Activity limitation symptom score	88.2 (17.0)	89.9 (16.7)	81.6 (19.6)	81.9 (19.0)	88.4 (18.0)	89.7 (17.5)	87.9 (13.2)	88.0 (13.3)
Constipation and swallowing symptom score	79.2 (20.0)	78.7 (20.7)	75.8 (20.9)	77.0 (19.8)	77.1 (21.2)	76.9 (21.3)	84.2 (16.9)	84.6 (16.5)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	49.2 (10.0)	49.3 (10.4)	46.5 (11.5)	46.6 (11.6)	49.9 (9.7)	50.4 (9.4)	47.6 (10.3)	47.5 (10.4)
Norm-based role physical	47.7 (11.8)	47.4 (12.1)	44.8 (12.1)	45.0 (12.1)	48.1 (11.3)	48.7 (10.7)	47.1 (10.4)	47.1 (10.4)
Norm-based bodily pain	48.5 (10.3)	48.8 (10.8)	45.3 (11.4)	45.3 (11.3)	48.4 (11.3)	49.0 (11.2)	46.5 (10.2)	46.5 (10.3)
Norm-based general health	46.3 (11.0)	47.4 (11.0)	40.7 (11.2)	40.7 (11.2)	47.2 (11.3)	48.2 (11.1)	42.5 (10.5)	42.6 (10.4)
Norm-based vitality	47.1 (11.9)	48.0 (12.1)	43.9 (12.4)	44.3 (12.2)	48.0 (11.9)	48.4 (11.9)	44.7 (11.4)	44.8 (11.4)
Norm-based social functioning	47.2 (11.5)	47.5 (12.1)	43.6 (12.7)	43.8 (12.6)	46.8 (12.3)	47.6 (12.0)	46.9 (10.5)	46.9 (10.5)
Norm-based role emotional	48.3 (12.3)	48.4 (12.5)	43.9 (14.2)	44.1 (14.2)	47.0 (12.6)	48.9 (11.7)	47.0 (11.4)	46.9 (11.4)
Norm-based mental health	48.7 (12.0)	49.7 (11.9)	44.5 (12.2)	44.7 (11.9)	48.3 (12.2)	49.2 (11.8)	47.1 (10.6)	47.1 (10.7)
EQ-5D, mean (SD)[a]	0.788 (0.233)	0.806 (0.239)	0.689 (0.301)	0.696 (0.299)	0.806 (0.245)	0.817 (0.240)	0.763 (0.231)	0.765 (0.229)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 48 Follow-up at the time point equivalent to 12 months after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	154	98	165	149	232	192	181	169
REFLUX QoL, mean (SD)[a]	84.6 (17.9)	88.3 (15.6)	73.4 (23.3)	73.1 (23.7)	83.3 (20.7)	86.0 (17.9)	79.2 (19.2)	79.4 (19.0)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	84.7 (17.5)	90.2 (14.0)	67.4 (25.8)	66.7 (25.8)	85.0 (19.4)	87.7 (16.5)	73.9 (20.7)	74.0 (20.8)
Wind and frequency symptom score	56.7 (21.0)	56.9 (21.7)	52.6 (23.3)	52.7 (23.5)	56.9 (22.5)	57.5 (22.1)	61.4 (21.9)	61.5 (22.0)
Nausea and vomiting symptom score	91.9 (14.4)	94.7 (11.8)	84.0 (18.6)	83.3 (18.8)	91.1 (16.5)	93.3 (13.8)	88.6 (15.4)	88.9 (14.4)
Activity limitation symptom score	90.7 (12.8)	93.3 (11.5)	82.2 (19.2)	81.6 (19.4)	90.8 (16.8)	92.4 (14.8)	87.3 (14.7)	87.4 (14.8)
Constipation and swallowing symptom score	79.3 (19.1)	80.2 (19.6)	74.5 (22.8)	75.2 (22.3)	78.5 (20.2)	79.1 (19.7)	83.6 (17.6)	83.8 (17.4)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	48.9 (10.3)	49.6 (10.3)	47.2 (11.0)	47.2 (10.9)	49.7 (10.8)	50.3 (10.5)	47.4 (10.5)	47.4 (10.6)
Norm-based role physical	46.7 (11.4)	47.4 (11.3)	45.8 (11.8)	46.0 (11.7)	49.0 (11.2)	49.6 (10.5)	46.8 (10.7)	46.8 (10.7)
Norm-based bodily pain	47.7 (10.4)	48.5 (10.7)	44.5 (10.9)	44.5 (10.9)	49.1 (11.3)	49.9 (11.1)	47.4 (9.9)	47.4 (10.0)
Norm-based general health	45.2 (11.1)	46.2 (11.8)	40.7 (11.2)	40.5 (11.1)	46.4 (10.8)	47.2 (10.6)	42.3 (10.1)	42.3 (10.1)
Norm-based vitality	46.9 (11.5)	47.6 (11.6)	44.2 (11.9)	44.4 (11.7)	47.3 (12.0)	48.0 (11.7)	45.1 (10.3)	45.2 (10.3)
Norm-based social functioning	46.9 (11.6)	47.8 (11.7)	45.2 (12.2)	45.4 (12.1)	46.9 (12.5)	47.8 (12.1)	46.6 (10.6)	46.6 (10.6)
Norm-based role emotional	46.4 (13.5)	47.2 (12.9)	44.2 (14.4)	44.4 (14.2)	47.3 (13.3)	48.1 (12.7)	46.2 (12.0)	46.1 (12.0)
Norm-based mental health	47.2 (11.7)	48.5 (11.6)	46.4 (12.1)	46.5 (12.2)	46.9 (12.0)	47.4 (12.0)	46.5 (10.9)	46.6 (10.9)
EQ-5D, mean (SD)[a]	0.754 (0.247)	0.777 (0.232)	0.709 (0.272)	0.710 (0.270)	0.791 (0.263)	0.803 (0.252)	0.741 (0.240)	0.743 (0.238)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 49 Follow-up at the time point equivalent to 2 years after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	128	86	142	136	203	182	156	153
REFLUX QoL, mean (SD)[a]	85.5 (17.3)	89.2 (15.1)	76.9 (22.8)	77.0 (22.9)	85.3 (19.0)	87.1 (17.5)	80.4 (19.1)	80.5 (19.1)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	83.1 (18.5)	88.2 (15.0)	71.8 (25.4)	71.5 (25.4)	85.7 (19.9)	88.2 (17.1)	75.2 (19.5)	75.1 (19.6)
Wind and frequency symptom score	57.1 (20.0)	57.1 (20.0)	54.9 (24.1)	55.4 (24.4)	56.1 (23.2)	56.9 (22.5)	61.1 (21.6)	61.3 (21.5)
Nausea and vomiting symptom score	92.4 (13.1)	94.3 (11.6)	86.3 (18.1)	86.3 (18.4)	91.9 (14.6)	93.0 (13.1)	89.3 (14.0)	89.5 (13.4)
Activity limitation symptom score	91.2 (12.1)	93.7 (10.5)	83.3 (20.6)	83.6 (20.2)	92.4 (14.7)	93.3 (13.5)	86.8 (15.6)	86.9 (15.6)
Constipation and swallowing symptom score	80.5 (19.5)	81.3 (20.2)	77.6 (22.5)	77.5 (22.9)	80.1 (21.0)	80.7 (20.7)	81.5 (17.6)	81.6 (17.7)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	48.4 (9.8)	48.9 (9.7)	46.7 (11.3)	47.0 (11.0)	49.0 (11.0)	49.6 (10.2)	46.2 (11.7)	46.2 (11.7)
Norm-based role physical	48.6 (10.6)	49.2 (10.0)	45.8 (12.2)	46.0 (12.0)	49.3 (10.8)	49.9 (10.1)	46.1 (11.1)	45.9 (11.2)
Norm-based bodily pain	47.6 (9.3)	48.1 (8.8)	44.8 (10.7)	45.1 (10.6)	47.7 (10.0)	48.2 (9.6)	45.7 (9.1)	45.9 (9.1)
Norm-based general health	44.6 (11.1)	45.2 (11.9)	41.3 (11.4)	41.3 (11.3)	46.5 (10.7)	46.9 (10.4)	41.6 (10.4)	41.7 (10.3)
Norm-based vitality	46.6 (10.7)	46.9 (10.9)	43.4 (11.6)	43.4 (11.3)	47.1 (11.6)	47.5 (11.4)	44.4 (10.6)	44.4 (10.5)
Norm-based social functioning	47.4 (11.4)	48.1 (11.2)	45.3 (12.0)	45.5 (11.7)	47.7 (12.1)	48.1 (11.8)	46.2 (11.6)	46.2 (11.6)
Norm-based role emotional	48.4 (11.8)	49.0 (11.6)	45.5 (14.1)	45.8 (13.8)	48.9 (11.8)	49.8 (10.7)	45.5 (12.4)	45.3 (12.4)
Norm-based mental health	47.9 (11.5)	47.9 (12.0)	45.6 (11.7)	45.9 (11.6)	48.0 (12.0)	48.8 (11.5)	45.6 (10.8)	45.5 (10.8)
EQ-5D, mean (SD)[a]	0.762 (0.272)	0.790 (0.244)	0.717 (0.313)	0.721 (0.308)	0.796 (0.257)	0.816 (0.233)	0.736 (0.235)	0.735 (0.237)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 50 Follow-up at the time point equivalent to 3 years after surgery: health status

Health status measure	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	132	92	134	133	196	175	159	156
REFLUX QoL, mean (SD)[a]	87.0 (15.0)	88.0 (15.3)	79.9 (20.1)	79.7 (20.1)	85.6 (18.2)	87.3 (17.2)	81.9 (16.4)	81.9 (16.5)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	85.4 (17.4)	88.4 (17.1)	74.8 (23.1)	74.6 (23.1)	85.7 (19.6)	87.3 (18.6)	77.4 (18.7)	77.2 (18.7)
Wind and frequency symptom score	59.5 (22.9)	57.9 (23.6)	56.1 (25.5)	55.8 (25.4)	55.3 (22.5)	55.9 (22.4)	62.0 (22.8)	62.1 (22.7)
Nausea and vomiting symptom score	94.0 (10.2)	95.4 (9.0)	89.2 (16.2)	89.1 (16.2)	92.1 (14.8)	93.0 (14.3)	90.3 (13.4)	90.4 (13.3)
Activity limitation symptom score	91.6 (13.2)	93.0 (13.3)	87.6 (16.7)	87.5 (16.7)	92.7 (12.1)	93.6 (11.6)	88.9 (12.9)	88.9 (13.0)
Constipation and swallowing symptom score	82.1 (16.8)	81.1 (17.0)	79.6 (20.1)	79.4 (20.1)	78.4 (21.7)	78.7 (21.4)	83.1 (17.4)	83.3 (17.4)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	49.1 (10.2)	49.8 (9.9)	47.8 (11.3)	47.8 (11.3)	49.3 (10.6)	49.6 (10.4)	46.9 (11.2)	46.9 (11.3)
Norm-based role physical	48.1 (10.9)	48.1 (11.1)	47.0 (11.4)	46.9 (11.4)	48.5 (11.3)	48.9 (10.9)	46.7 (11.4)	46.5 (11.4)
Norm-based bodily pain	47.4 (9.7)	46.3 (9.8)	46.3 (10.3)	46.3 (10.3)	48.1 (10.2)	48.6 (10.0)	46.4 (9.1)	46.3 (9.1)
Norm-based general health	45.3 (10.0)	45.8 (10.1)	42.4 (11.8)	42.3 (11.8)	46.2 (11.2)	46.6 (11.1)	41.8 (10.2)	41.8 (10.2)
Norm-based vitality	46.0 (11.5)	46.8 (11.2)	44.7 (12.7)	44.6 (12.7)	47.1 (11.7)	47.7 (11.6)	44.5 (10.3)	44.5 (10.3)
Norm-based social functioning	48.5 (10.4)	48.7 (10.6)	46.2 (11.9)	46.1 (11.9)	47.6 (12.4)	47.9 (12.3)	47.0 (11.3)	46.9 (11.3)
Norm-based role emotional	49.6 (9.9)	49.2 (10.4)	45.9 (13.3)	45.8 (13.4)	48.0 (12.9)	48.2 (12.8)	47.0 (11.8)	46.8 (11.8)
Norm-based mental health	49.5 (10.8)	49.7 (11.0)	46.1 (12.0)	46.0 (12.0)	47.9 (12.0)	48.4 (11.8)	46.6 (10.6)	46.5 (10.6)
EQ-5D, mean (SD)[a]	0.803 (0.231)	0.790 (0.252)	0.747 (0.262)	0.745 (0.262)	0.803 (0.249)	0.805 (0.251)	0.763 (0.231)	0.761 (0.232)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 51 Follow-up at the time point equivalent to 4 years after surgery: health status

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
Health status measure	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	126	88	129	127	168	152	142	139
REFLUX QoL, mean (SD)[a]	85.2 (18.2)	87.7 (17.7)	81.1 (20.7)	81.9 (19.4)	86.2 (16.8)	86.9 (16.3)	83.7 (17.2)	83.7 (17.2)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	83.1 (20.8)	87.0 (20.2)	77.4 (21.9)	77.5 (21.8)	85.7 (19.4)	87.5 (17.4)	79.2 (19.7)	79.0 (19.8)
Wind and frequency symptom score	58.1 (22.1)	57.0 (23.1)	53.6 (23.9)	53.2 (23.7)	55.3 (23.0)	55.7 (22.8)	62.4 (23.4)	62.5 (23.5)
Nausea and vomiting symptom score	91.0 (16.7)	93.6 (14.9)	89.2 (16.4)	89.4 (16.1)	92.9 (12.7)	94.0 (11.3)	91.4 (11.7)	91.6 (11.4)
Activity limitation symptom score	91.8 (13.1)	93.6 (12.4)	87.5 (17.2)	88.1 (15.3)	92.1 (13.7)	92.9 (12.6)	90.4 (12.9)	90.4 (12.9)
Constipation and swallowing symptom score	80.4 (19.4)	79.7 (20.0)	79.6 (20.9)	79.8 (20.6)	79.0 (21.9)	79.5 (20.9)	82.8 (18.1)	82.9 (18.1)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	47.9 (10.1)	48.5 (10.4)	47.5 (11.7)	47.7 (11.4)	49.5 (10.6)	50.0 (10.1)	47.2 (10.4)	47.1 (10.5)
Norm-based role physical	47.5 (11.9)	47.1 (12.4)	46.7 (12.2)	46.8 (12.0)	49.4 (10.6)	49.8 (10.2)	47.6 (10.3)	47.5 (10.4)
Norm-based bodily pain	46.1 (10.4)	46.4 (11.0)	46.3 (10.2)	46.4 (10.0)	47.8 (10.1)	48.1 (10.1)	47.9 (9.0)	47.9 (9.0)
Norm-based general health	44.6 (10.4)	45.5 (11.0)	42.2 (11.4)	42.4 (11.2)	46.5 (11.2)	47.0 (10.9)	42.2 (11.4)	42.1 (11.5)
Norm-based vitality	44.8 (10.7)	44.8 (10.9)	45.6 (11.7)	45.6 (11.4)	47.4 (11.6)	48.1 (11.5)	45.1 (11.0)	45.1 (11.0)
Norm-based social functioning	45.6 (12.7)	45.8 (12.9)	46.1 (11.9)	46.3 (11.6)	48.3 (11.1)	48.9 (10.6)	46.7 (11.3)	46.8 (11.4)
Norm-based role emotional	48.1 (12.4)	48.7 (12.4)	46.5 (14.1)	46.7 (13.8)	48.9 (11.7)	49.4 (11.3)	47.2 (11.5)	47.3 (11.5)
Norm-based mental health	47.5 (11.8)	49.1 (11.2)	47.0 (12.0)	47.2 (11.6)	48.7 (11.5)	49.1 (11.3)	47.3 (11.2)	47.4 (11.2)
EQ-5D, mean (SD)[a]	0.771 (0.244)	0.778 (0.264)	0.754 (0.272)	0.760 (0.258)	0.806 (0.254)	0.825 (0.229)	0.773 (0.213)	0.773 (0.215)

a Mean (SD) based on responders completing relevant section of the questionnaire.

TABLE 52 Follow-up at the time point equivalent to 5 years after surgery: health status

Health status measure	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Number randomised/allocated	178	111	179	169	261	218	192	189
Number of responders	127	90	119	116	176	158	136	133
REFLUX QoL, mean (SD)[a]	86.7 (13.8)	89.8 (11.7)	80.7 (20.3)	80.6 (20.4)	85.3 (17.3)	86.2 (17.1)	84.8 (15.2)	85.0 (15.3)
REFLUX symptom score, mean (SD)[a]								
General discomfort symptom score	85.0 (17.5)	89.6 (14.2)	75.3 (22.6)	75.0 (22.7)	85.7 (19.0)	86.7 (18.6)	78.8 (19.2)	78.8 (19.3)
Wind and frequency symptom score	58.8 (21.8)	58.7 (22.5)	56.4 (22.7)	55.8 (22.5)	54.0 (23.5)	53.9 (23.3)	65.3 (22.2)	65.7 (22.0)
Nausea and vomiting symptom score	92.5 (12.7)	94.9 (10.1)	89.6 (15.1)	89.9 (14.5)	93.2 (11.8)	93.8 (10.8)	92.1 (11.3)	92.4 (10.9)
Activity limitation symptom score	93.2 (11.4)	95.3 (9.5)	87.7 (18.5)	88.3 (16.6)	92.6 (13.6)	93.4 (13.2)	91.2 (11.9)	91.3 (11.9)
Constipation and swallowing symptom score	81.2 (18.7)	81.0 (18.9)	78.3 (20.4)	78.8 (19.8)	80.3 (19.9)	80.5 (19.8)	84.3 (17.9)	84.5 (18.0)
SF-36 scores, mean (SD)[a]								
Norm-based physical functioning	48.4 (9.6)	48.3 (9.9)	48.2 (11.1)	48.4 (11.0)	49.3 (10.6)	49.8 (10.2)	47.0 (12.0)	46.9 (12.1)
Norm-based role physical	47.3 (11.9)	47.3 (12.4)	47.7 (11.8)	47.9 (11.6)	49.7 (10.1)	50.1 (9.9)	47.9 (10.0)	47.8 (10.0)
Norm-based bodily pain	46.3 (10.3)	47.3 (11.1)	46.2 (10.9)	46.6 (10.7)	47.9 (10.5)	48.0 (10.5)	48.0 (8.8)	47.9 (8.9)
Norm-based general health	44.1 (10.3)	44.9 (10.6)	43.2 (11.5)	43.4 (11.4)	47.0 (10.8)	47.2 (10.6)	43.3 (9.2)	43.2 (9.3)
Norm-based vitality	45.3 (11.2)	46.0 (11.5)	46.4 (12.0)	46.5 (12.0)	47.7 (11.9)	47.8 (11.9)	45.2 (10.9)	45.2 (11.0)
Norm-based social functioning	47.4 (11.7)	48.0 (12.1)	47.0 (12.0)	47.3 (11.7)	48.4 (11.4)	48.7 (11.3)	48.2 (10.4)	48.1 (10.4)
Norm-based role emotional	48.6 (11.8)	49.0 (11.7)	47.5 (12.6)	47.7 (12.2)	49.8 (10.0)	50.1 (9.7)	48.2 (10.6)	48.1 (10.6)
Norm-based mental health	47.7 (11.9)	48.8 (11.8)	48.9 (11.5)	49.2 (11.2)	49.4 (11.6)	49.4 (11.6)	47.5 (10.3)	47.4 (10.3)
EQ-5D, mean (SD)[a]	0.774 (0.259)	0.777 (0.281)	0.761 (0.282)	0.770 (0.269)	0.800 (0.253)	0.807 (0.249)	0.794 (0.206)	0.793 (0.208)

a Mean (SD) based on responders completing relevant section of the questionnaire.

Appendix 5 Characteristics of the four randomised controlled trials of laparoscopic fundoplication compared with medical management

Anvari *et al.* trial[44–46]

Methods	**Randomisation:** computerised sequence generation **Allocation concealment:** apparently yes, although blocking used to ensure 1 : 1 randomisation ('blocking factor determined by data centre') **Blinding:** not possible; outcome assessment: at office visit (questionnaires before medical assessment) at 6 and 12 months, by telephone at 3 and 9 months **Follow-up:** 3, 6, 9 and 12 months and 3 years **Setting:** single centre in Canada (four experienced surgeons) **Inclusion criteria:** chronic symptoms of GORD requiring long-term therapy; dependent on PPIs for at least 12 months; adults aged 18–70 years; GORD symptom score of < 18 and a score of > 70 on visual analogue scale (VAS) (0–100) of symptom control at screening; % acid reflux > 4% at baseline **Exclusion criteria:** pregnancy, malignancy, aperistaltic esophagus, severe comorbidity and previous GORD surgery
Participants	**Sample size:** 216 (a priori) **Randomised:** 104; medical: 52 [50 received medication (96%)], surgical: 52 [51 received surgery (98%)] **Age, mean:** medical 42.1 years; surgical 42.9 years **Sex (M/F):** medical 26/26; surgical 29/23
Interventions	**Medical:** optimised PPI as per detailed symptom management algorithm **Surgical:** laparoscopic Nissen fundoplication. Comprised construction of 2.5- to 3-cm 360° wrap. Short gastric vessels divided routinely to achieve floppy wrap
Outcomes	**Primary outcome:** GERSS – includes heartburn, regurgitation, bloating, dysphagia and epigastric/retrosternal pain. Total scale score 0–60. Well controlled defined as score < 18 **Secondary outcomes:** oesophageal function: endoscopy, manometry and 24-hour pH; QoL: SF-36 (0–100), EQ-5D (0–1) and VAS 0–100 for patient satisfaction with symptom control. A score of 70 was considered the threshold for symptom control on the VAS
Type of trial design	On explanatory end of explanatory–pragmatic continuum
Clinical leadership	Upper gastrointestinal surgeon
Risk of bias	
Allocation concealment?	Probably concealed – explanation of randomisation and concealment given in methods, although blocking could have jeopardised this
Free of selective reporting?	One concern: heartburn-free days promoted to primary outcome at 3 years
Sequence generation?	Computerised sequence generation but blocked and size of block not stated
Incomplete outcome data addressed?	Some evidence to suggest differential loss to follow-up at 3 years: 8/52 vs 3/52; no responder analysis
Notes	Trial funded by the Canadian Institute of Health Research and Ontario Ministry of Health

LOTUS trial[47–50]

Methods	**Randomisation:** randomisation in blocks of four **Allocation concealment:** unclear **Blinding:** not possible; outcome assessment: primary outcome (treatment failure) dependent on clinical decision-making, which was not blinded **Follow-up:** 6 months and 1, 3 and 5 years **Setting:** 39 centres across 11 European countries **Inclusion criteria:** oesophagitis grade no more than Los Angeles grade B; GORD symptoms no more than mild; response to PPI in run-in phase **Exclusion criteria:** previous oesophageal, gastric or duodenal surgery; primary oesophageal disorders; inflammatory bowel disorders; any gastrointestinal absorption abnormality; other significant concomitant disease
Participants	**Sample size:** 550 – not clear if stated a priori **Randomised:** 554; medical: 266, surgical: 288 [248 received surgery (86%)] – specialist surgery **Age, mean (SD):** medical 45.4 (11.5) years; surgical 44.8 (10.9) years **Sex (M/F):** medical 199/67; surgical 199/89
Interventions	**Medical:** esomeprazole 20 mg once daily, which could be increased stepwise **Surgical:** laparoscopic anti-reflux surgery. Used crural repair and short floppy total fundoplication in standardised approach
Outcomes	**Primary outcome:** time to treatment failure **Secondary outcomes:** symptoms related to GORD (heartburn, acid regurgitation and dysphagia severity); other gastrointestinal symptoms (flatulence, diarrhoea, epigastric pain, bloating) from GSRS; endoscopy; QoL using QOLRAD; perioperative and postoperative mortality (<30 days); dysphagia requiring further treatment; serious adverse events; rate of conversion to open surgery
Type of trial design	Principally explanatory with some pragmatic features (calls itself 'exploratory')
Clinical leadership	Upper gastrointestinal surgeon
Risk of bias	
Allocation concealment?	Unclear; randomisation in blocks of four, otherwise not reported
Free of selective reporting?	No evidence of selective reporting, although QOLRAD data only reported in supplementary table at 5 years
Sequence generation?	Unclear; randomisation in blocks of four
Incomplete outcome data addressed?	Not fully: follow-up at 3 years: 204/288 vs 208/266; at 5 years: 180/288 (62.5%) vs 192/266 (72.2%). No data on 14% allocated surgery who did not have an operation
Notes	Trial funded by AstraZeneca R&D, with three authors employed by AstraZeneca

Mahon *et al.* trial[51–53]

Methods	**Randomisation:** 'computerised randomisation' – no details **Allocation concealment:** unclear, not reported **Blinding:** not possible **Follow-up:** 3 months and 1 year; separate follow-up of participants from one centre at 7 years **Setting:** two UK centres (two experienced surgeons) **Inclusion criteria:** GORD for at least 6 months, dependent on PPIs for at least 3 months and aged >16 to <70 years **Exclusion criteria:** significant oesophageal dysmotility and morbid obesity (BMI > 35 kg/m^2)
Participants	**Sample size:** a priori apparently 215 although basis not clear **Randomised**: 217; medical: 108, surgical: 109 (apparently all received surgery) **Age, median (range):** medical 47 (35–57) years; surgical 48 (39–56) years **Sex (M : F ratio):** medical 1 : 2.6; surgical 1 : 1.9
Interventions	**Medical:** one of four different PPI regimens, aiming to abolish symptoms **Surgical:** laparoscopic Nissen fundoplication. Used crural repair and short floppy wrap of 3 cm; division of short gastric vessels as deemed necessary
Outcomes	PGWI, GSRS, dysphagia, DeMeester score, operation time, length of stay, conversion to open surgery, reoperation rate, mortality rate, lower oesophageal sphincter pressure, postoperative complications, % time pH <4, cost, patient satisfaction only at 7 years (scale 1–3)
Type of trial design	At explanatory end of explanatory–pragmatic continuum
Clinical leadership	Upper gastrointestinal surgeon
Risk of bias	
Allocation concealment?	Unclear, not reported
Free of selective reporting?	Unclear, primary outcome not clearly prespecified
Sequence generation?	'Computerised randomisation'
Incomplete outcome data addressed?	Among 108 in medical group, well-being scores were available for 108 at baseline and 96 at one year; equivalent figures among 109 in surgical group were 104 and 99, respectively
Notes	Trial partially funded by Jansen Pharmaceutics; economic evaluation funded by Ethicon Endo-Surgery. All participants in medical group offered surgery at 1 year: 54/92 (59%) underwent surgery

REFLUX trial[1–3]

Methods	**Randomisation:** computer-generated sequence **Allocation concealment:** yes **Blinding:** not possible; outcome assessment by patient-completed postal questionnaires **Follow-up:** 3 months and annually for 5 years **Setting:** 21 UK centres **Inclusion criteria:** GORD symptoms for >12 months requiring PPI; evidence of GORD (endoscopy and/or pH monitoring) **Exclusion criteria:** BMI >40 kg/m^2; Barrett's esophagus >3 cm; paraoesophageal hernia; oesophageal stricture
Participants	**Sample size:** 600 (sample size recalculated from 600 to 392 after advice from DMC) **Randomised:** 357; medical: 179, surgical: 178 [111 received surgery (62%)] – by, or supervised by, experienced surgeon **Age, mean (SD):** medical 45.9 (11.9) years; surgical 46.7 (10.3) years **Sex (M/F):** medical 120/59; surgical 116/62
Interventions	**Medical:** best medical management after review. Lansoprazole was predominant PPI at study entry; omeprazole and lansoprazole most commonly reported at follow-up **Surgical:** laparoscopic surgery. Type of fundoplication was left to discretion of surgeon and all surgical techniques considered as a single policy
Outcomes	**Primary outcome:** REFLUX questionnaire score (heartburn, acid reflux, wind, eating and swallowing, bowel movements, sleep, work, physical and social activity) **Secondary outcomes:** QoL: EQ-5D and SF-36; serious morbidity; mortality; patient costs; NHS costs
Type of trial design	Pragmatic on explanatory–pragmatic continuum. Also included parallel, non-randomised preference groups
Clinical leadership	Upper gastrointestinal surgeon and gastroenterologist partnerships
Risk of bias	
Allocation concealment?	Allocation conducted by trials unit independent of all clinical teams
Free of selective reporting?	ITT and PP analysis presented as prespecified
Sequence generation?	Computerised randomisation
Incomplete outcome data addressed?	Adjusted treatment received and PP analyses reported in addition to ITT. Follow-up at 12 months: 154/178 (87%) vs 164/179 (92%)
Notes	Trial funded by NIHR HTA programme

Appendix 6 Search strategies for economic evaluation review

Gastro-oesophageal reflux disease search terms used in a recent Cochrane Review and adapted for use in the systematic review described in *Chapter 5*.[57]

Economic evaluation search

The Cochrane Library (includes NHS Economic Evaluation Database)

http://onlinelibrary.wiley.com/o/cochrane/cochrane_search_fs.html
Searched: 19 April 2011.

#1 MeSH descriptor Gastroesophageal Reflux explode all trees (1356)
#2 (gastroesophageal near/3 reflux):ti,ab,kw (1764)
#3 (gastro near/3 oesophageal near/3 reflux):ti,ab,kw (657)
#4 (gastro near/3 esophageal near/3 reflux):ti,ab,kw (657)
#5 (gord):ti,ab,kw (103)
#6 (gerd):ti,ab,kw (413)
#7 MeSH descriptor Duodenogastric Reflux explode all trees (50)
#8 (duodenogastric near/3 reflux):ti,ab,kw (58)
#9 MeSH descriptor Bile Reflux explode all trees (22)
#10 (bile near/3 reflux):ti,ab,kw (78)
#11 (acid near/3 reflux):ti,ab,kw (281)
#12 MeSH descriptor Dyspepsia explode all trees (864)
#13 (dyspep*):ti,ab,kw (2165)
#14 (belch* or burp*):ti,ab,kw 100
#15 MeSH descriptor Eructation explode all trees (18)
#16 (eructation):ti,ab,kw (52)
#17 MeSH descriptor Heartburn explode all trees (255)
#18 (heartburn or indigestion):ti,ab,kw (985)
#19 MeSH descriptor Esophagitis explode all trees (583)
#20 (esophagitis or oesophagitis):ti,ab,kw (1273)
#21 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20), from 2005 to 2011 (1367)

1367 results made up of:

- Economic evaluations (NHS EED): 85
- Cochrane reviews (CDSR): 54
- Other systematic reviews (DARE): 59
- Technology assessments (HTA): 13
- Clinical trials (Cochrane Central Register of Controlled Trials, CENTRAL): 1147
- Methods studies (Cochrane Methodology Register): 9.

A total of 10 of the CDSR records were pre 2005 and so they were deleted. In addition to the 85 NHS EED records, all CDSR, DARE and HTA records were also saved to EndNote library reflux.enl (marked CDSR, DARE, HTA or NHS EED in the Custom 4 field) in case they are useful (Thomson Reuters, CA, USA).

Quality-of-life searches

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations

Ovid MEDLINE(R) and Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations – 1948 to present.
Searched: 20 April 2011 via OVID interface.

1. exp gastroesophageal reflux/ (18,908)
2. (gastroesophageal adj3 reflux).tw. (11,313)
3. (gastro adj3 oesophageal adj3 reflux).tw. (3146)
4. (gastro adj3 esophageal adj3 reflux).tw. (898)
5. gord.tw. (562)
6. gerd.tw. (4147)
7. exp duodenogastric reflux/ (1511)
8. (duodenogastric adj3 reflux).tw. (813)
9. exp bile reflux/ (649)
10. (bile adj3 reflux).tw. (895)
11. (acid adj3 reflux).tw. (2044)
12. exp dyspepsia/ (6549)
13. dyspep$.tw. (9171)
14. (belch$ or burp$).tw. (771)
15. exp eructation/ (257)
16. eructation.tw. (173)
17. exp heartburn/ (1395)
18. (heartburn or indigestion).tw. (3760)
19. exp esophagitis/ (8478)
20. (esophagitis or oesophagitis).tw. (10,081)
21. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 (44,775)
22. exp life tables/ (10,574)
23. "quality of life"/ (89,134)
24. health status/ (47,125)
25. exp health status indicators/ (150,921)
26. (utilit$ approach$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab. (1099)
27. (health measurement$ scale$ or health measurement$ questionnaire$).ti,ab. (31)
28. (standard gamble$ or categor$ scal$ or linear scal$ or linear analog$ or visual scal$ or magnitude estimat$).ti,ab. (3759)
29. (time trade off$ or rosser$ classif$ or rosser$ matrix or rosser$ distress$ or hrqol).ti,ab. (5160)
30. (index of wellbeing or quality of wellbeing or qwb).ti,ab. (150)
31. (rating scale$ or multiattribute$ health ind$ or multi attribute$ health ind$).ti,ab. (26,439)
32. (health utilit$ index or health utilit$ indices).ti,ab. (484)
33. (multiattribute$ theor$ or multi attribute$ theor$ or multiattribute$ analys$ or multi attribute$ analys$).ti,ab. (9)
34. (health utilit$ scale$ or classification of illness state$ or 15d or 15 d or 15 dimension).ti,ab. (2878)
35. (health state$ utilit$ or 12d or 12 d or 12 dimension).ti,ab. (2009)
36. well year$.ti,ab. (20)
37. (multiattribute$ utilit$ or multi attribute$ utilit$).ti,ab. (152)
38. health utilit$ scale$.ti,ab. (7)
39. (qol or 5d or 5-d or 5 dimension or quality of life or eq-5d or eq5d or eq 5d or euroqol).ti,ab. (114,629)
40. (qualy or qaly or qualys or qalys or quality adjusted life year$).ti,ab. (4631)
41. life year$ gain$.ti,ab. (1289)
42. willingness to pay.ti,ab. (1517)
43. (hye or hyes or health$ year$ equivalent$).ti,ab. (58)

44. (person trade off$ or person tradeoff$ or time tradeoff$ or time trade off$).ti,ab. (776)
45. theory utilit$.ti,ab. (7)
46. life table$.ti,ab. (6627)
47. health state$.ti,ab. (2838)
48. (sf36 or sf 36).ti,ab. (9654)
49. (short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).ti,ab. (4429)
50. (6d or 6-d or 6 dimension).ti,ab. (4590)
51. or/22-50 (366,981)
52. 21 and 51 (3337)
53. limit 52 to yr="2005 - 2011" (1726)

A total of 1726 results saved to EndNote library reflux.enl (marked MEDLINE in the Custom 4 field).

EMBASE

EMBASE – 1996 to week 15 2011.
Searched: 20 April 2011 via OVID interface.

1. exp gastroesophageal reflux/ (24,724)
2. (gastroesophageal adj3 reflux).tw. (10,538)
3. (gastro adj3 oesophageal adj3 reflux).tw. (2717)
4. (gastro adj3 esophageal adj3 reflux).tw. (867)
5. gord.tw. (670)
6. gerd.tw. (5466)
7. exp duodenogastric reflux/ (993)
8. (duodenogastric adj3 reflux).tw. (292)
9. exp bile reflux/ (509)
10. (bile adj3 reflux).tw. (505)
11. (acid adj3 reflux).tw. (1939)
12. exp dyspepsia/ (15,437)
13. dyspep$.tw. (7897)
14. (belch$ or burp$).tw. (706)
15. exp eructation/ (315)
16. eructation.tw. (100)
17. exp heartburn/ (5566)
18. (heartburn or indigestion).tw. (3660)
19. exp esophagitis/ (12,163)
20. esophagitis.tw. (6246)
21. oesophagitis.tw. (1696)
22. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 (51,561)
23. life tables/ (1840)
24. exp "quality of life"/ (158,965)
25. health status/ (53,662)
26. health survey/ (104,329)
27. (utilit$ approach$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab. (1100)
28. (health measurement$ scale$ or health measurement$ questionnaire$).ti,ab. (35)
29. (standard gamble$ or categor$ scal$ or linear scal$ or linear analog$ or visual scal$ or magnitude estimat$).ti,ab. (2771)
30. (time trade off$ or rosser$ classif$ or rosser$ matrix or rosser$ distress$ or hrqol).ti,ab. (6267)
31. (index of wellbeing or quality of wellbeing or qwb).ti,ab. (139)
32. (rating scale$ or multiattribute$ health ind$ or multi attribute$ health ind$).ti,ab. (25,950)
33. (health utilit$ index or health utilit$ indices).ti,ab. (550)

34. (multiattribute$ theor$ or multi attribute$ theor$ or multiattribute$ analys$ or multi attribute$ analys$).ti,ab. (9)
35. (health utilit$ scale$ or classification of illness state$ or 15d or 15 d or 15 dimension).ti,ab. (2773)
36. (health state$ utilit$ or 12d or 12 d or 12 dimension).ti,ab. (1630)
37. well year$.ti,ab. (7)
38. (multiattribute$ utilit$ or multi attribute$ utilit$).ti,ab. (132)
39. health utilit$ scale$.ti,ab. (5)
40. (qol or 5d or 5-d or 5 dimension or quality of life or eq-5d or eq5d or eq 5d or euroqol).ti,ab. (131,090)
41. (qualy or qaly or qualys or qalys or quality adjusted life year$).ti,ab. (5406)
42. life year$ gain$.ti,ab. (1495)
43. willingness to pay.ti,ab. (1724)
44. (hye or hyes or health$ year$ equivalent$).ti,ab. (40)
45. (person trade off$ or person tradeoff$ or time tradeoff$ or time trade off$).ti,ab. (785)
46. theory utilit$.ti,ab. (7)
47. life table$.ti,ab. (3428)
48. health state$.ti,ab. (2931)
49. (sf36 or sf 36).ti,ab. (12,136)
50. (short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).ti,ab. (5063)
51. (6d or 6-d or 6 dimension).ti,ab. (3389)
52. or/23-51 (364,882)
53. 22 and 52 (4601)
54. limit 53 to yr="2005 - 2011" (2906)

A total of 2906 results saved to EndNote library reflux.enl (marked EMBASE in the Custom 4 field).

Results of literature search

EndNote library records were deduplicated as far as possible.

Source	Results	Results after deduplication
NHS EED	85	85
CDSR	44	44
DARE	59	56
HTA	13	12
MEDLINE	1726	1640
EMBASE	2906	1825
Total	4833	3662

Appendix 7 Within-trial cost-effectiveness analysis: health-related quality-of-life and cost-effectiveness results

TABLE 53 Within-trial cost-effectiveness analysis: health-related quality-of-life and cost-effectiveness results

Study	Grant *et al*. 2008[1]	Goeree *et al*. 2011[46]
Trial	REFLUX (multicentre UK)	Anvari (single centre in Canada)
Follow-up	Within-trial cost-effectiveness analysis over 1 year	Within-trial cost-effectiveness analysis over 3 years
Number of patients	318[a]	104
Perspective	UK NHS	Societal perspective
Price year	2006 UK pounds	2009 Canadian dollars (2010 tested in sensitivity analysis)
HRQoL instrument	EQ-5D	HUI (primary instrument); SF-6D and EQ-5D (tested in sensitivity analysis) QoL improved over time across all utility instruments; however, the QALYs gained estimated with EQ-5D were less than half of those estimated with HUI3 and SF-6D
Difference in mean QALYs	0.066 (95% CI 0.026 to 0.107)	0.109 (SD 0.784)
Difference in mean costs	£1280 (£1054 to £1468)	C$3205 (SD C$16,828)
ICER	£19,000 per QALY gained	C$29,400 per QALY gained (utilities from HUI3); C$76,310 per QALY gained (utilities from EQ-5D)
Probability of surgery being cost-effective	When k = £20,000, probability = 46%; when k = £30,000, probability = 86%	Laparoscopic Nissen fundoplication has the highest probability of being the most cost-effective treatment when k is >C$30,000

a The REFLUX economic analysis included both ITT and PP analysis. Results presented in this table are based on the ITT analysis.

Appendix 8 Validation of the multiple imputation

TABLE 54 Predictors of missingness at the 95% confidence level

	Predictors of missingness ($p < 0.05$)		
Follow-up	Variable	Coefficient	Pseudo-R^2
Year 1	EQ-5D at baseline	2.2842	0.0673
	EQ-5D at 3 months	–3.7987	
Year 2	EQ-5D at baseline	1.4209	0.0230
Year 3	EQ-5D at baseline	–3.4594	0.1681
	EQ-5D at 3 months	2.7446	
	EQ-5D at year 2	2.0889	
Year 4	–[a]	–[a]	0.0288
Year 5	EQ-5D at baseline	–7.4267	0.1358
	EQ-5D at year 3	3.1675	

a For year 4, no coefficient was statistically significant at the 95% confidence level. Pseudo-R^2 obtained with constant only.

Note: Only coefficients for the variables significant at the 95% confidence level are shown, despite all models tested including a similar set of variables: demographics (age, sex, BMI), ITT allocation, PP status, costs for the previous years and EQ-5D scores for the previous follow-up points.

The existence of predictors for missingness at the 95% confidence level indicates that data may not be MCAR and therefore that the multiple imputed data set is more reliable than the complete case.

Figures 24 and *25* compare the distribution of total costs and total QALYs, respectively, across the first 10 imputed data sets and the original data (imputation number 0). The distribution is similar, providing some assurance that the multiple imputation strategy was successful.

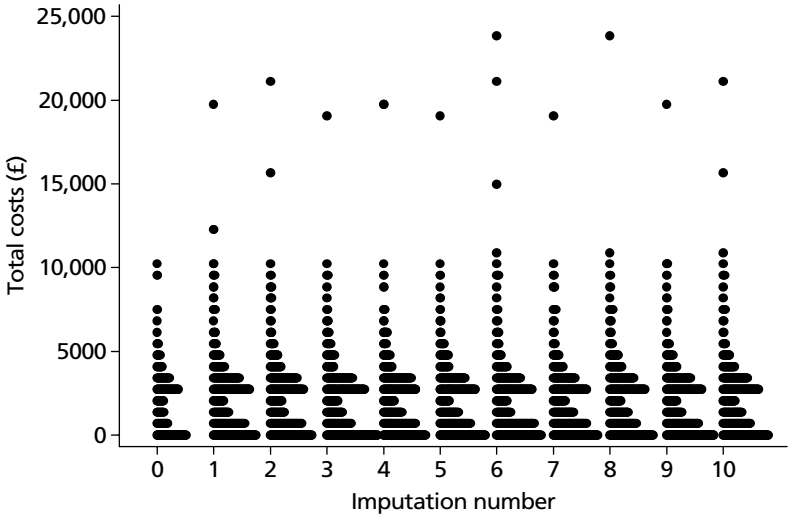


FIGURE 24 Distribution of total costs across the first 10 imputed data sets and for the original data set (imputation number 0).

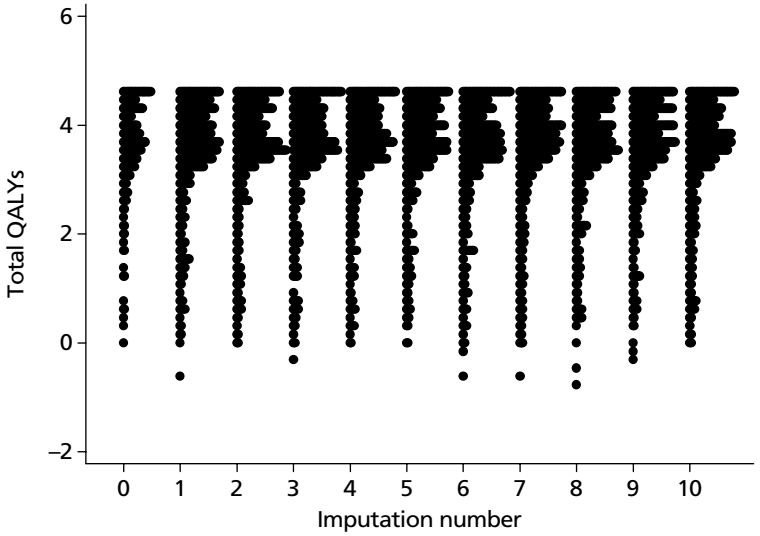


FIGURE 25 Distribution of total QALYs across the first 10 imputed data sets and for the original data set (imputation number 0).

Appendix 9 Costs and health-related quality of life for allocation according to per protocol at 1 year: structural sensitivity analysis

TABLE 55 Total mean costs and total mean QALYs for the medical management and surgery groups according to ITT and PP at 1 year for the complete case

		PP at 1 year		
	Treatment allocated (ITT)	Medical management	Surgery	Total
Total costs	Medical management	£1201.61	£3718.18	£1316.00
Total QALYs		3.5665	2.6076	3.5229
Number of patients		84	4	88
Total costs	Surgery	£989.06	£3525.38	£2981.89
Total QALYs		3.7016	3.7447	3.7354
Number of patients		18	66	84
Total costs	Total	£1164.10	£3536.40	£2129.57
Total QALYs		3.5904	3.6797	3.6268
Number of patients		102	70	172

Note: cells highlighted by shading refer to PP at 1 year groups considered for the incremental analysis.

TABLE 56 Health-related quality of life (EQ-5D) according to PP analysis

Completed questionnaires returned at each time point			Mean (SD) EQ-5D		
Surgery ($n = 178$[a])	Medical management ($n = 179$[a])	Follow-up	Surgery	Medical management	Difference in mean EQ-5D (surgery – medical management) (95% CI)[b,c]
108	162	Baseline	0.7184 (0.2394)	0.7266 (0.2553)	–0.0082 (–0.0691 to 0.0528)
108	143	3 months	0.8059 (0.2393)	0.6910 (0.3068)	0.1148 (0.0446 to 0.1851)
102	153	Year 1	0.7773 (0.2323)	0.7064 (0.2703)	0.0709 (0.0065 to 0.1353)
83	129	Year 2	0.7903 (0.2442)	0.7170 (0.3133)	0.0733 (–0.0068 to 0.1532)
89	127	Year 3	0.7897 (0.2521)	0.7563 (0.2492)	0.0336 (–0.0356 to 0.1018)
88	122	Year 4	0.7785 (0.2636)	0.7550 (0.2678)	0.0236 (–0.0498 to 0.0969)
87	110	Year 5	0.7771 (0.2812)	0.7654 (0.2782)	0.0117 (–0.0674 to 0.0908)

a *n* refers to number of patients originally randomised to each trial arm.
b CIs estimated using OLS regression.
c Unadjusted for baseline EQ-5D.

TABLE 57 Costs associated with resource use for PP analysis

Returned questionnaires in each year			Mean (SD) resource-use cost (£) according to PP at 1 year		Incremental mean cost (surgery–medical management) (95% CI[a]) (£)
Surgery	Medical management	Year	Surgery	Medical management	
104	154	1[b]	3241.78 (1263.80)	361.28 (668.12)	2880.50 (2634.08 to 3126.91)
86	133	2	82.57 (373.46)	159.80 (366.76)	–77.23 (–177.98 to 23.51)
92	128	3	79.80 (349.15)	289.27 (913.00)	–209.47 (–406.78 to –12.15)
88	124	4	96.49 (368.53)	314.70 (1362.96)	–218.21 (–512.14 to 75.72)
90	112	5	43.17 (133.98)	233.60 (645.33)	–190.43 (–326.93 to –53.93)
Cost category					
Surgery in year 1			2780.73 (1756.83)	0[c]	2780.73 (2704.69 to 2856.76)
Reflux-related hospital night admissions			403.07 (1305.30)	315.49 (833.69)	87.58 (–242.21 to 417.37)
Reflux-related hospital day admissions			159.78 (423.32)	231.64 (608.67)	–71.86 (–236.56 to 92.84)
Reflux-related GP visits			130.93 (940.68)	193.31 (465.70)	–62.37 (–176.00 to 51.26)
Medication			87.69 (217.69)	383.01 (525.63)	–295.32 (–424.17 to –166.47)

a CIs estimated using OLS regression.
b Refers to the patients who returned the final year 1 questionnaire.
c By definition, none of the medical management patients in PP underwent surgery.

Appendix 10 Protocol

THE PLACE OF MINIMAL ACCESS SURGERY AMONGST PEOPLE WITH GASTRO-OESOPHAGEAL REFLUX DISEASE (GORD)

A UK COLLABORATIVE STUDY FUNDED BY THE NHS R&D HTA PROGRAMME

PROTOCOL

VERSION 6a - November 2003

THE PLACE OF MINIMAL ACCESS SURGERY AMONGST PEOPLE WITH GASTRO-OESOPHAGEAL REFLUX DISEASE (GORD) A UK COLLABORATIVE STUDY

(Known as the REFLUX Trial)

PROTOCOL SUMMARY

AIM	To identify the optimal place within the NHS for minimal access surgery amongst people with GORD, whose symptoms would otherwise be managed with long-term medical therapy.
DESIGN	Multicentre, pragmatic randomised trial (with parallel non-randomised preference groups).
PATIENT ELIGIBILTY	• Documented evidence of GORD (endoscopy and/or manometry/24h pH monitoring) • Symptoms for more than 12 months and currently requiring maintenance proton pump inhibitor (PPI) therapy for reasonable symptom control • Received care from a participating clinician • Suitable for either policy (ASA grade I or II) • Recruiting doctor uncertain which management policy is better • Give informed consent to either random allocation of management or follow-up after preferred management
RECRUITMENT	Based on surgeon-physician 'partnership' in at least 15 centres.
INTERVENTIONS	A laparoscopic surgery based policy compared with a continued medical management policy.
OUTCOME MEASUREMENT	**Primary –** Disease specific outcome and NHS costs **Secondary –** Patient costs and Health-related quality of life (EQ5D, SF36)
ORGANISATION	• All whole-hearted contributors part of the GORD Trialist Group (with group authorship of main reports) • Conduct overseen by Steering Group • Trial Office in Aberdeen responsible for day-to-day non-clinical co-ordination • Sessional research nurses in each clinical centre • Health economic evaluation and outcome measure assessment jointly led from York and Aberdeen
FUNDING	NHS R&D Health Technology Assessment Programme

TABLE OF CONTENTS

See Grant *et al.*[1] for details of appendices.

1. OUTLINE OF THE TRIAL

Aim

The aim is to identify the optimal place within the NHS of minimal access surgery amongst people with gastro-oesophageal reflux disease (GORD). Its focus is people whose symptoms would otherwise be managed with long-term medical therapy. The background and justification are summarised in Appendix I.

Objectives

- To evaluate the clinical effectiveness, cost-effectiveness, and safety of a policy of relatively early laparoscopic surgery compared with continued medical management amongst people with GORD judged suitable for both policies.
- To explore factors which may influence the relative performance of the two policies, such as patient preference, surgeon experience, pre-enrolment symptoms and signs, underlying pathology, type of operative procedure used or choice of therapy, and time since surgery.
- To explore the impact that various policies for using laparoscopic surgery would have on the NHS and society in respect of the costs or savings that they would imply for (a) those providing surgical care (in secondary care settings), (b) those providing long-term medical management (usually in primary care settings), and (c) those with GORD.

Design

The study will have two complementary components:

A A randomised trial (with parallel non-randomised preference groups) comparing a laparoscopic surgery based policy with a continued medical management policy to assess their relative clinical effectiveness.

B An economic evaluation of laparoscopic surgery for GORD to compare the cost-effectiveness of the two management policies, to identify the most efficient provision of future care, and to describe the resource impact that various policies for fundoplication would have on the NHS.

The rationale for the study design is described in Appendix II.

2. THE RANDOMISED TRIAL (WITH PARALLEL PREFERENCE GROUPS)

Centre eligibility

Clinical centres will be based on local partnerships between surgeons with experience of laparoscopic fundoplication and the gastroenterologists, with whom they share the secondary care of patients with GORD. Centres will be eligible if they include:

1. a surgeon who has performed at least 50 laparoscopic fundoplication operations
2. one or more gastroenterologists who agree to collaborate with the surgeon in the trial.

Patient eligibility

Inclusion criteria

1. Documented evidence of GORD (based on endoscopy and/or manometry/24hr pH monitoring)
2. Symptoms for more than 12 months and currently requiring maintenance proton pump inhibitor (PPI) therapy for reasonable symptom control (Patients who are intolerant to PPIs and therefore require Histamine Receptor Antagonists (H_2RAs) therapy to control their symptoms will also be included)
3. Care provided by a participating clinician
4. Suitable for either policy (including ASA grade I or II)
5. Recruiting doctor uncertain which management policy is better
6. Informed consent either to random allocation of management or to follow-up after preferred management

Exclusion criteria

1. Morbidly obese (BMI >40 kg/m^2)
2. Barrett's oesophagus of more than 3 cm or have evidence of dysplasia
3. Paraoesophageal hernia
4. Oesophageal stricture

Although there is no formal age limit, it will be younger patients with GORD who will be eligible, who are expected to be aged between 18 and 65 years .

Health technology policies being compared

Laparoscopic surgery policy:

Most of those allocated to this policy will have surgery. Deferring or declining will remain an option, however, even after trial entry, particularly amongst those recruited by a gastroenterologist and referred to a surgeon for consideration of surgery within the trial. Participants who have not had manometry/pH studies will undergo these tests before surgery to exclude achalasia.

The surgery will be performed either by a surgeon who has undertaken more than 50 laparoscopic fundoplications or by a less experienced surgeon working under the supervision of an experienced surgeon. It is recommended that crural repair be routine and non-absorbable, synthetic sutures (not silk) be used for the repair. The type of fundoplication used will be left to the discretion of an experienced surgeon. For the purposes of the main comparisons, the different surgical techniques for laparoscopic fundoplication will be considered as parts of a single policy. The study design will, however, allow indirect comparisons between techniques.

It is expected that local policies for thromboembolism prophylaxis will include a suitable anticoagulant (such as heparin or tinzaparin) plus surgical stockings or pneumatic compression.

Medical therapy policy:

Most of those allocated to the medical therapy policy will continue 'best medical management' (appropriate PPI), as recommended by the clinician responsible for care. Management should conform to the principles of the Genval Workshop Report (see Appendix III). While all the recommendations of this workshop cannot be summarised here, they include stepping down antisecretory medication in most patients to the lowest dose that maintains acceptable symptom control. Patients who have had severe oesophagitis should not be managed on the basis of symptoms alone, however. While it is expected that most trial participants allocated medical management will continue to be managed in this way, surgery should be considered if a clear indication for it subsequently develops.

Outcome measurements

Primary:

a 'Disease-specific' outcome to include the need for changes in treatment, reflux and other gastro-intestinal symptoms, and the side effects and complications of both therapies.

b NHS costs including treatments, investigations, consultations and other contacts with the health service.

Secondary:

c Health-related quality of life – EQ5D and SF36.

d Patient costs including loss of earnings, reduction in activities, and the cost of prescriptions and travel to health care.

Other:

e Other serious morbidity, such as operative complications

f Mortality

The instrument for collecting this information are in Appendix IV. The ways in which these data will be displayed in the final report are illustrated in Appendix V.

Sample size and statistical analysis

A sample size of 600 will identify a difference between the two randomised groups of less than 0.25 of the standard deviation of the disease-specific instrument, EQ5D or SF36 with 80% power using a significance level of 5%. Based on the same arguments, about 300 people will be recruited to each arm of the preference study.

The cost savings of a surgical policy will largely depend on the number of patients managed surgically who no longer require PPI treatment. A trial with 300 surgically managed patients will estimate this proportion to within about 5% with 95% statistical confidence.

A single principal analysis is planned within the current time frame when all participants have been followed-up for at least 12 months after surgery (or an equivalent time if managed medically). Standard statistical techniques will be used with analysis by intention to treat and 95% confidence intervals. Secondary analyses will explore differential effects within pre-stated sub-groups, characterised by initial symptom severity and surgeon's preferred operative procedure; 99% confidence intervals will be

generated for such analyses to reflect their exploratory nature. The issue of continued surgeon 'learning' will also be investigated using curve fitting techniques.

3. THE ECONOMIC EVALUATION

The economic evaluation is described in detail in Appendix VI. It will have three components: a within-trial cost-effectiveness study; a detailed assessment of the preferences of patients with GORD; and an outside-trial cost-effectiveness analysis based on decision modelling.

4. PRACTICAL ARRANGEMENTS

Each clinical centre will be supported by a part-time research nurse.

Identification of potential participants

Potential participants will be identified in three ways:

- Retrospective case-note review
- Prospective identification of current case
- Referral from general practice

These are summarised in Figure 1. The actual approach used will vary between centres, but case note review is likely to be the principal method.

As a general rule, potentially eligible participants will be booked for an outpatient appointment. They will be sent a brief letter, together with a copy of the information leaflets in advance, letting them know that the trial is likely to be discussed with them (Appendix VII). At the appointment, the clinician will review the person's symptoms and current treatment regimen, and assess eligibility for the trial following the completion of a Patient Assessment Form (Appendix VIII). If eligibility is confirmed, the person will be invited to see the research nurse who will describe the study and discuss any issues that arise. This is summarised in Figure 2. The nurse will also give a supplementary information leaflet that describes the operation in more detail (Appendix IX). Information will also be sent to the general practitioner (GP) in case the participant consults them to discuss the trial (Appendix X); a specific clinic letter will follow from the consultant.

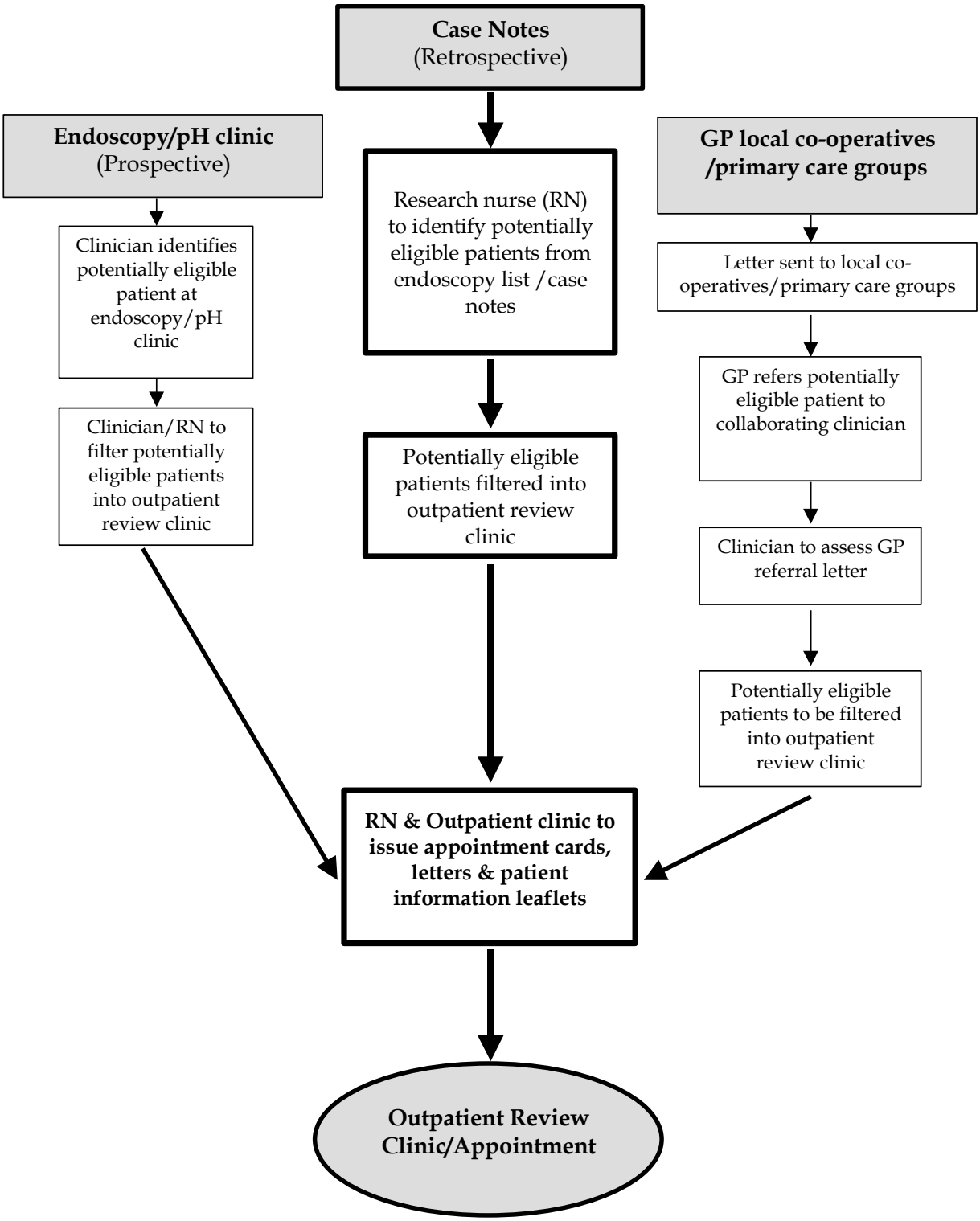


Figure 1. Flowchart describing sources for patient identification

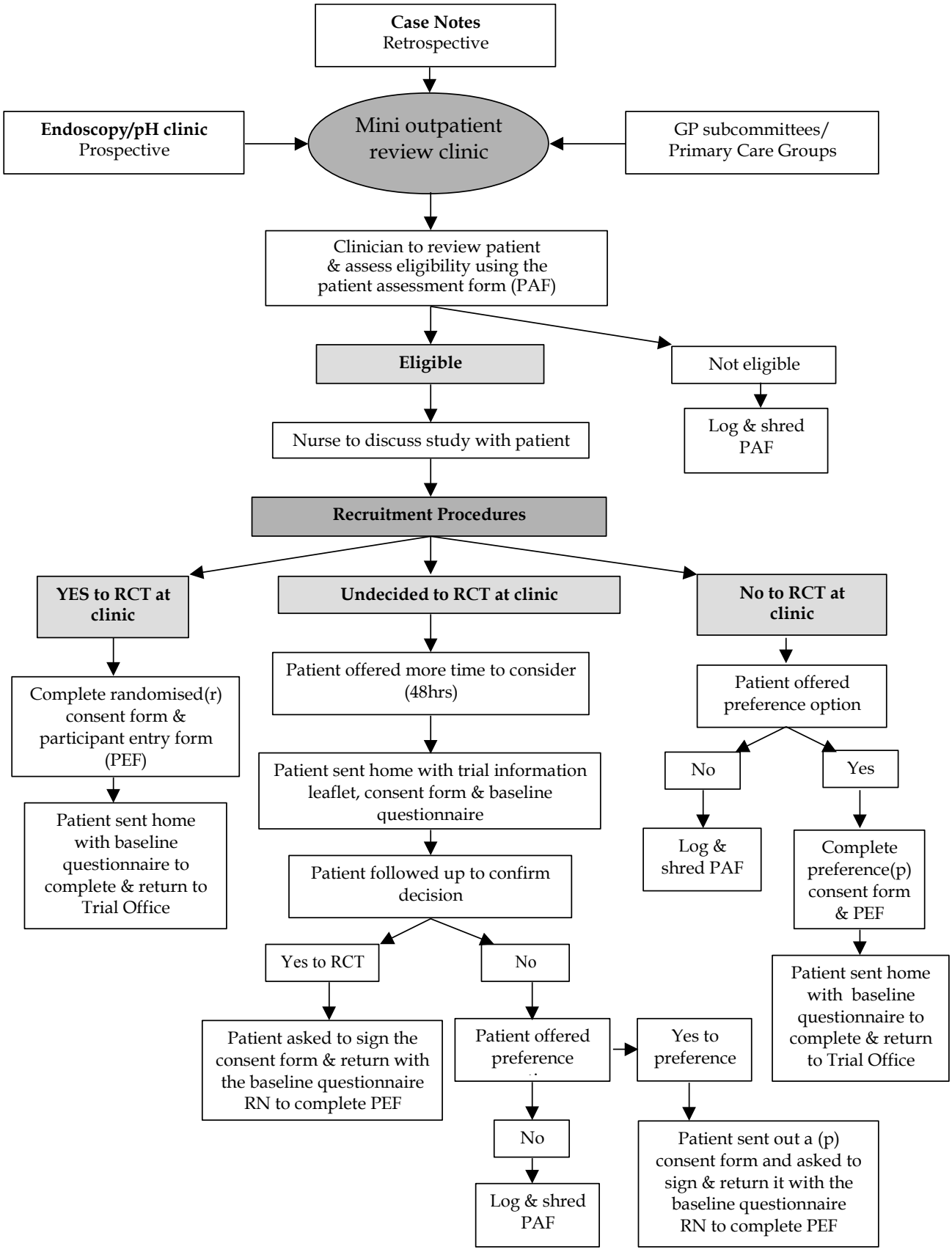


Figure 2. Flowchart describing patient recruitment

Consent to participate

The randomised trial:

Some potential participants will make a decision about participation at this appointment. Those who wish to participate in the randomised trial will be asked to sign a consent form (Appendix XI). On this, they will confirm that they have been given the information they require and that the study has been explained to them. They will also confirm that they understand that they will be sent questionnaires from the Trial Office at participant-specific time intervals after joining the study. (This will be at a time equivalent to around three months and 12 months after surgery.) They will also be told that it is anticipated that further follow-up will be performed periodically thereafter for some years.

The preference study:

A person who does not want to take part in the randomised trial because of a strong preference for one type of treatment management will be asked to take part in the preference arm of the study. Those who wish to participate in the preference study will be given a preference information leaflet and asked to sign a consent form (Appendix XII). In addition to the details collected on the randomised consent form, they will confirm their preferred treatment allocation.

Any person who is uncertain will be given at least 48 hours to consider participation. A research nurse will then phone them to find out their decision and make arrangements as appropriate for them to sign a randomised trial or a preference study consent form.

One copy of the consent form will be given to the participant, another will be filed in the patient's hospital case notes, and the third will be posted to the Trial Office.

Information to be collected at trial entry

Once a participant has agreed to join the trial, the research nurse will record basic identifying and descriptive information on a standard form (Appendix XIII). This information will be sent to the Trial Office.

The participant will take home a baseline questionnaire to complete, and will be asked to return it in a pre-paid envelope to the Trial Office.

Study registration (and treatment allocation when randomised)

The entry procedure will distinguish between those who have agreed to randomisation and those who have agreed to participate in the preference part of the study.

The treatment allocation for participants consenting to the randomised arm of the trial will be computer-generated in the Trial Office. The allocation will be stratified by centre, with balance in respect of other key prognostic variables – age (18-50 y or 51-65 y), sex (M or F), and BMI (≤28 or >29 kg/m^2) - by a process of minimisation.

A letter will be sent from the Trial Office to each participant (Appendix XIV), their GP (Appendix XV) and the local research nurse, confirming the treatment allocation and whether they are taking part in the randomised- or preference-arm of the trial. A letter will also be sent to the respective collaborating surgeon or gastroenterologist with respect to the treatment the participant is allocated.

Clinical management

Clinical management will be left to the discretion of the clinician responsible for care. A summary of the different clinical management pathways is illustrated in Figure 3.

Participants who are allocated to the surgical arm, will be invited to a consultation with the collaborating surgeon. (Participants who have not already had manometry/pH studies will be booked to undergo these tests prior to this consultation.) During this consultation, the surgeon will confirm that there is no contra-indication to surgery and discuss the operation in more detail with the participant, before arranging a date for the operation. The intra-operative details will be recorded by the surgeon on specially designed study forms (Appendix XVI).

All other in-hospital data collection will be the responsibility of the local study nurse.

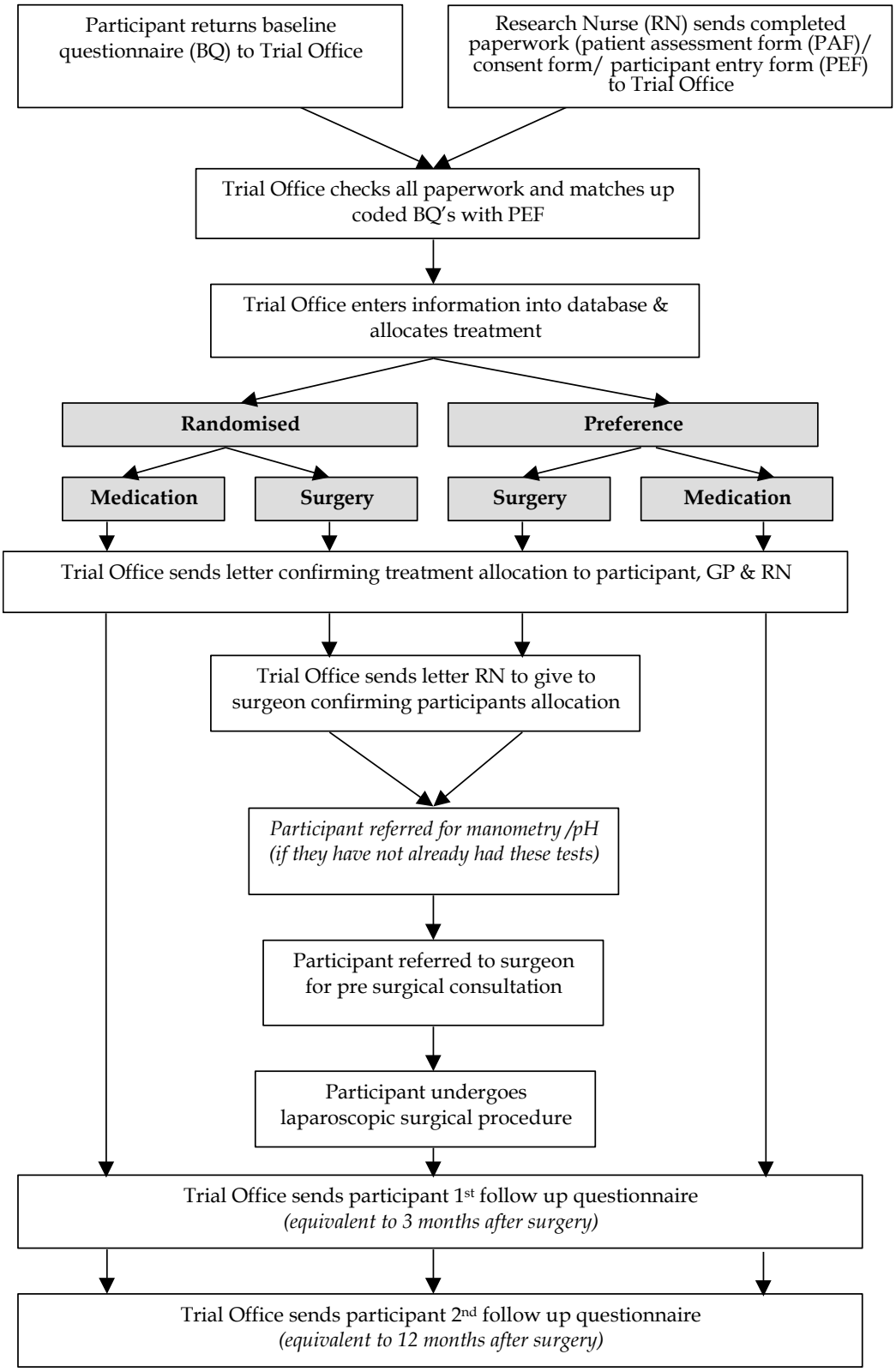


Figure 3. Flowchart showing clinical management post recruitment

Follow-up in the trial

Follow-up by postal questionnaire will be performed twice, at participant-specific time intervals after joining the study. (This will be at a time equivalent to around three and 12 months after surgery). When necessary, clarification of clinical management will be sought through the research nurses (while they are in post) and then subsequently through the recruiting doctor or general practitioner. While it is anticipated that further follow-up will be performed periodically thereafter for some years (dependent on funding being available at that time) these subsequent assessments are not part of this protocol.

Data collection after trial entry

All data will be sent to the Trial Office in Aberdeen for processing. Staff in Aberdeen will work closely with the research nurses to secure as complete and accurate data as possible. A random 10% sample of data will be double-entered to check accuracy. Extensive range and consistency checks will further enhance the quality of the data.

Organisation

Local organisation

The trial is designed to limit the extra work for collaborating clinicians to tasks that only they can do. Research nurses will facilitate the trial locally, and the central organisation will take responsibility for data management and participant follow-up.

Clinical collaborators (gastroenterologist and/or surgeon) will:

1. establish the trial locally (e.g. identifying a 'partnering' clinician or surgeon if not already agreed; facilitating local research ethics committee approval; identifying and appointing a local research nurse; and ensuring that all clinical staff involved in the care of patients with GORD are informed about the trial)
2. take responsibility for clinical aspects of the trial locally (e.g. if any particular concerns emerge)
3. notify the Trial Office of any unexpected clinical events that might be related to trial participation
4. provide support and supervision for all aspects of the work of the local research nurse
5. represent the centre at REFLUX trial collaborators' meetings

Research nurses will:

1. keep local staff informed about the trial and its progress
2. keep regular contact with the local gastroenterologist(s) and surgeon
3. maintain regular contact with the Trial Office
4. identify potential participants and log whether or not they are recruited to the trial (including the preference groups) - with reasons for non-participation
5. arrange for the initial letter of invitation and information leaflet to be sent to potential participants prior to an out-patient assessment.
6. assist the participating clinicians (e.g. at assessment clinics) to give additional information and seek consent to study entry
7. ensure that the baseline data describing participants are collected and sent back to the Trial Office
8. facilitate later follow-up by, for example, helping with local tracing
9. provide support for participants in other ways if there are difficulties
10. represent the centre at trial nurse meetings and collaborators' meetings

5. TRIAL CO-ORDINATION

Trial Offices

The main Trial Office is within the Health Services Research Unit in Aberdeen and gives day-to-day support to the clinical centres. This Office is responsible for all central co-ordination of the trial, including centre and research nurse support, study entry and randomisation, postal follow-up, data processing and statistical analysis.

The economic evaluation and the outcome development work is based in the Centre for Health Economics and the Department of Health Sciences and Clinical Evaluation, respectively, both within the University of York.

The Steering Group

The trial is co-ordinated by a Steering Group (listed in Appendix XVII). The Steering Group, in consultation with the Collaborative Group (see below), will take responsibility for any major decisions, such as the need to close recruitment early to one or more parts of the study or to change the protocol for any reason.

The Collaborative Group

The Collaborative Group is made up of the surgeons, gastroenterologists and research nurses contributing to the trial, members of the Steering Group, and representatives from the Trial Offices.

The Data Monitoring Committee

A data monitoring committee will be established. It will be independent of the trial organisers. During the period of recruitment to the trial, interim analyses will be supplied, in strict confidence, to the data monitoring committee, together with any other analyses that the committee may request. This may include analyses of data from other comparable trials. In the light of these interim analyses, the data monitoring committee will advise the Steering Group if, in its view, the trial has provided both (a) proof beyond reasonable doubt[1] that for all or some types of patients one intervention is clearly indicated in terms of clinical- and cost-effectiveness, and (b) evidence that might reasonably be expected to influence materially the care of people with GORD by clinicians who know the results of this and comparable trials. The Steering Group can then decide to consult the Collaborative Group about whether or not to modify intake into the trial or to report results early. Unless this happens, however, the Steering Group, the Collaborative Group and Trial Offices (except those who supply the confidential analyses) will remain ignorant of the interim results considered by the committee.

The frequency of interim analyses will depend on the judgement of the chairman of the committee, in consultation with the Steering Group.

6. FINANCE

The trial is supported by a grant from the Health Technology Assessment Programme of the NHS Executive Research and Development Programme.

Note:

[1] Appropriate criteria for proof beyond reasonable doubt cannot be specified precisely. A difference of at least three standard deviations in the interim analysis of a major endpoint may be needed to justify halting, or modifying, such a study prematurely. If this criteria were to be adopted, it would have the practical advantage that the exact number of interim analyses would be of little importance, and so no fixed schedule is proposed (Peto R et al *Br J Cancer* 1976; 34: 584-612).

7. A STUDY OF FACTORS IMPACTING ON PATIENTS DECISION TO PARTICIPATE IN THE REFLUX TRIAL (APPENDIX XVIII)

During the recruitment phase of the trial, it is anticipated that a CSO research fellow will undertake supplementary site visits to explore the patients' perspective in relation to trial recruitment. A small number of centres will be purposively selected using qualitative methods (non-participation observation and in-depth interviews). It is proposed that the selected centres will reflect varying recruitment rates.

It is expected that, subject to clinician and patient consent, the research fellow would sit-in and observe reflux clinics where patients are approached to join the study. The researcher would aim to supplement the observational work by interviewing some of the patients (again, subject to consent) about their experience of trial recruitment and factors impacting on their decision to join the trial or not.

It is hoped this small but very useful complementary study nested in the REFLUX trial, will help identify factors impacting on patient recruitment and enable us to look at ways of addressing these issues to facilitate improved future trial recruitment.

8. PUBLICATION

The success of the trial depends entirely on the whole-hearted collaboration of a large number of people. For this reason, chief credit for the trial will be given, not to the committees or central organisers, but to all those who have whole-heartedly collaborated in the trial. The trial's publication policy is described in detail in Appendix XIX. The results of the trial will be reported first to the trial collaborators. The main report will be drafted by the Steering Group, and circulated to all the clinical collaborators for comment. The final version will be agreed by the Steering Group before submission for publication, on behalf of the collaboration. To safeguard the integrity of the study, reports of sub-studies will not be submitted for publication without prior discussion with the Steering Group. Once the main report has been published, a lay summary will be sent to participants who have indicated that they would like to receive one.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library

Laparoscopic Antireflux Surgery vs Esomeprazole Treatment for Chronic GERD

The LOTUS Randomized Clinical Trial

Jean-Paul Galmiche, MD, FRCP
Jan Hatlebakk, MD, PhD
Stephen Attwood, MD, PhD
Christian Ell, MD, PhD
Roberto Fiocca, MD, PhD
Stefan Eklund, MD, PhD
Göran Långström, PhD
Tore Lind, MD, PhD
Lars Lundell, MD, PhD
for the LOTUS Trial Collaborators

Context Gastroesophageal reflux disease (GERD) is a chronic, relapsing disease with symptoms that have negative effects on daily life. Two treatment options are long-term medication or surgery.

Objective To evaluate optimized esomeprazole therapy vs standardized laparoscopic antireflux surgery (LARS) in patients with GERD.

Design, Setting, and Participants The LOTUS trial, a 5-year exploratory randomized, open, parallel-group trial conducted in academic hospitals in 11 European countries between October 2001 and April 2009 among 554 patients with well-established chronic GERD who initially responded to acid suppression. A total of 372 patients (esomeprazole, n=192; LARS, n=180) completed 5-year follow-up.

Interventions Two hundred sixty-six patients were randomly assigned to receive esomeprazole, 20 to 40 mg/d, allowing for dose adjustments; 288 were randomly assigned to undergo LARS, of whom 248 actually underwent the operation.

Main Outcome Measure Time to treatment failure (for LARS, defined as need for acid suppressive therapy; for esomeprazole, inadequate symptom control after dose adjustment), expressed as estimated remission rates and analyzed using the Kaplan-Meier method.

Results Estimated remission rates at 5 years were 92% (95% confidence interval [CI], 89%-96%) in the esomeprazole group and 85% (95% CI, 81%-90%) in the LARS group (log-rank $P=.048$). The difference between groups was no longer statistically significant following best-case scenario modeling of the effects of study dropout. The prevalence and severity of symptoms at 5 years in the esomeprazole and LARS groups, respectively, were 16% and 8% for heartburn ($P=.14$), 13% and 2% for acid regurgitation ($P<.001$), 5% and 11% for dysphagia ($P<.001$), 28% and 40% for bloating ($P<.001$), and 40% and 57% for flatulence ($P<.001$). Mortality during the study was low (4 deaths in the esomeprazole group and 1 death in the LARS group) and not attributed to treatment, and the percentages of patients reporting serious adverse events were similar in the esomeprazole group (24.1%) and in the LARS group (28.6%).

Conclusion This multicenter clinical trial demonstrated that with contemporary antireflux therapy for GERD, either by drug-induced acid suppression with esomeprazole or by LARS, most patients achieve and remain in remission at 5 years.

Trial Registration clinicaltrials.gov Identifier: NCT00251927

JAMA. 2011;305(19):1969-1977 www.jama.com

GASTROESOPHAGEAL REFLUX disease (GERD) is a highly prevalent disorder caused by the reflux of gastric contents into the esophagus. It is a chronic, relapsing disease that negatively affects patients' health-related quality of life and reduces work productivity.[1-3] Consequently, a long-term management plan is required for each individual patient. Maintenance treatment with proton pump inhibitor (PPI) therapy may be an option, offering high rates of symptom resolution and healing of esophagitis.[4-6] However, some patients are reluctant to take long-term medication and may prefer to have antireflux surgery. A number of controlled studies have been undertaken comparing open antireflux surgery and laparoscopic antireflux surgery (LARS)[7] or open antireflux surgery and pharmaceutical treatment,[8,9] but few stud-

See also Patient Page.

CME available online at www.jamaarchivescme.com and questions on p 2018.

Author Affiliations: Department of Gastroenterology and Hepatology, Nantes University, and Centre d'Investigation Clinique INSERM, Nantes, France (Dr Galmiche); Institute of Medicine, Haukeland University Hospital, University of Bergen, Bergen, Norway (Dr Hatlebakk); Department of Surgery, North Tyneside General Hospital, North Shields, Tyne and Wear, England (Dr Attwood); Department of Gastroenterology, Dr Horst Schmidt Hospital, Wiesbaden, Germany (Dr Ell); Department of Surgical and Morphological Sciences, Anatomic Pathology Division, University of Genova, Genova, Italy (Dr Fiocca); AstraZeneca Research and Development, Mölndal, Sweden (Drs Eklund, Långström, and Lind); and Department of Surgery, Karolinska University Hospital, Huddinge, Sweden (Dr Lundell).
A complete list of the LOTUS Trial Collaborators appears at the end of this article.
Corresponding Author: Jean-Paul Galmiche, MD, FRCP, Department of Gastroenterology and Hepatology, Nantes University, CIC INSERM, Place Alexis Ricordeau, 44093 Nantes, France (jeanpaul.galmiche@chu-nantes.fr).

ies have compared pharmaceutical treatment with LARS, particularly over a longer term. Additionally, most of these comparisons included relatively small sample sizes and did not use optimized drug dosing or carefully controlled laparoscopic surgical techniques.[10-13]

The LOTUS (Long-Term Usage of Esomeprazole vs Surgery for Treatment of Chronic GERD) trial compared maintenance therapy provided by esomeprazole (dose-adjusted when required) with standardized LARS in patients who responded well to acid-suppressive therapy. Herein, we report the final results of the 5-year follow-up for the LOTUS trial.

METHODS

Study Design and Objectives

The LOTUS trial was an exploratory randomized, open, parallel-group, multicenter study conducted in 11 European countries between October 2001 and April 2009. The primary objective was to evaluate maintenance therapy with esomeprazole (in patients tested to be responsive to the medication) vs LARS performed by experts.

Patients were aged 18 to 70 years and had chronic symptomatic GERD. The diagnosis of GERD was established on the basis of typical clinical history and presence of esophageal mucosal breaks at endoscopy, classified by Los Angeles grade, and/or pathological pH-metry. Assessments included endoscopy with biopsy, 24-hour pH-metry and symptom response to esomeprazole. The participating centers had to be either academic units or affiliated with a university; each center participated in training sessions to ensure that operative procedures were conducted or supervised by a consultant surgeon who specialized in this type of laparoscopic upper gastrointestinal tract surgery; and surgical techniques were standardized[14] across centers. All patients had to be eligible for both LARS and pharmaceutical therapy and were randomized in blocks of 4 consecutive patient numbers to either treatment. Participants were not permitted to switch treatment groups if they requested the alternative treatment; patients had to leave the study to receive the alternative treatment. Protocol approval for this trial was obtained from local ethics committees. Written informed consent was obtained from all patients.

Because sustained resolution of reflux symptoms with esomeprazole treatment occurs in approximately 70% of patients with GERD,[15] a 3-month run-in period was required to verify the clinical response to esomeprazole, 40 mg/d, and only those who responded were randomized. Partial responders or patients refractory to treatment were excluded. This 3-month run-in period also allowed baseline assessments. Patients were required to have no more than Los Angeles grade B esophagitis at baseline and no more than mild heartburn or regurgitation at the end of 3 months of esomeprazole treatment to permit randomization. Symptom severity was classified as none, mild (awareness of symptoms but easily tolerated), moderate (discomfort sufficient to cause interference with normal activities), or severe (incapacitating, with inability to perform normal activities).

Responders were randomly assigned to undergo LARS or to receive esomeprazole, 20 mg once per day, increased stepwise to 40 mg once per day then 20 mg twice per day in case of incomplete control of heartburn and regurgitation. Full details of the protocol are described in the report of the interim 3-year results.[16] Patients visited the clinic 6 months after randomization and every 6 months thereafter. Follow-up endoscopy was planned at 1, 3, and 5 years. At endoscopy, the esophagus, cardia region, stomach, and duodenum were examined and biopsies were repeated.[17] Patients underwent pH-metry at baseline and again at 6 months and 5 years.[18]

Symptoms related to GERD were assessed at every visit, during which the investigator asked standardized questions about heartburn, acid regurgitation, epigastric pain, bloating, flatulence, diarrhea, and dysphagia severity. In addition, patients in the LARS group were asked about other gastrointestinal symptoms such as ability to vomit and ability to belch. Health-related quality of life and patient-reported symptoms were assessed by administering the validated Quality of Life in Reflux and Dyspepsia (QOLRAD) and Gastrointestinal Symptom Rating Scale (GSRS) questionnaires to patients at randomization and annually thereafter. Translations of the questionnaires into different languages were performed according to proposed guidelines and involved several independent translators.

During the follow-up period, patients in both treatment groups who experienced moderate to severe recurrent GERD symptoms for at least 3 consecutive days were instructed to contact the clinic. They were then questioned about their symptom control and need for other regular medication and were offered endoscopy.

Treatment End Points and Statistical Analyses

The main analysis was conducted using the intention-to-treat population comprising all randomized patients. Including patients randomized to undergo surgery but not operated on had little influence on the primary analysis because they were censored early.

The primary end point in this study, time to treatment failure, was defined as follows for the 2 study treatments.

In the esomeprazole group, the need for escalation in treatment for control of reflux disease was assessed at clinic visits by asking "Do you have sufficient control of your heartburn and acid regurgitation?" If the answer was no and the patient stated a need for other regular drug therapy, the dose of esomeprazole was increased to 40 mg once per day for 8 weeks and could be adjusted to 20 mg twice per day for a further 8 weeks if symptoms had not resolved. If this proved insufficient to control symptoms, the patient was classified as having had treatment failure.

The same questions were asked at clinic visits about symptom control in the LARS group, and if the answer was no and was backed up by a need for treatment with acid-suppressive drugs, the patient was classified as having had treatment failure. Patients were also classified as having had treatment failure if they had postoperative symptoms requiring medical action, perioperative death, postoperative death within 30 days of surgery, dysphagia requiring further treatment, or any other requirement to reoperate for symptom control. In the case of functional esophageal postoperative stenosis, 1 dilatation was allowed.

Time to treatment failure/censoring was defined as number of days between randomization and last visit for all participants within 5 years after randomization, regardless of reason for discontinuation or reason for visit. For patients who never returned for a visit, time of censoring was set as 0. As an exploratory analysis, the Kaplan-Meier method was used to estimate the proportion of patients in remission over time and, as specified in the study protocol, the log-rank test was used to test the statistical significance of the observed difference between the treatment groups. A per-protocol analysis was also performed on the primary end point and included all randomized patients except those with major protocol violations.[16]

In addition, to test the robustness of our main analysis, best- and worst-case outcomes scenarios were analyzed with censored patients considered to have had either treatment failures or treatment successes and after excluding censored patients.

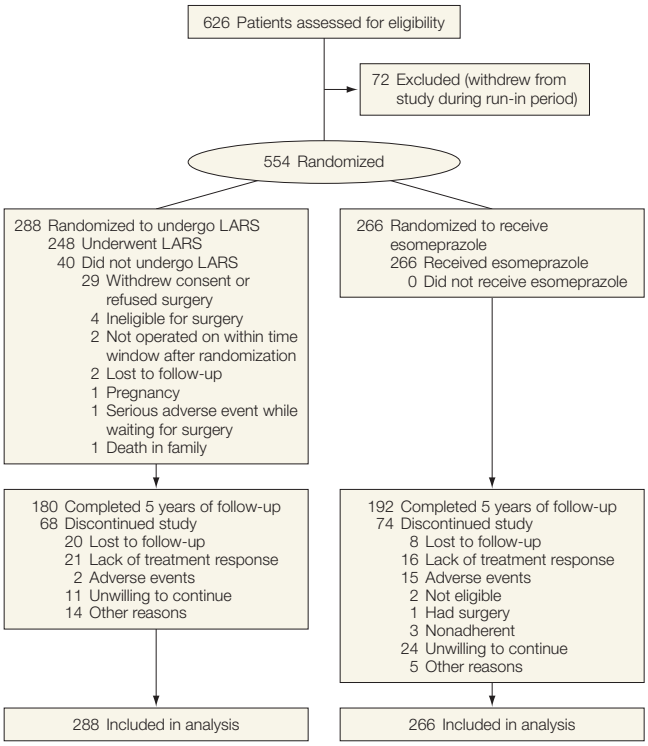
Secondary variables were presented descriptively and analyzed only for the intention-to-treat population, without any analysis of missing data. There were no adjustments for multiple comparisons because of the exploratory character of the study.

In a post hoc analysis, severity of GERD symptoms (none = 0; mild = 1; moderate = 2; and severe = 3) reported at 5 years was compared between treatments using a 2-sided Wilcoxon rank sum test. The safety population included all patients who received at least 1 dose of study drug and from whom postdose data were available.

This study was not designed as a superiority or equivalence trial but, rather, was an exploratory study to estimate the efficacy of LARS and PPI treatment in PPI responders. The sample size was determined by assuming that the true rate of treatment success (ie, patients who did not experience treatment failure within 5 years) would be at least 70% for both treatments. With 275 patients in each group, the true difference between the treatments was estimated not to differ from the observed difference by more than 8 percentage points with a probability of 95%. Thus, the sample size was derived to give a specific length of the confidence interval (CI) between the proportions of treatment success in the 2 treatment groups. The computation is based on the normal approximation for a continuity-corrected interval.[19] In an exploratory analysis, the log-rank test was used to test for the superiority of the observed difference between the treatment groups.

Statistical analyses were performed using SAS version 8.2 (SAS Institute Inc, Cary, North Carolina).

Figure 1. Patient Flow in the LOTUS Trial



LARS indicates laparoscopic antireflux surgery.

RESULTS

Study Population

A total of 626 patients completed enrollment for the study, of whom 554 were randomized, 288 to undergo LARS (40 of whom were not operated on) and 266 to receive esomeprazole. The reasons for the 40 patients who were not operated on were as follows: 29 withdrew consent or refused surgery; 4 were considered ineligible for surgery; 2 were not operated on within the time window after randomization; 2 were lost to follow-up; 1 was pregnant; 1 had a serious adverse event while waiting for surgery; and 1 had a death in the family. The demographic characteristics of

these 40 patients did not differ substantially from the other randomized patients in the study. The flow of patients included in the study and reasons for withdrawal at each stage are summarized in FIGURE 1. Of the 248 patients in the LARS group, 180 (73%) completed the 5-year follow-up visit and 68 discontinued the study before the 5-year visit, 33 of whom met the primary end point of treatment failure. In the esomeprazole group, 192 of the 266 patients (72%) completed 5-year follow-up and 74 discontinued, 19 of whom had treatment failure. Thus, the total discontinuation rate at 5 years for participants randomized to the LARS group (including the 40 who did not undergo surgery) was 108 of 288 (38%) and for participants randomized to the esomeprazole group was 74 of 266 (28%). In violation of the protocol, 1 participant in the esomeprazole was operated on by a surgeon who was not aware of the study at a time when the investigator was on vacation. This patient was withdrawn from the study subsequently.

Demographic characteristics and GERD disease history for participants in each treatment group are presented in TABLE 1. The 2 groups were well matched with regard to both demographics and history and current symptoms of GERD.

Table 1. Patient Demographics and Baseline Characteristics[a]

Characteristics	Laparoscopic Antireflux Surgery (n = 288)	Esomeprazole (n = 266)
Age, mean (SD), y	45 (10.9)	45 (11.5)
Male	199 (69)	199 (75)
Body mass index, mean (SD)[b]	27 (3.7)	27 (4.4)
Current smokers	81 (28)	58 (22)
Alcohol use	168 (58)	176 (66)
Previous upper gastrointestinal tract surgery	5 (1.7)	6 (2.3)
History of reflux symptoms, y		
<1	7 (2.4)	3 (1.1)
1-5	97 (34)	91 (34)
>5	184 (64)	172 (65)
Duration of verified reflux disease, y		
<1	84 (29)	80 (30)
1-5	146 (51)	135 (51)
>5	56 (19)	50 (19)
Heartburn severity		
None	102 (35)	92 (35)
Mild	72 (25)	61 (23)
Moderate	70 (24)	65 (24)
Severe	44 (15)	48 (18)
Regurgitation severity		
None	132 (46)	125 (47)
Mild	62 (22)	52 (20)
Moderate	70 (24)	66 (25)
Severe	24 (8)	23 (9)
Los Angeles grade of esophagitis		
No esophagitis	135 (47)	130 (49)
Grade A	79 (27)	56 (21)
Grade B	64 (22)	71 (27)
Grade C	10 (3.5)	10 (3.8)
Grade D	1 (0.3)	0
Abnormal 24-h esophageal pH	209 (73)	200 (75)
Endoscopic suspicion of esophageal metaplasia	32 (11.1)	28 (10.5)
Hiatal hernia	204 (71)	188 (71)
Helicobacter pylori–positive status	30 (10.4)	39 (14.3)

[a] Data are expressed as No. (%) of participants unless otherwise indicated.
[b] Body mass index is calculated as weight in kilograms divided by height in meters squared.

Treatment Efficacy

Time to Treatment Failure. Time to treatment failure, the primary efficacy variable, is presented as Kaplan-Meier plots for the intention-to-treat population in FIGURE 2. At 5 years, an estimated 85% (95% CI, 81%-90%) in the LARS group and an estimated 92% (95% CI, 89%-96%) in the esomeprazole group remained in remission (log-rank $P = .048$). There were 33 treatment failures in the LARS group (29 patients required other treatment to control reflux symptoms, 1 needed more than 1 dilatation, and 3 had postfundoplication adverse events including 1 gastric perforation and 2 with severe flatulence, bloating, and diarrhea) compared with 19 treatment failures in the esomeprazole group (all failures of symptom resolution). The results of the per-protocol analysis were similar: 85% (95% CI, 80%-90%) remission in the LARS group and 94% (95% CI, 91%-98%) remission in the esomeprazole group at 5 years (ie, 30 vs 12 treatment failures respectively; $P = .004$).

When best- and worst-case scenario case analyses were applied, the remission rates were 88.5% (95% CI, 84.1%-91.9%) in the LARS group and 92.9% (95% CI, 88.9%-95.5%) in the esomeprazole group, for a treatment difference of 4.3% (95% CI, –0.9% to 8.5%) when all censored patients were considered to have successful treatment. Corresponding rates when all censored cases were considered treatment failures were 61.5% (95% CI, 55.5%-67.1%) in the LARS group and 71.8% (95% CI, 65.9%-77.0%) in the esomeprazole group, for a treatment difference of 10.3% (95% CI, 2.2%-18.5%). When all censored patients were excluded from the analysis, the estimated remission rates were 84.3% (95% CI, 78.5%-88.8%) in the LARS group and 91.0% (95% CI, 86.0%-94.3% in the

esomeprazole group, for a mean treatment difference of 6.7% (95% CI, –0.1% to 13.4%).

The percentages of patients in the esomeprazole group who required an increased dose of esomeprazole to control their symptoms were similar for each year during the study; at 5 years, 23.1% of patients were receiving an increased dose (16.5% were taking 40 mg once per day and 6.6% were taking 20 mg twice per day).

GERD and Postoperative Symptoms. The prevalence and severity of GERD symptoms reported by patients at each clinic visit throughout the study is shown in FIGURE 3. The esomeprazole group showed similar levels of heartburn and regurgitation from baseline up to 5 years, while both symptoms decreased in the LARS group after randomization. At 5 years, acid regurgitation was significantly worse in the esomeprazole group than in the LARS group (13% vs 2%, respectively; $P<.001$), although there was no significant difference between the groups in the severity of heartburn (16% vs 8%; $P=.14$), epigastric pain (18% vs 18%; $P=.55$), or diarrhea (15% vs 16%; $P=.25$). At 5 years, dysphagia remained significantly more common in the LARS group than in the esomeprazole group (11% vs 5%, respectively; $P<.001$), as did bloating (40% vs 28%, respectively; $P<.001$) and flatulence (57% vs 40%, respectively; $P<.001$).

Endoscopic Findings. At 5 years, esophagitis of Los Angeles grades A, B, or C was observed in 12, 5, and 1 patients in the LARS group and in 16, 7, and 2 patients in the esomeprazole group, respectively. The percentage of patients in the esomeprazole group with hiatal hernia remained consistent over 5 years and was present in 62% at 5 years compared with 6% in the LARS group. The presence of stricture decreased in both treatment groups throughout the study, with 5 reported during the run-in period (3 in the esomeprazole group and 2 in the LARS group) and 2 after operation in the LARS group.

Endoscopic suspicion of esophageal metaplasia was present in 11.1% (32/288) of the LARS group and in 10.5% (28/266) of the esomeprazole group at entry, and its prevalence at 5 years remained stable in both study groups (13.6% [22/162] and 9.3% [17/183], respectively).

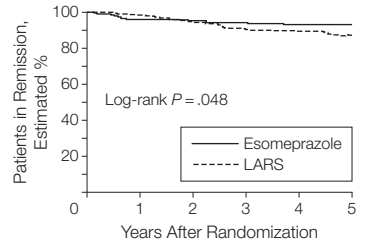
pH-Metry. Complete pH data were available for approximately 70% of the participants who were still in follow-up at 5 years. Baseline intraesophageal acid exposure was similar for the 2 treatment groups; the median percentage of time that pH was below 4 (upright plus recumbent) was 8.6% in the LARS group and 8.8% in the esomeprazole group. At 5 years, exposure time had decreased to 0.7% in the LARS group and to 1.9% in the esomeprazole group. The mean percentage of time with raised intragastric pH (>4) increased from 12.1% at baseline to 62.1% at 5 years in the esomeprazole group, while in the LARS group it remained fairly stable, decreasing slightly from 12.4% at baseline to 11.4% at 5 years.

Health-Related Quality of Life. QOLRAD scores on the food and drink and vitality dimensions as well as scores on the GSRS reflux dimension were the most abnormal at entry and the most sensitive to change with treatment. The mean scores for all dimensions improved in both groups and remained close to values observed in a healthy population (eTable 1; available online at http://www.jama.com).

Safety

There was no perioperative mortality and only 3% of patients had in-hospital morbidity. Serious adverse events were reported by 28.6% of patients who underwent LARS (n=248) and by 24.1% of the esomeprazole group (n=266) over 5 years (TABLE 2). Five patients had serious adverse events during the study that led to death either during the study (3 patients in the esomeprazole group, 1 of whom had pneumonia and 2 of whom had pancreatic carcinoma) or after the study (1 patient in the LARS group who had a malignant lung neoplasm and 1 patient in the esomeprazole group who had a fall that led to traumatic brain injury and femur and pelvic fracture). Further details of serious adverse events are shown in eTable 2. Laboratory variables monitored throughout the study are summarized in Table 2. Mean gastrin and chromogranin levels were elevated in patients treated with esomeprazole, as expected after long-term acid suppression. They appeared to stabilize after 3 years. No clinically relevant changes were noted in other laboratory variables.

Figure 2. Time to Treatment Failure



No. at risk

Esomeprazole	266	228	217	205	199	181
LARS	288	231	216	202	192	168

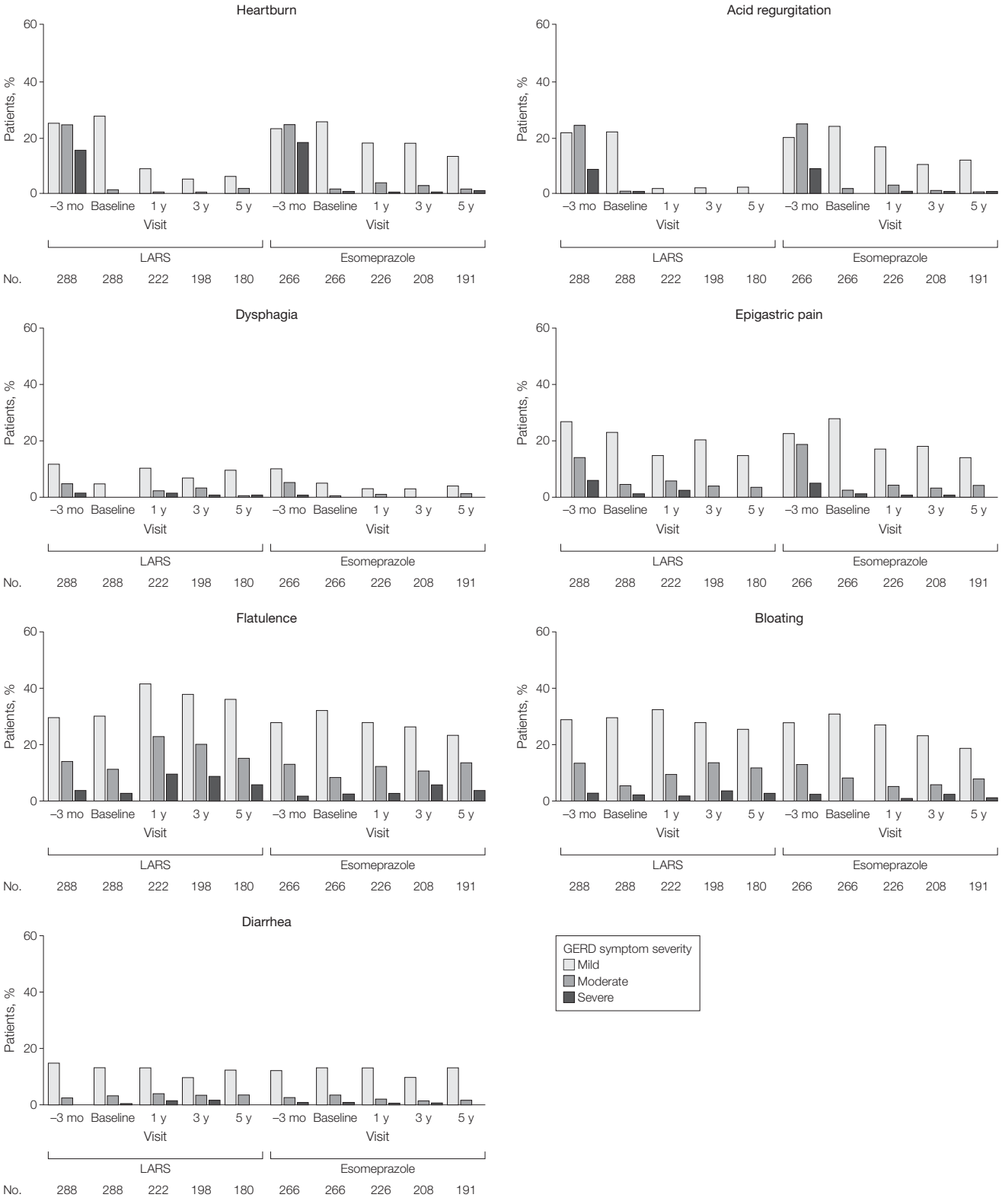
LARS indicates laparoscopic antireflux surgery.

COMMENT

This large, multicenter randomized trial demonstrated that with modern forms of antireflux therapy, either by drug-induced acid suppression or after LARS, most patients remain in remission for at least 5 years. In an exploratory analysis, the estimated remission rates at 5 years were higher in the esomeprazole group (92%; 95% CI, 89%-96%) than in the LARS group (85%; 95% CI, 81%-90%; log-rank $P=.048$). There was more regurgitation with esomeprazole than with LARS. In contrast, dysphagia, bloating, and flatulence were more common after LARS vs with esomeprazole. Both treatments were well tolerated, with no surgery-related mortality and similar safety profiles for both.[20]

The high remission rates reported in this trial are at variance with previous randomized studies comparing long-term medical therapy vs antireflux surgery. There may be several reasons for these apparent discrepancies. With respect to drug therapy, earlier trials used

Figure 3. Symptoms Reported by Each Treatment Group Throughout the Study as Mild, Moderate, or Severe



See definitions of mild, moderate, or severe symptoms in "Methods" section of text. GERD indicates gastroesophageal reflux disease; LARS, laparoscopic antireflux surgery.

drugs such as antacids, prokinetics, or histamine$_2$ receptor antagonists that are now known to be of limited efficacy. Proton pump inhibitors are more potent acid-suppressive agents, reducing the intensity of esophageal acid exposure. In the present study, patients were treated with esomeprazole, which suppresses gastric acidity more effectively than omeprazole and other PPIs.[21,22] Moreover, in our study, patients whose reflux symptoms were not adequately controlled by a standard maintenance regimen (ie, esomeprazole, 20 mg/d) were allowed to increase the dosage to 40 mg once per day and then to 20 mg twice per day. Dinnertime or split dosing can improve breakthrough nocturnal symptoms for some patients.[23] Dose escalation or split dosing applied in the LOTUS study may have contributed to the improved remission rate (92%) compared with that reported in the SOPRAN study at 5 years (57%),[9] in which patients received omeprazole and were not actively dose-titrated to the same extent. Most likely, LARS outcomes were better than reported in earlier studies because we recruited participating centers and surgeons with demonstrable expertise and standardized surgical technique, which has been shown to improve outcomes in other studies.[24,25] The outcome from this approach was manifested by the absence of mortality and the very low morbidity rate in the LARS group. Only 1 patient had dysphagia requiring more than 1 endoscopic dilatation. One recent meta-analysis suggested better outcomes for LARS compared with open surgery, but the need for reoperation may be more frequent after LARS.[7] In our experience, most patients (98%) did not experience long-term complications from LARS. The final endoscopic assessment did not show anatomical deterioration and hiatal repair was maintained, with only 5.6% of the LARS group having hiatal hernia after 5 years compared with 62.3% in the esomeprazole group, confirming similar observations from the SOPRAN study.[9] Our LARS group showed slight deterioration in symptom control between 3 years (estimated remission rate, 90%) and 5 years (estimated remission rate, 85%), while the esomeprazole group remained more stable. Better long-term symptom control in the esomeprazole group might have been related to dose escalation.

Long-term acid suppression has been associated with complications.[26-28] The serious adverse events reported in this study[20] (eTable 2) were similar between the LARS and esomeprazole groups, apart from slightly more cardiovascular complications in the esomeprazole group. However, there were no specific serious adverse events that were judged by the investigators to be attributable to acid-supressive therapy alone. Two hip fractures occurred during the study, 1 in the LARS group and 1 in the esomeprazole group that was caused by a serious fall that also resulted in femur fracture, brain trauma, and death. The few hip fractures we observed suggest that fractures are rare with PPI and that previous observational studies might have overestimated the risk of these events.[29]

Our trial has several limitations. First, we enrolled only PPI responders; our results do not generalize to patients who initially are partially or completely refractory to PPI therapy.[30] These poor responders are a heterogeneous group of patients with many underlying causes for their nonresponsiveness to treatment. The most common cause is the absence of actual reflux disease, with symptoms being caused by nonreflux conditions. Assessing the role of surgery in nonresponders requires specific investigations such as pH impedancemetry to better classify patients. The choice of long-term PPI maintenance therapy or LARS in patients who initially respond to acid suppression is relevant to clinical practice.

Second, 14% of participants randomized to receive surgery were not operated on for various reasons. Despite our efforts, we were unable to follow up this patient cohort, who did not differ from

Table 2. Safety Assessments

	Baseline		3-Year Follow-up		5-Year Follow-up	
	LARS	Esomeprazole	LARS	Esomeprazole	LARS	Esomeprazole
Serious adverse events[a]						
No. of patients with a serious adverse event	NA	NA	54	38	71	64
No. of patients with a fatal serious adverse event	NA	NA	0	1	1[b]	4[b]
Blood variables, mean						
Hemoglobin, g/L	149.4	149.5	149.4	150.4	149.0	150.0
Vitamin B_{12}, pmol/L	329.4	332.2	325.6	339.4	313.0	335.8
Serum gastrin, pg/mL	70.2	65.6	51.3	159.5	54.0	164.4
Chromogranin A, ng/mL	91.3	81.2	38.0	206.1	45.8	216.3
Alkaline phosphatase, u/L	71.7	71.8	69.1	71.6	67.9	68.3
Calcium, mmol/L	2.3	2.3	2.3	2.3	2.2	2.3
Vitamin D, nmol/L	49.5	50.1	53.6	57.9	50.2	50.6
Homocysteine, µmol/mL	11.7	11.5	11.2	11.5	13.2	12.5

Abbreviations: LARS, laparoscopic antireflux surgery; NA, not applicable.
[a] See also eTable 2. Total at 5 years is cumulative.
[b] One patient in each group died after the end of the study, but the serious adverse event started during the study.

participants at baseline but who declined surgery. For this reason, we performed a sensitivity analysis with best- and worst-case scenarios assuming that all participants not completing the study after randomization all either had treatment response or treatment failure. The results of this were similar to our overall findings. The number of participants randomized to receive surgery who did not undergo operation was considerably lower than the 38% of participants reported in the large UK REFLUX trial.[13] When treatment failures were excluded, the dropout rate during the 5-year duration was consistent with rates observed in other studies of chronic conditions and better than in other previous antireflux surgery clinical trials.[8,12,13]

Third, this study was not designed as a superiority or equivalence trial but, rather, as an exploratory study to estimate the efficacy of antireflux surgery and PPI treatment in PPI responders. At the time the study was designed, there were no good estimates for long-term treatment efficacy of esomeprazole (or other PPIs) in this patient population, and the 70% estimate of success with surgery was based on results with nonlaparoscopic procedures. We therefore selected a more pragmatic strategy for sample size determination by estimating the size of the CI for a given difference in efficacy. Nonetheless, we did prespecify that the treatment success rates in each group would be compared using log-rank tests for the superiority of the observed difference between the treatment groups.

In summary, most patients with GERD who are initially responsive to PPIs achieve and remain in remission at 5 years with contemporary antireflux therapy using either LARS or esomeprazole in a dose-escalating manner when required.

Author Contributions: Dr Lundell had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Galmiche and Hatlebakk contributed equally to the article's content.
Study concept and design: Galmiche, Hatlebakk, Attwood, Fiocca, Eklund, Lind, Lundell.
Acquisition of data: Galmiche, Hatlebakk, Attwood, Ell, Eklund, Lundell.
Analysis and interpretation of data: Galmiche, Hatlebakk, Attwood, Ell, Fiocca, Eklund, Långström, Lind, Lundell.
Drafting of the manuscript: Galmiche, Hatlebakk, Attwood, Ell, Eklund, Långström, Lind, Lundell.
Critical revision of the manuscript for important intellectual content: Galmiche, Hatlebakk, Attwood, Ell, Fiocca, Eklund, Långström, Lind, Lundell.
Statistical analysis: Långström.
Obtained funding: Galmiche, Hatlebakk, Lundell.
Administrative, technical, or material support: Galmiche, Hatlebakk, Attwood, Ell, Fiocca, Långström, Lind, Lundell.
Study supervision: Galmiche, Hatlebakk, Attwood, Lind, Lundell.
Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Galmiche reported that he is a consultant for several biomedical companies (AstraZeneca, Janssen-Cilag, Given Imaging, Xenoport, and Norgine) and that his institution receives grants for research from AstraZeneca, Given Imaging, and Mauna Kea Technology. He has prepared educational presentations for Shire Movetis. Dr Attwood reported that he has received honoraria for speaking at meetings sponsored by AstraZeneca. Dr Ell reported that he receives grants for research from several biomedical companies (AstraZeneca, Fujinon, Erbe, and Hitachi). Dr Fiocca reported that he has received travel and related expenses for attending study-associated meetings and his institution received a grant from AstraZeneca for central histological analyses. Drs Eklund, Långström, and Lind reported that they are employees of AstraZeneca. Dr Lundell reported that he receives consultancy and lecture fees from several biomedical companies including AstraZeneca and that his institution (Karolinska University Hospital) receives grants for research on his behalf.
LOTUS Trial Collaborators: *Austria*: Johannes Miholic (country coordinator), Univ Klinik für Chirurgie, Vienna; Rainer Hubmann, Jan Danis (country coordinator), Krankenhaus, Linz; *Belgium*: Jan Tack, Toni Lerut (country coordinator), UZ Gasthuisberg, Leuven; Hubert Piessevaux, UCL St-Luc, Jacques Devière, Clinique Universitaire Bruxelles Hôpital Erasme, Michel Buset, Centre Hospitalier Universitaire St-Pierre, and Cristiano Chioccioli, Clinique St-Jean, Brussels; Danny De Looze, UZ Gent, Gent; Jean-Claude Demoulin, Cliniques St-Joseph, and Edouard Louis, Centre Hospitalier Universitaire Sart Tilman, Liège; Jean-Michel Ghilain, Jean-Marc Maisin, CH Jolimont-Lobbes (Jolimont), Haine-Saint-Paul; *Denmark*: Peter Funch-Jensen (country coordinator), Århus Universitetshospital, Århus; Jørn Nielsen, Regionshospitalet Viborg, Viborg; Lars Christensen, Regionshospitalet Herning, Herning; Henning Antonsen, Kolding Sygehus, Kolding; Karsten Lauritsen, Odense Universitetshospital, Odense; Per Jess, Hillerød Hospital, Hillerød; Lene Wallin, Glostrup Hospital, Glostrup; Lars Naver, Hvidovre Hospital, Hvidovre; *France*: Jean-Paul Galmiche (country coordinator), Eric Letessier, CHU de Nantes–Hôtel Dieu, Nantes; Frank Zerbib, Groupe Hospitalier Saint André, Bordeaux; Bruno Bonaz, Richard Bost, CHU de Grenoble, Grenoble; Jean-Charles Delchier, Hôpital Henri Mondor, Creteil; *Germany:* Martin Fein, Jörn Maroske, Würzburg, Universitätsklinikum, Würzburg; Christian Ell, Dr Horst Schmidt Kliniken GmbH, Wiesbaden; Arnulf Hölscher, Uniklinik und Poliklinik für Visceral und Gefäßchirurgie, Köln; Carsten Zornig, Israelitisches Krankenhaus Hamburg, Hamburg; Joachim Helmut Schneider, Klinikum der Eberhard Karls Universität, Tübingen; Thomas Hüttl, München, Ludwig Maximilians Universität LMU, Chirurgie Poliklinikum, München; Markus Büchler, Universitätsklinikum Heidelberg, Heidelberg; Ursula Wehrmann, Universitätsklinikum Carl-Gustav-Carus an der TU Dresden, Dresden; Matthias Kemen, Herne, Ev Kkh, Chirurgie, Herne; Peter Layer, Hamburg Israelitisches Krankenhaus, Hamburg; Karl-Hermann Fuchs (country coordinator), St Markus Krankenhaus, Allgemeine Chirurgie Frankfurt/Main, Frankfurt; Solveig Kemen, Schwerpunktpraxis Gastroenterologie Praxis Kemen & Schmidt-Heinevetter, Bochum; Dieterich Hüppe, Internistische Praxisgemeinschaft, Herne; *Iceland*: Margrét Oddsdóttir (country coordinator), Bjarni Thjodleifsson (country coordinator), Landspitali, University Hospital, Reykjavik; *Italy*: Luigi Bonavina (country coordinator), Policlinico S Donato, San Donato Milanese; Ermanno Ancona, Policlinico Universitario di Padova, Padova; Mario Morino, Ospedale Le Molinette, Torino; Giusto Pignata, Alessandro Giurissa, Ospedale di Monfalcone, Monfalcone; Mauro Rossi, Ospedale S Chiara, Pisa; Riccardo Rosati, Istituto Clinico Humanitas, Rozzano; Renzo Cestari, Spedali Civili di Brescia, Brescia; *Netherlands*: Hein Gooszen (country coordinator), Universitair Medisch Centrum Utrecht, Utrecht; *Norway*: Jan Hatlebakk (country coordinator), Helse Bergen HF Haukeland Universitetssykehus, and Solomon Tefera, Haraldsplass Diakonale Sykehus AS, Bergen; Gjermund Johnsen, St Olav's Hospital HF, Trondheim; Jørgen Jahnsen, Aker Universitetssykehus HF, and Olav Sandstad, Ullevål Universitetssykehus HF, Oslo; Asbjørn Stallemo, Sørlandet Sykehus HF, Kristiansand; Jon Florholmen, Universitetssykehuset Nord-Norge HF, Tromsø; Geir Tollåli, Nordlandssykehuset HF, Bodø; *Sweden*: Cecilia Engström, Bengt Liedman, Lars Lundell (country coordinator), SU/Sahlgrenska Universitetssjukhus, Göteborg; *United Kingdom*: Chris Babbs, Stephen Attwood (country coordinator), Hope Hospital, Salford.
Funding/Support: The study was funded in total by AstraZeneca Research and Development, Mölndal, Sweden.
Role of the Sponsor: AstraZeneca contributed to the design of the study; collection, management, analysis, and interpretation of the data; and preparation, review, and approval of the manuscript, in full collaboration with the independent statistician and the study steering committee. Madeline Frame, BSc, PhD, a medical scientist affiliated with AstraZeneca, provided medical writing support to the principal author in terms of drafting the methods and results and incorporating edits from all authors. All writing was done in close collaboration with the first author and the steering committee and they gave major input to the introduction, interpretation, and discussion sections.
Independent Statistical Analysis: Hans Wedel, PhD, emeritus professor in Epidemiology and Biostatistics at the Nordic School of Public Health, Gothenburg, Sweden, undertook an independent statistical analysis, of the data on behalf of AstraZeneca, for which he was reimbursed at standard rates. The results of his analyses are included in this article. His position as an academic faculty member working independent of AstraZeneca has been verified by the dean of the faculty, Anders Foldspang.
Online-Only Material: eTables 1 and 2 are available online at http://www.jama.com.
Additional Contributions: We thank all study site personnel who contributed to the study. This article is dedicated to Ola Junghard, PhD, our colleague and statistician for many years, who passed away in November 2009.

REFERENCES

1. Kulig M, Leodolter A, Vieth M, et al. Quality of life in relation to symptoms in patients with gastro-oesophageal reflux disease—an analysis based on the ProGERD initiative. *Aliment Pharmacol Ther*. 2003; 18(8):767-776.
2. Wahlqvist P, Reilly MC, Barkun A. Systematic review: the impact of gastro-oesophageal reflux dis-

ease on work productivity. *Aliment Pharmacol Ther*. 2006;24(2):259-272.
3. Camilleri M, Dubois D, Coulie B, et al. Prevalence and socioeconomic impact of upper gastrointestinal disorders in the United States: results of the US Upper Gastrointestinal Study. *Clin Gastroenterol Hepatol*. 2005;3(6):543-552.
4. McQuaid KR, Laine L. Early heartburn relief with proton pump inhibitors: a systematic review and meta-analysis of clinical trials. *Clin Gastroenterol Hepatol*. 2005;3(6):553-563.
5. Gralnek IM, Dulai GS, Fennerty MB, Spiegel BMR. Esomeprazole versus other proton pump inhibitors in erosive esophagitis: a meta-analysis of randomized clinical trials. *Clin Gastroenterol Hepatol*. 2006;4(12):1452-1458.
6. Edwards SJ, Lind T, Lundell L. Systematic review: proton pump inhibitors (PPIs) for the healing of reflux oesophagitis—a comparison of esomeprazole with other PPIs. *Aliment Pharmacol Ther*. 2006;24(5):743-750.
7. Peters MJ, Mukhtar A, Yunus RM, et al. Meta-analysis of randomized clinical trials comparing open and laparoscopic anti-reflux surgery. *Am J Gastroenterol*. 2009;104(6):1548-1561.
8. Spechler SJ, Lee E, Ahnen D, et al. Long-term outcome of medical and surgical therapies for gastroesophageal reflux disease: follow-up of a randomized controlled trial. *JAMA*. 2001;285(18):2331-2338.
9. Lundell L, Miettinen P, Myrvold HE, et al; Nordic GERD Study Group. Comparison of outcomes 12 years after antireflux surgery or omeprazole maintenance therapy for reflux esophagitis. *Clin Gastroenterol Hepatol*. 2009;7(12):1292-1298, quiz 1260.
10. Mahon D, Rhodes M, Decadt B, et al. Randomized clinical trial of laparoscopic Nissen fundoplication compared with proton-pump inhibitors for treatment of chronic gastro-oesophageal reflux. *Br J Surg*. 2005;92(6):695-699.
11. Anvari M, Allen C, Borm A. Laparoscopic Nissen fundoplication is a satisfactory alternative to long-term omeprazole therapy. *Br J Surg*. 1995;82(7):938-942.
12. Mehta S, Bennett J, Mahon D, Rhodes M. Prospective trial of laparoscopic Nissen fundoplication vs proton pump inhibitor therapy for gastroesophageal reflux disease: 7-year follow-up. *J Gastrointest Surg*. 2006;10(9):1312-1316.
13. Grant AM, Wileman SM, Ramsay CR, et al; REFLUX Trial Group. Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ*. 2008;337:a2664.
14. Attwood SE, Lundell L, Ell C, et al; LOTUS Trial Group. Standardization of surgical technique in antireflux surgery: the LOTUS trial experience. *World J Surg*. 2008;32(6):995-998.
15. Castell DO, Kahrilas PJ, Richter JE, et al. Esomeprazole (40 mg) compared with lansoprazole (30 mg) in the treatment of erosive esophagitis. *Am J Gastroenterol*. 2002;97(3):575-583.
16. Lundell L, Attwood S, Ell C, et al; LOTUS Trial Collaborators. Comparing laparoscopic antireflux surgery with esomeprazole in the management of patients with chronic gastro-oesophageal reflux disease: a 3-year interim analysis of the LOTUS trial. *Gut*. 2008;57(9):1207-1213.
17. Fiocca R, Mastracci L, Engström C, et al; LOTUS Trial Collaborators. Long-term outcome of microscopic esophagitis in chronic GERD patients treated with esomeprazole or laparoscopic antireflux surgery in the LOTUS trial. *Am J Gastroenterol*. 2010;105(5):1015-1023.
18. Hatlebakk JG, Zerbib F, Bonavina L, et al. Control of acid reflux 5 years after laparoscopic anti-reflux surgery (LARS) compared to long-term treatment with esomeprazole 20-40 mg daily (ESO) in the LOTUS study. *Gut*. 2010;59(suppl 3):A87.
19. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York, NY: Wiley; 1981.
20. Ell C, Lundell LR, Seiler CM, et al. Safety profile in GERD patients 5 years after laparoscopic antireflux surgery or long-term treatment with esomeprazole 20-40 mg daily in the LOTUS study. *Gastroenterology*. 2010;138(5)(suppl 1):S-654.
21. Miner P Jr, Katz PO, Chen Y, Sostek M. Gastric acid control with esomeprazole, lansoprazole, omeprazole, pantoprazole, and rabeprazole: a 5-way crossover study. *Am J Gastroenterol*. 2003;98(12):2616-2620.
22. Röhss K, Lind T, Wilder-Smith C. Esomeprazole 40 mg provides more effective intragastric acid control than lansoprazole 30 mg, omeprazole 20 mg, pantoprazole 40 mg and rabeprazole 20 mg in patients with gastro-oesophageal reflux symptoms. *Eur J Clin Pharmacol*. 2004;60(8):531-539.
23. Hatlebakk JG, Katz PO, Kuo B, Castell DO. Nocturnal gastric acidity and acid breakthrough on different regimens of omeprazole 40 mg daily. *Aliment Pharmacol Ther*. 1998;12(12):1235-1240.
24. Luostarinen MES, Isolauri JO. Surgical experience improves the long-term results of Nissen fundoplication. *Scand J Gastroenterol*. 1999;34(2):117-120.
25. Oelschlager BK, Quiroga E, Parra JD, Cahill M, Polissar N, Pellegrini CA. Long-term outcomes after laparoscopic antireflux surgery. *Am J Gastroenterol*. 2008;103(2):280-287.
26. Moayyedi P, Cranney A. Hip fracture and proton pump inhibitor therapy: balancing the evidence for benefit and harm. *Am J Gastroenterol*. 2008;103(10):2428-2431.
27. Targownik LE, Lix LM, Leung S, Leslie WD. Proton-pump inhibitor use is not associated with osteoporosis or accelerated bone mineral density loss. *Gastroenterology*. 2010;138(3):896-904.
28. Lodato F, Azzaroli F, Turco L, et al. Adverse effects of proton pump inhibitors. *Best Pract Res Clin Gastroenterol*. 2010;24(2):193-201.
29. Yang YX, Lewis JD, Epstein S, Metz DC. Long-term proton pump inhibitor therapy and risk of hip fracture. *JAMA*. 2006;296(24):2947-2953.
30. Zaninotto G, Attwood SE. Surgical management of refractory gastro-oesophageal reflux. *Br J Surg*. 2010;97(2):139-140.

Surg Endosc (2011) 25:2547–2554
DOI 10.1007/s00464-011-1585-5

A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for the treatment of patients with chronic gastroesophageal reflux disease (GERD): 3-year outcomes

Mehran Anvari · Christopher Allen · John Marshall · David Armstrong · Ron Goeree · Wendy Ungar · Charles Goldsmith

Received: 25 May 2010 / Accepted: 10 January 2011 / Published online: 22 April 2011

Abstract
Background A randomized controlled trial (RCT) investigated patients with gastroesophageal reflux disease (GERD) who were stable and symptomatically controlled with long-term medical therapy to compare ongoing optimized medical therapy with laparoscopic Nissen fundoplication (LNF).
Methods Of the 180 patients eligible for randomization, 104 gave informed consent, and 3 withdrew from the study immediately after randomization. The patients randomized to medical therapy received optimized treatment with proton pump inhibitors (PPIs) using a standardized management protocol based on best evidence and published guidelines. The surgical patients underwent LNF by one of four surgeons using a previously published technique. The patients underwent symptom evaluation using the GERD symptom scale (GERSS) and the global visual analog scale (VAS) for overall symptom control. They had 24-h esophageal pH monitoring at baseline and after 3 years. The medical patients were evaluated receiving PPI, and the surgical patients were evaluated not receiving PPI.
Results For the 3-year follow-up assessment, 93 patients were available. At 3 years, surgery was associated with more heartburn-free days, showing a mean difference of -1.35 days per week ($p = 0.0077$) and a lower VAS score ($p = 0.0093$) than medical management. Surgical patients reported improved quality of life on the general health subscore of the Medical Outcomes Survey Short Form 36 (SF-36) at 3 years, with a mean difference of -12.19 ($p = 0.0124$). The groups did not differ significantly in terms of GERSS or acid exposure on 24-h esophageal pH monitoring at 3 years. There were six treatment failures (11.8%) in the surgical group and eight treatment failures (16%) in the medical group by 3 years.
Conclusions For patients whose GERD symptoms are stable and controlled with PPI, continuing medical therapy and laparoscopic antireflux surgery are equally effective, although surgery may result in better symptom control and quality of life.

Keywords Gastroesophageal reflux disease · GERD · Laparoscopic antireflux surgery · Laparoscopic Nissen fundoplication · Medical therapy · Proton pump inhibitors

M. Anvari (✉)
Department of Surgery, St. Joseph's Healthcare, McMaster University, 50 Charlton Ave East, Hamilton, ON L8N 4A6, Canada
e-mail: anvari@mcmaster.ca

C. Allen
Firestone Institute of Respiratory Health, McMaster University, Hamilton, ON, Canada

J. Marshall · D. Armstrong
Department of Medicine, McMaster University, Hamilton, ON, Canada

R. Goeree · C. Goldsmith
Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada

W. Ungar
Institute for Clinical Evaluative Sciences, Hospital for Sick Children, Toronto, ON, Canada

The medical treatment of choice for gastroesophageal reflux disease (GERD) is administration of proton pump inhibitors (PPIs) [1, 2]. Although PPIs are effective in the treatment of esophagitis and the control of GERD symptoms, studies have demonstrated that up to 50% of patients continue to experience abnormal (>4%) acid reflux

according to 24-h pH testing [3], and up to 37% experience a relapse of symptoms during a 5-year follow-up period [4]. The latter group of patients requires titration with increasing doses of PPIs for symptom control, and some eventually require referral for more definitive surgical intervention. Treatment with PPIs also is associated with adverse drug effects, drug interaction, and long-term treatment [1].

Since the initial report of laparoscopic Nissen fundoplication (LNF) in 1991 [5], antireflux surgery has gained popularity as an alternative treatment for GERD patients [1, 2, 6, 7]. Since then, LNF has emerged as the gold standard for the surgical treatment of GERD patients [1, 8].

The relative merits of surgery still are debated due to a paucity of data from well-designed randomized controlled trials (RCTs) [9–11]. In particular, 11-year follow-up data on a limited number of patients from an earlier randomized controlled trial suggested a high recurrence of dyspeptic symptoms that required resumption of antisecretory medication [12]. This study aimed to compare optimized medical therapy with LNF for the treatment of patients with GERD who were stable and symptomatically controlled with long-term medical therapy.

Methods

The methodologic details for this two-arm, parallel, nonblinded prospective RCT were described in a previously published paper reporting 1-year follow-up data [13].

Eligibility

Patients were recruited from a combined medical–surgical GERD clinic for the diagnosis and management of GERD. Patients eligible for inclusion in this study included males and females (ages, 18–70 years) who had chronic symptoms of GERD requiring long-term therapy; prior long-term treatment using PPIs with a minimum duration of 1 year and an expected future duration of at least 2 more years; symptoms controlled before the study, defined as a GERD symptom scale (GERSS) score lower than 18 (without cough) and a score of 70 or more on a 1–100 global visual analog scale (VAS) at screening (with medication); and a baseline level of acid reflux pH exceeding 4% in 24 h (without medication). Detailed inclusion and exclusion criteria were published previously [13]. All the patients provided written informed consent.

Interventions

Patients were treated according to a management algorithm [13]. Those randomized to the medical treatment group received the same PPI dose they had taken previously to control their symptoms. Patients' symptoms were assessed at baseline. If a patient's reflux symptoms had been well controlled based on VAS and GERSS for the preceding month, the regimen was reduced by one step and the effect reassessed 1 month later. If a patient's symptoms were poorly controlled, the regimen was increased by one step. This approach provided a standardized treatment algorithm consistent with good clinical practice and was applied to the patients in both groups.

Laparoscopic Nissen fundoplication was performed by one of four laparoscopic surgeons, each with experience managing more than 50 cases. All the patients underwent repair of the esophageal hiatus and construction of a 2.5- to 3.0-cm 360° wrap. The short gastric vessels were divided routinely to achieve a floppy wrap. No nasogastric tubes or bougies were used.

Primary outcomes

GERSS

The GERSS is a published and validated instrument [14, 15]. Patients were asked to rate the following GERD symptoms: heartburn, regurgitation, bloating, dysphagia, and epigastric/retrosternal pain from the previous month. Each symptom was scored based on severity and frequency, with values varying from 0 to 12, yielding a total scale score of 0–60. Patients were considered to have controlled reflux symptoms with either medical or surgical therapy if they reported a total symptom score lower than 18.

Heartburn-free days

To calculate heartburn-free days per week, patients were asked to report the number of days in the past week that they experienced moderate to severe heartburn.

Secondary outcomes

Esophageal acid exposure

A 24-h esophageal pH monitoring study was performed according to a standard technique [6] using an ambulatory Digitrapper (Synetics Medical, Stockholm, Sweden). The pH probe was positioned 5 cm above the lower esophageal sphincter (LES), as determined previously during manometry. Gastroesophageal reflux was defined as a drop in esophageal pH to below 4, and the percentage of reflux in 24 h (percentage of time that esophageal pH was below 4) was calculated for each patient.

The baseline (pretreatment) pH studies were performed after cessation of all antisecretory medication for 5 days. The 1- and 3-year studies were performed with the medical group receiving medication and the surgical group not receiving medication.

LES pressure

Esophageal manometry was performed with a seven-lumen sleeve-sidehole catheter. The basal pressure in the LES region was measured with the sleeve sensor (Dent Sleeve, Adelaide, Australia) in relation to the gastric pressure [6].

Patient symptom control score (VAS)

All patients were asked to rate their satisfaction with symptom control on a global VAS of 0 (no relief at all) to 100 (complete symptom relief). A value of 70 was considered the threshold for symptom control. This value was derived empirically from clinical observations and systematic comparison with the GERD symptom scale [14, 16].

Health-related quality of life

The Medical Outcomes Survey Short Form-36 (SF-36) was used to capture generic health-related quality of life (HRQOL) over the 4-week period preceding the study visit. The SF-36 was completed every 3 months until completion of the study.

Follow-up evaluation

Physiologic measures of effectiveness, including esophageal manometry and 24-h pH, were evaluated at baseline (with patients not receiving medication) and at 12 months and 3 years (with patients in the medical group receiving PPI but surgical patients not receiving medications). Symptom scores (VAS and GERSS), quality-of-life assessments, and review of medications were completed by telephone or during clinic visits 1 month after surgery or 1 month after randomization for the medical group, then every 3 months up to 3 years, or in the event of premature withdrawal from the study, at the time of discontinuation.

Treatment failure

The algorithm for dealing with treatment failure was published previously [13]. Patients in the surgical group with recurrent reflux symptoms were treated with titrating doses of PPI. If medical therapy was inadequate to control reflux symptoms and objective evidence indicated failure of the antireflux barrier, the patient was offered a redo LNF.

Patients in the medical group who were inadequately controlled with maximal doses of PPI were considered treatment failures. These patients were reinvestigated to assess acid breakthrough or nonacid reflux and offered surgery if objective evidence indicated failure of antisecretory medication and correlation of symptoms with episodes of either acid or nonacid reflux.

Adverse events and safety

All adverse events were scrutinized by a Safety Monitoring Board (SMB). The SMB consisted of a gastrointestinal surgeon and a gastroenterologist not involved with performance of the study. The SMB used clearly specified criteria to determine the seriousness of the adverse event and its relationship to the medical or surgical intervention. All serious adverse events were reported immediately to the investigators, and appropriate therapy or the continued participation of the patient was discussed. Any serious adverse events were reported to the institutional research ethics committees and the patient's referring physician. Patients who discontinued the study prematurely returned for a final visit and underwent the same procedures required for the 12-month assessment.

The study was approved by the Research Ethics Board (REB) of St Joseph's Healthcare, and all the requirements of the REB, including regular progress reports, were met.

Statistical analysis

Sample size calculations and randomization methods were described previously [13]. In brief, conventional methods were used to calculate a sample size based on the primary effectiveness measure, GERSS. A sample size of 216 was calculated at a statistical significance level of $\alpha = 0.05$ and powers of 0.8 and 0.9. Randomization was stratified according to previous PPI use (once per day or 2 or more times per day) and *Helicobacter pylori* status (absent or present). Block randomization was used to ensure 1:1 treatment allocation.

At baseline, each group consisted of 52 patients with mostly complete data. Fewer data were available at the 3 year follow-up visit. Because we were concerned about the possible biases created by these missing data, we used multiple imputation with SAS version 9.1 (SAS, Cary, NC, USA) and implemented 10 copies of replacements for missing data using all the data we had available on the patients, including data at baseline, 3 months, 6 months, 9 months, 12 months, and 15 months, to generate the missing data either at baseline or at 12 or 36 months.

Ten copies of the replacement data were used to generate the missing values for each patient, and these missing values then were used to create the inference data shown in

Table 1 Comparison between medicine and surgery groups at 3 years follow-up

	Medical therapy	Surgery	All patients	Estimated difference between groups (at 3 years), with baseline as covariate	
				Estimate [95% CI]	*p* value
GERSS score	9.05 ± 10.40	6.21 ± 8.66	7.58 ± 9.58	2.66 [−1.11, 6.43]	0.1660
HB-free days	5.98 ± 1.82	6.81 ± 0.66	6.41 ± 1.41	−1.35 [−2.35, −0.36]	0.0077
% Time pH < 4	4.29 ± 6.66	2.11 ± 3.84	3.20 ± 5.51	2.96 [−0.87, 6.79]	0.1301
LES (mmHg)	7.41 ± 6.54	13.63 ± 4.51	10.12 ± 6.50	−5.85 [−8.84, −2.85]	0.0002
VAS	81.95 ± 14.25	92.67 ± 11.49	87.50 ± 13.90	−10.16 [−17.82, −2.51]	0.0093
SF-36, GH score	71.41 ± 21.73	78.50 ± 19.76	75.04 ± 20.92	−12.19 [−21.72, −2.65]	0.0124

Data presented as mean ± standard deviation. Analysis included differences as medical minus surgery, stratification, blocking, and randomization

GERSS gastroesophageal reflux disease (GERD) symptom scale score, *HB* heartburn, *LES* lower esophageal sphincter, *VAS* visual analog scale, *GH* general health

Table 1. The descriptive statistics are listed as counts and percentages for rate data with means and standard deviations for continuously measured responses. In addition, the analysis at 36 months was performed without adjusting for baseline as the first set of *p* values, taking into account the stratification and blocking in the analysis using analysis of variance (ANOVA). All analyses were conducted using SAS version 9.1.

Results

Of the 1,666 patients screened, 180 were eligible for randomization, and 104 gave informed consent. Randomization assigned 52 patients (29 men and 23 women; mean age, 42.9 years) to the surgical group and 52 patients (26 men and 26 women; mean age, 42.1 years) to the medical group. The remaining 76 eligible patients refused to

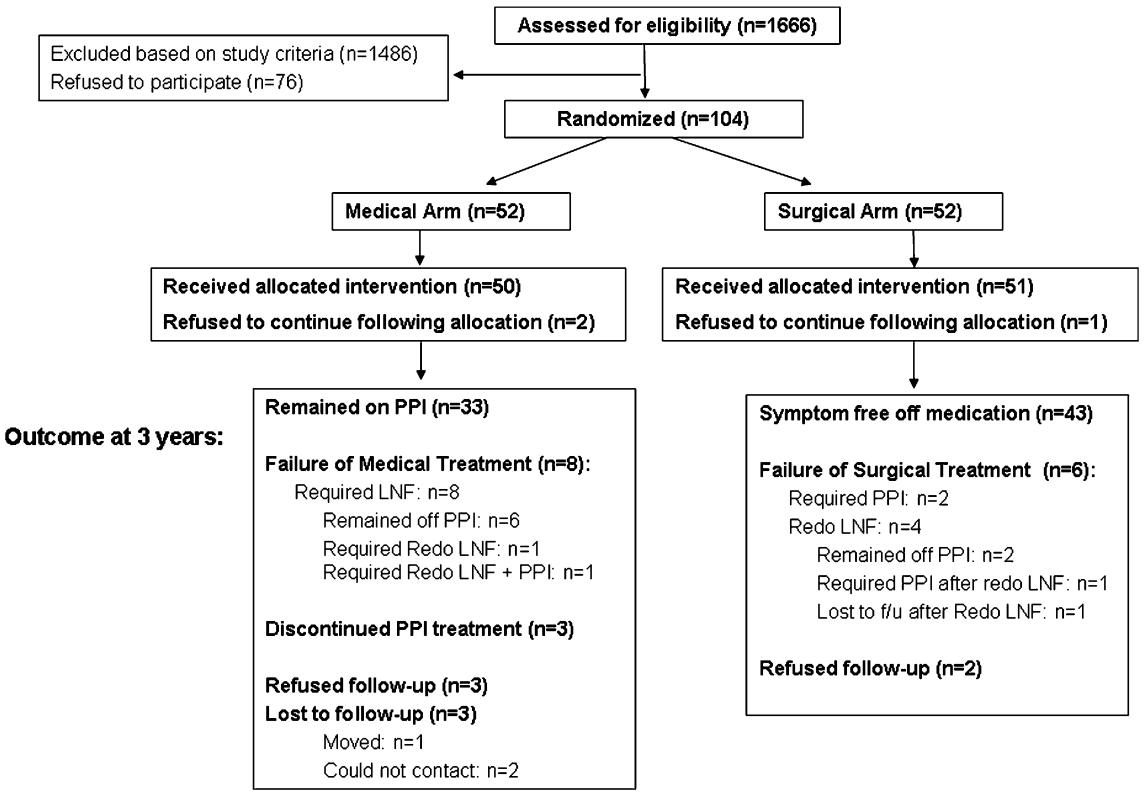


Fig. 1 Consort diagram

participate. Of the 104 randomized patients, 3 withdrew from the study immediately (2 medical patients and 1 surgical patient) (Fig. 1). Baseline and 1-year data have been described previously [13]. For the 3-year follow-up assessment, 93 patients were available.

Among the patients screened, the most common reasons for ineligibility were lack of symptom control (VAS score < 70 for 1,101 patients) and PPI usage less than 1 year (133 patients). Of the remaining patients screened and found to be ineligible for the study, 167 had a normal 24-h pH study, and 38 were not prepared to follow the protocol for 3 years.

Primary outcomes

GERSS

At the 3-year follow-up assessment, the two groups did not differ in terms of GERSS because both interventions were effective in reducing GERSS scores (Fig. 2). The mean GERSS score was well below 18 in both treatment groups at 3 years (6.21 ± 8.66 in the surgical group and 9.05 ± 10.40 in the medical group), indicating adequate control of reflux symptoms (Table 1). The improvement in GERSS from baseline (with no PPI use) was significant in both groups ($p < 0.0001$), but the two groups did not differ significantly in change from baseline at 3 years (Table 1).

Heartburn-free days

The surgical group improved significantly compared with medical group in terms of heartburn-free days. The surgical patients experienced a mean of 1.35 more heartburn-free days per week (95% confidence interval [CI], 0.36–2.35; $p = 0.0077$) (Fig. 3).

Secondary outcomes

Acid reflux

Compared with baseline, both groups improved with respect to percentage of acid reflux in the esophagus according to 24-h pH monitoring (Fig. 4). In the surgical group, the mean percentage of time that pH was less than 4 improved from a mean of 10.26 ± 11.61 at baseline (with no medication use) to 2.11 ± 3.84. The medical group improved from a mean baseline value of 9.46 ± 5.70 to a value of 4.29 ± 6.66. Although in the medical group, the mean acid reflux time in 24 h with medication was in the abnormal range, the groups did not differ significantly at 3 years in change from baseline ($p = 0.1301$) (Table 1).

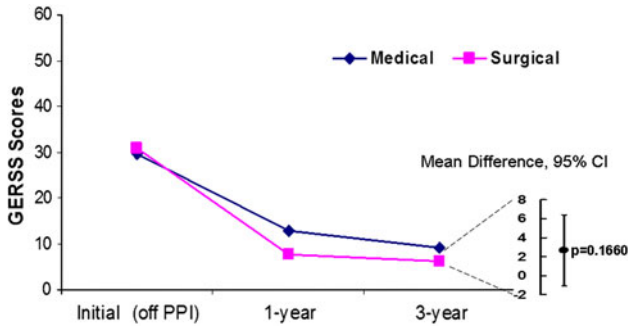


Fig. 2 Gastroesophageal reflux disease (GERD) symptom scale. *CI* confidence interval

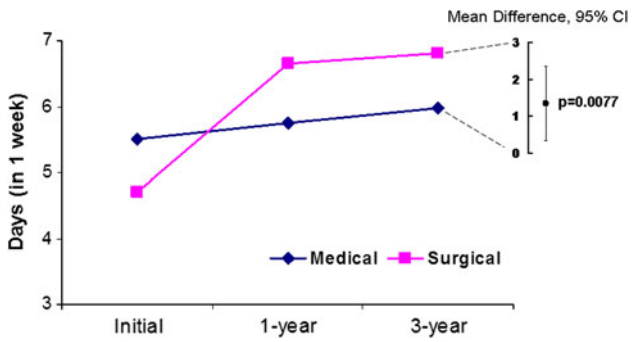


Fig. 3 Heartburn-free days. *CI* confidence interval

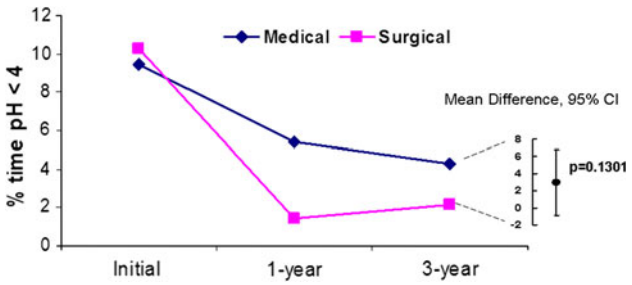


Fig. 4 The 24-h esophageal acid exposure. *CI* confidence interval

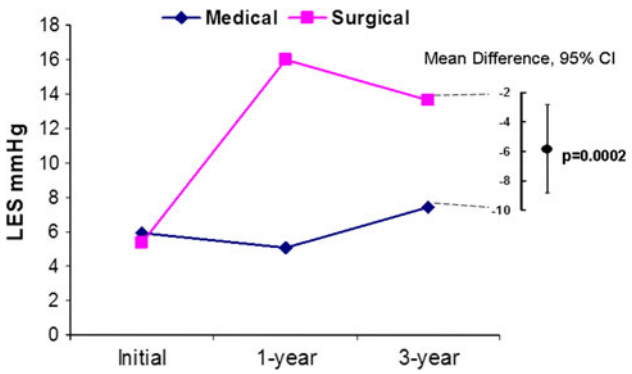


Fig. 5 Pressure recorded at the lower esophageal sphincter (LES). *CI* confidence interval

Esophageal manometry

At the initial assessment, the two groups did not differ in the pressure recorded at the LES. The mean pressure (mmHg) was 5.92 ± 4.80 in the medical group and 5.38 ± 3.54 in the surgical group. However the surgical group had significantly better LES pressure at 3 years than the medical group. As seen in Fig. 5, the mean difference in LES pressure between the groups was -5.85 (95% CI, -8.84, -2.85; $p = 0.0002$).

VAS and SF-36

As seen in Fig. 6, the medical patients controlled with medical therapy at baseline maintained their symptom control, as indicated by the VAS score at 3 years (81.95 ± 14.25 vs 82.60 ± 10.790 at baseline). The surgical group, however, showed a statistically significant improvement in global symptom control, with VAS scores increasing from a mean of 81.79 ± 12.59 at baseline to 92.67 ± 11.49 ($p = 0.0072$). The mean difference in scores was -10.16 (95% CI, -17.82, -2.51; $p = 0.0093$).

Quality of life

As seen in Fig. 7, the surgical group also experienced a significant improvement in quality of life according to the general health subscore of the SF-36 compared with the

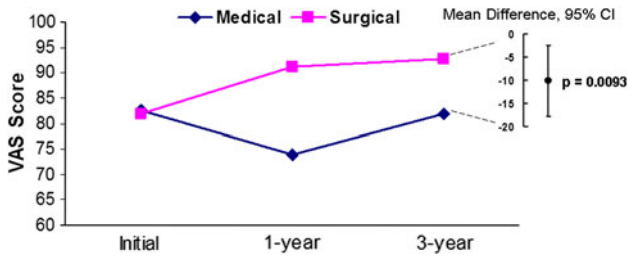


Fig. 6 Visual analog scale (VAS). *CI* confidence interval

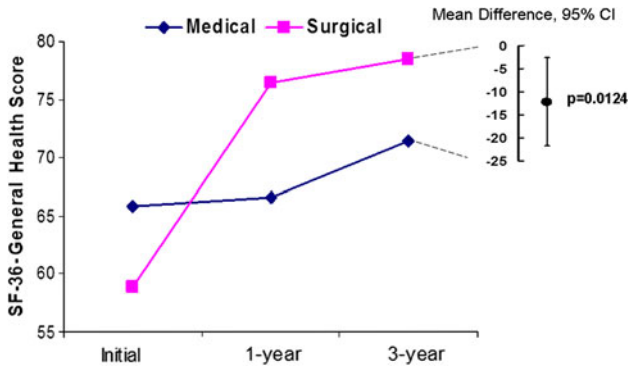


Fig. 7 Quality of life (SF-36)–general health. *CI* confidence interval

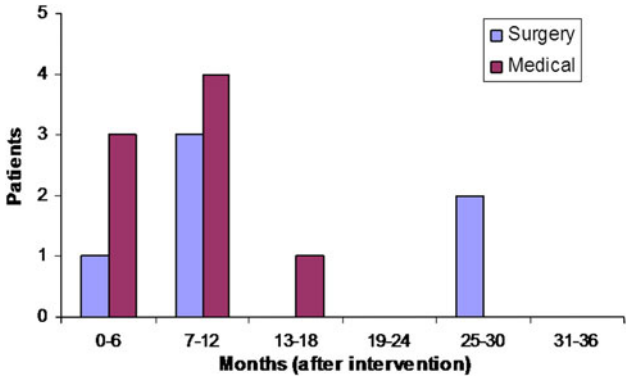


Fig. 8 Time of symptom recurrence

medical group, with a mean difference in scores of -12.19 (95% CI, -21.72, -2.65; $p = 0.0124$).

Treatment failure

At the 3-year follow-up assessment, six patients in the surgical group (11.8%) and eight patients (16%) in the medical group had failed their primary treatment. All six surgical patients who failed treatment resumed PPI therapy, and four patients had a redo LNF. Of these four patients, one required further PPI treatment and one could not be contacted after redo LNF.

As seen in Fig. 8, four of the six surgical patients who failed primary treatment reported symptom recurrence during the first year and two patients during the third year. Eight patients in the medical group required LNF after unsuccessful medical therapy. Among the eight medical failures, seven patients reported recurrent symptoms during year 1, and one patient reported symptoms during year 2. Two of these eight patients required a redo LNF, and one patient required further PPI therapy after the redo LNF.

Adverse events

All surgeries were completed laparoscopically with no intraoperative complications. Seven patients experienced minor postoperative complications. No major morbidities and no mortality occurred in the group. The mean hospital stay was 2.8 ± 1.3 days (range, 1–8 days). Four LNF patients reported dysphagia (GERSS > 6/12), and seven patients experienced postprandial bloating (GERSS > 6/12) at 3 months. Two patients required single dilation of the wrap for the aforementioned symptoms. No adverse effects were reported with medical therapy.

Post-3-year follow-up assessment

Patients will continue to be followed up on a once yearly basis with symptom assessment including VAS and

GERSS. At 5 years, full testing will be repeated. At the 3 year follow-up assessment, the medical patients who did not require surgery during the study were given the choice of having an LNF based on the data presented. At this writing, 15 of 36 medical patients (41.7%) have chosen to have surgery.

Discussion

This study is the first RCT to provide a long-term comparison of laparoscopic Nissen fundoplication with optimized medical therapy using PPIs. The study provides evidence that for GERD patients controlled with long-term PPI therapy, laparoscopic Nissen fundoplication may provide an effective alternative to medical therapy, with minimal morbidity and side effects, when performed by experienced surgeons.

Although surgery was no better than medical therapy with respect to the GERD symptom scale, it was found to be superior to medical therapy in increasing heartburn-free days. This may have been reflected in the improved rating of symptom control in the VAS score reported by the patients. Furthermore, LNF lowered the esophageal acid exposure into the normal range, whereas the medical therapy (despite symptomatic dose titration) reduced the acid reflux to levels still considered abnormal. This difference compared with baseline was not statistically significant, but it may have been due to low power.

Our results are similar to those of another RCT comparing laparoscopic antireflux surgery with PPI therapy [11], which demonstrated improved symptom response and health well-being in patients 1 year after surgery. The long-term results of this study are still awaited. For our study patients, most of the surgical failures occurred after the first year. Despite this, most of the patients chose revisional surgery over resuming long-term medical therapy. We found that treatment failures in both groups had a more complex course and that the patients were more likely to fail again irrespective of whether medical or surgical therapy was adopted. This finding was supported by an earlier observation by Oelschlager et al. [17], who found patients failing an antireflux surgery were more likely to fail the second repair than the original cohort.

The strength of our study was in its comprehensive evaluation of patients using a number of accepted subjective and objective physiologic variables. We had excellent patient compliance with the protocol and a close patient follow-up response. As a result, our dropout rate was relatively low compared with previous similar studies [4, 9].

The primary weakness of our study was its low recruitment rate, which resulted in our failure to recruit the desired number of patients. This was in great part due to our strict inclusion criteria, which required excellent symptom control (GERSS < 18 and VAS > 70) and continuous use of PPI for at least 1 year. We found that many patients receiving long-term medical therapy had less than optimal control of their symptoms, and many stopped and started the medication despite their experience of unwanted GERD symptoms during these periods. Thus, when these patients were offered a minimally invasive surgery that could provide an effective alternative to long-term PPI therapy, many elected surgery instead of agreeing to be randomized to either surgery or medical therapy. This was supported by the observation that at the 3 year follow-up assessment, more than 40% of the medical group chose to undergo surgery despite good control of their symptoms with PPI therapy.

One consequence of the small sample size was inability to detect a significant difference in the change from baseline in the esophageal acid exposure (pH < 4) between the two treatment groups. However, it should be noted that surgery produced normalization of esophageal acid exposure, whereas optimization of medical therapy based on symptom control still was associated with acid exposure time exceeding 4%.

This observation is supported by another study, which showed that symptomatic titration of PPI dosage does not necessarily lead to normalization of esophageal acid exposure [3]. In this study, Gerson et al. [3] observed normalization of pH in only 58% of GERD patients and 50% of Barrett's patients when the PPI dose was titrated against symptoms.

Over the past two decades, seven randomized trials have compared antireflux surgery with optimized medical therapy. Each of these studies found surgery to be superior to medical therapy in controlling symptoms and patient well-being. Nevertheless, debate about the relative merits of surgery persists. Recently published algorithms for the treatment of GERD patients do not offer surgery as an option even to patients with repeated failures of medical therapy [18]. Our study supports earlier studies in suggesting that LNF is an effective therapy over time and should be offered to all patients receiving long-term PPI therapy who are seeking an alternative.

Acknowledgments We thank the following individuals for their contributions to this study: Lianna Butler and Sonja DePauw (study coordination), Drs. Cliff Sample, Herawaty Sebajang, and Laurent Biertho (patient care), Drs. Margherita Cadeddu and Robert Goodacre (Safety Monitoring Board), Leslie Deluca, Meredine Haynes, Jennifer McGugan, and Denise Stys Norman (data collection and patient follow-up), Paula McKay, and Karen Barlow (manuscript preparation), and Dr. Gary Foster (data management and detailed statistical analysis).This study was supported by grants from the Canadian Institute of Health Research and the Ontario Ministry of Health and Long-term Care.

Disclosures Mehran Anvari, Christopher Allen, John Marshall, David Armstrong, Ron Goeree, Wendy Ungar, and Charles Goldsmith have no conflicts of interest or financial ties to disclose.

References

1. Hartmann J, Jacobi CA, Menenakos C, Ismail M, Braumann C (2008) Surgical treatment of gastroesophageal reflux disease and upside-down stomach using the Da Vinci robotic system: a prospective study. J Gastrointest Surg 12:504–509
2. Katelaris PH (2004) An evaluation of current GERD therapy: a summary and comparison of effectiveness, adverse effects and costs of drugs, surgery, and endoscopic therapy. Best Pract Res Clin Gastroenterol 18(Suppl):39–45
3. Gerson LB, Boparai V, Ullah N, Triadafilopoulos G (2004) Oesophageal and gastric pH profiles in patients with gastro-oesophageal reflux disease and Barrett's oesophagus treated with proton pump inhibitors. Aliment Pharmacol Ther 20:637–643
4. Lundell L, Miettinen P, Myrvold HE, Pedersen SA, Liedman B, Hatlebakk JG, Julkonen R, Levander K, Carlsson J, Lamm M, Wiklund I (2001) Continued (5-year) follow-up of a randomized clinical study comparing antireflux surgery and omeprazole in gastroesophageal reflux disease. J Am Coll Surg 192:172–179
5. Dallemagne B, Weerts JM, Jehaes C, Markiewicz S, Lombard R (1991) Laparoscopic Nissen fundoplication: preliminary report. Surg Laparosc Endosc 1:138–143
6. Anvari M, Allen C (2003) Five year comprehensive outcomes evaluation in 181 patients after laparoscopic Nissen fundoplication. J Am Coll Surg 196:51–57
7. Hinder RA, Filipi CJ, Wetscher G, Neary P, DeMeester TR, Perdikis G (1994) Laparoscopic Nissen fundoplication is an effective treatment for gastroesophageal reflux disease. Ann Surg 220:472–481; discussion 481–483
8. Gould JC, Melvin WS (2002) Computer-assisted robotic antireflux surgery. Surg Laparosc Endosc Percutan Tech 12:26–29
9. Spechler SJ (1992) Comparison of medical and surgical therapy for complicated gastroesophageal reflux disease in veterans. N Engl J Med 326:786–792
10. Sontag SJ, O'Connell S, Khandelwal S, Greenlee H, Schnell T, Nemchausky B, Chejfec G, Miller T, Seidel J, Sonnenberg A (2003) Asthmatics with gastroesophageal reflux: long-term results of a randomized trial of medical and surgical therapies. Am J Gastoenterol 98:987–999
11. Mahon D, Rhodes M, Decadt B, Hindmarsh A, Lowndes R, Beckingham I, Koo B, Newcombe RG (2005) Randomized clinical trial of laparoscopic Nissen fundoplication compared with proton pump inhibitors for treatment of chronic gastro-esophageal reflux. Br J Surg 92:695–699
12. Spechler SJ, Lee E, Ahnen D, Goyal RK, Hirano I, Ramirez F, Raufman JP, Sampliner R, Schnell T, Sontag S, Vlahcevic ZR, Young R, Williford W (2001) Long-term outcome of medical and surgical therapies for gastroesophageal reflux disease: follow-up of a randomized controlled trial. JAMA 285:2331–2338
13. Anvari M, Allen C, Marshall J, Armstong D, Goeree R, Ungar W, Goldsmith CH (2006) A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for treatment of patients with chronic gastroesophageal reflux disease: one-year follow-up. Surg Innov 4:238–249
14. Allen CJ, Parameswaran K, Belda J, Anvari M (2000) Reproducibility, validity, and responsiveness of a disease-specific symptom questionnaire for gastroesophageal reflux disease. Dis Esophagus 13:65–70
15. Maddern GJ, Chatterton BE, Collins PJ, Horowitz M, Shearman DJ, Jamieson GG (1985) Solid and liquid gastric emptying in patients with gastro-oesophageal reflux. Br J Surg 72:344–347
16. Anvari M, Allen C (2003) Surgical outcome in gastro-esophageal reflux disease patients with inadequate response to proton pump inhibitors. Surg Endosc 17:1029–1035
17. Oelschlager BK, Lal DR, Jensen E, Cahill M, Quiroga E, Pellegrini CA (2006) Medium- and long-term outcome of laparoscopic redo fundoplication. Surg Endosc 20:1817–1823
18. Tytgat GN, McColl K, Tack J, Holtmann G, Hunt RH, Malfertheiner P, Hungin AP, Batchelor HK (2008) New algorithm for the treatment of gastro-oesophageal reflux disease. Aliment Pharmacol Ther 27:249–256

Six to 12-year outcomes of magnetic sphincter augmentation for gastroesophageal reflux disease

Article *in* Scientific Reports · August 2020

DOI: 10.1038/s41598-020-70742-3

6 authors, including:

Davide Ferrari
University of Milan
12 PUBLICATIONS **53** CITATIONS
SEE PROFILE

Emanuele Asti
I.R.C.C.S. Policlinico San Donato
122 PUBLICATIONS **755** CITATIONS
SEE PROFILE

Veronica Lazzari
University of Milan
21 PUBLICATIONS **231** CITATIONS
SEE PROFILE

Stefano Siboni
University of Milan
79 PUBLICATIONS **513** CITATIONS
SEE PROFILE

SCIENTIFIC REPORTS
nature research

OPEN

Six to 12-year outcomes of magnetic sphincter augmentation for gastroesophageal reflux disease

Davide Ferrari[1], Emanuele Asti[2], Veronica Lazzari[2], Stefano Siboni[2], Daniele Bernardi[2] & Luigi Bonavina[1,2]✉

The magnetic sphincter augmentation (MSA) device has been proven safe and effective in controlling typical reflux symptoms and esophageal acid exposure for up to 6-year follow-up. Longer term outcomes have not been reported yet. A prospectively maintained database was reviewed to assess long-term safety and efficacy of the laparoscopic MSA procedure at a single referral center. Gastro-Esophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL), use of proton-pump inhibitors (PPI), and esophageal acid exposure were compared to baseline. Favorable outcomes were defined as ≥ 50% improvement of GERD-HRQL total score and PPI discontinuation. Between March 2007 and March 2020, 335 patients met the study inclusion criteria, and 124 of them were followed from 6 to 12 years after surgery (median 9 years, IQR 2). Mean total GERD-HRQL score significantly improved from 19.9 to 4.01 ($p < 0.001$), and PPI were discontinued by 79% of patients. The mean total percent time with pH < 4 decreased from 9.6% at baseline to 4.1% ($p < 0.001$), with 89% of patients achieving pH normalization. Independent predictors of a favorable outcome were age at intervention < 40 years (OR 4.17) and GERD-HRQL score > 15 (OR 4.09). We confirm long-term safety and efficacy of MSA in terms of symptom improvement, decreased drug dependency, and reduced esophageal acid exposure.

The global burden of gastroesophageal reflux disease (GERD) is enormous, with a pooled prevalence of 13.3% in community-based studies[1]. Symptoms and complications of GERD persist in up to 40% of patients treated with proton-pump inhibitors (PPI)[2,3], and fundoplication is largely underused because of the steep learning curve and reported variability in outcomes[4]. The aim of fundoplication is to restore lower esophageal sphincter (LES) function by remodeling the esophagogastric junction[5]. Both total (Nissen) and partial (Toupet) fundoplication procedures require mobilization of the gastric fundus to be wrapped around the distal esophagus[6]. To further enhance the antireflux barrier, a crural diaphragmatic repair is routinely added[7]. Currently, the fact that fundoplication is offered to less than 1% of the GERD population may have an impact on the progression of symptoms and the development of Barrett's esophagus[8].

The magnetic sphincter augmentation (MSA) procedure (Linx Reflux Management System, Ethicon, Johnson & Johnson, Shoreview, Mn, USA) was developed as a less disruptive and more standardized and reproducible laparoscopic surgical option for the treatment of GERD. The MSA device is composed by a variable number of interlinked titanium beads with a magnetic core inside. This ring-like system produces a magnetic force that augments the LES. The first feasibility trial and a large prospective nonrandomized study were published in 2008[9] and 2013[10], respectively. The MSA procedure has been granted approval for clinical use by the Food and Drug Administration in 2012. Previous reports from our group have shown feasibility, safety, and efficacy of the MSA procedure up to 6 years of follow-up[11,12]. We now provide the long-term outcomes of a cohort of patients followed for a minimum of 6 years.

[1]Department of Biomedical Sciences for Health, University of Milan, Milan, Italy. [2]Division of General and Foregut Surgery, IRCCS Policlinico San Donato, Piazza E. Malan, 1, 20097 San Donato Milanese, MI, Italy. ✉email: luigi.bonavina@unimi.it

Subjects and methods

The study was a single-center, retrospective, single-arm study, where patients served as their own controls. The study protocol was approved by the Internal Review Board of IRCCS Policlinico San Donato (HSD 2019-072), and the research was performed in accordance with the relevant regulations. Informed consent was obtained from all study participants. The prospectively maintained database was reviewed to assess long-term safety and efficacy of the MSA. All patients who underwent a MSA procedure between March 2007 and March 30, 2020 were included in the study. Data analysis was performed in the whole group of patients and in a cohort of individuals followed for 6- to 12 years. Gastro-Esophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL), use of proton-pump inhibitors (PPI), and esophageal pH monitoring parameters were compared to patients' own preoperative data. Favorable outcome of the MSA procedure was defined as ≥ 50% improvement in GERD-HRQL total score and PPI discontinuation.

Preoperative assessment and patient selection. Before surgery, all patients referred for surgical therapy of GERD were evaluated by a multidisciplinary team including gastroenterologists, dieticians, and clinical psycologists at our center. The diagnostic assessment included the foregut symptom questionnaire and the GERD-HRQL questionnaire, upper gastrointestinal endoscopy, barium swallow study, ambulatory esophageal pH monitoring, and esophageal manometry. The foregut symptom questionnaire gives a score for heartburn, regurgitation, dysphagia, and chest pain on a scale of 0 to 4 (grade 0, none; grade 1, less than once a week; grade 2, several times a week; grade 3, daily, affecting lifestyle; grade 4, always, markedly affecting lifestyle. The GERD-HRQL score consists of 10 questions that specifically address GERD symptoms. Each question has a score ranging from 0 to 5, and the total score ranges from 0 to 50[13]. Preoperative determination of hernia size was based on upper gastrointestinal endoscopy and barium swallow study. Endoscopy evaluated the presence of hiatus hernia, morphology of the gastroesophageal valve using the Hill classification, grade of esophagitis using the Los Angeles classification, and the presence of Barrett esophagus using the Prague classification. Ambulatory esophageal pH monitoring was performed using trans-nasal pH-impedance equipment or the Bravo wireless system (48–96 h pH study). Measurements collected from esophageal pH testing included the DeMeester score and its individual components. A standard or, more recently, a high-resolution manometry were performed[14]. Main parameters investigated were the resting pressure and length of the LES, the distal esophageal amplitude (DEA) and/or the distal contractile integral (DCI), and the percent of effective contractions. An esophageal amplitude of less than 40 mm Hg and a greater than 50% of non-transmitted swallows indicated ineffective esophageal motility. Initial criteria for patient selection were the following: persistent reflux symptoms despite optimal PPI therapy, abnormal esophageal acid exposure confirmed by ambulatory esophageal pH monitoring, hiatus hernia < 3 cm, esophagitis < grade B, body mass index < 35 kg/m^2, and absence of specific motility disorders. With further clinical experience and research, the criteria have been expanded to include patients with larger hiatus hernia, short Barrett's esophagus, and mild esophageal dysmotility. The MSA procedure was not offered to patients with recurrent GERD after failed fundoplication or other surgical/endoscopic procedures at the esophagogastric junction, and to those with known history of nickel allergy or eating disorders.

Surgical approach. The MSA device was implanted via laparoscopy as previously described[15]. Under general anesthesia, the esophago-gastric junction was exposed following incision of the peritoneal reflection. The posterior vagus nerve was identified and separated from the esophagus for a length of about 1 cm. No short gastric vessels were divided. The esophageal circumference was measured with an appropriate magnetic sizer inserted through the retroesophageal tunnel. A minimal or formal posterior crura repair was performed depending on the size of the hiatal defect and the degree of hiatus hernia. Over the study period and starting from 2014, modifications of the surgical technique occurred. First, formal mediastinal dissection became routine practice; second, a new generation MSA device was introduced for use in magnetic resonance up to 1.5 T; and, third, a new generation sizer device for measuring the esophageal circumference was introduced.

Postoperative assessment and follow-up. All patients underwent a comprehensive clinical evaluation including the foregut symptom and the GERD-HRQL questionnaires, use and dosage of PPI, esophageal pH measurements, upper gastrointestinal endoscopy, and esophageal manometry. The treatment was considered successful if, compared to baseline, at least a 50% reduction in the total GERD-HRQL score and PPI discontinuation or at least a 50% dose reduction was achieved.

Statistical analysis. Continuous data are reported as median ± interquartile range (IQR) or mean ± standard deviation (SD). Patients served as their own control, and pre- and postoperative data were compared using the two-tailed, paired Student's t test. Categorical demographic and baseline variables are reported as proportions or frequencies, and compared using Wilcoxon test for continuous outcomes and McNemar's test for paired samples. A p value < 0.05 was considered statistically significant. Parameters of the univariate analysis with $p < 0.05$ were included in a multivariate logistic regression test to determine what independent variables might predict the clinical success of the surgical procedure. A p value < 0.05 was considered statistically significant. Statistical analyses were performed using SPSS software 23.0 (IBM, Armonk, New York, U.S.).

Results

Between March 2007 and March 30, 2020, a total of 1,052 patients underwent laparoscopic surgery for GERD at our institution. A Toupet fundoplication was performed in 499 patients, Nissen fundoplication in 218, and MSA procedure in 335. The baseline demographic and clinical characteristics of patients who received MSA are

	F.U. < 6 years (n = 211)	F.U. 6–12 years (n = 124)
Age, years	46 (20)	44 (20.8)
Male, n (%)	139 (65.8)	83 (66.9)
BMI*, kg/m^2	25.4 (5)	23.9 (4.5)
Duration of symptoms, years	8 (11.3)	6 (7)
PPI† use, years	7 (7)	4 (6)
GERD-HRQL§ total score	19.5 (10)	21 (9.5)
Esophagitis, n (%)		
None	167 (79.1)	103 (83.1)
Grade A	22 (10.4)	11 (8.9)
Grade B	18 (8.5)	9 (7.2)
Grade C	2 (1.0)	1 (0.8)
Grade D	2 (1.0)	0 (0.0)
Barrett's esophagus, n (%)	10 (4.7)	4 (3.2)
Hiatal hernia length, n (%)		
None	57 (27.0)	18 (14.5)
1 cm	24 (11.4)	37 (29.8)
2 cm	70 (33.2)	44 (35.6)
3 cm	35 (16.6)	20 (16.1)
≥ 4 cm	25 (11.8)	5 (4.0)
Basal LESP‡, mmHg	14.2 (15.4)	15.1 (12)
DEA※, mmHg	66 (40)	63.0 (34.2)
IEM$^\delta$, n (%)	19 (9.0)	1 (0.8)
DeMeester score	24.8 (26.8)	31.3 (24.6)
% total time pH < 4	6.4 (6.8)	8 (6.6)

Table 1. Baseline patient characteristics (continuous variables expressed as median (IQR)). *BMI = Body Mass Index; †PPI = Proton Pump Inhibitors; §GERD-HRQL = Gastro-Esophageal Reflux Disease Health Related Quality of Life; ‡LESP = Lower Esophageal Sphincter Pressure; ※DEA = Distal Esophageal Amplitude; $^\delta$IEM = Ineffective esophageal motility.

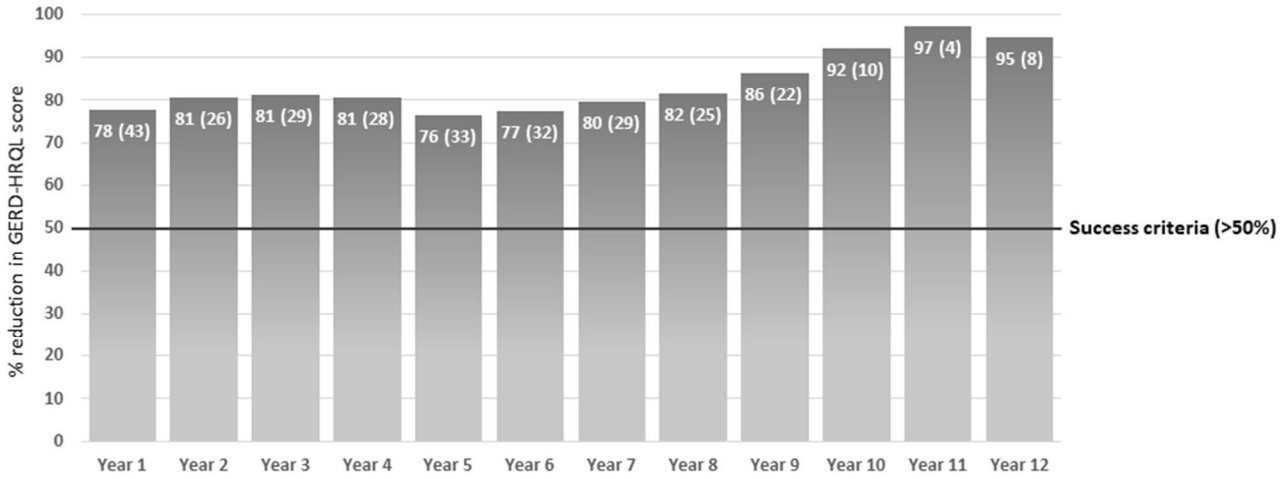


Figure 1. Average percent reduction (± SD) of total GERD-HRQL score per year over the follow-up.

listed in Table 1. Two patients died during the follow-up for unrelated reasons. Overall, there was more than 50% reduction in the total GERD-HRQL score compared to baseline in each year of follow-up (Fig. 1). Table 2 shows the median GERD-HRQL scores by question.

Postoperative adverse events and long-term safety profile. Adverse events were assessed from the time of implant through to the final visit. The rate of procedure-related adverse events was 11.6% (39/335) throughout the overall study period.

Eight patients (2.4%) required a single endoscopic pneumatic dilation due to persistent dysphagia at 11, 13, 21, 23, 28, 53, 60, and 65 months, respectively, after surgery. Thirty-one patients (9.2%) required laparoscopic

	Baseline	< 6 years	6–12 years
	n = 124	n = 211	n = 124
How bad is your heartburn?	4 (2)	2 (2)	2 (2)
Heartburn when lying down?	4 (2)	0 (0)	0 (0)
Heartburn when standing up?	3 (2)	0 (0)	0 (0)
Heartburn after meals?	4 (2)	1 (2)	0 (2)
Does heartburn change your diet?	2 (2)	0 (1)	0 (0)
Does heartburn wake you from sleep?	2 (3)	0 (0)	0 (0)
Do you have difficulty swallowing?	0 (0)	0 (0)	0 (0)
Do you have bloating and gassy feelings?	0 (2)	1 (2)	0 (1.5)
Do you have pain with swallowing?	0 (0)	0 (0)	0 (0)
If you take medication, does this affect daily life?	0 (1)	0 (0)	0 (0)
Total median GERD-HRQL* score	21 (9.5)	4 (5)	3 (5.5)

Table 2. Summary of median (IQR) GERD-HRQL scores by question. *GERD-HRQL = Gastro-Esophageal Reflux Disease Health Related Quality of Life.

	< 6 years (n = 28)	6–12 years (n = 3)
Erosion	6	0
Regurgitation	6	0
Heartburn	5	1
Dysphagia	5	1
"Foreign body" sensation	2	0
Odynophagia	1	0
Pharyngodinia	1	0
Chronic cough	1	0
Need of magnetic resonance study	1	1

Table 3. Main reasons for magnetic sphincter augmentation device removal.

device removal for various reasons (Table 3). The most common one-stage remedial procedure was a laparoscopic Toupet fundoplication (n = 18).

Long-term (6–12 year) outcomes. One-hundred-twenty-four patients, who were implanted between March 2007 and February 2014, had a minimum follow-up of 6 years. The median follow-up was 9 years (IQR 2). At the latest follow-up, 92 of 124 patients (74.2%) did not report any esophageal symptom (grade 0–1 for heartburn, regurgitation, dysphagia, and chest pain). The mean total GERD-HRQL score decreased from 19.9 at baseline to 4.01 at the latest follow-up ($p < 0.001$); 89% of patients met the criteria of favorable long-term outcome, Clinically significant improvement in GERD-HRQL is also reflected by the reported patient satisfaction, which was achieved in 92.7% of patients. The prevalence of grade 2–4 regurgitation significantly decreased from 59.6% at baseline to 9.6% postoperatively ($p < 0.01$). At the latest follow-up, complete or at least 50% reduction in the average daily dose of PPI was achieved by 79% and 89.5% of patients, respectively.

The majority of patients (86.3%) underwent upper gastrointestinal endoscopy after 6 years of follow-up. Hiatus hernia was found in 7 patients (6.5%), grade A esophagitis in 5 patients (4.7%), and incomplete intestinal metaplasia in 3 (2.8%). Four additional patients, who had been treated with radiofrequency ablation for short-segment Barrett's esophagus without dysplasia before the MSA procedure, were followed endoscopically for up to 8 years without recurrence of intestinal metaplasia. The Hill grade was measured in 45 patients before and after surgery. At the latest endoscopic follow-up, 41 patients (91%) retained their preoperative Hill grade I or improved, 3 (7%) remained stable, and in 1 (2%) patient the Hill grade worsened ($p < 0.01$) (Fig. 2).

Esophageal pH testing off PPI therapy showed that the mean percentage of time that pH was < 4 decreased from 9.7% at baseline to 4.2% at latest follow-up ($p < 0.001$). All the other pH monitoring variables were significantly reduced at 6–12 years compared with baseline (Table 4). Eighty-nine percent of patients who completed esophageal pH monitoring at 6- to 12 years follow-up achieved either normal esophageal acid exposure or had at least a 50% reduction compared to baseline. Sequential pH studies were performed in 37 patients at various time intervals since surgery (Fig. 3).

Table 5 shows the long-term results in a subgroup of patients with follow-up longer than 10 years. Overall patient satisfaction, based on the question "would you undergo the operation again or recommend it to a friend?", was 93.8%.

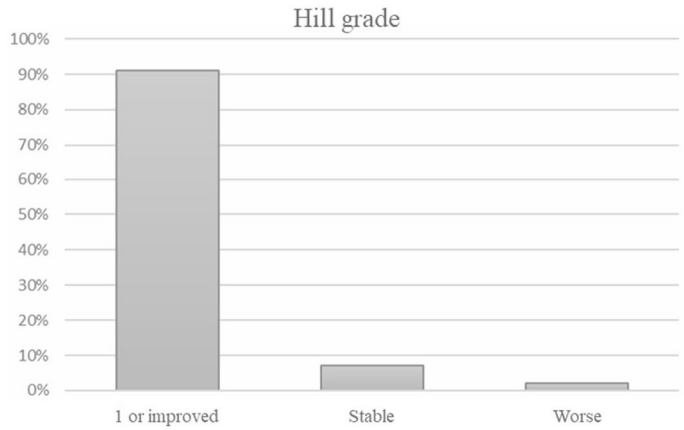


Figure 2. Changes of Hill grade classification in 45 patients after magnetic sphincter augmentation.

Measure	Baseline n = 124	6–12 years n = 91	*p*
Total time (%)			
pH < 4	9.7 (6.4)	4.2 (4.9)	< 0.001
Upright	9.7 (7.8)	4.6 (4.9)	< 0.001
Supine	8.3 (9.6)	3.3 (7.4)	< 0.001
Reflux episodes			
Total number	92.2 (92.2)	71.5 (67.7)	0.125
Number lasting > 5 min	6.1 (6.0)	4.3 (5.8)	0.036
Longest (minutes)	32.9 (34.2)	19.6 (31.5)	0.005
DeMeester score	40.7 (26.5)	16.3 (18.8)	< 0.001

Table 4. Esophageal pH measurements (mean ± SD) off proton-pump inhibitors.

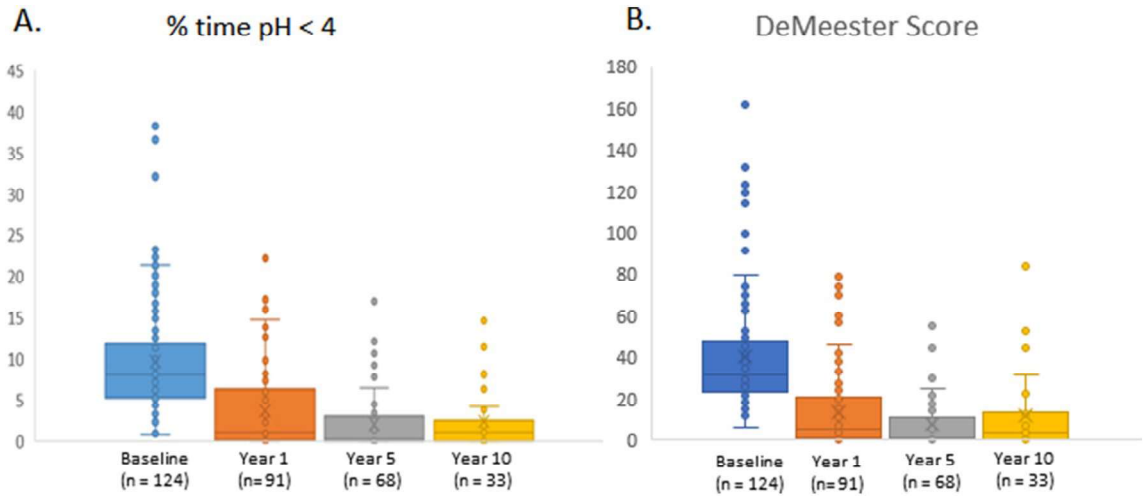


Figure 3. Median % time at pH < 4 (**A**) and median DeMeester score (**B**) in sequential pH studies compared to baseline.

Predictors of long-term clinical success. Eighty-one percent of patients had a successful clinical outcome, defined as GERD-HRQL score improvement > 50% and complete discontinuation of PPI use. At univariate analysis, age at intervention < 40 years, preoperative GERD-HRQL total score > 15, duration of symptoms, regurgitation, atypical symptoms and absence of generalized anxiety disorder were statistically significant as independent predictors of clinical success and were included in the multivariate logistic regression test

	n = 32
Median GERD-HRQL* score	2
Dysphagia	0
Ability to belch	32 (100)
Ability to vomit	29 (90.6)
Occasional PPI use	7 (21.8)
Daily PPI use	3 (9.4)
Overall patient satisfaction	30 (93.8)

Table 5. Long term results in 32 patients with follow-up > 10 years. *GERD-HRQL = Gastro-Esophageal Reflux Disease Health Related Quality of Life.

Variable	*p*
Age at intervention (< 40 y)	**0.036**
Sex, male	0.975
BMI* (< 25)	0.361
Disease duration	**0.036**
GERD-HRQL§ total score (> 15)	**0.013**
General anxiety disorder	**0.043**
Typical symptoms	0.353
Atypical symptoms	**0.019**
Regurgitation	**0.021**
Esophagitis	0.936
Hiatal hernia	0.054
Barrett's esophagus	0.489
LES basal pressure	0.600
LES overall length	0.442
Distal esophageal amplitude (< 43 mmHg)	0.468
DeMeester score	0.161
% total time pH < 4	0.409
Number of MSA‡ beads (> 14)	0.328

Table 6. Potential predictors of success at univariate logistic regression model. *BMI = Body Mass Index; §GERD-HRQL = Gastro-Esophageal Reflux Disease Health Related Quality of Life); ‡MSA = Magnetic Sphincter Augmentation. Statistically significant values/differences are indicated in bold.

Variable	OR (CI 95%)	B	SE	*p*
Age at intervention (< 40 years)	4.61 (1.29–16.45)	1.527	0.649	**0.019**
GERD-HRQL□ total score (> 15)	4.19 (1.39–12.63)	1.432	0.563	**0.011**
Duration of disease	0.97 (0.89–1.06)	−0.027	0.043	0.531
General Anxiety Disorder	0.35 (0.11–1.14)	−1.042	0.599	0.082
Atypical symptoms	2.86 (0.99–8.23)	1.051	0.539	0.051
Regurgitation	1.88 (0.68–5.19)	0.630	0.519	0.225

Table 7. Independent predictive variables of success at multivariate analysis. **OR,** odds ratio; **B,** logistic regression coefficient; **SE,** standard error. □GERD-HRQL = Gastro-Esophageal Reflux Disease-Health Related Quality of Life. Statistically significant values/differences are indicated in bold.

(Table 6). At multivariate analysis, independent predictive variables of successful outcome were confirmed to be age < 40 years and GERD-HRQL score > 15 (Table 7).

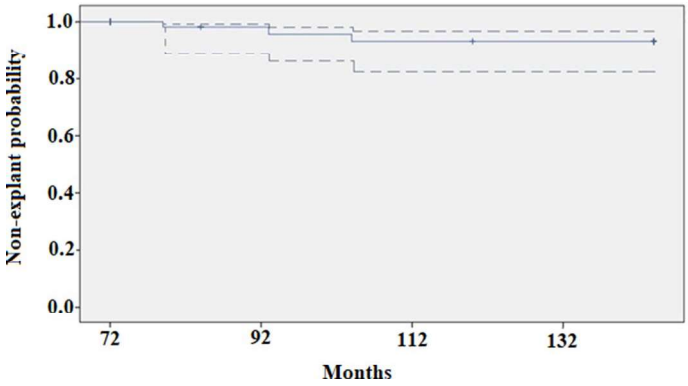


Figure 4. Kaplan–Meier estimate of explant-free probability over 6- to 12-year follow-up after magnetic sphincter augmentation (95% confidence interval indicated pointwise). [Graph created using SPSS software 23.0, URL = https://www.ibm.com/software/analytics/spss/register/].

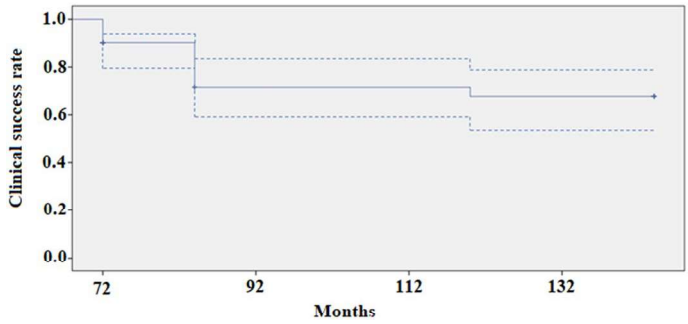


Figure 5. Kaplan–Meier estimate of clinical success (GERD-HRQL < 50% reduction or PPI discontinuation) over 6- to 12-year follow-up after magnetic sphincter augmentation (95% confidence interval indicated pointwise).[Graph created using SPSS software 23.0, URL = https://www.ibm.com/software/analytics/spss/register/].

Discussion

This 6- to 12-year follow-up study of a cohort of patients undergoing MSA confirms satisfactory and durable clinical outcomes over a median follow-up of 9 years. The present report corroborates the findings of two previous studies with up to 6-year follow-up documenting symptom relief, discontinuation of PPI, minimal side effects, and long-term safety[11,12].

The incidence of adverse events was low during the study time-frame, providing reasonable assurance that the risk of MSA complications does not increase with longer implant duration. The reasons for late device removal included dysphagia, continued reflux symptoms, and planned magnetic resonance imaging, but no erosions or migrations were observed. The overall reported rate of MSA device erosion is less than 0.5%; most events occurred within 4 years of the implant and have been managed electively without complications[16–18]. In our whole series, including 335 implants, most complications requiring MSA removal have occurred in patients implanted with a smaller (no.12 and 13) device. It should be noted that, during the study period, the sizer instrument has been replaced with a more user-friendly device in an attempt to improve the reproducibility of measurements. Furthermore, it has become clear over time that it is wiser to oversize by increasing 3 beads from the point of sizer release, and to use a larger MSA device to minimize dysphagia and decrease the likelihood of removal[19,20]. In our cohort of patients followed for 6–12 years, the overall estimated probability of MSA explant was 0.1 (Fig. 4).

The results of the present study show a 0.7 estimated probability of clinical success of at 6 to 12 years of follow-up (Fig. 5). The overall satisfaction rate of our patients was 92.5%. The prevalence of grade 2–4 regurgitation was significantly decreased ($p < 0.01$), and this is consistent with the one-year results of a recent randomized trial comparing the effect of MSA versus PPI[21]. Our multivariate analysis indicated that age < 40 years is an independent predictive variable of successful outcome. This is consistent with the study by Ayazi et al.[22] who found that male sex was also an independent predictive factor. The fact that MSA is more effective in younger and male patients is of particular interest because this operation may have a profound impact on the course of GERD if performed in an earlier disease stage[23]. In addition, a recent population-based study showed lower recurrence rates after fundoplication in young men who would otherwise require several decades of PPI therapy[24].

Although MSA was not directly compared with fundoplication in the present study, historical data from other clinical studies provide evidence that side effects typically associated with Nissen fundoplication, such

as persistent dysphagia, gas bloat, and inability to belch/vomit, are less frequent and severe after the MSA procedure[25,26]. Of interest, disease-specific quality of life was similar in a propensity-score matched analysis comparing MSA and Toupet fundoplication[27]. In the present study, gas bloat and inability to belch/vomit were reported by 4% and 1.6% of patients, respectively.

The present study also suggests that the effect of MSA on esophageal acid exposure is sustained over time, as demonstrated by sequential pH studies showing reflux control up to 12 years of follow-up. Further, in a recent study, we showed that ineffective esophageal motility detected by high-resolution manometry can reverse to normal at a median of 12 months after MSA, and the only factor significantly associated to postoperative dysphagia was the presence of preoperative dysphagia[14].

In recent years, it has become evident that the effectiveness of MSA can be enhanced by adding a formal crural repair[28,29]. The rationale behind this concept is that the extent of hiatus hernia can be underestimated both pre- and intraoperatively. Therefore, minimizing the amount of dissection performed and preserving the phreno-esophageal ligament may cause placement of the device below the true esophagogastric junction and may result in less effective reflux control. On the other hand, the MSA procedure is feasible even in large hernias based on our own clinical experience and other recent reports[30–34].

Finally, encouraging data support the hypothesis that intestinal metaplasia can reverse after MSA[35], especially in patients with short Barrett's segments and in those with normalized DeMeester score. Although experience with MSA before or after radiofrequency ablation for Barrett's esophagus is still very limited, we have successfully treated 4 patients who have been followed by endoscopy up to 8 years without recurrence.

The main limitations of this study are the retrospective design, the fact that there was no comparison treatment group, and the possible selection bias. However, despite the fact that criteria for patient selection, surgical technique, and type of device and sizer have evolved during the study period, the patient population of this study was homogeneous and indications for surgical therapy were consistent based on the preoperative pH study confirming GERD. Finally, this is the first report of a cohort of patients who completed the 6–12 year follow-up after MSA.

Conclusions

When offered as a first-line surgical option, MSA allows durable control of symptoms and esophageal acid exposure, and improves patient quality of life up to 12 years of follow-up without significant safety issues. A preoperative GERD-HRQL total score > 15 and age below 40 years are independent predictive factors of favorable outcome. Based on the above results and the high levels of patient satisfaction, MSA may represent a true paradigm shift that has the potential to fill the current therapy gap in GERD. A randomized clinical trial comparing MSA and either total or partial fundoplication could provide more robust and definitive conclusions.

Received: 9 May 2020; Accepted: 27 July 2020
Published online: 13 August 2020

References

1. Eusebi, L. H. *et al.* Global prevalence of, and risk factors for, gastro-oesophageal reflux symptoms: a meta-analysis. *Gut* **67**, 430–440 (2018).
2. El-Serag, H. B., Becher, A. & Jones, R. Systematic review: persistent reflux symptoms on proton pump inhibitor therapy in primary care and community studies. *Aliment Pharmacol. Ther.* **32**, 720–737 (2010).
3. Kahrilas, P. J. *et al.* Regurgitation is less responsive to acid suppression than heartburn in patients with gastroesophageal reflux disease. *Clin. Gastroenterol. Hepatol.* **10**, 612–619 (2012).
4. Watson, D. I., Baigrie, R. J. & Jamieson, G. G. A learning curve for laparoscopic fundoplication. Definable, avoidable, or a waste of time?. *Ann. Surg.* **224**, 198–203 (1996).
5. Pandolfino, J. E. *et al.* Restoration of normal distensive characteristics of the esophagogastric junction after fundoplication. *Ann. Surg.* **242**, 43–48 (2005).
6. Bona, D., Aiolfi, A., Asti, E. & Bonavina, L. Laparoscopic Toupet fundoplication for gastroesophageal reflux disease and hiatus hernia: proposal for standardization using the "critical view" concept. *Updates Surg.* **72**(2), 555–558 (2020).
7. Louie, B. E. *et al.* Length and pressure of the reconstructed lower esophageal sphincter is determined by both crural closure and Nissen fundoplication. *J. Gastrointest. Surg.* **17**, 236–243 (2013).
8. Inadomi, J. *et al.* Recent advances in Barrett's esophagus. *Ann. N. Y. Acad. Sci.* **1434**, 227–238 (2018).
9. Bonavina, L. *et al.* Magnetic augmentation of the lower esophageal sphincter: Results of a feasibility clinical trial. *J. Gastrointest. Surg.* **12**, 2133–2140 (2008).
10. Ganz, R. A. *et al.* Esophageal sphincter device for gastroesophageal reflux disease. *N. Engl. J. Med.* **368**, 719–727 (2013).
11. Saino, G., Bonavina, L., Lipham, J. C., Dunn, D. & Ganz, R. A. Magnetic sphincter augmentation for gastroesophageal reflux at 5 years: final results of a pilot study show long-term acid reduction and symptom improvement. *J. Laparoendosc. Adv. Surg. Tech.* **25**, 787–792 (2015).
12. Bonavina, L., Saino, G., Bona, D., Sironi, A. & Lazzari, V. One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. *J. Am. Coll. Surg.* **217**, 577–585 (2013).
13. Velanovich, V. The development of the GERD-HRQL symptom severity instrument. *Dis. Esophagus.* **20**, 130–134 (2007).
14. Riva, C. G. *et al.* High-resolution manometry findings after Linx procedure for gastro-esophageal reflux disease. *Neurogastroenterol. Motil.* **32**(3), e13750 (2020).
15. Bonavina, L. *et al.* Laparoscopic sphincter augmentation devices eliminates reflux symptoms and normalizes esophageal acid exposure: one- and 2-year results of a feasibility trial. *Ann. Surg.* **252**, 857–862 (2010).
16. Lipham, J. C., Taiganides, P. A., Louie, B. E., Ganz, R. A. & DeMeester, T. R. Safety analysis of first 1000 patients treated with magnetic sphincter augmentation for gastroesophageal re- flux disease. *Dis. Esophagus.* **28**, 305–311 (2015).
17. Alicuben, E. T. *et al.* Worldwide experience with erosion of the magnetic sphincter augmentation device. *J. Gastrointest. Surg.* **22**, 1442–1447 (2018).
18. Asti, E. *et al.* Removal of the magnetic sphincter augmentation device. Surgical technique and results of a single-center cohort study. *Ann. Surg.* **265**(5), 941–945 (2017).

19. Tatum, J. M. *et al.* Removing the magnetic sphincter augmentation device: operative management and outcomes. *Surg. Endosc.* **33**, 2663–2669 (2019).
20. Ayazi, S. *et al.* Magnetic sphincter augmentation and postoperative dysphagia; characterization, clinical risk factors, and management. *J. Gastrointest. Surg.* **24**, 39–49 (2020).
21. Bell, R. *et al.* Magnetic sphincter augmentation superior to proton pump inhibitors for regurgitation in a 1-year randomized trial. *Clin. Gastroenterol. Hepatol.* **18**, 1736–1743 (2020).
22. Ayazi, S. *et al.* Clinical outcomes and predictors of favorable result after laparoscopic magnetic sphincter augmentation: single institution experience with over 500 patients. *J. Am. Coll. Surg.* **230**, 733–743 (2020).
23. Bonavina, L., Fisichella, P. M., Gavini, S., Lee, Y. Y. & Tatum, R. P. Clinical course of gastroesophageal reflux disease and impact of treatment in symptomatic young patients. *Ann. N. Y. Acad. Sci.* https://doi.org/10.1111/nyas.14350 (2020).
24. Maret-Ouda, J., Wahlin, K., El-Serag, H. B. & Lagergren, J. Association between laparoscopic antireflux surgery and recurrence of gastroesophageal reflux. *JAMA* **318**(10), 939–946 (2017).
25. Reynolds, J. L. *et al.* Laparoscopic magnetic sphincter augmentation vs laparoscopic Nissen fundoplication: A matched-pair analysis of 100 patients. *J. Am. Coll. Surg.* **221**, 123–128 (2015).
26. Ganz, R. A. *et al.* Long-term outcomes of patients receiving a magnetic sphincter augmentation device for gastroesophageal reflux. *Clin. Gastroenterol. Hepatol.* **14**, 671–677 (2016).
27. Asti, E., Bonitta, G., Lovece, A., Lazzari, V. & Bonavina, L. Longitudinal comparison of quality of life in patients undergoing laparoscopic Toupet fundoplication versus magnetic sphincter augmentation. *Medicine.* **95**, 30 (2016).
28. Schwameis, K. *et al.* Crural closure improves outcomes of magnetic sphincter augmentation in GERD patients with hiatal hernia. *Sci. Rep.* **8**, 7319 (2018).
29. Tatum, J. M. *et al.* Minimal versus obligatory dissection of the diaphragmatic hiatus during magnetic sphincter augmentation surgery. *Surg. Endosc.* **33**(3), 782–788 (2019).
30. Asti, E. *et al.* Magnetic sphincter augmentation for gastroesophageal reflux disease: review of clinical studies. *Updates Surg.* **70**, 323–330 (2018).
31. Mihura Irribarra, M. M. *et al.* Does treatment of the hiatus influence the outcomes of magnetic sphincter augmentation for chronic GERD?. *J. Gastrointest. Surg.* **23**(6), 1104–1112 (2019).
32. Buckley, F. P. III., Bell, R. C. W., Freeman, K., Doggett, S. & Heidrick, R. Favorable results from a prospective evaluation of 200 patients with large hiatal hernias undergoing LINX magnetic sphincter augmentation. *Surg. Endosc.* **32**(4), 1762–1768 (2018).
33. Kuckelman, J. P., Phillips, C. J., Hardin, M. O. & Martin, M. J. Standard vs expanded indications foe esophageal magnetic sphincter augmentation for reflux disease. *JAMA Surg.* **152**(9), 890–891 (2017).
34. Rona, K. A. *et al.* Efficacy of magnetic sphincter augmentation in patients with large hiatal hernias. *Surg. Endosc.* **31**, 2096–2102 (2017).
35. Alicuben, E. T. *et al.* Regression of intestinal metaplasia following magnetic sphincter augmentation. *Surg. Endosc.* **33**, 76–579 (2019).

Acknowledgements

Supported by AIRES (Associazione Italiana Ricerca Esofago).

Author contributions

Study conception and design: L.B. Acquisition of data: V.L., S.S., D.B. Analysis and interpretation of data: D.F., E.A. Drafting of manuscript: D.F., L.B. Critical revision: L.B. All authors reviewed the manuscript and approved.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Clinical Gastroenterology and Hepatology 2016;14:671–677

ALIMENTARY TRACT

Long-term Outcomes of Patients Receiving a Magnetic Sphincter Augmentation Device for Gastroesophageal Reflux

Robert A. Ganz,[*] Steven A. Edmundowicz,[‡] Paul A. Taiganides,[§] John C. Lipham,[‖] C. Daniel Smith,[¶] Kenneth R. DeVault,[#] Santiago Horgan,[**] Garth Jacobsen,[**] James D. Luketich,[‡‡] Christopher C. Smith,[§§] Steven C. Schlack-Haerer,[‖‖] Shanu N. Kothari,[¶¶] Christy M. Dunst,[##] Thomas J. Watson,[***] Jeffrey Peters,[‡‡‡] Brant K. Oelschlager,[§§§] Kyle A. Perry,[‖‖‖] Scott Melvin,[‖‖‖] Willem A. Bemelman,[¶¶¶] André J. P. M. Smout,[###] and Dan Dunn[****]

*Minnesota Gastroenterology PA, Plymouth, Minnesota; [‡]Division of Gastroenterology, Washington University School of Medicine, St. Louis, Missouri; [§]Knox Regional Heartburn Treatment Center, Mount Vernon, Ohio; [‖]Department of Surgery, Keck School of Medicine, University of Southern California, Los Angeles, California; [¶]Department of Surgery, [#]Department of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, Florida; [**]Department of Surgery, University of California at San Diego, San Diego, California; [‡‡]Division of Thoracic Surgery, University of Pittsburgh, Pittsburgh, Pennsylvania; [§§]Southern Reflux Center at Albany Surgical Center, Albany, Georgia; [‖‖]Department of Gastroenterology, [¶¶]Department of Surgery, Gundersen Health System, LaCrosse, Wisconsin; [##]Gastrointestinal and Minimally Invasive Surgery Division, The Oregon Clinic, Portland, Oregon; [***]Department of Surgery, University of Rochester School of Medicine and Dentistry, Rochester, New York; [‡‡‡]University Hospitals, Cleveland, Ohio; [§§§]Department of Surgery, University of Washington, Seattle, Washington; [‖‖‖]Department of Surgery, The Ohio State University, Columbus, Ohio; [¶¶¶]Department of Surgery, [###]Division of Gastroenterology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [****]Division of Surgery, Abbott-Northwestern Hospital, Minneapolis, Minnesota

BACKGROUND & AIMS: **Based on results from year 2 of a 5-year trial, in 2012 the US Food and Drug Administration approved the use of a magnetic device to augment lower esophageal sphincter function in patients with gastroesophageal reflux disease (GERD). We report the final results of 5 years of follow-up evaluation of patients who received this device.**

METHODS: **We performed a prospective study of the safety and efficacy of a magnetic device in 100 adults with GERD for 6 months or more, who were partially responsive to daily proton pump inhibitors (PPIs) and had evidence of pathologic esophageal acid exposure, at 14 centers in the United States and The Netherlands. The magnetic device was placed using standard laparoscopic tools and techniques. Eighty-five subjects were followed up for 5 years to evaluate quality of life, reflux control, use of PPIs, and side effects. The GERD–health-related quality of life (GERD-HRQL) questionnaire was administered at baseline to patients on and off PPIs, and after placement of the device; patients served as their own controls. A partial response to PPIs was defined as a GERD-HRQL score of 10 or less on PPIs and a score of 15 or higher off PPIs, or a 6-point or more improvement when scores on vs off PPI were compared.**

RESULTS: **Over the follow-up period, no device erosions, migrations, or malfunctions occurred. At baseline, the median GERD-HRQL scores were 27 in patients not taking PPIs and 11 in patients on PPIs; 5 years after device placement this score decreased to 4. All patients used PPIs at baseline; this value decreased to 15.3% at 5 years. Moderate or severe regurgitation occurred in 57% of subjects at baseline, but only 1.2% at 5 years. All patients reported the ability to belch and vomit if needed. Bothersome dysphagia was present in 5% at baseline and in 6% at 5 years. Bothersome gas-bloat was present in 52% at baseline and decreased to 8.3% at 5 years.**

CONCLUSIONS: **Augmentation of the lower esophageal sphincter with a magnetic device provides significant and sustained control of reflux, with minimal side effects or complications. No new safety risks**

Abbreviations used in this paper: FDA, Food and Drug Administration; GERD, gastroesophageal reflux disease; LES, lower esophageal sphincter; PPI, proton pump inhibitor.

Most current article

1542-3565/$36.00
http://dx.doi.org/10.1016/j.cgh.2015.05.028

emerged over a 5-year follow-up period. These findings validate the long-term safety and efficacy of the magnetic sphincter augmentation device for patients with GERD. ClinicalTrials.gov no: NCT00776997.

Keywords: Heartburn; Hiatal Hernia; Clinical Trial; Surgery.

Gastroesophageal reflux disease (GERD) is a serious condition because of the potential for chronic symptoms and complications, and its association with Barrett's esophagus and esophageal adenocarcinoma.[1] Fundamental to the disease process is an incompetent lower esophageal sphincter (LES) that allows abnormal reflux of gastric content into the esophagus. The gastric refluxate, which contains varying concentrations of acid, pepsin, enzymes, and other content, influences symptoms and mucosal damage by its type, acidity, volume, and exposure time via contact with esophageal surfaces.[2,3] Acid-suppression therapy, in the form of proton pump inhibitors (PPIs), is the primary treatment for GERD. This drug class has been shown to be most effective for acid-related symptoms, such as heartburn and esophagitis, and less effective for regurgitation because PPI therapy does not address the underlying incompetency of the LES.[4] Acid-suppression therapy may change the composition of the refluxate, particularly its acidity, but it does not prevent abnormal reflux, leaving patients at risk for ongoing symptoms and progression of disease.[5]

Esophagogastric fundoplication addresses the incompetent LES by mobilizing the gastric fundus to form a fundic wrap around the distal esophagus, resulting in permanent loss of normal gastric fundic anatomy, and involves dissection of the phreno-esophageal ligaments, which secure the esophagus to the diaphragm.[6] Nissen fundoplication has long been associated with effective reflux control, albeit at the expense of inducing new side effects, such as gas-bloat, flatulence, and inability to vomit.[7,8] The placement of a magnetic sphincter augmentation device (LINX Reflux Management System; Torax Medical, Inc, Shoreview, MN) is the only antireflux procedure that mechanically restores competency to the reflux barrier without using the gastric fundus.[9,10] Based on 2-year results, the Food and Drug Administration (FDA) granted approval of the magnetic device for GERD in 2012.[11,12] The final results of the 5-year magnetic sphincter augmentation study are reported.

Methods

Study Design

This was a prospective, multicenter, single-arm study with predefined success criteria. Each patient served as his or her own control, with the treatment effect assessed by comparing follow-up assessments with baseline. This study purposely was designed to provide valid scientific evidence to support FDA approval. Both objective and subjective evaluations were performed to evaluate the treatment effect. The authors had access to the study data at all times and the co-authors reviewed and approved all manuscript drafts and the final manuscript.

Patients

A total of 14 centers (13 in the United States and 1 in The Netherlands) enrolled patients. Eligible patients were 18 to 75 years old, had at least a 6-month history of GERD, a partial response to daily PPIs, and pathologic esophageal acid exposure confirmed by pH monitoring. Exclusion criteria included evidence of hiatal hernia greater than 3 cm, esophagitis grade C or D according to the Los Angeles classification, body mass index higher than 35, Barrett's esophagus, or motility disorder. In the study, the Gastroesophageal Reflux Disease-Health Related Quality of Life (GERD-HRQL) questionnaire was administered both with and without PPIs before treatment to assess changes in the total score. Per the protocol, a subject was considered a partial PPI responder if the following inclusion criterion were met: subjects with persistent reflux symptoms and partial symptomatic improvement on PPI therapy shown by a GERD-HRQL score of 10 or less on PPI and 15 or higher off PPI, or subjects with a 6-point or more improvement when comparing their GERD-HRQL score on PPI and off PPI.

Study Procedures

The baseline screening, surgical technique, and follow-up evaluation were reported previously.[12] The magnetic device was placed by foregut surgeons using standard laparoscopic tools and techniques. A crural repair was performed at the surgeon's discretion; cruroplasty was performed in 34% of patients. The device uses magnetic attraction to create resistance to an abnormal opening of the LES to prevent reflux events, but still allows normal LES opening for swallowing food, belching, and vomiting (Figure 1).

Postapproval Evaluations

The efficacy end points after FDA approval were the same as before approval, with the exception that esophageal pH monitoring was performed after the procedure only at 1 year and these results were reported previously.[12] Quality of life was measured with the GERD-HRQL questionnaire.[13] Total scores range from

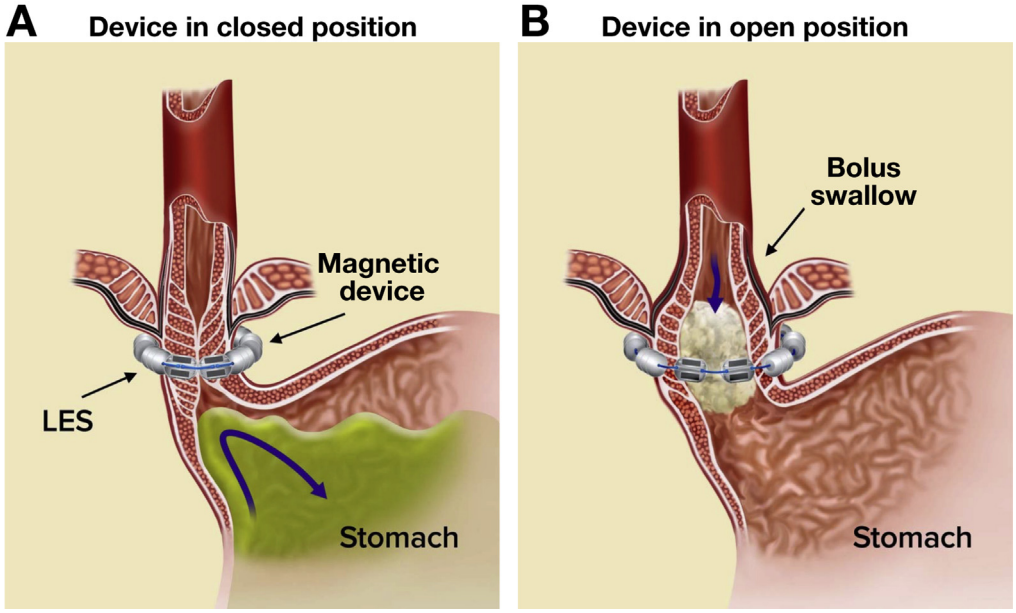


Figure 1. Magnetic sphincter augmentation. (*A*) In the closed position, the magnetic attraction of the beads augments the lower esophageal sphincter to prevent its opening and subsequent reflux of gastric content into the esophagus. (*B*) The device is shown in the open position, which allows for normal physiologic function such as transport of food, belching, and vomiting.

0 to 50, with higher scores indicating worse symptoms. For those patients who had resumed PPIs, the dose and frequency were recorded, and the PPI was stopped for at least 7 days before completing the questionnaires. Postapproval efficacy end points included the following: the number of patients achieving at least a 50% reduction in the GERD-HRQL score as compared with the baseline score without PPIs; and a reduction of at least 50% in the dose of PPIs compared with baseline. The efficacy end point was achieved if met by at least 60% of patients.

Additional side effects and reflux-related symptoms were actively queried before and after treatment with the Foregut Symptoms Questionnaire.[14] The 5-year evaluation included endoscopy to evaluate the presence of esophagitis, Barrett's esophagus, or device erosion. Chest radiography was used to confirm the device remained at the gastroesophageal junction.

Statistical Analysis

Postapproval analyses were performed on GERD-HRQL scores and use of PPIs per the predefined success criteria. Safety was monitored throughout the postapproval period as the rate and type of serious adverse events related to the device or implant procedure.

Continuous variables were summarized with the use of standard descriptive statistics (eg, mean, standard deviation, median, range). Categoric variables were summarized via frequency distributions. The 2-tailed, paired Student *t* test or the Wilcoxon signed rank test were used to compare before and after implant values for continuous outcomes and the McNemar test was used to assess changes in binary outcomes from before to after implant. Differences were considered significant at a *P* value of less than .05.

Results

Patient Characteristics

The study population consisted of 100 patients, 52% males and 48% females. The median age at the time of implant was 53 years (range, 18–75 y), with a median body mass index of 28 (range, 20–35). The median duration of reflux symptoms was 10 years (range, 1–40 y). The median duration of treatment with PPIs was 5 years (range, <1 to 20 y). Each patient had confirmed pathologic esophageal acid exposure without PPIs (median percentage of time pH was less than 4 was 10.9%; range, 4.8%–25.4%). Eighty-five patients had a follow-up evaluation at 5 years; a few analyses were performed on 84 patients if data were incomplete, and 82 patients completed endoscopy at the 5-year mark. A consolidated standards for the reporting of trials diagram is provided in the Supplementary Appendix.

Efficacy Success Criteria at 5 Years

A 50% or greater reduction in GERD-HRQL score was achieved in 83% of patients at 5 years (70 of 84; 95% confidence interval, 73–91) (Figure 2). A reduction of 50% or more in the average daily dose of PPIs occurred in 89.4% of patients at 5 years (76 of 85 patients; 95% confidence interval, 81–95) (Figure 2). Per the predefined criteria, long-term efficacy was maintained.

Additional Reflux Analyses

Additional analyses, using the GERD-HRQL and Foregut Symptoms Questionnaire, were performed (84 patients were available for these analyses). Patients with moderate or severe heartburn had a decrease from 89%

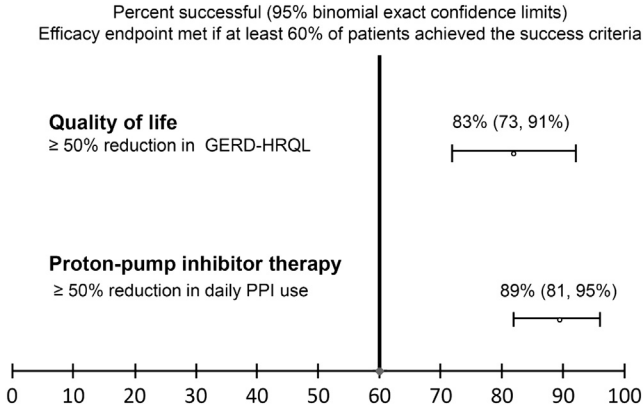


Figure 2. Summary of efficacy end points at 5 years. At 5 years, the end points of quality of life and reduction in PPI use was met if at least 60% of patients achieved the success criteria. For quality of life, 83% achieved at least a 50% reduction in total GERD-HRQL score, with a lower-bound confidence interval of 73%. For reduction in proton pump inhibitor use, 89% achieved at least a 50% reduction in daily proton pump inhibitor use, with a lower-bound confidence interval of 81%.

to 11.9%, and moderate or severe regurgitation decreased from 57% to 1.2% without use of PPIs at baseline and 5 years. Patient dissatisfaction before treatment was 95%, and decreased to 7.1% at 5 years (Figure 3) ($P < .001$ for all comparisons with baseline).

Daily use of PPIs was 100% at baseline and decreased to 15.3% at 5 years. At 5 years, 75.3% of patients reported complete cessation of PPIs, and 9.4% reported PPI use only as needed. Thus, 84.7% were either completely off PPIs or reported use as needed at 5 years after surgery. Patients who required double-dose PPIs decreased from 36% at baseline to 2.4% at 5 years. Of the patients reporting dissatisfaction at 5 years, all but 1 (5 of 6) reported daily use of PPIs (Figure 4).

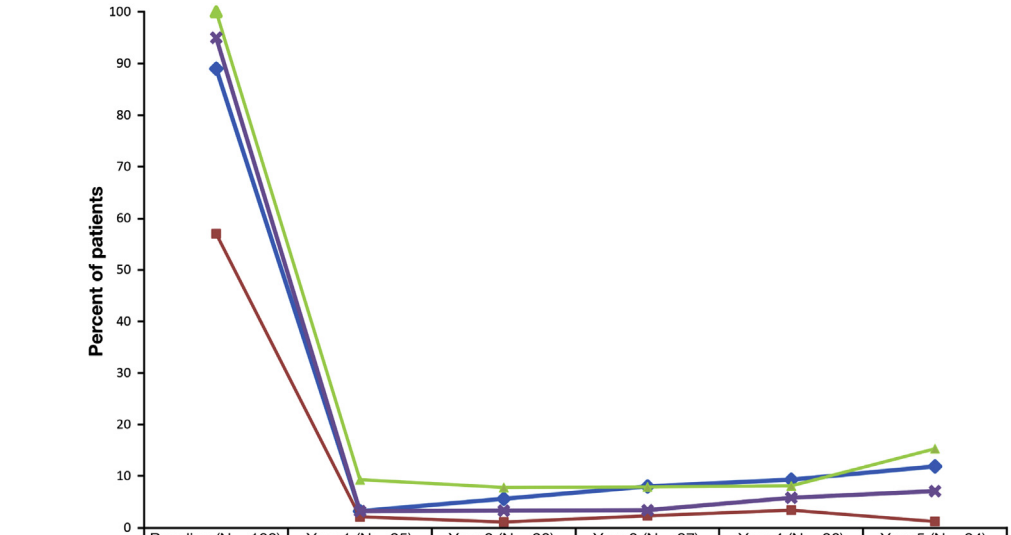
Comparing the total GERD-HRQL scores at 5 years without PPIs with scores with and without PPIs at baseline, the median total score at baseline was 27 without PPIs and 11 with PPIs, and the score decreased to 4 after surgery at 5 years ($P < .001$ for all comparisons with baseline) (Figure 5). The median GERD-HRQL score for patients reporting any use of PPIs within 30 days of the 5-year follow-up period was 7 (after PPIs were discontinued for at least 1 week) and 2.5 for patients reporting no PPIs.

Esophagitis

Healing of esophagitis occurred in 76.5% (26 of 34) of patients evaluated at 5 years. Among the 8 patients with ongoing esophagitis, 6 patients had grade A and the other patients had grade B. Of the patients without esophagitis at baseline and evaluated at 5 years, 90% (43 of 48) continued to have no esophagitis. Among the 5 patients with de novo esophagitis, 4 patients had grade A and 1 patient had grade B. No patients developed Barrett's esophagus during the study.

Other Symptoms

Other symptoms commonly associated with antireflux surgery or reflux disease were minimal at 5 years compared with baseline (Figure 6). All patients reported the ability to belch and vomit (if needed). Patients reporting bothersome swallowing was 5% at baseline and 6% at 5 years ($P = .739$); symptoms of bloating/gas decreased from 52% at baseline to 8.3% at 5 years ($P < .001$). In addition, per the Foregut Symptom Questionnaire, patients reported less diarrhea ($P = .103$),



	Baseline (N = 100)	Year 1 (N = 95)	Year 2 (N = 90)	Year 3 (N = 87)	Year 4 (N = 86)	Year 5 (N = 84)
Heartburn	89	3.2	5.6	8	9.3	11.9
Regurgitation	57	2.1	1.1	2.3	3.4	1.2
PPI dependence	100	9.3	7.8	7.9	8.1	15.3
Dissatisfaction	95	3.2	3.3	3.4	5.8	7.1

$P < .001$ for comparision between baseline and all follow-ups

Figure 3. Reflux control before and after magnetic sphincter augmentation. Change in baseline compared with 1 to 5 years after magnetic sphincter augmentation for moderate-severe heartburn, moderate-severe regurgitation, proton-pump inhibitor dependence, and dissatisfaction ($P < .001$ between baseline and follow-up evaluation for all comparisons).

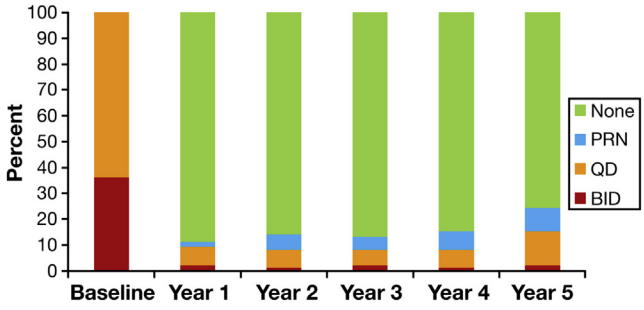


Figure 4. Use of PPIs at baseline through year 5. PPI use was categorized as none, as needed (PRN), once a day (QD), and twice a day (BID) at each visit, based on the prior 30 days.

constipation ($P = .008$), and nausea/vomiting ($P = .003$) after treatment.

Safety

Since the last report at 3 years, no new safety concerns have emerged.[12] No device erosions, migrations, or malfunctions occurred in this study. Device removal occurred in 7 patients. In 4 of the patients, the device was removed at 21, 31, 93, and 1807 days after implantation because of persistent dysphagia, with resolution after removal. One patient had the device removed at 357 days owing to intermittent vomiting of unknown cause starting 3 months after placement, without relief after removal. The device was removed in 1 patient at 489 days because of persistent reflux symptoms and another device was removed at 1062 days because of persistent chest pain. Three patients subsequently underwent uneventful Nissen fundoplication after device removal.

Discussion

Persistent symptoms of heartburn and regurgitation warrant careful monitoring. Ignoring persistent reflux symptoms can lead to severe complications, such as esophageal strictures, Barrett's esophagus, and esophageal adenocarcinoma.[15] Consideration of other treatments may become necessary when medical therapy fails to control symptoms. In the current study, we enrolled patients who had been treated with PPIs for several years, who still did not obtain adequate reflux control. These patients sought surgical treatment, but elected to forego conventional fundoplication surgery, instead opting for esophageal sphincter augmentation using a magnetic device.

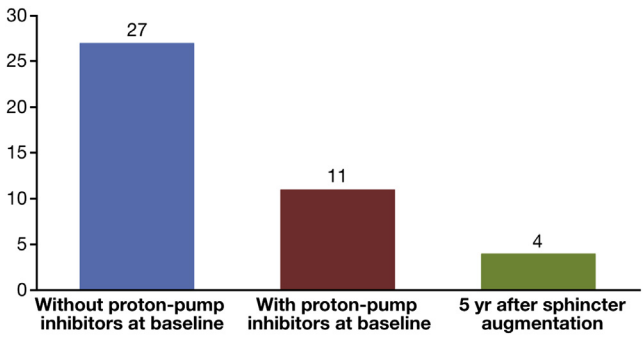


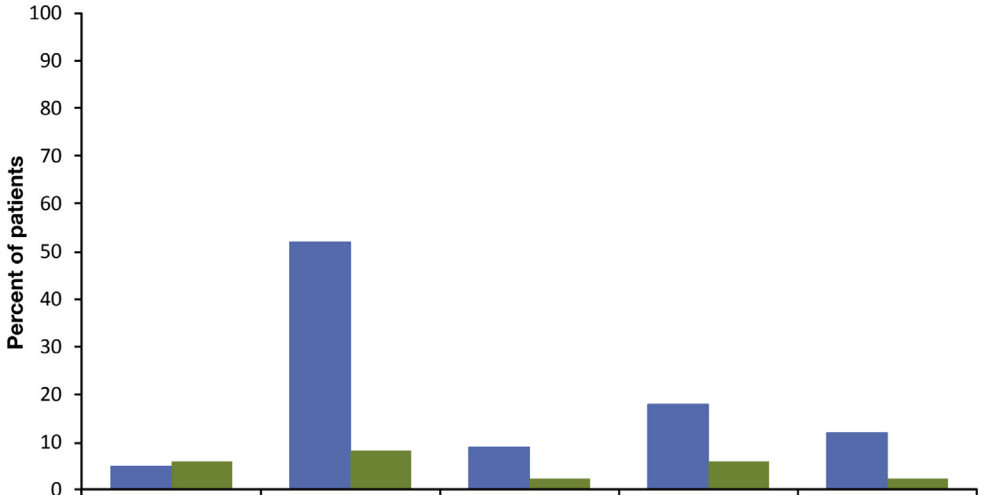
Figure 5. Median total GERD-HRQL score. Median score from the GERD-HRQL measures at baseline without and with proton pump inhibitors, as compared with magnetic sphincter augmentation at 5 years. Higher scores indicate worse symptoms. There was significant improvement in the median score at 5 years, as compared with baseline, both without and with proton pump inhibitor use.

The majority of patients reported moderate or severe regurgitation at baseline in this study, symptoms for which PPIs largely are ineffective.[16,17] Five years after magnetic sphincter device placement, moderate or severe regurgitation was eliminated in all but 1 of 57 patients. Without a procedure to correct an incompetent LES, it is unlikely that continued medical therapy would have improved these reflux symptoms, and the severity and frequency of the symptoms may have worsened.[18]

In this study, side effects commonly associated with Nissen fundoplication largely were absent, consistent with other studies of the magnetic device.[12,19–21] Louie et al[22] provided a rationale for the difference in side-effect profile between the magnetic device and Nissen fundoplication: Nissen fundoplication reduces the total number of reflux episodes to less than what is considered normal, creating a "super-normal" sphincter that is highly effective at preventing reflux but to the point of preventing or decreasing venting of ingested air under normal circumstances. The inability to vent (ie, belch) and the reduced number of normal reflux episodes after Nissen fundoplication likely is associated with the side effects of bloating and flatulence. In contrast, the magnetic device results in more normal sphincter function.[22]

The rate of laparoscopic device removal was 7% over a 5-year period and has been reported to be 3% in another study spanning 6 years of clinical experience.[21] These rates are less than the expected range for reoperation after laparoscopic fundoplication at 5 years.[23–26] Reoperation rates for transoral fundoplication at 5 years are not available, but revision rates from shorter-term experiences have been reported to be between 11.5% and 52.6%.[27–30] Importantly, all device removals in our study were performed electively with no procedure-related complications, whereas reoperations after laparoscopic fundoplication often are associated with a higher rate of complications and morbidity.[31,32] It appears that the severity of complications and risks associated with reoperation are less after magnetic sphincter augmentation than fundoplication.[33–35]

Importantly, in this long-term study, no device erosions occurred. Other investigators[36] have reported an erosion rate of much less than 1%. The magnetic sphincter augmentation device was engineered specifically to minimize the risk of device erosion and overcome the problems of previous barrier devices.[37] The device provides sphincter augmentation by means of



	Difficulty swallowing	Bloating/gassy	Diarrhea	Constipation	Nausea/ vomiting
Baseline	5	52	9	18	12
Year 5	6.0	8.3	2.4	6	2.4

Figure 6. Other symptoms after magnetic sphincter augmentation. Comparison at baseline and 5 years after magnetic sphincter augmentation for other symptoms experienced by reflux patients, such as difficulty swallowing ($P = .739$), bloating/gassy feeling ($P < .001$), diarrhea ($P = .103$), constipation ($P = .008$), and nausea/vomiting ($P = .003$).

dynamic, expansible magnetic forces, not bulk or compression. The device mirrors physiologic LES opening and closing in that it opens for food bolus transport (its opening area can increase more than the esophagus), and shows progressively less force the larger the bolus, as opposed to alternative rigid designs that use fixed diameters around the esophagus, allowing for little or no device distention. Long-term clinical experience confirms that the magnetic sphincter augmentation device design is not prone to cause esophageal wall erosions.

This study had some limitations. Per the FDA-approved protocol, esophageal pH testing and manometry were not performed beyond 1 year. Manometry data have been reported previously with no significant change in any manometric parameter.[12] Esophageal pH results at 1 year showed that the majority of the patients had normalization of esophageal acid exposure along with symptomatic improvement and discontinuation of PPIs.[12] Longer-term pH data would have strengthened our conclusions. In addition, the study did not have a comparison group. Instead the trial design allowed for assessment of long-term outcomes via rigorous follow-up compliance, which is important in a chronic disease state such as GERD. This study describes a long and complete multicenter follow-up evaluation of a novel GERD-device cohort, used predefined success criteria, and purposely was designed to provide valid scientific evidence to support FDA approval. Both objective (pH and manometry) and subjective evaluations were performed to evaluate the treatment effect. Consistent with clinical practice, the extended 5-year follow-up period focused on both maintenance and durability of symptom improvement, and discontinuation of PPI therapy, and the objective clinical evidence of reduction in esophageal acid exposure was corroborated by the subjective findings. The consistent results of this 5-year study provide confidence that when used as indicated, there is a high probability that magnetic sphincter augmentation will improve the antireflux barrier and provide durable clinical benefits.

In conclusion, this study showed that patients with chronic GERD and failed long-term PPI therapy benefited from surgical intervention with magnetic sphincter augmentation. Long-term safety and efficacy have been validated for this procedure. It should be considered a first-line therapy for patients and physicians seeking a fundic-sparing antireflux procedure.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at http://dx.doi.org/10.1016/j.cgh.2015.05.028.

References

1. Lagergren J, Reinhold B, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340:825–831.
2. Woodland P, Sifrim D. The refluxate: the impact of its magnitude, composition and distribution. Best Pract Res Clin Gastroenterol 2010;24:861–871.
3. Pearson JP, Parikh S, Orlando RC, et al. Review article: reflux and its consequences – the laryngeal, pulmonary and oesophageal manifestations. Conference held in conjunction with the 9th International Symposium on Human Pepsin (ISHP) Kingston-upon-Hull, UK, 21-23 April 2010. Aliment Pharmacol Ther 2011;33:1–71.
4. Kahrilas P, Jonsson A, Denison H, et al. Regurgitation is less responsive to acid suppression than heartburn in patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol 2012;10:612–619.
5. Blonski W, Vela MF, Castell DO. Comparison of reflux frequency during prolonged multichannel intraluminal impedance and pH

monitoring on and off acid suppression therapy. J Clin Gastroenterol 2009;43:816–820.
6. Spivak H, Lelcuk S, Hunter JG. Laparoscopic surgery of the gastroesophageal junction. World J Surg 1999;23:356–367.
7. Humphries L, Hernandez J, Clark W, et al. Causes of dissatisfaction after laparoscopic fundoplication: the impact of new symptoms, recurrent symptoms, and the patient experience. Surg Endosc 2013;27:1537–1545.
8. Kessing BF, Broeders JA, Vinke N, et al. Gas-related symptoms after antireflux surgery. Surg Endosc 2013;27:3739–3747.
9. Woodland P, Amarasinghe G, Sifrim D. Emerging therapeutic options in GERD. Best Pract Res Clin Gastroenterol 2013; 27:455–467.
10. Bonavina L, DeMeester TR, Ganz RA. LINX™ Reflux Management System: magnetic sphincter augmentation in the treatment of gastroesophageal reflux disease. Expert Rev Gastroenterol Hepatol 2012;6:667–674.
11. Summary of safety and effectiveness data: LINX Reflux Management System. Available at: http://www.accessdata.fda.gov/crdh_docs/pdf10/P100049b.pdf. Accessed: January 21, 2015.
12. Ganz RA, Peters JH, Horgan S, et al. Esophageal sphincter device for gastroesophageal reflux disease. N Engl J Med 2013; 368:719–727.
13. Velanovich V. Comparison of generic (SF-36) vs. disease-specific (GERD-HRQL) quality of life scales for GERD. J Gastrointest Surg 1998;2:141–145.
14. DeMeester TR, Bonavina L, Albertucci M. Nissen fundoplication for gastroesophageal reflux disease. Ann Surg 1986;204:9–20.
15. Fennerty MB. The continuum of GERD complications. Cleve Clin J Med 2003;70(Suppl 5):S33.
16. Kahrilas PJ, Jonsson A, Denison H, et al. Impact of regurgitation on health-related quality of life in gastroesophageal reflux disease before and after short-term potent acid suppression therapy. Gut 2014;63:720–726.
17. Clayton SB, Rife CC, Singh ER, et al. Twice-daily proton pump inhibitor therapy does not decrease the frequency of reflux episodes during nocturnal recumbency in patients with refractory GERD. Dis Esophagus 2012;25:682–686.
18. Malfertheiner P, Nocon M, Vieth M, et al. Evolution of gastro-oesophageal reflux disease over 5 years under routine medical care – the ProGERD study. Aliment Pharmacol Ther 2012; 35:154–164.
19. Riegler M, Schoppman SF, Bonavina L, et al. Magnetic sphincter augmentation and fundoplication for GERD in clinical practice: one-year results of a multicenter, prospective observational study. Surg Endosc 2015;29:1123–1129.
20. Lipham JC, DeMeester TR, Ganz RA, et al. The LINX® reflux management system: confirmed safety and efficacy now at 4 years. Surg Endosc 2012;26:2944–2999.
21. Bonavina L, Saino G, Bona D, et al. One hundred consecutive patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease: 6 years of clinical experience from a single center. J Am Coll Surg 2013;217:577–585.
22. Louie BE, Farivar AS, Schultz D, et al. Short-term outcomes using magnetic sphincter augmentation versus nissen fundoplication for medically resistant gastroesophageal reflux disease. Ann Thorac Surg 2014;98:498–505.
23. Broeders JA, Roks DJ, Ali UA, et al. Laparoscopic anterior 180-degree versus Nissen fundoplication for gastroesophageal reflux disease: systematic review and meta-analysis of randomized clinical trials. Ann Surg 2013;257:850–859.
24. Baigrie RJ, Cullis SN, Ndhluni AJ, et al. Randomized double-blind trial of laparoscopic Nissen fundoplication versus anterior partial fundoplication. Br J Surg 2005;92:819–823.
25. Cao W, Watson DI, Lally CJ, et al. Ten-year clinical outcome of a prospective randomized clinical trial of laparoscopic Nissen versus anterior 180 (degrees) partial fundoplication. Br J Surg 2008;95:1501–1505.
26. Watson DI, Jamieson GC, Pike GK, et al. Five-year follow-up of a randomized clinical trial of laparoscopic total versus partial anterior 180 degrees fundoplication. Br J Surg 1999;86:123–130.
27. Hoppo T, Immanuel A, Schuchert M, et al. Transoral incisionless fundoplication 2.0 procedure using EsophyX™ for gastroesophageal reflux disease. J Gastrointest Surg 2010;14:1895–1901.
28. Witterman B, Strijkers R, de Vries E, et al. Transoral incisionless fundoplication for treatment of gastroesophageal reflux disease in clinical practice. Surg Endosc 2012;26:3307–3315.
29. Muls V, Eckhardt AJ, Marchese M, et al. Three-year results of a multicenter prospective study of transoral incisional fundoplication. Surg Innov 2013;20:321–330.
30. Demyttenaere SV, Bergman S, Pham T, et al. Transoral incisionless fundoplication for gastroesophageal reflux disease in an unselected patient population. Surg Endosc 2010;24:854–858.
31. Furnee JB, Draaisma WA, Broeders IAMJ, et al. Surgical reintervention after failed antireflux surgery: a systematic review of the literature. J Gastrointest Surg 2009;13:1539–1549.
32. Symons NR, Purkayastha S, Dillemans B, et al. Laparoscopic revision of failed antireflux surgery: a systematic review. Am J Surg 2011;202:336–343.
33. Richter JE. Gastroesophageal reflux disease treatment: side effects and complications of fundoplication. Clin Gastroenterol Hepatol 2013;11:465–471.
34. Titus JM, Mason DP, Raymond DP, et al. Esophagopulmonary fistula and left lung abscess after transoral incisionless fundoplication. Ann Thorac Surg 2013;96:689–691.
35. Hawa E, El-Bakush A, Nugent K. Esophageal perforation and bilateral empyema following endoscopic EsophyX transoral incisionless fundoplication. Clin Endosc 2014;47:560–563.
36. Lipham JC, Taiganides PA, Louie BE, et al. Safety analysis of first 1000 patients treated with magnetic sphincter augmentation for gastroesophageal reflux disease. Dis Esophagus 2015; 28:305–311.
37. Ganz RA, Gostout CJ, Grudem J, et al. Use of a magnetic sphincter for the treatment of GERD: a feasibility study. Gastrointest Endosc 2008;67:287–294.

Reprint requests
Address requests for reprints to: Robert A. Ganz, MD, FASGE, Minnesota Gastroenterology PA, 15700 37th Avenue North, Suite 300, Plymouth, Minnesota 55446. e-mail: rganz@mngastro.com; fax: (612) 870-5837.

Conflicts of interest
These authors disclose the following: Paul Taiganides, John Lipham, C. Daniel Smith, Santiago Horgan, and Shanu N. Kothari have served as consultants to Torax Medical and have received honoraria.

Funding
The study was funded by Torax Medical, Inc. This study was designed cooperatively by the sponsor (Torax Medical, Inc, Shoreview, MN), the study investigators, and the Food and Drug Administration.

Surgical Endoscopy
https://doi.org/10.1007/s00464-022-09015-1

2021 SAGES ORAL

S148: Long-term patient-reported outcomes of laparoscopic magnetic sphincter augmentation versus Nissen fundoplication: a 5-year follow-up study

Sean M. O'Neill[1] · Anahita D. Jalilvand[2] · Jennifer S. Colvin[3] · Kelly R. Haisley[4] · Kyle A. Perry[4]

Received: 2 September 2021 / Accepted: 3 January 2022

Abstract

Background Laparoscopic magnetic sphincter augmentation (MSA) has emerged as an alternative to laparoscopic Nissen fundoplication (LNF) for the management of symptomatic gastroesophageal reflux disease (GERD). While short-term outcomes of MSA compare favorably to those of LNF, direct comparisons of long-term outcomes are lacking. We hypothesized that the long-term patient-reported outcomes of MSA would be similar to those achieved with LNF.

Methods We tested this hypothesis in a retrospective cohort undergoing primary LNF or MSA between March 2013 and July 2015. The primary outcome was GERD-Health Related Quality of Life (GERD-HRQL) score at long-term follow-up relative to baseline. Secondary outcomes included dysphagia and bloating scores, proton-pump inhibitor (PPI) cessation, reoperations, and overall satisfaction with surgery.

Results 70 patients (25 MSA, 45 LNF) met criteria for study inclusion. MSA patients had lower baseline BMI (median: 27.1 [IQR: 22.7–29.9] versus 30.4 [26.4–32.8], $p=0.02$), lower total GERD-HRQL (26 [19–32] versus 34 [25–40], $p=0.02$), and dysphagia (2 [0–3] versus 3 [1–4], $p=0.02$) scores. Median follow-up interval exceeded 5 years (MSA: 68 [65–74]; LNF: 65 months [62–69]). Total GERD-HRQL improved from 26 to 9 after MSA ($p<0.001$) and from 34 to 7.5 after LNF ($p<0.01$); these scores did not differ between groups ($p=0.68$). Dysphagia (MSA: 1 [0–2]; LNF: 0 [0–2], $p=0.96$) and bloating (MSA: 1.5 [0.5–3.0]; LNF: 3.0 [1.0–4.0], $p=0.08$) scores did not show any statistically significant differences. Device removal was performed in 4 (16%) MSA patients and reoperation in 3 (7%) LNF patients. Eighty-nine percent of LNF patients reported satisfaction with the procedure, compared to 70% of MSA patients ($p=0.09$).

Conclusions MSA appears to offer similar long-term improvement in disease-specific quality of life as LNF. For MSA, there was a trend toward reduced long-term bloating compared to LNF, but need for reoperation and device removal may be associated with patient dissatisfaction.

Keywords GERD · Magnetic sphincter augmentation · Anti-reflux surgery · Fundoplication · GERD-Health Related Quality of Life · Patient-reported outcomes

Approximately 13% of all people worldwide experience heartburn or regurgitation, symptoms of gastroesophageal reflux disease (GERD), every week [1]. Proton pump inhibitors (PPIs) provide symptomatic relief in most cases, but up to 40% of patients experience symptoms that persist despite optimal high-dose therapy [2]. Laparoscopic anti-reflux surgery (LARS), including laparoscopic Nissen fundoplication (LNF), plays a key role in the treatment of GERD symptoms for these patients. The symptomatic benefit of LARS for objectively diagnosed, PPI-recalcitrant GERD has been demonstrated in multiple studies [3–5], but it is estimated that relatively few eligible patients are referred for surgical

✉ Kyle A. Perry
Kyle.Perry@osumc.edu

1 Division of Minimally Invasive Surgery, University of Michigan, Ann Arbor, MI, USA

2 Department of Surgery, The Ohio State University Wexner Medical Center, Columbus, OH, USA

3 Department of Surgery, University of Cincinnati, Cincinnati, OH, USA

4 Center for Minimally Invasive Surgery, The Ohio State University Wexner Medical Center, N729 Doan Hall, 410 W 10th Ave, Columbus, OH, USA

Published online: 18 January 2022

intervention due to the risk of complications, technical variability in the procedure, and local referral patterns [6, 7].

As a result, magnetic sphincter augmentation (MSA) has been developed as a relatively less technically demanding procedure that may be able to offer more consistent results for patients with symptomatic GERD [8, 9]. Several studies, most notably, the original device trial registries, have demonstrated disease-specific improvements in quality of life after MSA [10, 11] for up to 12 years post-placement.

Other studies have compared MSA against laparoscopic fundoplication for up to 3 years post-placement [12]. However, no studies to date have directly compared cohorts of MSA and LNF patients as long as 5 years. Therefore, we hypothesized that the improvements in disease-specific quality of life seen after MSA and LNF would persist at long-term follow-up in a cohort of patients with symptomatic, PPI-refractory GERD.

Materials and methods

Between March 3, 2013 and July 14, 2015, patients undergoing either laparoscopic Nissen fundoplication (LNF) or Magnetic Sphincter Augmentation (MSA) with the Linx device (Torax Medical, Inc.) were enrolled in an IRB-approved, prospectively maintained single-institution registry at the Ohio State University Wexner Medical Center. Patients were considered for laparoscopic anti-reflux surgery with MSA or LNF if they presented with symptoms of reflux despite optimal PPI therapy, were > 18 and < 85 years old, had objective evidence of reflux with pH probe testing, and demonstrated normal esophageal motility (distal contractile amplitude > 35 mmHg or normal peristalsis on high-resolution esophageal manometry). Patients were excluded if they had a history of previous anti-reflux esophageal or gastric surgery, hiatal hernia > 5 cm or paraesophageal hernia as seen on upper endoscopy or barium esophagram, or endoscopically visible Barrett's esophagus. The decision to perform LNF or MSA was determined through shared decision making between the patient and primary surgeon. At our center, patients with medically refractory GERD and BMI > 40 are routinely referred for bariatric surgery rather than LARS. In rare cases in which a patient is unable to undergo a gastric bypass, we will selectively perform LARS. In this cohort, 1 patient with a BMI of 42.7 underwent LNF.

Patient variables including age, sex, body mass index (BMI), American Society of Anesthesiologists (ASA) Class, predominant symptoms, primary diagnosis (Gastroesophageal Reflux Disease (GERD) or Laryngopharyngeal Reflux Disease (LPRD)), PPI use and frequency, preoperative workup (pH testing, endoscopy, esophagram), patient-reported outcome questionnaires, and clinical outcomes (PPI use and frequency, endoscopic dilations, reoperations, and device removals) were collected in the database. Symptoms and disease-specific quality of life were measured using the gastroesophageal reflux disease health-related quality of life (GERD-HRQL) scale [13] and the Gastroesophageal Reflux Symptom Scale (GERSS) [14]. Patients were also asked if they were satisfied with the results of their treatment and if they would choose the same procedure again. These assessments were collected preoperatively in the clinic, postoperatively in the clinic setting 1–2 months after surgery, and on a long-term basis via telephone interview, with a median interval of 67 months (range 51–85 months). Up to three attempts over an approximately 2 week period were made to contact patients for long-term follow-up.

All operations were completed laparoscopically. The minimal dissection technique was employed for the first 5 MSA cases, but this was then abandoned as we felt that visualization of the posterior esophagus and exposure for hiatal closure was challenging. In all other cases for both MSA and LNF, a circumferential esophageal and mediastinal dissection was performed to achieve 3 cm of intra-abdominal esophagus, with reduction of any hiatal hernia if present. Cruroplasty was performed routinely with interrupted braided nonabsorbable polyester suture, using pledgets only during LNF cases. MSA was performed by dissecting between the posterior vagus nerve and the esophagus at the GE junction, and using a standardized sizing device to choose the appropriate diameter. LNF was performed in routine fashion, with division of the short gastric vessels and construction of a 2 cm circumferential wrap above the GE junction over a 56 French bougie. MSA patients were fed a regular diet immediately with the expectation of same-day discharge, and LNF patients were advanced to a full liquid diet prior to discharge on postoperative day 1.

Primary outcomes were GERD-HRQL composite score, GERSS composite score, and PPI use. Secondary outcomes included dysphagia and bloating symptoms (sub-items of the GERD-HRQL scale), need for endoscopic dilation, need for device removal or revisional operation, and patient satisfaction and willingness to choose the same procedure again. Of note, one patient had an MSA device removal less than 1 year postoperatively, and was unable to be contacted for long-term follow-up. Calculation of device removal and reoperation rates used the original cohort as the denominator. Analyses were performed using STATA/SE 15.1 (StataCorp, College Station, TX). Univariate results are displayed as mean ± standard deviation, median [interquartile range] or number (%) as indicated. Bivariate comparisons of outcomes between patients undergoing MSA and LNF were made using Pearson's Chi-square test or Fisher's exact test as appropriate for categorical variables. For continuous variables, Student's T-test was used for normally distributed variables and the Wilcoxon rank-sum test for nonparametric comparisons of symptom scales. Longitudinal within-group

outcomes were compared using the Wilcoxon matched pairs signed rank test between baseline and postoperative, between postoperative and long-term, and between baseline and long-term measurements. Sensitivity analyses were performed to include only patients with a minimum of 5 years (60 months) of follow-up. p values of $<=0.05$ were taken to be statistically significant.

Results

A total of 70 patients (25 MSA, 45 LNF) underwent surgical intervention and were included in the study cohort. Baseline demographic data, preoperative testing, and symptom scores are shown in Table 1. LNF patients were

Table 1 Baseline characteristics

Patient characteristics	MSA[a] ($n=25$)	LNF[b] ($n=45$)	p[i]
Age, years, mean ± [SD]	51.6 [13.6]	48.5 [13.5]	0.36
Gender, female, *N* (%)	52%	76%	**0.04**
BMI[c], kg/m^2, median [IQR]	27.1 [22.7–29.9]	30.4 [26.4–32.8]	**0.01**
ASA[d] class, *N* (%)			
1 or 2	15 (60%)	26 (58%)	0.86
3 or 4	10 (40%)	19 (42%)	0.86
Diagnosis, *N* (%)			
Gastroesophageal reflux	24 (96%)	43 (96%)	0.93
Laryngopharnygeal reflux	1 (4%)	2 (4%)	0.93
Baseline PPI[e] use, *N* (%)	23 (92%)	42 (93%)	0.84
Baseline PPI dose, *N* (%)			
None	2 (10%)	3 (7%)	0.83
Daily	11 (40%)	14 (31%)	0.28
Twice daily	12 (50%)	28 (62%)	0.25
Hiatal hernia			
None	18 (72%)	26 (58%)	0.24
Small (<5 cm)	7 (28%)	19 (42%)	0.24
Large (>5 cm)	0 (0%)	0 (0%)	1.00
Esophagitis, *N* (%)			
None	24 (96%)	38 (86%)	0.21
A or B	1 (4%)	5 (11%)	0.31
C or D	0 (0%)	1 (2%)	0.45
Barrett's esophagus	0 (0%)	0 (0%)	1.00
% Time pH<4, median [IQR]	8.9 [5.7–11.9]	10.7 [7.0–13.6]	0.12
DeMeester Score, median [IQR]	31.7 [21.6–42.5]	37.1 [27.6–56.5]	0.08
LES[f] pressure, mmHg, median [IQR]	17.0 [13.3–20.4]	17.6 [10.3–22.9]	0.92
Distal Esophageal Pressure, mmHg, median [IQR]	73 [57–80]	67 [44–122]	0.93
Peristalsis, % of swallows, median [IQR]	90 [80–100]	90 [80–100]	0.89
Baseline GERD-HRQL[g] Overall Score, median [IQR]	26 [19–32]	34 [25–40]	**0.02**
Baseline GERD-HRQL: Dysphagia, median [IQR]	2 [0–3]	3 [1–4]	**0.02**
Baseline GERD-HRQL: Bloating, median [IQR]	3 [3–4]	4 [3–5]	0.16
Baseline GERSS[h], median [IQR]	33 [23–44]	48 [34–60]	**0.008**

Statistically significant p values are in bold

[a]MSA: Magnetic sphincter augmentation

[b]LNF: Laparoscopic Nissen fundoplication

[c]BMI: Body mass index

[d]ASA: American Society of Anesthesiologists

[e]PPI: Proton pump inhibitor

[f]LES: Lower esophageal sphincter

[g]GERD-HRQL: Gastroesophageal reflux disease health-related quality of life

[h]GERSS: Gastroesophageal Reflux Symptom Scale

[i]Chi-square or Fisher's exact test used for categorical variables. Wilcoxon Rank-sum or Student's *T*-test used for continuous variables

more likely to be female (76% vs 52%, $p = 0.04$) and had higher median BMI (30.4 vs 27.1, $p = 0.01$), higher median total GERD-HRQL scores (34 vs 26, $p = 0.02$), higher dysphagia-specific GERD-HRQL item scores (3 vs 2, $p = 0.02$), and higher total GERSS scores (48 vs 33, $p = 0.008$). There were no significant differences in age, ASA class, presence of hiatal hernia or esophagitis, pH testing or esophageal motility.

Short-term postoperative results (Table 2) showed reductions in PPI use, improvements in total GERD-HRQL score, dysphagia- and bloating-specific HRQL scores, and total GERSS scores. There were no significant differences between MSA and LNF patients.

Long-term follow-up results are shown in Table 3. Response rate was 80% for both MSA and LNF patients. Median follow-up interval was 68 months for MSA (range 57–85) and 65 months for LNF (range 51–85). There were no significant differences between MSA and LNF patients in regard to follow-up interval, PPI use or dose, total GERD-HRQL score less than 10, median total HRQL-GERD score, dysphagia, bloating, total GERSS score, satisfaction, or willingness to have the operation performed again. A majority of both MSA and LNF patients at long-term follow-up expressed a high degree of satisfaction with the procedure (MSA: 70%, LNF: 89%, $p = 0.15$) and that they would choose the same operation again (MSA: 75%, LNF:

Table 2 Postoperative outcomes

Postoperative outcomes	MSA ($n = 25$)	LNF ($n = 40$)	p
Follow-up interval, months, median [IQR]	1.0 [0.8–2.0]	1.9 [1.0–2.0]	0.41
PPI use, *N* (%)	1 (4%)	1 (2%)	0.73
PPI dose, *N* (%)			
As needed	0 (0%)	0 (0%)	1.00
Daily	1 (4%)	1 (2%)	0.73
Twice daily	0 (0%)	0 (0%)	1.00
GERD-HRQL Overall Score, median [IQR]	9.5 [2.5–15.5]	6.5 [3.0–8.0]	0.25
GERD-HRQL: Dysphagia, median [IQR]	2.0 [0.5–3.0]	1.0 [0.0–3.0]	0.27
GERD-HRQL: Bloating, median [IQR]	2.0 [0.5–4.0]	2.0 [1.0–4.0]	0.73
GERD-HRQL Overall Score < 10, *N* (%)	12 (48%)	27 (68%)	0.12
GERSS overall score	13.0 [2.0–26.0]	7.0 [2.0–15.0]	0.18
Satisfied with procedure, *N* (%)	12 (48%)	16 (40%)	0.53

Table 3 Long-term outcomes

Long-term follow-up outcomes	MSA ($n = 20$)	LNF ($n = 36$)	p
Follow-up Interval, months, median [IQR]	68 [65–74]	65 [62–69]	0.08
PPI use, *N* (%)	8 (40%)	12 (33%)	0.62
PPI dose, *N* (%)			
None	12 (60%)	24 (67%)	0.62
As needed	1 (5%)	1 (3%)	0.67
Daily	6 (30%)	5 (14%)	0.15
Twice daily	1 (5%)	6 (17%)	0.21
GERD-HRQL Overall Score, median [IQR]	9.0 [4.0–14.0]	7.5 [2.5–14.0]	0.68
GERD-HRQL: Dysphagia, median [IQR]	1.0 [0.0–2.0]	0.0 [0.0–2.0]	0.96
GERD-HRQL: Bloating, median [IQR]	1.5 [0.5–3.0]	3.0 [1.0–4.0]	0.08
GERD-HRQL Overall Score < 10, *N* (%)	11 (55%)	19 (53%)	0.87
GERSS Overall Score, median [IQR]	10.0 [4.0–23.5]	11.0 [4.0–27.0]	0.89
Satisfied with procedure, *N* (%)	14 (70%)	31 (89%)	0.15
Would do again, *N* (%)	15 (75%)	31 (89%)	0.19
Stricture requiring dilation, *N* (%)	5 (25%)	13 (36%)	0.39
Total dilations performed, median [range]	0 [0–1]	0 [0–8]	0.40
Redo operation, *N* (%)[a]	5 (20%)	3 (7%)	0.32
Device removed, *N* (%)[a]	4 (16%)	–	
Postoperative month of reoperation, median [range]	34 [8–60]	56 [48–64]	0.10

[a]Percentage uses original cohort as the denominator

89%, $p = 0.19$), and these results did not significantly differ between procedures.

There were no significant differences in the rate of endoscopic dilation or revisional surgery (Table 3). Patients with MSA had a total of 5 reoperations (20%), including 1 repair of a recurrent hiatal hernia (without device removal) and 4 (16%) instances of device removal. These included 1 removal due to patient demand at 8 months, 1 removal for persistent bloating at 22 months, 1 removal for erosion at 45 months, and 1 for dysphagia with recurrent reflux at 60 months; only the patient with dysphagia underwent LNF at the time of removal. Of note, one patient demanded MSA device removal due to a perceived nickel allergy, despite the satisfactory resolution of reflux symptoms (total GERD-HRQL reduced from 50 to 7, GERSS reduced from 70 to 10). Patients with LNF had 3 (7%) reoperations, two for recurrent reflux symptoms (at 48 and 56 months) and one conversion for dysphagia (at 64 months) after failing management by endoscopic dilation.

When including only patients with a full 60 months of follow-up, data were available for 76% ($n = 19/25$) of MSA and 73% ($n = 33/45$) LNF patients. There were no significant changes to the long-term results when limiting the sample to the full 5-year cohort.

Table 4 Longitudinal comparisons of baseline, postoperative, and long-term outcomes

	Preoperative (PRE) $n=25$	Postoperative (PO) $n=25$	Long-term (LT) $n=20$	PRE to PO p	PO to LT p	PRE to LT p
Months of follow-up, median [IQR]	–	1 [1–2]	68 [65–74]	–	–	–
PPI use, *N* (%)	23 (92%)	1 (4%)	8 (40%)	**< 0.001**	**0.008**	**0.002**
PPI dose, *N* (%)						
As needed	2 (10%)	0 (0%)	1 (5%)	0.32	1.00	0.32
Daily	11 (40%)	1 (4%)	6 (30%)	**< 0.001**	**0.01**	0.17
Twice daily	12 (50%)	0 (0%)	1 (5%)	**< 0.001**	0.32	**0.002**
GERD-HRQL Overall Score < 10, N (%)	2 (8%)	12 (48%)	11 (55%)	**0.002**	0.76	**0.007**
GERD-HRQL Overall Score, median [IQR]	26.0 [19.0–32.0]	9.5 [2.5–15.5]	9.0 [4.0–14.0]	**0.002**	0.84	**0.001**
GERD-HRQL: Dysphagia, median [IQR]	2.0 [0.0–3.0]	2.0 [0.5–3.0]	1.0 [0.0–2.0]	0.58	0.06	0.10
GERD-HRQL: Bloating, median [IQR]	3.0 [3.0–4.0]	2.0 [0.5–4.0]	1.5 [0.5–3.0]	**0.05**	0.55	**0.02**
GERSS Overall Score	33.0 [23.0–44.0]	13.0 [2.0–26.0]	10.0 [4.0–23.5]	**0.001**	0.23	**< 0.001**
Satisfied with procedure, *N* (%)	–	12 (48%)	14 (70%)	–	0.53	–
Would do again, *N* (%)	–	–	15 (75%)	–	–	–
Redo operation, *N* (%)[a]	–	0 (0%)	5 (20%)	–	**0.03**	–
Device removal, *N* (%)[a]	–	0 (0%)	4 (16%)	–	**0.05**	–
	Preoperative (PRE) $n=45$	Postoperative (PO) $n=40$	Long Term (LT) $n=36$	PRE to PO p	PO to LT p	PRE to LT p
Months of follow-up, median [IQR]	–	2 [1–2]	65 [62–69]	–	–	–
PPI use, *N* (%)	42 (93%)	1 (2%)	12 (33%)	**< 0.001**	**0.001**	**< 0.001**
PPI dose, *N* (%)						
As Needed	3 (7%)	0 (0%)	1 (3%)	0.32	1.00	0.32
Daily	14 (31%)	1 (2%)	5 (14%)	**< 0.001**	**0.03**	**0.01**
Twice Daily	28 (62%)	0 (0%)	6 (17%)	**< 0.001**	**0.01**	**< 0.001**
GERD-HRQL Overall Score < 10, *N* (%)	2 (4%)	27 (68%)	19 (53%)	**< 0.001**	0.07	**0.001**
GERD-HRQL Overall Score, median [IQR]	34.0 [25.0–40.0]	6.5 [3.0–8.0]	7.5 [2.5–14.0]	**< 0.001**	0.30	**< 0.001**
GERD-HRQL: Dysphagia, median [IQR]	3.0 [1.0–4.0]	1.0 [0.0–3.0]	0.0 [0.0–2.0]	**0.003**	0.91	**0.001**
GERD-HRQL: Bloating, median [IQR]	4.0 [3.0–5.0]	2.0 [1.0–4.0]	3.0 [1.0–4.0]	**0.01**	0.13	**0.01**
GERSS Overall Score	48.0 [34.0–60.0]	7.0 [2.0–15.0]	11.0 [4.0–27.0]	**< 0.001**	**0.02**	**< 0.001**
Satisfied with procedure, *N* (%)	–	16 (40%)	31 (89%)	–	**< 0.001**	–
Would do again, *N* (%)	–	–	31 (89%)		–	–
Redo operation, *N* (%)[a]	–	0 (0%)	3 (7%)	–	0.08	–

Statistically significant p values are in bold

Statistical comparisons performed with Wilcoxon matched signed rank test between preoperative and postoperative (PRE to PO), postoperative and long-term (PO to LT), and preoperative and long-term (PRE to LT) values

[a]Percentage uses original cohort as the denominator

Within-group longitudinal outcomes are shown in Table 4. From baseline to postoperative assessment, both MSA and LNF patients had significant improvements in PPI use, bloating, and total GERD-HRQL and GERSS scores, while only LNF patients had a significant improvement in dysphagia. At long-term follow-up, PPI use in both groups had increased, but remained below baseline levels. Median total GERD-HRQL and bloating scores in both groups did not change between postoperative and long-term follow-up, and remained significantly improved relative to baseline levels. For dysphagia, postoperative improvements seen in LNF patients persisted to long-term follow-up, while in MSA patients, there was a trend toward improvement between postoperative and long-term follow-up, but this did not reach statistical significance. Total GERSS scores remained significantly improved relative to baseline in both groups. Both MSA and LNF patients tended toward increasing satisfaction with their condition over time, although this effect was only statistically significant among LNF patients.

Discussion

Five-year outcomes have been reported for both MSA and LNF separately, and up to 3 years of comparative data between MSA and LNF have recently been published through a manufacturer-funded registry [12]. However, this analysis is the first to our knowledge to compare outcomes for both procedures at 5 years postoperatively.

The major finding is that the postoperative improvement in GERD-specific quality of life seen after both MSA and LNF are sustained for a median of 5.5 years (and up to 7 years at the high end of the range in both groups). However, we report a higher rate of MSA device removal (16%) than that seen in the original trial registry with median 9 years of follow-up (9.2%) [11], and in the largest single-center series to date with 10 months of follow-up (6.7%) [15]. This may be a reflection of small sample size, as our cohort of 25 MSA placements is less than that of both Ferrari et al. [11] ($N=335$ placements, 124 patients with long-term follow-up) and Ayazi et al. [15] ($N=553$ placements). Ferrari et al. [11] reported the majority of device removals occurring at less than 6 years post placement, whereas one of the first trials in the USA reported no device removals after 5 years [10]. In our experience, MSA device removal was relatively straightforward and did not result in any postoperative complications. A complete profile of the need for device removal over time is not yet clear, but it is apparent that a significant proportion of removals will occur after the first year. In regard to the reasons for MSA device removal, we saw a variety, including recurrent symptoms, interval development of hiatal hernia with dysphagia, erosion, and patient request. Patient selection is important, as we did have one patient demand device removal at 8 months despite objectively excellent resolution of symptoms. Only 1 of the 4 patients undergoing device removal had a simultaneous fundoplication performed, for dysphagia and recurrent reflux (DeMeester score 29). The remaining 3 patients have not subsequently gone on to fundoplication, and any recurrent reflux symptoms have been managed medically. It is notable that we no longer employ the minimal dissection technique first advocated for MSA, and now routinely perform a circumferential dissection and cruroplasty in all MSA cases.

We observed baseline differences in gender, BMI and GERD-HRQL in a non-randomly assigned cohort. The decision to perform LNF or MSA was reached through shared decision making with the patient. Factors commonly discussed included procedure complexity, reversibility, and insurance coverage and cost. Although not the focus of this analysis, MSA patients frequently must self-pay for the procedure, while LNF is typically covered by insurance. Although not reaching statistical significance in this cohort, LNF patients also tended to have a higher rate of hiatal hernia. It may be the case that patients with relatively less severe symptoms are more inclined to pursue a potentially reversible, less surgically complex procedure, even at greater out-of-pocket cost, while those patients with relatively more severe disease and larger hiatal hernias may feel more comfortable with a well-established procedure, or are simply unable or unwilling to pay higher out-of-pocket costs. A significant proportion of patients in both the MSA (40%) and LNF (33%) groups were using PPIs again at long-term follow-up, although only 5% of MSA and 17% of LNF patients required twice-daily dosing, which is consistent with findings from previous long-term studies [11]. In regard to specific postoperative symptoms, MSA has been associated with more severe dysphagia initially relative to LNF [16], and LNF with worse bloating relative to MSA [17, 18]. In this study, MSA patients did not experience a statistically significant improvement in dysphagia relative to baseline, although the trend was toward improvement between the postoperative and long-term follow-up time points (median score 2.0 to 1.0, $p=0.06$). In regard to bloating, median point estimates suggest more bloating with LNF relative to MSA, but this effect did not reach statistical significance (MSA 1.5 vs LNF 3.0, $p=0.08$). Gains in patient satisfaction over time did not reach statistical significance in the MSA group, which may simply reflect a small sample size and 16% device removal rate. Alternatively, this could suggest that LNF patients experience more initial unpleasant effects (e.g., inability to belch, flatulence) that lessen over time, while the experience of having the MSA device changes less as time goes on.

Overall, our results are consistent with previous comparative analyses of MSA and laparoscopic anti-reflux surgery with more than 12 months of follow-up data. Asti et al. [16]

compared 135 MSA placements to 103 Toupet fundoplications performed between 2007 and 2014, with at least 12 months of follow-up for all patients, and up to 80 months of follow-up for 24 MSA and 10 Toupet patients. They found that overall GERD-HRQL scores improved from 21 to 3 with MSA and 20 to 3 with Toupet, with these improvements remaining stable for the duration of the follow-up period. Dysphagia was significantly more common in MSA patients at 3 months, but this difference disappeared at the 12 month mark. There was no significant difference in use of PPIs or gas-related symptoms. The MSA device was removed in 7 patients (5%) for several reasons (dysphagia 3, recurrent heartburn 3, erosion 1) and re-do fundoplication was performed in 4 patients (4%), all for recurrent heartburn.

Reynolds et al. [17] performed 62 MSA placements and 117 Nissen fundoplications between 2010 and 2013, then selected 50 pairs of patients using propensity score matching to generate a cohort with no significant differences in baseline demographic or disease-specific variables. The MSA technique employed in that study was the minimal dissection approach, with selective cruroplasty performed in less than half of patients with hiatal hernias > 2 cm. GERD-HRQL scores decreased from 20 to 4 with MSA and from 19 to 4 with Nissen at 12 month follow-up. They reported no statistically significant differences in rates of PPI use or dysphagia; Nissen patients experienced significantly more frequent symptoms of severe gas bloat. No device removals or reoperations occurred within the first postoperative year.

Warren et al. [18] compared 201 MSA placements to 214 Nissen fundoplications at 12 months, and found similar improvements in GERD-HRQL; a propensity-matched subset of 114 pairs showed higher rates of mild dysphagia and resumption of PPI use with MSA, with more gas bloat symptoms in the Nissen group. Two device removals were reported, one for erosion, and one for persistent reflux.

A manufacturer-sponsored comparative observational study has reported results at 12 months [19] and most recently at 36 months [12]. This non-randomized trial reported only on patients completing 12 months of follow-up, which included 202 MSA and 47 laparoscopic fundoplication (LF) patients. Fundoplication patients were significantly older, had more hiatal hernias > 3 cm, greater incidence of Barrett's esophagus, and higher rates of severe GERD (6% in MSA vs 62% in LF). At 12 months, both groups had markedly improved GERD-HRQL scores, but discontinuation of PPIs was only 63% in the LF group, significantly lower than 82% in the MSA group, and lower than rates reported in contemporaneous trials. The MSA device was removed in 8 (4%) of cases, for either pain, dysphagia, or persistent reflux. At three years of follow-up, the groups included 465 MSA patients and 166 LF patients, with only 9% of the MSA patients experiencing severe GERD at baseline, compared to 82% of the LF patients. Improvement in GERD-HRQL scores was similar over 3 years. PPI discontinuation was more comparable—76% in the MSA group and 80% in the LF group. 12 devices were removed in the MSA group; LF patients underwent 11 reoperations.

Our study has several limitations. It is a single-institution, retrospective, observational comparison, and thus at risk of selection bias and threats to external validity. Although we noted baseline differences in gender, BMI, and total GERD-HRQL, there was no significant difference in long-term disease-specific quality of life. Therefore, although LNF patients consequently demonstrated a greater reduction in overall GERD-HRQL scores from baseline to long-term in our cohort, we can only say with certainty that this reflects differences in selection rather than differences in the relative effectiveness of the procedures. Additionally, sensitivity analyses of long-term total GERD-HRQL after using propensity score matching on gender, BMI, and baseline total GERD-HRQL showed no significant average treatment effect of MSA relative to LNF. Ayazi et al. reported male sex and higher GERD-HRQL score as independent predictors of success with MSA [15]; therefore, the gender disparity seen in our study may reflect selection that is accounting for this effect. We also report on a relatively smaller population than that described in larger series, and thus several effects may not have reached statistical significance. However, this study does report a higher rate of follow-up than other long-term studies. At 3 years, Bonavina et al. reported 61% follow-up for MSA ($n = 278/457$) and 49% for LNF ($n = 80/163$) [12]. In contrast, we report 80% long-term follow-up for both MSA and LNF, and 76% and 73%, respectively, for MSA and LNF among a cohort with a minimum of 5 years of follow-up. As a single-institution study, it is unclear to what degree these results apply to a broader population.

In conclusion, for patients with symptomatic reflux and objective evidence of GERD, both MSA and LNF produce significant and durable improvements in disease-specific quality of life for at least 5 years postoperatively, and patients largely remain satisfied with these procedures.

Funding None.

Declarations

Disclosures Drs. O'Neill, Jalilvand, Colvin, Haisley, and Perry have no conflicts of interest or financial ties to disclose.

References

1. Eusebi LH, Ratnakumaran R, Yuan Y, Solaymani-Dodaran M, Bazzoli F, Ford AC (2018) Global prevalence of, and risk factors

for, gastro-oesophageal reflux symptoms: a meta-analysis. Gut 67(3):430–440
2. Kahrilas PJ, Jonsson A, Denison H, Wernersson B, Hughes N, Howden CW (2012) Regurgitation is less responsive to acid suppression than heartburn in patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol 10(6):612–619
3. Galmiche JP, Hatlebakk J, Attwood S, Ell C, Fiocca R, Eklund S, Långström G, Lind T, Lundell L, LOTUS Trial Collaborators (2011) Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD: the LOTUS randomized clinical trial. JAMA 305(19):1969–1977
4. Hatlebakk JG, Zerbib F, Bruley des Varannes S, Attwood SE, Ell C, Fiocca R, Galmiche JP, Eklund S, Långström G, Lind T, Lundell LR, LOTUS Study Group (2016) Gastroesophageal acid reflux control 5 years after antireflux surgery, compared with long-term esomeprazole therapy. Clin Gastroenterol Hepatol 14(5):678–85.e3
5. Spechler SJ, Hunter JG, Jones KM, Lee R, Smith BR, Mashimo H, Sanchez VM, Dunbar KB, Pham TH, Murthy UK, Kim T, Jackson CS, Wallen JM, von Rosenvinge EC, Pearl JP, Laine L, Kim AW, Kaz AM, Tatum RP, Gellad ZF et al (2019) Randomized trial of medical versus surgical treatment for refractory heartburn. N Engl J Med 381(16):1513–1523
6. Liu JY, Finlayson SR, Laycock WS, Rothstein RI, Trus TL, Pohl H, Birkmeyer JD (2003) Determining an appropriate threshold for referral to surgery for gastroesophageal reflux disease. Surgery 133(1):5–12
7. Finlayson SR, Birkmeyer JD, Laycock WS (2003) Trends in surgery for gastroesophageal reflux disease: the effect of laparoscopic surgery on utilization. Surgery 133(2):147–153
8. Bonavina L, DeMeester T, Fockens P, Dunn D, Saino G, Bona D, Lipham J, Bemelman W, Ganz RA (2010) Laparoscopic sphincter augmentation device eliminates reflux symptoms and normalizes esophageal acid exposure: one- and 2-year results of a feasibility trial. Ann Surg 252(5):857–862
9. Asti E, Aiolfi A, Lazzari V, Sironi A, Porta M, Bonavina L (2018) Magnetic sphincter augmentation for gastroesophageal reflux disease: review of clinical studies. Updates Surg 70(3):323–330
10. Ganz RA, Edmundowicz SA, Taiganides PA, Lipham JC, Smith CD, DeVault KR, Horgan S, Jacobsen G, Luketich JD, Smith CC, Schlack-Haerer SC, Kothari SN, Dunst CM, Watson TJ, Peters J, Oelschlager BK, Perry KA, Melvin S, Bemelman WA, Smout AJ et al (2016) Long-term outcomes of patients receiving a magnetic sphincter augmentation device for gastroesophageal reflux. Clin Gastroenterol Hepatol 14(5):671–677
11. Ferrari D, Asti E, Lazzari V, Siboni S, Bernardi D, Bonavina L (2020) Six to 12-year outcomes of magnetic sphincter augmentation for gastroesophageal reflux disease. Sci Rep 10(1):13753
12. Bonavina L, Horbach T, Schoppmann SF, DeMarchi J (2021) Three-year clinical experience with magnetic sphincter augmentation and laparoscopic fundoplication. Surg Endosc 35(7):3449–3458
13. Velanovich V, Vallance SR, Gusz JR, Tapia FV, Harkabus MA (1996) Quality of life scale for gastroesophageal reflux disease. J Am Coll Surg 183:217–224
14. Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, Goldsmith C (2006) A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for treatment of patients with chronic gastroesophageal reflux disease: one-year follow-up. Surg Innov 13:238–249
15. Ayazi S, Zheng P, Zaidi AH, Chovanec K, Salvitti M, Newhams K, Hoppo T, Jobe BA (2020) Clinical outcomes and predictors of favorable result after laparoscopic magnetic sphincter augmentation: single-institution experience with more than 500 patients. J Am Coll Surg 230(5):733–743
16. Asti E, Bonitta G, Lovece A, Lazzari V, Bonavina L (2016) Longitudinal comparison of quality of life in patients undergoing laparoscopic Toupet fundoplication versus magnetic sphincter augmentation: Observational cohort study with propensity score analysis. Medicine 95(30):e4366
17. Reynolds JL, Zehetner J, Wu P, Shah S, Bildzukewicz N, Lipham JC (2015) Laparoscopic magnetic sphincter augmentation vs laparoscopic Nissen fundoplication: a matched-pair analysis of 100 patients. J Am Coll Surg 221(1):123–128
18. Warren HF, Reynolds JL, Lipham JC, Zehetner J, Bildzukewicz NA, Taiganides PA, Mickley J, Aye RW, Farivar AS, Louie BE (2016) Multi-institutional outcomes using magnetic sphincter augmentation versus Nissen fundoplication for chronic gastroesophageal reflux disease. Surg Endosc 30(8):3289–3296
19. Riegler M, Schoppman SF, Bonavina L, Ashton D, Horbach T, Kemen M (2015) Magnetic sphincter augmentation and fundoplication for GERD in clinical practice: one-year results of a multicenter, prospective observational study. Surg Endosc 29(5):1123–1129

Stand 25.01.2023

Gemeinsamer Bundesausschuss

Beschlussentwurf

des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung gemäß § 137e des Fünften Buches Sozialgesetzbuch: Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Vom T. Monat JJJJ

Der Gemeinsame Bundesausschuss (G-BA) hat in seiner Sitzung am T. Monat JJJJ folgende Richtlinie zur Erprobung beschlossen:

I. Die Richtlinie zur Erprobung wird wie folgt gefasst:

„Richtlinie des Gemeinsamen Bundesausschusses zur Erprobung der magnetischen Ösophagus-Sphinkter-Augmentation zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit (Erprobungs-Richtlinie MSA bei GERD)

§ 1 Zielsetzung

[1]Um den Gemeinsamen Bundesausschuss (G-BA) in die Lage zu versetzen, eine abschließende Bewertung des Nutzens der magnetischen Ösophagus-Sphinkter-Augmentation (MSA) zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit (GERD), die für eine laparoskopischen Fundoplicatio (LF) geeignet sind, durchzuführen, sollen im Wege der Erprobung die hierfür nach den §§ 135 und 137c des Fünften Buches Sozialgesetzbuch (SGB V) in Verbindung mit den Vorgaben der Verfahrensordnung des G-BA (VerfO) notwendigen Erkenntnisse für die Bewertung des Nutzens der Methode gewonnen werden. [2]Die für die Beantwortung dieser Frage in ihrer Konkretisierung nach § 2 notwendige Studie soll durch eine unabhängige wissenschaftliche Institution (UWI) nach Maßgabe dieser Richtlinie entworfen, durchgeführt und ausgewertet werden. [3]Die Ausgestaltung des Studiendesigns ist – soweit nicht im Folgenden näher bestimmt – von der UWI auf der Basis des Standes der wissenschaftlichen Erkenntnisse vorzunehmen und zu begründen. [4]Bei der Erstellung des Studienprotokolls ist das Wirtschaftlichkeitsprinzip zu beachten.

§ 2 Fragestellung

Mit der Erprobungsstudie soll nachgewiesen werden, dass bei Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen GERD die MSA im Vergleich zur LF bezüglich des Endpunkts gesundheitsbezogene Lebensqualität nicht unterlegen ist

GKV-SV
und ein Vorteil bezüglich der Krankenhausverweildauer besteht

.

§ 3 Population

[1]In die Erprobungsstudie einzuschließen sind erwachsene Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen GERD infolge einer Schwäche oder Insuffizienz des unteren ösophagealen Sphinkters, die trotz maximaler konservativer Refluxtherapie entweder weiterhin unter GERD-Symptomen leiden oder bereits eine GERD-assoziierte Komplikation im oberen Gastrointestinaltrakt erlitten haben. [2]Die weiteren Einschlusskriterien und konkreten Ausschlusskriterien (z. B. Komorbiditäten, Kontraindikationen für die Prüf- oder Vergleichsintervention) sind so festzulegen, dass eine Übertragbarkeit der Ergebnisse auf die Zielpopulation gemäß Satz 1 ermöglicht wird.

§ 4 Intervention und Vergleichsintervention

(1) Die Prüfintervention besteht in der MSA: über einen laparoskopischen Eingriff wird ein der Größe des unteren Ösophagus-Sphinkters (UÖS) entsprechendes, flexibles, ringförmiges Implantat um den UÖS gelegt, welches über seine magnetischen Anziehungskräfte die Verschlussfunktion des UÖS in physiologischer Weise unterstützt.

(2) Die Vergleichsintervention ist die laparoskopische Fundoplicatio.

§ 5 Endpunkte

(1) Primärer Endpunkt ist die gesundheitsbezogene Lebensqualität.

GKV-SV
Koprimärer Endpunkt ist die Krankenhausverweildauer.

Die genaue Operationalisierung der Endpunkte ist im Rahmen der konkreten Studienplanung festzulegen.

(2) [1]Als sekundäre Endpunkte sind zu erheben:

- GERD-bezogene Symptome (insbesondere Sodbrennen und Regurgitationen

PatV
und Schlafqualität

)

- postoperative Morbidität (insbesondere Dysphagie und die Rückkehr zu normalen Alltagsaktivitäten)

DKG, KBV, PatV
– Krankenhausverweildauer

- Re-Hospitalisierungen
- unerwünschte Ereignisse (insbesondere die Unfähigkeit aufzustoßen oder zu erbrechen.

[2]Die Operationalisierung der Endpunkte sowie die Erhebung und die Operationalisierung weiterer Endpunkte sind jeweils zu begründen.

(3) Sofern vorhanden, sind für alle Endpunkte validierte Erhebungsinstrumente zu verwenden.

§ 6 Studientyp und Beobachtungszeitraum

(1) [1]Die Erprobungsstudie ist als randomisierte, kontrollierte Studie (RCT) zu konzipieren und durchzuführen. [2]Die Studie soll multizentrisch durchgeführt werden.

(2) Die Endpunkterhebung ist zu verblinden.

(3) Die patientenindividuelle (Nach-)Beobachtungszeit ist so zu bestimmen, dass die Gewinnung hinreichender Informationen zu Langzeiteffekten der Intervention sichergestellt ist, und soll mindestens 12 Monate ab Randomisierung betragen.

(4) Die Art und Anzahl weiterer therapeutischer Interventionen mit Bezug zur GERD oder mit möglichem Einfluss auf die zu erfassenden Endpunkte sind zu dokumentieren.

§ 7 Anforderungen an die Qualität der Leistungserbringung im Rahmen der Erprobung

Es ist in jedem Studienzentrum sicherzustellen, dass die Behandlung gemäß dem Studienprotokoll unter Berücksichtigung aller erforderlichen, anerkannten, nach ethischen und wissenschaftlichen Gesichtspunkten aufgestellten Regeln für die Durchführung von klinischen Studien erfolgt.

§ 8 Anforderungen an die Durchführung, die wissenschaftliche Begleitung und die Auswertung der Erprobung

(1) Im Auftrag an die unabhängige wissenschaftliche Institution ist diese – unabhängig davon, ob die Erprobung durch den G-BA oder Hersteller oder Unternehmen durchgeführt wird – insbesondere zu verpflichten,

a) ein Studienprotokoll zu erstellen und dieses sowie gegebenenfalls die Amendments unverzüglich nach Fertigstellung an den G-BA zur weitergehenden Information zu übersenden,

b) die Konformität des Studienprotokolls mit den Vorgaben der Erprobungs-Richtlinie und bei Abweichungen gegenüber diesen Vorgaben eine Begründung bei Übersendung des Studienprotokolls darzulegen,

c) die Studie in einem einschlägigen, von der World Health Organization akkreditierten Register klinischer Studien zu registrieren und den Eintrag regelmäßig zu aktualisieren und den G-BA hierüber zu informieren,

d) zur Durchführung der Erprobung nach den Anforderungen der Richtlinie und nach Maßgabe des Auftrags, einschließlich der datenschutzkonformen Erhebung, Speicherung und Nutzung der Daten und der Einholung von erforderlichen Genehmigungen,

e) Bericht zu erstatten an den G-BA bei Abweichungen von den Vorgaben in der Erprobungs-Richtlinie,

f) zur Auswahl der Leistungserbringer, Festsetzung und Auszahlung der angemessenen Aufwandsentschädigung an diese,

g) zur Auswertung der Studie,

h) unverzüglich nach Abschluss der Studie den Studienbericht, der entsprechend der International Council for Harmonisation (ICH)-E3-Richtlinie zu erstellen ist, zusammen mit dem statistischen Analyseplan an den G-BA zu übermitteln,

i) dem G-BA das Recht einzuräumen, ihm auf seine Kosten eine nachträgliche Datenauswertung zur Verfügung zu stellen und

j) dem G-BA das Recht zur Veröffentlichung zumindest der Synopse des Studienberichts sowie weitergehender für seine Entscheidung relevanter Informationen aus dem Studienbericht und aus den nachträglichen Datenauswertungen einzuräumen.

(2) [1]Wird die Studie vom G-BA durchgeführt, ist die unabhängige wissenschaftliche Institution in diesem Fall zu verpflichten, an den G-BA zu festgelegten Meilensteinen Bericht zu erstatten. [2]Außerdem ist die unabhängige wissenschaftliche Institution in Ergänzung der Verpflichtung nach Absatz 1 Buchstabe j zu beauftragen, dass sie die Studienergebnisse spätestens 3 Monate nach Abnahme des Studienberichts durch den G-BA zur Veröffentlichung in einer Fachzeitschrift mit wissenschaftlichem Begutachtungsprozess einreicht und dem G-BA das Recht einräumt, im Anschluss an deren Veröffentlichung oder nach Ablauf eines Jahres nach Einreichung der Studienergebnisse den Studienbericht zu veröffentlichen. [3]Die wissenschaftliche Institution arbeitet vertrauensvoll mit der mit dem Projektmanagement beauftragten Stelle zusammen und hat dieser die zur Ausübung ihrer Aufgabe erforderlichen Informationen und Unterlagen zur Verfügung zu stellen.

(3) [1]Wird die Studie durch Medizinprodukthersteller oder Unternehmen durchgeführt, sind diese verpflichtet, die Anforderungen dieser Richtlinie an die Durchführung und Auswertung der Erprobung zu beachten. [2]Um sicherzustellen, dass eine durchgeführte Studie den Anforderungen dieser Richtlinie entspricht und geeignet ist, die notwendigen Erkenntnisse des Nutzens der Methode zu gewinnen, haben die durchführenden Medizinprodukthersteller und Unternehmen dem G-BA das Studienkonzept zur Prüfung vorzulegen und zu erklären, dass der Vertrag mit der unabhängigen wissenschaftlichen Institution den Anforderungen nach Absatz 1 entspricht und eine Einflussnahme durch den Sponsor auf das Ergebnis der Studie vertraglich ausgeschlossen ist. [3]Bei positivem Ergebnis der Überprüfung bescheinigt der G-BA die Konformität des vorgelegten Studienkonzepts mit den Anforderungen dieser Richtlinie und dass damit die im Rahmen der Erprobung erbrachten Leistungen von der GKV übernommen werden; andernfalls teilt er die bestehenden Defizite mit.“

II. Die Richtlinie tritt am Tag nach der Veröffentlichung im Bundesanzeiger in Kraft.

Die Tragenden Gründe zu diesem Beschluss werden auf den Internetseiten des G-BA unter www.g-ba.de veröffentlicht.

Berlin, den T. Monat JJJJ

Gemeinsamer Bundesausschuss
gemäß § 91 SGB V
Der Vorsitzende

Prof. Hecken

Stand 25.01.2023

Gemeinsamer Bundesausschuss

Tragende Gründe

zum Beschlussentwurf des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung gemäß § 137e des Fünften Buches Sozialgesetzbuch:
Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Vom T. Monat JJJJ

Inhalt

1. Rechtsgrundlage

Gemäß § 137e Absatz 7 des Fünften Buches Sozialgesetzbuch (SGB V) können unabhängig von einem Beratungsverfahren nach § 135 oder § 137c SGB V Hersteller eines Medizinprodukts, auf dessen Einsatz die technische Anwendung einer neuen Untersuchungs- oder Behandlungsmethode maßgeblich beruht und Unternehmen, die in sonstiger Weise als Anbieter einer neuen Methode ein wirtschaftliches Interesse an einer Erbringung zulasten der Krankenkassen haben, beim Gemeinsamen Bundesausschuss (G-BA) beantragen, dass dieser eine Richtlinie zur Erprobung der neuen Methode nach § 137e Absatz 1 SGB V beschließt.

Der G-BA regelt in der Richtlinie nach § 137e Absatz 1 Satz 1 SGB V die in die Erprobung einbezogenen Indikationen und die sächlichen, personellen und sonstigen Anforderungen an die Qualität der Leistungserbringung im Rahmen der Erprobung. Er legt zudem Anforderungen an die Durchführung, die wissenschaftliche Begleitung und die Auswertung der Erprobung fest (§ 137e Absatz 2 Satz 1 und 2 SGB V).

2. Eckpunkte der Entscheidung

2.1 Hintergrund

Der G-BA hat am 18. März 2022 den Antrag auf Erprobung der magnetischen Ösophagus-Sphinkter-Augmentation (MSA) zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit (GERD), die für eine laparoskopischen Fundoplicatio (LF) geeignet sind, positiv beschieden. Danach weist die Methode das Potenzial einer erforderlichen Behandlungsalternative im Vergleich zur LF auf, welches auf den Erkenntnissen zur Reduktion postoperativer Komplikationen hinsichtlich der postoperativen Unfähigkeit, aufzustoßen und zu erbrechen sowie auf einer Reduktion der postoperativen Morbidität, gemessen an der Krankenhausverweildauer. Der G-BA hat in gleicher Sitzung am 18. März 2022 das Beratungsverfahren über eine Richtlinie zur Erprobung der MSA zur Behandlung von Patientinnen und Patienten mit GERD, die für eine LF geeignet sind, eingeleitet.

In der Folge hat sich der Unterausschuss Methodenbewertung (UA MB) versichert, dass keine weiteren abgeschlossenen oder laufenden Studien vorlagen, die grundsätzlich geeignet wären, derzeit oder in naher Zukunft den Nachweis des Nutzens dieser Methode zu liefern.

2.2 Zu § 1 Zielsetzung

Die in **Satz 1** formulierte Zielsetzung dieser Erprobungs-Richtlinie verdeutlicht, dass die entsprechend den Vorgaben dieser Erprobungs-Richtlinie zu konzipierende Erprobungsstudie geeignet sein muss, die in § 2 konkretisierte Fragestellung zu beantworten. Damit wird dem G-BA eine Bewertung des Nutzens dieser Methode auf einem für eine spätere Richtlinienentscheidung ausreichend sicheren Erkenntnisniveau erlaubt.

Mit **Satz 2** wird vorgeschrieben, dass eine unabhängige wissenschaftliche Institution (UWI) mit der Planung, Durchführung und Auswertung einer Studie beauftragt werden soll, die den Vorgaben dieser Erprobungs-Richtlinie entspricht.

Die UWI wird mit **Satz 3** verpflichtet, aus Gründen der Objektivierbarkeit und Nachvollziehbarkeit sämtliche Festlegungen der Parameter des Studiendesigns nach wissenschaftlichen Kriterien zu treffen; damit wird sichergestellt, dass die Zielsetzung nach § 1 Satz 1 erreicht werden kann.

Das Wirtschaftlichkeitsprinzip ist gemäß § 1 **Satz 4** bereits bei der Erstellung des Studienprotokolls zu beachten, da sich die späteren Studienkosten unmittelbar oder mittelbar aus den im Studienprotokoll spezifizierten Eckdaten und Parametern (z. B. der benötigten Patientenzahl, der Studiendauer, der Anzahl der Studienzentren, der Studienvisiten und der Qualitätssicherung) ergeben. Darüber hinaus ist gemäß 2. Kapitel § 25 Absatz 3 Spiegelstrich 3 VerfO neben der fachlichen Eignung sowie der Geeignetheit des Angebots der angebotene Preis der wissenschaftlichen Begleitung und Auswertung ein Kriterium für die Beauftragung der UWI.

2.3 Zu § 2 Fragestellung

Mit der hier definierten Fragestellung adressiert der G-BA die am 18. März 2022 im Rahmen der Potenzialbescheidung festgestellte Erkenntnislücke. Die potenzialbegründenden Studien ließen erkennen, dass die MSA zur Behandlung von erwachsenen Patientinnen und Patienten mit einer anhand pH-Metrie nachgewiesenen GERD infolge einer Schwäche oder Insuffizienz des unteren ösophagealen Sphinkters, die trotz maximaler konservativer Refluxtherapie entweder weiterhin unter GERD-Symptomen leiden oder bereits eine GERD-assoziierte Komplikation im oberen Gastrointestinaltrakt erlitten haben, ein hinreichendes Potenzial für eine Erprobung gemäß § 137e SGB V bietet. Die Studienergebnisse sind mit der Erwartung verbunden, dass die MSA einer LF bezüglich des Endpunkts gesundheitsbezogene Lebensqualität nicht unterlegen ist, und sie zudem Vorteile durch eine möglicherweise geringere postoperative Komplikationsrate mit besserem Erhalt der Fähigkeit aufzustoßen und zu erbrechen sowie eine möglicherweise geringere postoperative Morbidität mit kürzerer Krankenhausverweildauer aufweist. Mit der geplanten Erprobungsstudie soll die benötigte Erkenntnissicherheit im Sinne der Zielsetzung von § 1 erreicht werden.

Die Begründungen zu den einzelnen Komponenten der Fragestellung (Population, Intervention, Vergleichsintervention, Endpunkt) sind in den nachfolgenden Kapiteln abgebildet.

2.4 Zu § 3 Population

Zu Satz 1

Die Studienpopulation baut auf der im Erprobungsantrag durch die Antragstellerin definierten Patientenpopulation auf. Da es sich bei der MSA um einen invasiven Eingriff handelt, sind Patientinnen und Patienten erst nach maximalen verhaltenspräventiven Maßnahmen (Gewichtsreduktion, Nahrungsumstellung, Schlafen mit erhöhtem Kopfende des Bettes, etc.) und einer maximalen medikamentösen Refluxtherapie, die zu keiner Verbesserung der GERD-Symptome führte, oder die ein fortgeschrittenes Stadium der Erkrankung und damit zusammenhängende Funktionsdefekte und Komplikationen im oberen Gastrointestinaltrakt aufweisen, einzuschließen.

Zu Satz 2

Bei der Studienplanung sind die weiteren Einschlusskriterien und die konkreten Ausschlusskriterien, wie z. B. Begleiterkrankungen, Kontraindikationen für die Prüf- oder Vergleichsintervention oder andere Einflussfaktoren auf den Endpunkt gesundheitsbezogene Lebensqualität festzulegen. Dabei ist darauf zu achten, dass die Übertragbarkeit der Ergebnisse auf die Zielpopulation (gemäß § 3 Satz 1) nicht gefährdet wird.

2.5 Zu § 4 Intervention und Vergleichsintervention

Zu Absatz 1

Für das für die Prüfintervention eingesetzte Medizinprodukt muss eine Verkehrsfähigkeit vorliegen und die Studienpopulation von der Zweckbestimmung umfasst sein. Die Intervention besteht in einer Behandlung mit der magnetischen Ösophagus-Sphinkter-Augmentation, deren wesentliche Prozessschritte im zweiten Halbsatz aufgeführt werden.

Zu Absatz 2

Als Vergleichsintervention erfolgt eine LF, da gemäß Leitlinien[1,2] eine chirurgische Behandlung empfohlen wird, wenn die doppelte Protonenpumpeninhibitoren (PPI)-Gabe bzw. eine maximale konservativ-medikamentöse Therapie ausgereizt bzw. kontraindiziert ist.

2.6 Zu § 5 Endpunkte

Zu Absatz 1

Der primäre Endpunkt ist die gesundheitsbezogene Lebensqualität , erfasst zum Beispiel mit dem krankheitsspezifischen Fragebogen Health Related Quality of Life (HRQL) für Patientinnen und Patienten mit GERD (GERD-HRQL). Damit schließt sich der G-BA den Erkenntnissen aus der potenzialbegründenden Evidenz an, wonach gezeigt werden soll, dass die MSA und LF hinsichtlich der gesundheitsbezogenen Lebensqualität vergleichbar sind.

GKV-SV
In dem koprimären Endpunkt kürzere Krankenhausverweildauer soll sich der Vorteil der MSA im Vergleich zur LF zeigen.

Zu Absatz 2

Die gewählten sekundären Endpunkte ergänzen den primären

GKV-SV
und koprimären

Endpunkt durch weitere patientenrelevante Endpunkte und dienen zur weiteren Beurteilung möglicher Effekte, die auch die methodenimmanenten Vorteile umfassen. Es soll vor allem gezeigt werden, dass die MSA gegenüber der LF auch Vorteile in Bezug auf die postoperative Morbidität und die unerwünschten Ereignisse (insbesondere die Unfähigkeit aufzustoßen oder zu erbrechen) hat. Die Operationalisierung der einzelnen Endpunkte wie auch die Festlegung zusätzlicher Endpunkte obliegt der UWI, die diese jeweils zu begründen hat.

1 Hunt R, Armstrong D, Katelaris P, Afihene M, Bane A, Bhatia S, et al. World Gastroenterology Organisation Global Guidelines: GERD Global Perspective on Gastroesophageal Reflux Disease. J Clin Gastroenterol 2017;51(6):467-478.

2 Pauwels A, Boecxstaens V, Andrews CN, Attwood SE, Berrisford R, Bisschops R, et al**.** How to select patients for antireflux surgery? The ICARUS guidelines (international consensus regarding preoperative examinations and clinical characteristics assessment to select adult patients for antireflux surgery). Gut 2019;68(11):1928-1941.

Zu Absatz 3

Grundsätzlich sind, wo immer möglich, in der betreffenden Indikation validierte Instrumente zur Erhebung der Endpunkte einzusetzen. Von besonderer Bedeutung ist dies bei subjektiven Endpunkten, d. h. solchen, die auf Befragung von Studienteilnehmenden, an der Behandlung beteiligten Personen oder Dritten beruhen.

2.7 Zu § 6 Studientyp und Beobachtungszeitraum

Zu Absatz 1

In Satz 1 ist geregelt, dass die Erprobungsstudie als eine randomisierte, kontrollierte Studie (RCT) zu konzipieren und durchzuführen ist, da jedenfalls dieser Studientyp ein ausreichend sicheres Erkenntnisniveau für eine spätere Methodenbewertung bietet. Die Studie soll multizentrisch durchgeführt werden, da die Aussagekraft multizentrischer Studien im Allgemeinen höher ist als bei monozentrischen Studien. Das liegt vornehmlich daran, dass der Einfluss lokaler Besonderheiten auf das Ergebnis reduziert wird. Zudem können schneller höhere Patientenzahlen rekrutiert werden. Weitere Konkretisierungen des Designs sind von der UWI vorzunehmen und zu begründen.

Zu Absatz 2

Um mögliche Verzerrungen des Studienergebnisses zu vermeiden, die aufgrund der Kenntnis der Gruppenzuordnung entstehen können, sind die Personen, die die Endpunkte erheben, gegen die Intervention bzw. Vergleichsintervention zu verblinden.

Eine Verblindung der behandelnden Personen ist aufgrund des Charakters der Intervention bzw. Vergleichsintervention nicht möglich. Auf eine Verblindung der Studienteilnehmenden könnte verzichtet werden, weil eine eventuelle Durchführung von Diagnostik per Magnetresonanztomografie nach LF problemlos, nach MSA jedoch nur eingeschränkt möglich ist, so dass eine Entblindung für Notfallsituationen ermöglicht werden müsste, was organisatorisch und rechtlich zu erheblichem Mehraufwand führen würde.

Zu Absatz 3

Dieser Absatz regelt, dass eine ausreichend lange patientenindividuelle (Nach-)Beobachtungszeit für die Studie einzuplanen ist, um hinreichende Informationen zu den Effekten der Intervention zu erhalten. Nach Einschätzung des G-BA ist dies jedenfalls nicht bei einer patientenindividuellen Beobachtungszeit von weniger als 12 Monaten gewährleistet. Die zur Potenzialbewertung herangezogenen Studien berichteten Daten zu patientenrelevanten Endpunkten zum Erhebungszeitpunkt von rund einem Jahr (Skubleny 2017[3]) und sogar nach drei Jahren (Bonavina 2021[4]).

Zu Absatz 4

Um eine mögliche Verzerrung bzw. das Ausmaß der Verzerrung auf den Effekt in beiden Gruppen abschätzen zu können, ist die Art und Anzahl weiterer therapeutischer Interventionen mit Bezug zur Grunderkrankung oder mit möglichem Einfluss auf die zu

3 Skubleny D, Switzer NJ, Dang J et al. LINX((R)) magnetic esophageal sphincter augmentation versus Nissen fundoplication for gastroesophageal reflux disease: a systematic review and meta-analysis. Surg Endosc 2017; 31(8): 3078-3084. https://dx.doi.org/10.1007/s00464-016-5370-3.

4 Bonavina L, Horbach T, Schoppmann SF et al. Three-year clinical experience with magnetic sphincter augmentation and laparoscopic fundoplication. Surg Endosc 2021; 35(7): 3449-3458. https://dx.doi.org/10.1007/s00464-020-07792-1.

erfassenden Endpunkte zu dokumentieren. Hierzu gehören beispielsweise die Einnahme von PPI und eine stationäre Behandlung.

2.8 Zu § 7 Anforderungen an die Qualität der Leistungserbringung im Rahmen der Erprobung

Bei der Durchführung von Erprobungsstudien des G-BA mit Medizinprodukten soll die Gute Klinische Praxis gemäß ISO 14155 (Klinische Prüfung von Medizinprodukten an Menschen - Gute Klinische Praxis) angewendet werden.

Die Gute Klinische Praxis ist ein internationaler ethischer und wissenschaftlicher Standard für Planung, Durchführung, Dokumentation und Berichterstattung von klinischen Studien am Menschen. Die Einhaltung dieses Standards schafft öffentliches Vertrauen, dass die Rechte, die Sicherheit und das Wohl der Prüfungsteilnehmerinnen und -teilnehmer gemäß der Deklaration von Helsinki geschützt werden und die bei der klinischen Studie erhobenen Daten glaubhaft sind.

2.9 Zu § 8 Anforderungen an die Durchführung, die wissenschaftliche Begleitung und die Auswertung der Erprobung

Zu Absatz 1

Absatz 1 beschreibt die notwendigen Inhalte des Auftrags an die UWI. Die in Absatz 1 aufgeführten Auftragsinhalte gelten sowohl für die durch Hersteller oder Unternehmen als auch durch den G-BA beauftragte wissenschaftliche Begleitung und Auswertung der Erprobung. Nur bei Vorliegen eines den Anforderungen dieses Absatzes genügenden Vertrages mit der UWI ist die Erprobung als konform mit der Erprobungs-Richtlinie anzusehen und kann damit als Erprobung im Sinne des § 137e SGB V gewertet und im Leistungsanteil von der GKV finanziert werden.

Nach Buchstabe a) soll die Übersendung des Studienprotokolls und der Amendements die rasche Abklärung von Zweifelsfragen ermöglichen; eine Gesamtprüfung auf Konformität des Studienprotokolls mit den Vorgaben der Erprobungs-Richtlinie wird vom G-BA nicht von Amts wegen vorgenommen.

In Buchstabe b) wird die UWI verpflichtet, die Konformität des Studienprotokolls mit den Vorgaben der Erprobungs-Richtlinie gegenüber dem G-BA zur weitergehenden Information mit Übersendung des Studienprotokolls darzulegen. Zeitgleich hat die wissenschaftliche Institution Abweichungen von den Vorgaben zu begründen. Dies eröffnet nicht die Möglichkeit, von der Erprobungs-Richtlinie abzuweichen.

Nach Buchstabe c) ist die Studie in einem einschlägigen, von der World Health Organization (WHO) akkreditierten Register klinischer Studien zu registrieren und der Eintrag regelmäßig zu aktualisieren. Der G-BA ist hierüber zu informieren. Zu den akkreditierten Registern zählen derzeit insbesondere das Deutsche Register Klinischer Studien (DRKS) und das ClinicalTrials.gov; eine vollständige Übersicht findet sich auf der Homepage der WHO (https://www.who.int/clinical-trials-registry-platform/network/data-providers). Durch die Registrierung wird der weltweite Überblick über laufende Studien unterstützt, der für die Transparenz der Studiendurchführung und auch für den G-BA insbesondere bei Methodenbewertungen wichtig ist.

Nach Buchstabe e) ist vorgesehen, dass Abweichungen von den Vorgaben der Erprobungs-Richtlinie im Laufe der Erprobung durch die UWI dem G-BA mitzuteilen sind.

Nach den Buchstaben f) und g) ist die UWI verpflichtet, die Leistungserbringer auszuwählen, die angemessene Aufwandsentschädigung festzusetzen und an diese auszuzahlen sowie die Studie auszuwerten.

Nach Buchstabe h) ist nach Abschluss der Studie der Studienbericht zusammen mit dem statistischen Analyseplan an den G-BA ohne schuldhaftes Zögern zu übermitteln. Es wird zwingend vorgegeben, dass dieser entsprechend der International Council for Harmonisation (ICH)-E3-Richtlinie zu erstellen ist.

Gemäß Buchstabe i) ist dem G-BA die Möglichkeit einzuräumen, auf eigene Kosten Datenauswertungen bei der UWI durchführen zu lassen. Die Datenhoheit verbleibt bei den durch Unternehmen und Hersteller durchgeführten Erprobungen grundsätzlich bei diesen Sponsoren. Da jedoch gesichert sein muss, dass die Bewertung der Studie durch den G-BA dadurch nicht beeinträchtigt wird, muss er die durch den Studienbericht nicht eindeutig beantworteten relevanten Fragen aufklären können.

Synopse im Sinne des Buchstaben j) meint eine der ICH-E3-Leitlinie Annex I entsprechende Übersicht zu den wesentlichen Eckdaten und Ergebnissen der Studie. Durch die in Buchstabe j) vorgesehene Regelung sichert der G-BA die Verwertbarkeit der Erprobungsstudie, weil die Qualität der Studie sowie Einzelfragen unter Umständen nur mit den angeforderten Daten oder deren spezifischer Auswertung geprüft werden können. Der G-BA geht davon aus, dass die Studienergebnisse zeitnah nach der Übermittlung des Studienberichts an den G-BA zur Veröffentlichung in einer referenzierten Fachzeitschrift eingereicht werden.

Zu Absatz 2

Absatz 2 legt erweiterte Verpflichtungen für die UWI fest, die gelten, wenn die Beauftragung der UWI durch den G-BA erfolgt:

Die UWI hat über die vereinbarten Meilensteine dem G-BA gegenüber Bericht zu erstatten. Über Absatz 1 Buchstabe j) hinausgehend, hat der G-BA im Auftrag mit der UWI festzulegen, dass diese die Studienergebnisse spätestens 3 Monate nach Abnahme des Studienberichts zur Veröffentlichung in einer Fachzeitschrift mit wissenschaftlichem Begutachtungsprozess einreicht. Sie hat dem G-BA im Anschluss an deren Veröffentlichung oder nach Ablauf eines Jahres nach Einreichung der Studienergebnisse das Recht zur Veröffentlichung des Studienberichts einzuräumen. Satz 3 legt fest, dass die wissenschaftliche Institution vertrauensvoll mit der mit dem Projektmanagement beauftragten Stelle zusammenzuarbeiten und dieser die zur Ausübung ihrer Aufgabe erforderlichen Informationen und Unterlagen zur Verfügung zu stellen hat. Die Verpflichtung ist ebenso im Vertrag mit der UWI zu regeln.

Zu Absatz 3

Absatz 3 stellt klar, dass die beteiligten Hersteller und Unternehmen sämtliche Anforderungen der Erprobungs-Richtlinie zu beachten haben, damit ihre Studie als Erprobung im Sinne des § 137e SGB V gewertet und im Leistungsanteil von der GKV finanziert wird.

Die Regelungen sehen vor, dass Medizinprodukthersteller und Unternehmen gehalten sind, in Abstimmung mit dem G-BA sicherzustellen, dass die Vorgaben nach § 137e Absatz 2 Satz 2 SGB V beachtet werden. Dem G-BA ist daher nach Absatz 3 Satz 2 das Studienkonzept und eine Erklärung, dass der Vertrag mit der UWI den Anforderungen nach Absatz 1 entspricht und eine Einflussnahme durch den Sponsor auf das Ergebnis der Studie vertraglich ausgeschlossen ist, vor Beauftragung einer UWI in deutscher Sprache vorzulegen. Damit erfolgt nicht erst nach Studienabschluss eine Prüfung der Konformität von Inhalt der Erprobungs-Richtlinie und Studiendurchführung und die Finanzierung im Leistungsanteil von der GKV wird bestätigt. Der G-BA bescheinigt nach positivem Prüfergebnis die Konformität. Weisen die vorgelegten Unterlagen hingegen noch Defizite auf, weil die Studie ausweislich der vorgelegten Unterlagen

den Anforderungen der Richtlinie nach § 137e Absatz 1 Satz 1 SGB V nicht entspricht oder nicht geeignet ist, die notwendigen Erkenntnisse des Nutzens der Methode zu gewinnen, wird dies dem vorlegenden Unternehmen oder Hersteller mitgeteilt, das beziehungsweise der daraufhin die verbesserten Unterlagen erneut zur Prüfung einreichen kann.

3. Würdigung der Stellungnahmen

Wird nach dem Stellungnahmeverfahren ergänzt.

4. Bürokratiekostenermittlung

Durch den vorgesehenen Beschluss entstehen keine neuen bzw. geänderten Informationspflichten für Leistungserbringer im Sinne von Anlage II zum 1. Kapitel VerfO und dementsprechend keine Bürokratiekosten.

5. Schätzung der Studienkosten entsprechend 2. Kapitel § 22 Absatz 2 Satz 4 VerfO

Die folgenden Ausführungen zur Fallzahlschätzung sind nicht als verbindliche Kalkulation, sondern als näherungsweise Schätzung der benötigten Fallzahlen zu verstehen (auf Basis der oben aufgeführten Überlegungen zum Studiendesign).

Zur Schätzung der Fallzahl ist die Nichtunterlegenheitsfragestellung zum Endpunkt gesundheitsbezogene Lebensqualität maßgeblich und hierbei insbesondere die Wahl der Nichtunterlegenheitsgrenze. Es ist davon auszugehen, dass eine Studiengröße, die zum Nachweis einer zwischen MSA und LF vergleichbaren gesundheitsbezogenen Lebensqualität ausreicht, auch hinreichend sicher zeigen kann, dass die MSA gegenüber der LF Vorteile in Bezug auf die postoperative Morbidität und die unerwünschten Ereignisse (insbesondere die Unfähigkeit auf-zustoßen oder zu erbrechen) hat.

Für die Prüfung auf Nichtunterlegenheit wird die standardisierte Mittelwertdifferenz Hedges'g angewendet. Die Nichtunterlegenheitsgrenze von 0,25 wird für den Endpunkt gesundheitsbezogene Lebensqualität (gemessen mittels GERD-HRQL) als adäquat betrachtet. Hieraus ergibt sich als grobe Approximation eine Fallzahl von ca. 400 Patientinnen und Patienten.

Die Ausführungen zur Fallzahlschätzung sind nicht als verbindliche Kalkulation, sondern als näherungsweise Schätzung der benötigten Fallzahl zu verstehen. Eine konkrete Fallzahlkalkulation und resultierende Kostenschätzung kann erst im Rahmen der genauen Studienplanung durch die UWI erfolgen.

Im Ergebnis von Informationen der Koordinierungszentren für Klinische Studien, dem Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen sowie dem DLR Projektträger (Projektmanagement für Erprobungen des G-BA) schätzt der G-BA die Kosten pro Teilnehmer auf Basis der Studiengröße und des studienbezogenen Mehraufwands (s. nachstehende Tabelle).

Studiengröße (n)	studienbezogener Mehraufwand		
	gering	normal	hoch
klein (< 100)	8.000 €	10.000 €	12.000 €
mittel (100 bis < 500)	4.000 €	5.500 €	7.000 €
groß (≥ 500)	2.000 €	3.000 €	4.000 €

Entsprechend der o. g. beispielhaften Fallzahlschätzung handelt es sich um eine mittlere Studie (100 bis < 500). Der studienbezogene Mehraufwand wird als normal (hier etwa 5.500 € je Studienteilnehmer) eingeschätzt. Auf der Basis dieser Annahmen lassen sich geschätzte Studienkosten von 2,2 Millionen € berechnen.

6. Verfahrensablauf

Datum	**Gremium**	**Beratungsgegenstand/ Verfahrensschritt**
18.03.2022	Plenum	Einleitung des Beratungsverfahrens zur Erprobungs-Richtlinie gemäß § 137e SGB V
11.08.2022	UA MB	Beschluss zur Ankündigung des Beratungsverfahrens im Bundesanzeiger und Freigabe des Fragenkatalogs zur strukturierten Einholung von ersten Einschätzungen anlässlich der Ankündigung des Beratungsverfahrens (gemäß 2. Kapitel § 6 VerfO)
22.08.2022		Ankündigung des Beratungsverfahrens im Bundesanzeiger
23.02.2023	UA MB	Einleitung des Stellungnahmeverfahrens
	UA MB	Mündliche Anhörung und Würdigung der Stellungnahmen
	UA MB	Beratung der Beschlussunterlagen und Beschlussempfehlung für das Plenum
	Plenum	Abschließende Beratung und Beschlussfassung

7. Fazit

Der Gemeinsame Bundesausschuss beschließt die Richtlinie zur Erprobung der magnetischen Ösophagus-Sphinkter-Augmentation zur Behandlung von Patientinnen und Patienten mit Gastroösophagealer Refluxkrankheit.

Berlin, den T. Monat JJJJ

Gemeinsamer Bundesausschuss
gemäß § 91 SGB V
Der Vorsitzende

Prof. Hecken

Gemeinsamer Bundesausschuss

Stellungnahme zum Beschlussentwurf des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung:

Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Bitte klicken Sie hier und geben dann den Namen der stellungnehmenden Organisation ein.
Bitte klicken Sie hier und fügen das Datum Ihrer Stellungnahme ein

Zu §1 Zielsetzung	
Stellungnahme / Änderungsvorschlag	**Begründung**
Bitte nutzen Sie nach Möglichkeit für inhaltlich voneinander abgrenzbare Aspekte Ihrer Stellungnahme bzw. Änderungsvorschläge jeweils gesonderte Tabellenzeilen und fügen bei Bedarf weitere Tabellenzeilen hinzu. Vielen Dank.	**Bitte fügen Sie hier eine entsprechende Begründung ein.**

Zu §2 Fragestellung	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §3 Population	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §4 Intervention und Vergleichsintervention	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §5 Endpunkte	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §6 Studientyp und Beobachtungszeitraum	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §7 Anforderungen an die Qualität der Leistungserbringung im Rahmen der Erprobung	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §8 Anforderungen an die Durchführung, die wissenschaftliche Begleitung und die Auswertung der Erprobung	
Stellungnahme / Änderungsvorschlag	**Begründung**

Voraussichtliche Teilnahme an der mündlichen Anhörung

Bitte klicken Sie hier und geben dann den Namen der stellungnehmenden Organisation ein.		
Die Anhörung findet voraussichtlich im 2. Quartal 2023 statt		
Teilnahmeoptionen	**Einladung**	**Ihre Rückmeldung zur Teilnahme**
Wir nehmen teil.	**Eine gesonderte Einladung wird Ihnen zugesandt**	**Bitte klicken Sie hier und geben dann "Wir nehmen teil." ein**
Wir können derzeit nicht sagen, ob wir an der Anhörung teilnehmen.	**Eine gesonderte Einladung wird Ihnen zugesandt**	**Bitte klicken Sie hier und geben dann "Wir nehmen teil." ein**
Wir nehmen nicht teil. Auch bei Terminänderungen für diese Anhörung möchten wir nicht teilnehmen.	**Sie werden nicht zur Anhörung eingeladen.**	**Bitte klicken Sie hier und geben dann "Wir nehmen nicht teil." ein**

Stellungnahme zum Beschlussentwurf des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung:

Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Deutsche Gesellschaft für Allgemein- und Viszeralchirurgie (DGAV) Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)
17. März 2023

Zu §1 Zielsetzung	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §2 Fragestellung	
Stellungnahme / Änderungsvorschlag	**Begründung**
Die Krankenhausverweildauer kann nicht als primärer Endpunkt der Studie dienen	**In der aktuellen Krankenhausbehandlung wird die laparoskopische Fundoplicatio mit der DRG G19C und einer Mindestverweildauer von 2 Tagen vergütet. Die laparoskopische Implantation der magnetischen Sphinkteraugmentation wird mit der nächst höheren DRG G19B vergütet, die eine Mindestverweildauer von 3 Tagen beinhaltet.** **Daher ist schon durch das DRG System der Krankenhausaufenthalt über die Mindestverweildauer vorgegeben und kann nicht als Studienendpunkt dienen. Die Krankenhausverweildauer bei Patienten mit magnetischer Sphinkteraugmentation ist allein systembedingt deutlich länger.**
	Zahlen aus den Zentren für Refluxchirurgie zeigen, dass die mittlere Verweildauer der Patienten im Bereich der Mindestverweildauer liegt.
Kosten	**Noch nicht in die Studienbetrachtung mit einbezogen sind die zusätzlichen Kosten des Implantates. Aktuell werden diese über die höhere DRG mit einem CM Faktor von 2,1 (8 398€) im Vergleich zur**

	laparoskopischen Fundoplicatio mit einem CM Faktor von 1,38 (5 519€) finanziert.

Zu §3 Population	
Stellungnahme / Änderungsvorschlag	**Begründung**
OP Indikation anhand der aktuellen Leitlinie übernehmen	**Anfang März ist die neue S2k-Leitlinie zur GERD veröffentlicht worden. Die Indikation zur Operation sollte nach diesen Leitlinien erfolgen.** **Eine maximale Ausschöpfung der konservativen Therapie ist nicht immer erforderlich.**
	Eine nachgewiesene Schwäche des unteren Ösophagussphinkters ist nicht erforderlich.
Schweregrad der Refluxerkrankung mit einbeziehen	**Für den operativen Erfolg der GERD ist der Schweregrad der Refluxerkrankung präoperativ entscheidend. Je schwerwiegender die Erkrankung hinsichtlich entzündlicher Veränderungen der Speiseröhre oder Größe der vorhandenen Hiatushernie, desto schwieriger ist die Operation und desto höher ist die Rezidivgefahr.** **Daher ist eine Vergleichbarkeit der Patientenkollektive von entscheidender Bedeutung.**

Zu §4 Intervention und Vergleichsintervention	
Stellungnahme / Änderungsvorschlag	**Begründung**
Spezifizierung der Fundoplicatio	**Die laparoskopische Fundoplicatio kann in 2 Varianten durchgeführt werden. Es gibt die 360° Manschette nach Nissen und die 270° Manschette nach Toupet. Es gibt keine eindeutige Evidenz für die Überlegenheit einer dieser Verfahren, daher wird sie in den Leitlinien äquivalent angewendet.** **Allerdings unterscheiden sie sich im Nebenwirkungsprofil, so dass dies mit einbezogen werden sollte.**

Zu §5 Endpunkte	
Stellungnahme / Änderungsvorschlag	**Begründung**
Primärer Endpunkt sollten die GERD bezogenen Symptome sein	**Die gesundheitsbezogene Lebensqualität ist sicher entscheidend bei der Behandlung der GERD, aber neben den refluxtypischen Symptomen auch von weiteren Faktoren abhängig. Da es primär um die Behandlung der Refluxerkrankung geht, sollten diese Symptome auch im Fokus stehen.** **Die Lebensqualität sollte sekundärer Endpunkt sein.**
Die Krankenhausverweildauer kann kein Endpunkt sein	**Wie oben schon erwähnt, ist DRG abhängig bei der Implantation des MSA ein Krankenhausaufenthalt von mindestens 3 Tagen erforderlich, bei der laparoskopischen Fundoplicatio ein Aufenthalt von mindestens 2 Tagen.**
Instrumente der Lebensqualität	**Neben dem HRQL sollte auch der vielfach angewandte Gastrointestinal Quality of Life Index (GIQLI) angewandt werden. Dieses Instrument ist validiert und wird in vielen Publikationen genutzt.**
Weitere sekundäre Endpunkte	**Die postoperative Komplikation des Implantates durch Arrosion, Dislokation oder Infektion ist zwingend mit aufzunehmen.**
	Die Revisionsrate, wie z.B. die Explantation des Implantates, das Slipping (Implantat oder Manschette) oder die Rezidivhernie, ist ebenfalls ein wichtiger sekundärer Endpunkt.
Definition der unerwünschten Ereignisse sollte objektivierbar sein	**Die unerwünschten Ereignisse sollten objektivierbar sein. Die Unfähigkeit aufzustoßen oder zu erbrechen sind sehr schwer klinisch nachzuvollziehen. Beispielsweise muss nicht jeder Mensch im nahen postoperativen Verlauf erbrechen und die Differenzierung zwischen nicht aufstoßen können und vermehrten Blähungen (ein häufiges Symptom nach der Operation) fällt fast allen Patienten sehr schwer.**

Zu §6 Studientyp und Beobachtungszeitraum	
Stellungnahme / Änderungsvorschlag	**Begründung**

Studiendauer mehr als 12 Monate	**Eine Studiendauer von mindestens 12 Monaten Nachbeobachtung ist sicher das Minimum. Eine längere Nachbeobachtung wäre sinnvoll, da Komplikationen und Revisionen erst später auftreten.**
Verblindung	**Eine Verblindung ist sicher sehr sinnvoll.** **Allerdings können die Patienten nicht verblindet werden, da sie bei der laparoskopischen MSA Implantation einen Implantatausweis erhalten. Dieser ist zwingend erforderlich, weil z.B. die Untersuchung mittels MRT eingeschränkt ist. Insbesondere bei der Beurteilung der Lebensqualität kann das natürlich eine Rolle spielen.**

Zu §7 Anforderungen an die Qualität der Leistungserbringung im Rahmen der Erprobung	
Stellungnahme / Änderungsvorschlag	**Begründung**

Zu §8 Anforderungen an die Durchführung, die wissenschaftliche Begleitung und die Auswertung der Erprobung	
Stellungnahme / Änderungsvorschlag	**Begründung**

Voraussichtliche Teilnahme an der mündlichen Anhörung

Deutsche Gesellschaft für Allgemein- und Viszeralchirurgie (DGAV) Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)
Die Anhörung findet voraussichtlich im 2. Quartal 2023 statt

Teilnahmeoptionen	**Einladung**	**Ihre Rückmeldung zur Teilnahme**
Wir nehmen teil.	**Eine gesonderte Einladung wird Ihnen zugesandt**	Wir nehmen teil.
Wir können derzeit nicht sagen, ob wir an der Anhörung teilnehmen.	**Eine gesonderte Einladung wird Ihnen zugesandt**	**Bitte klicken Sie hier und geben dann "Wir nehmen teil." ein**
Wir nehmen nicht teil. Auch bei Terminänderungen für diese Anhörung möchten wir nicht teilnehmen.	**Sie werden nicht zur Anhörung eingeladen.**	**Bitte klicken Sie hier und geben dann "Wir nehmen nicht teil." ein**

Stellungnahme zum Beschlussentwurf des Gemeinsamen Bundesausschusses über eine Richtlinie zur Erprobung:

Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit

Johnson & Johnson Medical GmbH
20.03.2023

Zu §1 Zielsetzung	
Stellungnahme / Änderungsvorschlag	**Begründung**
Einverstanden, keine Änderungswünsche	----------

Zu §2 Fragestellung	
Stellungnahme / Änderungsvorschlag	**Begründung**
GKV-Vorschlag aufnehmen: „und ein Vorteil bezüglich der Krankenhausverweil-dauer besteht“	Wir schließen uns dem Vorschlag des GKV-SV an. Die Krankenhausverweildauer ist ein wichtiger patientenrelevanter Endpunkt und sollte als ko-primärer Endpunkt aufgenommen werden.

Zu §3 Population	
Stellungnahme / Änderungsvorschlag	**Begründung**
Einverstanden, keine Änderungswünsche	----------

Zu §4 Intervention und Vergleichsintervention	

Stellungnahme / Änderungsvorschlag	**Begründung**
Einverstanden, keine Änderungswünsche	----------

Zu §5 Endpunkte	
Stellungnahme / Änderungsvorschlag	**Begründung**
(1) ... Als ko-primärer Endpunkt (im Sinne einer hierarchisch geordneten Hypothesentestung) soll die Überlegenheit hinsichtlich der Krankenhausverweil-dauer getestet werden. ...	Wir schließen uns dem Vorschlag des GKV-SV an. Die Krankenhausverweildauer ist ein wichtiger patientenrelevanter Endpunkt und sollte als ko-primärer Endpunkt aufgenommen werden. Nach Auswertung der InEK Daten [InEK DatenBrowser - Unterjährige Datenlieferung DRG Januar bis Dezember 2022] beträgt die mittlere Verweildauer in Deutschland für die laparoskopische Fundoplicatio (OPS 5-448.42) 6,3 Tage, für die MSA (OPS 5-429.p1) 4,3 Tage. Dies sollte in der Erprobungsstudie verifiziert werden. Die Hypothesentestung kann hierarchisch geordnete erfolgen. Dabei wird im ersten Schritt die Nichtunterlegenheit der MSA gegenüber der LF hinsichtlich des primären Endpunktes (gesundheitsbezogene Lebensqualität) zum Signifikanzniveau von 5% getestet. Sofern sich hierbei die Nichtunterlegenheit zeigt, wird im zweiten Schritt die Überlegenheit der MSA gegenüber der LF hinsichtlich des ko-primären Endpunktes (Krankenhausverweildauer) ebenfalls zum Signifikanzniveau von 5% getestet.

Zu §6 Studientyp und Beobachtungszeitraum	
Stellungnahme / Änderungsvorschlag	**Begründung**
Absatz (3) ersetze „ab Randomisierung“ durch „nach Operation“	Der Zeitraum zwischen Randomisierung und Operation ist selten genau spezifiziert und kann durch logistische Schwierigkeiten oder Patientenpräferenz (z.B. „Ich möchte vor der OP noch gerne in Urlaub fahren“) zu einer starken Verzerrung des aktuellen und eigentlich relevanten Nachverfolgungszeitraum führen.

Zu §7 Anforderungen an die Qualität der Leistungserbringung im Rahmen der Erprobung	
Stellungnahme / Änderungsvorschlag	**Begründung**
Satz 2 zu ergänzen: „Ferner ist sicherzustellen, dass die an der Studie beteiligten Chirurgen hinreichend Erfahrung in der Antirefluxchirurgie haben und mit beiden Verfahren (MSA und LF) bestens vertraut sind.“	Die Einführung neuer Verfahren geht immer mit einer gewissen Lernkurve einher. Bei einem Vergleich eines neuen Verfahrens (hier die MSA) mit einer etablierten Methode (LF) besteht die Gefahr einer Verzerrung, wenn die Ergebnisse der neuen Methode im Wesentlichen durch die Lernkurve der jeweiligen Operateure bestimmt wird, die der etablierten Methode jedoch die langjährige klinische Praxis widerspiegelt. Dieses Verzerrungsrisiko gilt es zu vermeiden.

Zu §8 Anforderungen an die Durchführung, die wissenschaftliche Begleitung und die Auswertung der Erprobung	
Stellungnahme / Änderungsvorschlag	**Begründung**
Einverstanden, keine Änderungswünsche	----------

Voraussichtliche Teilnahme an der mündlichen Anhörung

Bitte klicken Sie hier und geben dann den Namen der stellungnehmenden Organisation ein.		
Die Anhörung findet voraussichtlich im 2. Quartal 2023 statt		
Teilnahmeoptionen	**Einladung**	**Ihre Rückmeldung zur Teilnahme**
Wir nehmen teil.	**Eine gesonderte Einladung wird Ihnen zugesandt**	Wir nehmen teil
Wir können derzeit nicht sagen, ob wir an der Anhörung teilnehmen.	**Eine gesonderte Einladung wird Ihnen zugesandt**	**----------**
Wir nehmen nicht teil. Auch bei Terminänderungen für diese Anhörung möchten wir nicht teilnehmen.	**Sie werden nicht zur Anhörung eingeladen.**	**----------**

Wortprotokoll

Gemeinsamer Bundesausschuss

einer Anhörung zum Beschlussentwurf des Gemeinsamen Bundesausschusses über eine Erprobungs-Richtlinie gemäß § 137e SGB V: Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit (ER-21-004)

Vom 27. April 2023

Vorsitzende:	Frau Dr. Lelgemann
Beginn:	11:21 Uhr
Ende:	11:47 Uhr
Ort:	Videokonferenz des Gemeinsamen Bundesausschuss Gutenbergstraße 13, 10587 Berlin

Teilnehmer der Anhörung

Deutsche Gesellschaft für Allgemein- und Viszeralchirurgie (DGAV)
Frau Prof. Dr. Jessica Leers

Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS)
Frau Prof. Dr. Jessica Leers

Johnson & Johnson Medical GmbH (JNJ)
Herr Dr. Thomas Kersting
Herr Dr. Thomas Andreae

Beginn der Anhörung: 11:21 Uhr

(Die angemeldeten Teilnehmer sind der Videokonferenz beigetreten.)

Frau Dr. Lelgemann (Vorsitzende): Einen schönen guten Morgen! Ich darf Sie ganz herzlich zur Anhörung zu unserer Erprobungs-Richtlinie gemäß § 137e SGB V: Magnetische Ösophagus-Sphinkter-Augmentation bei Gastroösophagealer Refluxkrankheit begrüßen. Wenn ich richtig informiert bin, war das ein Antrag auf Erprobung, der gestellt worden ist. Wir sind jetzt in der mündlichen Anhörung zu unserem Entwurf einer Erprobungsstudie, der Ihnen hier vorliegt.
Aus diesem Anlass begrüße ich heute hier für die Deutsche Gesellschaft für Allgemein- und Viszeralchirurgie (DGAV) Frau Prof. Dr. Jessica Leers. Sie ist ebenfalls heute hier vertreten für die Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS). – Guten Morgen!
Zudem begrüße ich für die Johnson & Johnson Medical GmbH (JNJ) Herrn Dr. Thomas Kersting und Herrn Dr. Thomas Andreae. – Guten Morgen!

Die üblichen Vorbemerkungen zu dieser Anhörung: Wir bedanken uns für Ihre schriftlichen Stellungnahmen, die wir erhalten und gewürdigt haben und die auch – zumindest bei mir – einige Fragen generiert haben, es ist also nicht erforderlich, alles noch einmal darzustellen.

Des Weiteren mache ich Sie darauf aufmerksam, dass wir von dieser Anhörung ein Wortprotokoll erzeugen, das hinterher Bestandteil der zentralen Dokumentation wird, die wir veröffentlichen.

Nun erteile ich Ihnen, Frau Prof. Dr. Leers, für die beiden eben genannten Fachgesellschaften das Wort.

Frau Prof. Dr. Leers (DGAV & DGVS): Vielen Dank für die Möglichkeit, hier angehört zu werden. Als Vertreterin beider Fachgesellschaften freue ich mich, das hier ein bisschen begleiten zu können. Grundsätzlich befürworte ich solch eine Studie, aber nicht mit den primären Endpunkten, die im Studiendesign festgelegt waren. Das muss sicherlich noch einmal neu evaluiert werden. Aber grundsätzlich ist solch eine Studie, um die Effektivität dieses Magnetbandes zu prüfen, sicherlich eine gute Sache.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank für die Stellungnahme, wir kommen sicher gleich noch einmal darauf zurück. – Dann Herr Dr. Kersting oder Herr Dr. Andreae für die Johnson & Johnson Medical GmbH.

Herr Dr. Kersting (JNJ): Vielen Dank. Wir haben zu dem Entwurf, den der G-BA vorgelegt hat, relativ wenig gesagt. Ein wichtiger Punkt für uns war – ich schließe jetzt an das an, was Frau Kollegin Leers gesagt hat –, dass wir dem Vorschlag, einen koprimären Endpunkt zur Verweildauer einzuführen, zustimmen, wir das für richtig halten, das hatten wir auch in unserer ersten Einschätzung bereits so vorgetragen. Insofern würde das von uns unterstützt.

Dann haben wir noch auf die Frage, wann der Zeitraum zwischen Randomisierung und Operation auseinandergehalten wird, und in Abs. 3 des Beschlusses den Zeitpunkt „ab Randomisierung zu erfassen“ durch den Zeitpunkt „ab Operation“ oder „nach Operation zu erfassen“ hingewiesen.

Ein wichtiger Punkt, der sich nur partiell abbildet in dem bisherigen Entwurf, ist der, dass wir sehr viel Wert darauf legen würden, dass diese Untersuchung an Zentren durchgeführt wird, an denen die Chirurgen über die entsprechende Erfahrung verfügen. Das heißt, es müssten auf jeden Fall durch die UWI später sichergestellt werden bei der Auswahl der Zentren, die das machen, dass das nachgewiesenermaßen keine Beginner sind, weder für die eine noch für die andere Methode. Das ist also ein ganz wichtiges Kriterium, gerade weil es eine neue

Methode ist, dass also da wirklich erfahrene Chirurgen zu Werke gehen, das müsste in der Beschreibung der Studie sichergestellt werden. Das waren eigentlich die wesentlichen Punkte.

Ansonsten sind wir mit dem Vorschlag, wie er unterbreitet worden ist, einverstanden. Wir haben in den ersten Anhörungen sehr genau darauf hingewiesen, dass wir die Durchführung einer RCTs für ein schwieriges Unterfangen halten. Wir wissen ja auch aus vorhergehenden Erprobungsstudien, dass, wenn weniger invasive Methoden mit Operationen verglichen werden, es oftmals sehr schwierig ist, diese Studie tatsächlich durchzuführen. Gleichwohl halten wir es für richtig, dass jetzt eine finale Erprobung den Nutzen bestätigt, der bereits im Potenzialbescheid dargelegt worden ist.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank, Herr Dr. Kersting. – Möchten Sie ergänzen, Herr Andreae?

Herr Dr. Andreae Kersting (JNJ): Dem ist nichts hinzuzufügen. Danke schön.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank. – Dann eröffne ich die Runde für Fragen aus dem Kreis des Unterausschusses.

PatV: Vielen Dank. Prof. Leers, Sie haben gerade den primären Endpunkt angesprochen und hierbei eine Änderung vorgeschlagen. Sie stellen auf die GERD-bezogenen Symptome ab. Können Sie noch einmal näher ausführen, welche Symptome dies konkret sind? Wie werden diese dann erfasst?

Frau Prof. Dr. Leers (DGAV & DGVS): Klassische Symptome sind Sodbrennen, Hochlaufen von Flüssigkeiten, gelegentlich Schluckstörungen, aber vor allen Dingen das Sodbrennen. Es geht natürlich auch um die Symptome nach der Operation, die dann andere sein können, in der Studienbeschreibung sind bereits die Unfähigkeit aufzustoßen und Blähungen angemerkt worden. Das ist insofern schwierig, weil beispielsweise die Unfähigkeit aufzustoßen – das zu differenzieren für einen Patienten in einem Gespräch, ob er nicht einfach nur viel Luft im Bauch hat oder ob die nicht hochkommt, das ist extrem schwierig.

Ich halte es für ganz wichtig, dass es möglichst objektive Kriterien gibt. Das ist bei Symptomen natürlich immer schwieriger, aber wir haben Symptom-Scores hinsichtlich der Schluckstörungen, da gibt es evaluierte und validierte Scores, auch was das Sodbrennen angeht. Es ist sicherlich ganz wichtig, da möglichst objektivierbare Symptome zu nehmen: um den Erfolg oder Misserfolg zu evaluieren.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank, Frau Prof. Leers.– Gibt es weitere Fragen?

GKV-SV: Wir wollten die Gelegenheit nutzen, mit den Experten ins Gespräch zu kommen, damit Sie noch einmal erörtern, worin sie persönlich den Vorteil der einen Methode im Vergleich zu der anderen sehen, damit man eine bessere Vorstellung davon erhält, wenn von der Fachgesellschaft die Krankenhausverweildauer nicht als der ideale Endpunkt angesehen wird, was man besser operationalisieren kann. Wenn Sie sagen, dass Ihre Antwort das schon erfasst, können Sie auch kurz antworten, ansonsten noch einmal die Bitte, das hier für uns etwas ausführlicher darzustellen.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank. – Frau Prof. Leers.

Frau Prof. Dr. Leers (DGAV & DGVS): Die Mindestverweildauer ist ja DRG-abhängig. Um die komplette DRG zu bekommen, habe ich Vorgaben. Und im Moment ist es so, dass die DRG 19c eine Mindestverweildauer von zwei Tagen beinhaltet. Das ist die DRG, die ich codiere für eine laparoskopische Fundoplicatio, also die Technik, die standardisiert seit 50 Jahren bereits durchgeführt wird. Um einen klinischen Eindruck zu erhalten: Grundsätzlich kann ich manche Patienten tatsächlich schon nach einem Tag möglicherweise klinisch relevant entlassen. Das bedeutet aber einen erheblichen Abschlag in der Finanzierung.

Bei dem LINX-Band ist es so, dass man Zusatzkosten für das Material hat, und da die DRG 19b zum Tragen kommt und da die Mindestverweildauer im Krankenhaus, um den kompletten

DRG-Satz zu bekommen, drei Tage beträgt, ergibt es meines Erachtens überhaupt keinen Sinn, die Mindestverweildauer, die uns durch die DRGs mehr oder weniger als Ziel vorgegeben ist, als primären Endpunkt zu nehmen.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank, Frau Prof. Leers, das hat uns überzeugt: Geld steuert unser System, das ist so. Ich glaube, die Frage zielte auf Folgendes ab: Die Krankenhausverweildauer hatten wir ja genommen, weil die Hypothese im Raum steht, dass dieses Verfahren deutlich weniger invasiv ist als das, was bisher – also laparoskopische Fundoplicatio – durchgeführt wird. Und wir haben uns überlegt: Sollte uns diese DRG-Vereinbarung zum Beispiel daran zweifeln lassen, dass das wirklich so viel weniger invasiv ist? Das ist zumindest meine Frage, die ich Ihnen gern stellen würde. Denn das ist ja im Prinzip das Versprechen dieser neuen Methode und da wollten wir Sie noch einmal um Ihre Einschätzung bitten.

Frau Prof. Dr. Leers (DGAV & DGVS): Aus meiner Erfahrung würde ich diese These nicht bestätigen. Ich bin zwar eine absolute Expertin in diesem Bereich, aber für mich ist die Fundoplicatio nicht im Wesentlichen weniger invasiv als das LINX-Band.

Frau Dr. Lelgemann (Vorsitzende): Nein, unsere Frage ist umgekehrt. Also das Versprechen der neuen Methode ist ja, dass die magnetische Ösophagus-Augmentation weniger invasiv sei als die Fundoplicatio.

Frau Prof. Dr. Leers (DGAV & DGVS): Entschuldigung. Ich halte das LINX-Band für nicht wesentlich weniger invasiv als die Fundoplicatio. Das ist eine These. Aber das ist natürlich etwas, was ich so auch nicht belegen kann.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank. Diese Einschätzung ist für uns noch mal wichtig, auch was die Studienplanung anbelangt. Wir haben das ja bisher als Nicht-Unterlegenheitsstudie geplant, von daher ist das für uns ganz entscheidend. – Wer möchte weitere Fragen stellen oder ergänzen? – Herr Kersting.

Herr Dr. Kersting (JNJ): Wir hatten in unserer Kommentierung zur letzten Anfrage bereits dazu Stellung genommen. Wir haben auch in unserer ersten Stellungnahme bereits etwas dazu gesagt. Die aktuellen Zahlen zeigen uns, dass die Fundoplicatio im Moment mit einer Verweildauer von 6,3 Tagen zu Buche schlägt und dass für die MSA-Eingriffe 4,3 Tage zum Tragen kommen. Das heißt, die untere Verweildauergrenze spielt im Moment überhaupt keine Rolle, weder bei dem einen noch bei dem anderen Eingriff.

Tatsächlich wissen wir, dass die Entlassung von Patienten mit Fundoplicatio nicht im teilambulanten oder tagesstationären Bereich erfolgen kann. Insofern erscheint uns dieser Hinweis, der vom GKV-Spitzenverband aufgegriffen worden ist, richtig: dass auch die Verweildauer beschrieben wird. Wenn man das jetzt in einer Studie überprüfen will, dann wäre es vielleicht richtig, dass sich die Vergütung an einem anderen Level orientiert als an dem, was üblicherweise für eine DRG erstattet wird. Ob das möglich ist oder nicht, lasse ich mal dahingestellt. Der G-BA hat da eventuell gewisse Möglichkeiten, das nach der Gesetzeslage zu regeln. Aber das wäre eine Möglichkeit, das auszuhebeln.

Tatsächlich ist es so, dass alle Studien, die bisher vorgelegt und herangezogen worden sind in der Bewertung des G-BAs für den Potenzialbescheid und für die Methoden, die darin zum Ausdruck kommen, dass dieser Eingriff weniger invasiv ist und deutlich kürzere Operationszeiten nach sich zieht und zumindest noninferior ist im Ergebnis zu dem, was die Fundoplicatio ergibt. Insofern muss ich Frau Leers widersprechen, weil das nicht den Stand, der sich im Antrag bzw. im Entscheid des G-BAs niedergeschlagen hat, widerspiegeln würde.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank, Herr Prof. Kersting, für die Ergänzungen. Sie wissen aber auch, dass wir keine Möglichkeiten haben, hier eine andere Vergütung zu machen, das sieht die gesetzliche Grundlage nicht vor, nur um hier keine falschen Erwartungen zu wecken. – Frau Leers.

Frau Prof. Dr. Leers (DGAV & DGVS): Eine Frage: Woher kommt die mittlere Verweildauer? Ich überblicke leider nicht alle Zahlen deutschlandweit, da bin ich von den Gesellschaften tatsächlich aber auch dabei, das aufzuarbeiten. Aber die Zahlen, die ich überblicke, ist die mittlere Verweildauer bei einer Fundoplicatio von 2,4 Tagen. Das sind die Zahlen aus dem Zweckverband Rheinland, der über 200 Kliniken beschreibt.

Herr Dr. Kersting (JNJ): Die Zahlen, die ich zitiert habe, sind die InEK-Jahreszahlen für 2022.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank. Aber unabhängig von diesen Zahlen hatten wir ja Frau Prof. Leers nach ihrer Einschätzung gefragt.

GKV-SV: Ich würde gern noch einmal nach den konkreten Vorteilen – jetzt unabhängig von der Krankenhausverweildauer – dieser verschiedenen Interventionen Sie beide befragen. Worin sehen Sie den konkreten Vorteil?

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank. – Frau Prof. Leers.

Frau Prof. Dr. Leers (DGAV & DGVS): Ich sehe noch keinen Vorteil in dieser Methode. Das ist eben mein kritischer Punkt. Aber ich begleite diese Methode schon seit sehr vielen Jahren, war auch bei der ersten Studiengebung in Amerika vor Ort und verschließe mich natürlich nicht neuen Techniken. Deswegen kann man das grundsätzlich – und muss man es auch – evaluieren, gerade in solch einem Feld. Aber bislang haben mich die Daten nicht überzeugt, dass sie gegen eine Fundoplicatio, die gut gemacht ist, wesentliche Vorteile hat.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank, Frau Prof. Leers. – Herr Dr. Andreae.

Herr Dr. Andreae (JNJ): Die Vorteile: In den Studien hat sich gezeigt, dass es bezüglich Schluckbeschwerden und Aufstoßen, Regurgitation nach der Behandlung wohl Vorteile für das LINX-System, für die MSA, im Vergleich zur Fundoplicatio gibt.

In den Symptomen nach dem GERD-HRQL-Score sind die Methoden eigentlich gleichwertig. Wie gesagt: Es gibt Hinweise, dass die Behandlungszeit und die Krankenhausverweildauer für die MSA-Methode geringer sind, wobei eine Studie, die große europäische Registerstudie mit vielen Patienten aus Deutschland, in der Publikation darauf hingewiesen hat, dass aufgrund des Vergütungssystems die Daten für Deutschland verzerrt sind und die Patienten unnötig lange im Krankenhaus verblieben sind, um den Abschlag in der Vergütung zu vermeiden. Nach amerikanischen Studien, wonach sich die Vergütung nicht nach der Verweildauer richtet, gibt es da ganz klare Differenzen. Von daher wäre es eigentlich interessant, das auch in Deutschland zu erarbeiten. Aber dass unser System das nicht hergibt, ist eigentlich schade.

Frau Dr. Lelgemann (Vorsitzende): Herzlichen Dank.

PatV: Sie haben gesagt, dass die Krankenhausverweildauer etwas ist, was man als Vorteil benennen könnte. Ist es denn so – auch vielleicht an Prof. Leers jetzt eher gerichtet –, dass man weniger unerwünschte Ereignisse hat, oder woraus ist das begründet? Das würde uns noch einmal interessieren, weil das vielleicht dann der richtige Aspekt ist, wo man die Vorteile eventuell sehen könnte. Das ist jetzt wahrscheinlich hypothetisch, aber deswegen machen wir ja eine Erprobung.

Frau Dr. Lelgemann (Vorsitzende): Vielen Dank. – Frau Leers.

Frau Prof. Dr. Leers (DGAV & DGVS): Grundsätzlich – auch für die Patientenvertretung zur Einschätzung – reden wir hier über einen Eingriff, der extrem sicher ist und wenig Nebenwirkungen hat. Insofern ist es schwierig, Vorteile herauszuarbeiten.

Der wesentlichste Unterschied ist die Mobilisation, also die Freilegung eines Teils des Magens, der bei der Fundoplicatio erforderlich ist. Das ist beim LINX-Band nicht erforderlich. Alle anderen Schritte sind gleich. Das heißt, man muss eine Laparoskopie durchführen. Man muss die Region freilegen, wie wir sagen, also präparieren, damit man das gut einsehen kann.

Mittlerweile ist auch die Empfehlung, den Zwerchfellbruch immer mit zu versorgen. Das heißt, diese Schritte sind bei beiden Methoden gleich und dann erfolgt der Weg in zwei Richtungen: einmal diese Freilegung eines Teils des Magens, um aus dem Magen die Manschette zu machen, die dann auf der anderen Seite durch das LINX-Band erfolgt. Man darf dabei nicht vergessen, dass es sich hierbei um einen Fremdkörper handelt, der dann implantiert wird, das sind die Unterschiede: die Freilegung des Magens auf der einen Seite und auf der anderen Seite aber ein Fremdkörper.

Frau Dr. Lelgemann (Vorsitzende): Herzlichen Dank, Frau Prof. Leers, für die klare Schilderung dessen, was zu erfolgen hat. Das ist für uns sicher sehr hilfreich. – Ich habe noch eine fachliche Frage, da Sie ja auch die Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten vertreten: Es ist sicherlich davon auszugehen, dass alle Patienten nach OP – welche Methode auch immer angewendet wird – wirklich beschwerdefrei sind. Es wird wahrscheinlich – aber das ist jetzt eine Wissensfrage – einen nicht unerheblichen Teil von Patientinnen und Patienten geben, die auch hinterher noch zum Beispiel auf Protonenpumpenhemmer angewiesen sind. Wäre daher auch zu erfassen, wie die Gabe von Protonenpumpenhemmern nach der OP ist, also auch im Rahmen unserer Studie? Wäre das sinnvoll?

Frau Prof. Dr. Leers (DGAV & DGVS): Das ist ein ganz schwieriges Thema. Die Einnahme von Protonenpumpenhemmern bedeutet nicht immer, dass beispielsweise ein Reflux-Rezidiv besteht, von dem wir ja ausgehen. Sondern wir haben auch aus vielen anderen Studien in diesem Bereich – mit dann OP-Technik offen, laparoskopisch, wie auch immer – gesehen, dass Patienten dazu neigen, bei Problemen jedweder Art ihre altbekannten Protonenpumpenhemmer zu nehmen. Das ist ja sowieso ein grundsätzliches Problem: dass die Protonenpumpenhemmer viel zu häufig eingenommen werden. Das heißt, dass die einfache Einnahme der Protonenpumpenhemmer sicherlich kein gutes Mittel ist, um das zu evaluieren, weil eben beispielsweise auch bei Verdauungsstörungen der ein oder andere Patient die Protonenpumpenhemmer nimmt.

Das ist vor allen Dingen die Aufgabe – da komme ich zu Ihrer ersten Frage – die wirklich gute Selektion der Patienten und eine gute standardisierte Voruntersuchung. Da bin ich ganz froh, dass wir in den Leitlinien das jetzt verankern konnten: die Standardisierung dieses Verfahrens – zumindest, was die präoperative Evaluation angeht. Dann haben wir schon ein sehr hohen Erfolgsanteil. Das ist eben das Entscheidende. Aber das gilt natürlich nicht für alle Patienten. Das ist unsere chirurgische Aufgabe und deswegen gehe ich mit der Stellungnahme von Herrn Kersting ein, dass man dafür erfahrene Chirurgen und erfahrene Zentren nehmen muss, die da bereits eine gute Indikation stellen. Also wir sind eine chirurgische Klinik und von den 100 % Patienten, die wir sehen, operieren wir vielleicht 30 %, die anderen 70 bleiben konservativ. Aber diese 30 % haben ein sehr hohes Ansprechen.

Frau Dr. Lelgemann (Vorsitzende): Herzlichen Dank für diese Zahlen, auch für die Beantwortung der Frage nach der medikamentösen Therapie – das kann man sich gut vorstellen –,insbesondere für die Einschätzung dessen, wie viele Patientinnen und Patienten überhaupt für diese OP infrage kommen.

Gibt es weitere Fragen aus dem Kreis des Unterausschusses? Gibt es weitere Anmerkungen von Ihnen? – Da das nicht der Fall ist, möchte ich mich ganz herzlich bedanken, insbesondere bei Ihnen, Frau Prof. Leers. Herzlichen Dank, dass Sie teilgenommen haben. Vielen Dank für die exzellenten Darstellungen und die Beantwortung unserer Fragen! Schöne Grüße nach München! Vielen Dank an Sie, Herr Dr. Kersting! Vielen Dank, Herr Dr. Andreae! Machen Sie es gut!

Schluss der Anhörung: 11:47 Uhr