

Anwendungsbegleitende Datenerhebung zu Valoctocogen Roxaparvovec: Prüfung des Studienprotokolls (Version 2.0) und des statistischen Analyseplans (Version 2.0)

3. Addendum zum Projekt A22-20

A horizontal bar composed of 18 rectangular segments of varying shades of blue and grey. The word 'ADDENDUM' is written in white capital letters on a dark blue segment that spans across the middle of the bar.

ADDENDUM

Projekt: A23-102

Version: 1.0

Stand: 30.11.2023

IQWiG-Berichte – Nr. 1683

DOI: 10.60584/A23-102

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Anwendungsbegleitende Datenerhebung zu Valoctocogen Roxaparvovec: Prüfung des Studienprotokolls (Version 2.0) und des statistischen Analyseplans (Version 2.0) –

3. Addendum zum Projekt A22-20

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

20.10.2023

Interne Projektnummer

A23-102

DOI-URL

<https://dx.doi.org/10.60584/A23-102>

Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Im Mediapark 8

50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

Zitiervorschlag

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Anwendungsbegleitende Datenerhebung zu Valoctocogen Roxaparvec: Prüfung des Studienprotokolls (Version 2.0) und des statistischen Analyseplans (Version 2.0); 3. Addendum zum Projekt A22-20 [online]. 2023 [Zugriff: TT.MM.JJJJ]. URL: <https://dx.doi.org/10.60584/A23-102>.

Schlagwörter

Valoctocogen Roxaparvec, Hämophilie A, Nutzenbewertung, Register, Forschungsdesign, Peer Review – Forschung

Keywords

Valoctocogene Roxaparvec, Hemophilia A, Benefit Assessment, Registries, Research Design, Peer Review – Research

An dem Addendum beteiligte Mitarbeiterinnen und Mitarbeiter des IQWiG

- Ulrike Seay
- Lars Beckmann
- Simone Johner
- Ulrike Lampert
- Claudia-Martina Messow
- Volker Vervölgyi

Inhaltsverzeichnis

| | Seite |
|---|--------------|
| Abkürzungsverzeichnis..... | vi |
| 1 Hintergrund..... | 1 |
| 2 Prüfung der Unterlagen zur Planung der AbD von Valoctocogen Roxaparvovec | 2 |
| 2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen | 2 |
| 2.2 Anmerkungen zum Studienprotokoll und SAP | 2 |
| 2.2.1 Fragestellung gemäß PICO: Patientenpopulation | 2 |
| 2.2.2 Fragestellung gemäß PICO: Outcome | 3 |
| 2.2.3 Fragestellung gemäß PICO: Outcome, Mortalität | 3 |
| 2.2.4 Fragestellung gemäß PICO: Outcome, Blutungen | 3 |
| 2.2.5 Fragestellung gemäß PICO: Outcome, Gelenkfunktion | 4 |
| 2.2.6 Fragestellung gemäß PICO: Outcome, patientenberichtete Endpunkte (PROs) und Gelenkfunktion | 5 |
| 2.2.7 Fragestellung gemäß PICO: Outcome, unerwünschte Ereignisse (UEs)..... | 6 |
| 2.2.8 Fragestellung gemäß PICO: Outcome, ergänzende Informationen | 7 |
| 2.2.9 Datenquelle / Studiendesign: Allgemein..... | 8 |
| 2.2.10 Datenquelle: Erhebung von Baselinedaten..... | 8 |
| 2.2.11 Datenquelle: Definitionen und Operationalisierung..... | 9 |
| 2.2.12 Datenquelle: Confounder | 10 |
| 2.2.13 Datenquelle: Meldezeitpunkte..... | 11 |
| 2.2.14 Datenquelle: Vollständigkeit der Daten | 11 |
| 2.2.15 Datenquelle: Source Data Verification | 12 |
| 2.2.16 Studiendesign: Estimand | 13 |
| 2.2.17 Studiendesign: Rekrutierung | 13 |
| 2.2.18 Studiendesign: Zuordnung zu den Behandlungsgruppen | 14 |
| 2.2.19 Studiendesign: Fallzahlplanung | 15 |
| 2.2.20 Studiendesign: Beendigung der Studienteilnahme..... | 15 |
| 2.2.21 Studiendesign: Abbruchkriterien | 16 |
| 2.2.22 Studiendesign: Zwischenanalysen..... | 17 |
| 2.2.23 Auswertung der Daten: verschobene Hypothesengrenze | 17 |
| 2.2.24 Auswertung der Daten: Auswertungspopulation | 19 |
| 2.2.25 Auswertung der Daten: Sensitivitätsanalysen | 19 |
| 2.2.26 Auswertung der Daten: Subgruppenanalysen | 20 |

3. Addendum zum Projekt A22-20

| | | |
|----------|--|-----------|
| 2.2.27 | Auswertung der Daten: Propensity-Score-Verfahren | 21 |
| 2.2.28 | Auswertung der Daten: binäre Endpunkte | 25 |
| 2.2.29 | Auswertung der Daten: Blutungen..... | 25 |
| 2.2.30 | Auswertung der Daten: patientenberichtete Endpunkte (PROs) | 26 |
| 2.2.31 | Auswertung der Daten: unerwünschte Ereignisse (UE)..... | 27 |
| 2.2.32 | Auswertung der Daten: Umgang mit fehlenden Werten..... | 28 |
| 3 | Literatur | 31 |

Abkürzungsverzeichnis

| Abkürzung | Bedeutung |
|------------------|--|
| AAV5 | Adeno-assoziiertes Virus Serotyp 5 (AAV5) |
| AbD | Anwendungsbegleitende Datenerhebung |
| ABR | annualisierte Blutungsrate |
| AIC | Akaike Information Criterion |
| ATE | Average Treatment Effect |
| BPI-SF | Brief Pain Inventory – short form |
| COP | Clinical Operations Plan |
| DHR | Deutsches Hämophileregister |
| FCS | Fully Conditional Specification |
| G-BA | Gemeinsamer Bundesausschuss |
| Haemo-QoL-A | Hemophila-specific Quality of Life Questionnaire for Adults |
| HJHS | Hemophilia Joint Health Score |
| IPTW | Inverse Probability of Treatment Weighting |
| IQWiG | Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen |
| ITT | Intention to treat |
| MI | Multiple Imputation |
| MMRM | Mixed Model repeated Measures (gemischtes Modell mit Messwiederholungen) |
| PICO | Patient, Intervention, Comparator, Outcome |
| PS | Propensity Score |
| PSM | PS-Matching |
| pU | pharmazeutischer Unternehmer |
| RR | Relatives Risiko |
| SAP | statistischer Analyseplan |
| SGB | Sozialgesetzbuch |
| SMD | standardisierte Mittelwertdifferenz |
| SMRW | Standardized Mortality Ratio Weighting |
| SP | Studienprotokoll |
| SUE | schwerwiegendes unerwünschtes Ereignis |
| UE | unerwünschtes Ereignis |

1 Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) am 04.07.2023 mit der Prüfung des Studienprotokolls (SP) und des statistischen Analyseplans (SAP) (jeweils Version vom 19. Oktober 2023) zur anwendungsbegleitenden Datenerhebung (AbD) zu Valoctocogen Roxaparvovec beauftragt.

In seiner Sitzung am 02.02.2023 hat der G-BA beschlossen, eine AbD und Auswertungen nach § 35a Absatz 3b Satz 1 SGB V für den Wirkstoff Valoctocogen Roxaparvovec in der Indikation „Behandlung von schwerer Hämophilie A (kongenitalem Faktor-VIII-Mangel) bei erwachsenen Patienten ohne Faktor-VIII-Inhibitoren in der Vorgeschichte und ohne nachweisbare Antikörper gegen Adeno-assoziiertes Virus Serotyp 5 (AAV5).“ zu fordern [1,2]. Dem Beschluss liegt unter anderem das vom IQWiG erstellte Konzept für eine AbD zu Valoctocogen Roxaparvovec zugrunde (AbD-Konzept A22-20 vom 29.09.2022) [3].

Zur Prüfung, ob die Anforderungen des G-BA an die AbD und an Auswertungen durch die vom pharmazeutischen Unternehmer (pU) erstellten Unterlagen zum SP und SAP umgesetzt worden sind, hat der G-BA dem IQWiG diese Unterlagen übermittelt [4,5] und mit der Prüfung dieser Unterlagen beauftragt. Neben dem G-BA-Beschluss zu Valoctocogen Roxaparvovec sollen die Inhalte der diesbezüglichen Beratung des pU zur Studienplanung der AbD berücksichtigt werden (Beratungsanforderung 2023-B-065 [6]).

Auf Basis der Ergebnisse dieser Prüfung (Addendum A23-64 zum Konzept A22-20 [7]) hat der G-BA mit Beschluss vom 21.09.2023 [8,9] formuliert, bei welchen Angaben im vorgelegten SP und SAP Anpassungsbedarf besteht. Der pU hat am 19.10.2023 überarbeitete Unterlagen beim G-BA eingereicht. Der G-BA hat dem IQWiG diese Unterlagen am 20.10.2023 übermittelt.

2 Prüfung der Unterlagen zur Planung der AbD von Valoctocogen Roxaparvovec

2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen

Der pU hat das SP und den SAP (jeweils in der Version vom 19.10.2023) auf Basis der Anforderungen des G-BA [8,9] überarbeitet. Die vorliegende Prüfung des SPs und SAP bezieht sich ausschließlich auf die vom G-BA benannten erforderlichen Punkte.

In den vom pU überarbeiteten Dokumenten fehlt eine systematische Übersicht der Anpassungen in SP und SAP entsprechend der vom G-BA geforderten Anpassungen ebenso wie eine generelle Übersicht zur Änderungshistorie.

Ein Großteil der vom pU durchgeführten Anpassungen steht unter dem Vorbehalt, dass diese im Deutschen Hämophileregister (DHR) umgesetzt werden. Der Gesamtdatensatz für das DHR wurde bisher nicht angepasst. Dies muss vor dem Start der AbD vollumfänglich sichergestellt sein.

Des Weiteren plant der pU vor Beginn der AbD eine Literatursuche nach potenziell relevanten Confoundern durchzuführen. Allerdings wurde Anhang 2 zum SP, in dem die Methodik der Identifikation und Festlegung von Confoundern beschrieben wird, bisher nicht überarbeitet. Die dort beschriebene Methodik ist, wie im Addendum A23-64 [7] erläutert, nicht sachgerecht.

Der pU legt mit der aktuellen Version von SP und SAP (vom 19.10.2023) einen vorläufigen Clinical Operations Plan (COP) vor, wohl um die Forderung des G-BA nach einem Study Monitoring Plan umzusetzen. Eine umfassende Einschätzung des COP auf Basis der vorliegenden Version ist nicht möglich, da er zu weiten Teilen nur aus den geplanten Zwischenüberschriften besteht.

Diese verbleibenden Mängel können nur in einer weiteren Revision der Studienunterlagen und nur mit sehr viel Aufwand behoben werden. Ein Start der AbD auf Basis des vorliegenden SP und SAP ist nicht sinnvoll.

2.2 Anmerkungen zum Studienprotokoll und SAP

2.2.1 Fragestellung gemäß PICO: Patientenpopulation

Auflage des G-BA

Die Ausschlusskriterien sind um das Vorliegen aktiver Infektionen (akut oder unkontrolliert chronisch) und das Vorliegen einer bekannten signifikanten Leberfibrose oder -zirrhose zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die im SP benannten Ausschlusskriterien sind um das Vorliegen akuter oder unkontrollierter chronischer hepatischer Infektionen und / oder einer bekannten signifikanten Leberfibrose oder -zirrhose ergänzt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Umsetzung der Auflage des G-BA ist unvollständig. Die Ausschlusskriterien sind um das Vorliegen jeglicher aktiver Infektionen (akut oder unkontrolliert chronisch) zu ergänzen. Darüber hinaus sind die Ausschlusskriterien in der im SAP aufgeführten Variablenliste der Ein- und Ausschlusskriterien zu ergänzen.

2.2.2 Fragestellung gemäß PICO: Outcome

Auflage des G-BA

Im SP ist festzulegen, dass alle Endpunkte in beiden Studienarmen ab dem Indexdatum erhoben werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Alle Endpunkte in beiden Studienarmen sollen ab dem Indexdatum erhoben und ausgewertet werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommene Änderung ist sachgerecht.

2.2.3 Fragestellung gemäß PICO: Outcome, Mortalität

Auflage des G-BA

Die Erhebung der Gesamtmortalität ist in Tabelle 8 des SPs zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In Tabelle 8 des SPs ist die Erhebung der Gesamtmortalität festgelegt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommene Änderung ist sachgerecht.

2.2.4 Fragestellung gemäß PICO: Outcome, Blutungen

Auflage des G-BA

Der Anlass einer Bedarfsbehandlung ist im Falle von schweren und lebensbedrohlichen Blutungen soweit möglich konkreter zu erheben, die Auswahl der Angaben „Verdacht auf Blutung“ und „unbekannter Grund“ sind für diesen Fall zu streichen.

Die Operationalisierung von schweren und lebensbedrohlichen Blutungen ist im SP eindeutig zu hinterlegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP sind bei den Angaben zur Operationalisierung der behandelten Blutungen jeglichen Schweregrades die Optionen „Verdacht auf Blutung“ und „unbekannter Grund“ gestrichen. Nur spontan auftretende behandelte Blutungen sollen erhoben werden.

Für die Definition schwerer und lebensbedrohlicher Blutungen gibt der pU an, die vom DHR genutzte Definition zu verwenden. Schwere Blutungen sind demnach als Blutungen definiert, die Schmerzen, Schwellungen und / oder Bewegungseinschränkungen verursachen und nicht innerhalb von 24 Stunden abklingen. Lebensbedrohliche Blutungen sind definiert als schwere Blutungen, die ein besonderes Risiko für den Patienten darstellen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Streichung der Optionen „Verdacht auf Blutung“ und „unbekannter Grund“ bei den Angaben zum Anlass einer Bedarfsbehandlung bei Blutungen jeglichen Schweregrades ist nicht sachgerecht. Nur im Falle von schweren und lebensbedrohlichen Blutungen sind diese Optionen zu streichen. Darüber hinaus sollte der Anlass für eine Bedarfsbehandlung im Falle von schweren und lebensbedrohlichen Blutungen soweit möglich konkret erhoben werden, bspw. durch Hinzunahme des Datenfeldes „traumatische Blutungen“.

Die Operationalisierung der schweren und lebensbedrohlichen Blutungen erscheint sachgerecht, allerdings ist eine abschließende Prüfung nicht möglich, da die vom DHR genutzten Definitionen weder aus dem Gesamtdatensatz noch aus dem Handbuch des DHR [10,11] hervorgehen.

2.2.5 Fragestellung gemäß PICO: Outcome, Gelenkfunktion

Auflage des G-BA

Die Angaben zur Erhebung der Gelenkfunktion sind im SP zu vereinheitlichen. Eine Erhebung 1-mal jährlich wird als ausreichend erachtet.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP und im SAP ist einheitlich die 1-mal jährliche Erhebung der Gelenkfunktion angegeben.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind sachgerecht.

2.2.6 Fragestellung gemäß PICO: Outcome, patientenberichtete Endpunkte (PROs) und Gelenkfunktion

Auflage des G-BA

Im SP sind geeignete Toleranzfenster für die Erhebung der patientenberichteten Endpunkte (PROs) und der Gelenkfunktion, welche nicht aneinandergrenzen, festzulegen. Es sind darüber hinaus geeignete Maßnahmen im SP zur Vermeidung fehlender Werte zu beschreiben.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU plant weiterhin die Erhebungen der PROs zu Baseline und zu den Monaten 6, 12, 18, 24, 30 und 36 mit einem Toleranzfenster von ± 3 Monaten für jeden Erhebungszeitpunkt nach Baseline. Die Gelenkfunktion soll 1-mal jährlich erhoben werden, ebenfalls mit einem Toleranzfenster von ± 3 Monaten für jeden Erhebungszeitpunkt nach Baseline.

Der pU geht davon aus, dass die Source Data Verification dazu beiträgt, den Anteil fehlender Werte zu reduzieren. Aufgrund des nicht interventionellen Designs der AbD seien fehlende oder nicht plausible Werte nicht zu vermeiden.

Des Weiteren soll bei den Site Monitoring Visits auf das Problem fehlender oder unvollständiger Dateneingabe und Query Status eingegangen werden. Dateneingabefelder zu demografischen Informationen, Vorgeschichte, Therapieadministration und klinischen Endpunkten sollen für Studienteilnehmende im DHR verpflichtend gemacht werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Zum einen grenzen die Toleranzfenster für die Erhebung der PROs weiterhin aneinander, zum anderen sind sie mit einem Zeitraum von ± 3 Monaten zu breit. Wie schon in der 1. Protokollprüfung erläutert [7], ist dieses Vorgehen nicht sachgerecht. Da die Gelenkfunktion nur 1-mal jährlich erhoben werden soll, besteht das Problem der aneinandergrenzenden Toleranzfenster für diesen Endpunkt nicht.

Die Source Data Verification ist kein geeignetes Instrument zur Vermeidung fehlender Werte, sie dient in erster Linie der Sicherstellung der Richtigkeit der Registerdaten. Die beschriebenen Site Monitoring Visits und die Einführung von verpflichtenden Feldern in der Datenbank werden als geeignet und zur Erfüllung der Auflage des G-BA als sachgerecht erachtet. Allerdings bleibt unklar ob die Site Monitoring Visits in allen oder nur in ausgewählten Studienzentren durchgeführt werden sollen (siehe auch Abschnitt 2.2.14). Es ist in den Studienunterlagen klar zu stellen, dass die Site Monitoring Visits in allen Studienzentren durchgeführt werden.

2.2.7 Fragestellung gemäß PICO: Outcome, unerwünschte Ereignisse (UEs)

Auflage des G-BA

Die Erhebung der schwerwiegenden UEs (SUEs) ist bezogen auf alle Ereignisse, die zur Hospitalisierung oder zum Tod führen (Gesamtrate), festzulegen und nicht auf diejenigen Ereignisse einzuschränken, die im Zusammenhang mit der Behandlung der Hämophilie stehen.

Für spezifische UEs ist im SP zusätzlich zur Gesamtrate auch die Erfassung der Ereignisse, welche zur Hospitalisierung oder zum Tod führen, festzulegen.

Für das spezifische UE „thromboembolische Ereignisse“ ist die Erfassung aller Ereignisse unabhängig von einem möglichen Zusammenhang mit der Behandlung der Hämophilie im SP festzulegen.

Die Definition des spezifischen UE „schwere Leberschäden“ ist im SAP und SP in einer Form zu hinterlegen, welche die Erfassung von symptomatischen Leberschäden sicherstellt.

Für das spezifische UE „maligne Neubildungen“ ist im SP zu spezifizieren, anhand welcher Maßnahmen bzw. Definitionen eine einheitliche Dokumentation dieser Ereignisse sichergestellt wird.

Es ist festzulegen, dass die Endpunkte zu den Nebenwirkungen in beiden Studienarmen bis zum Studienende erhoben werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP ist ergänzt, dass in die Auswertung der SUEs aller Ereignisse, die zur Hospitalisierung oder zum Tod führen (Gesamtrate) eingehen. Der Begriff unerwünschte Ereignisse (adverse events) wurde für diesen Endpunkt durch medizinische Ereignisse (medical events) ersetzt.

Die Forderung, dass für spezifische UE im SP zusätzlich zur Gesamtrate auch Ereignisse, welche zur Hospitalisierung oder zum Tod führen erfasst werden sollen, wird im SP nicht adressiert.

Zur Erhebung des spezifischen UE thromboembolische Ereignisse ist geplant, sowohl solche Ereignisse zu erfassen, die nur im Zusammenhang mit der Behandlung der Hämophilie stehen, als auch Ereignisse unabhängig von der Ursache. Die thromboembolischen Ereignisse unabhängig von der Ursache sollen aber laut SP und SAP nicht ausgewertet bzw. verglichen werden. Lediglich die im Anhang des SAP präsentierten Tabellentemplates sehen vergleichende Analysen auch für thromboembolische Ereignisse jeglicher Ursache vor.

Zur Operationalisierung des spezifischen UEs schwere Leberschäden zitiert der pU verschiedene Quellen und geht davon aus, dass die Erfassung der schweren Leberschäden im DHR konsistent zu den Definitionen in diesen Quellen sein werde.

Für das spezifische UE maligne Neubildungen erwartet der pU, dass das DHR dahin gehend angepasst wird, dass zusätzliche Ereignisse, wie maligne Neubildungen erfasst werden können. Es wird davon ausgegangen, dass diese konsistent z. B. zu den Vorgaben des international World Federation of Hemophilia Gene Therapy Registry erhoben werden.

Im SP sowie im SAP ist festgelegt, dass die Nebenwirkungen bis zum Studienende erhoben werden sollen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die gemäß Beschluss des G-BA festgelegten erforderlichen Anpassungen im SP sind nur zum Teil umgesetzt. Es ist unklar, mit welchem Ziel der Begriff UEs (adverse events) durch den Begriff medizinische Ereignisse (medical events) ersetzt wurde. Die Ersetzung des Begriffs ist rückgängig zu machen. Die Ergänzung, dass in die Auswertung der SUEs alle Ereignisse, die zur Hospitalisierung oder zum Tod führen (Gesamtrate) eingehen ist sachgerecht. Für die spezifischen UEs muss dies gemäß den Vorgaben des G-BA ebenfalls ergänzt werden. Auch ist gemäß den Vorgaben des G-BA im SP und SAP zu ergänzen, dass thromboembolische Ereignisse unabhängig von einem möglichen Zusammenhang mit der Behandlung der Hämophilie nicht nur erfasst, sondern auch vergleichend ausgewertet werden. Das Vorgehen des pU, für die Definition spezifischer UE auf externe Quellen zu verweisen ist nicht ausreichend. Für das spezifische UE schwere Leberschäden bleibt unklar, wie die Erfassung ausschließlich symptomatischer Leberschäden sichergestellt wird, eine spezifische Definition dieser Ereignisse ist in das SP aufzunehmen. Auch für das spezifische UE maligne Neubildungen ist eine Definition im SP zu hinterlegen. Darüber hinaus sollten Maßnahmen beschrieben werden, wie sichergestellt wird, dass die Erhebung in den Studienzentren einheitlich erfolgt.

Die Erhebung der Nebenwirkungen bis zum Studienende ist sachgerecht (siehe auch Abschnitt 2.2.18).

2.2.8 Fragestellung gemäß PICO: Outcome, ergänzende Informationen

Auflage des G-BA

Für den Endpunkt „Zeitpunkt der Wiederaufnahme einer prophylaktischen Behandlung“ ist zu definieren, dass jegliche Wiederaufnahme einer prophylaktischen Behandlung dokumentiert wird.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP ist ergänzt, dass jegliche Wiederaufnahme einer prophylaktischen Behandlung dokumentiert werden soll.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommene Änderung ist sachgerecht.

2.2.9 Datenquelle / Studiendesign: Allgemein

Auflage des G-BA

Alle relevanten zu erhebenden Daten und zugehörige Auswertungen müssen vor Beginn der AbD a priori festgelegt und im SP und SAP beschrieben werden. Aus den Studienunterlagen ist zu streichen, dass Definitionen, Operationalisierungen oder Auswertungen erst anhand der beobachteten Daten oder im Verlauf der AbD festgelegt werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP ist ergänzt, dass erwartet wird, dass zum einen alle erforderlichen Anpassungen im DHR vor dem Start der AbD erfolgen werden (geplante Implementierung vor Januar 2024), und zum anderen, dass alle für die AbD erforderlichen Variablen Pflichtfelder im DHR werden. Eventuell im Studienverlauf auftretende Änderungen der a priori Definitionen oder hinzukommende Variablen sollen im SAP, in den Zwischenanalysen und im finalen Bericht beschrieben werden. In den Studienunterlagen wurde gestrichen, dass Definitionen, Operationalisierungen oder Auswertungen erst anhand der beobachteten Daten oder im Verlauf der AbD festgelegt werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das vom pU geplante Vorgehen erscheint zunächst sachgerecht. Allerdings fehlen an einigen Stellen noch klare Entscheidungskriterien bezüglich der geplanten Vorgehensweise. So bleibt im SAP beispielsweise offen, unter welchen Voraussetzungen im Propensity-Score(PS)-Verfahren stabilisierte Gewichte verwendet werden und unter welchen Voraussetzungen die Gewichte gestutzt werden (siehe Abschnitt 2.2.27), wann eine Verteilung als schief betrachtet wird, wann Schätzer als instabil betrachtet werden und unter welchen Voraussetzungen Subgruppenanalysen als machbar angesehen werden (siehe Abschnitt 2.2.26.) Auch die systematische Identifikation und Festlegung von Confoundern fehlt bislang (siehe auch Abschnitt 2.2.12). Die Einschätzung der Umsetzung der erforderlichen Anpassungen auf Basis der vorliegenden Unterlagen ist nicht möglich, sondern kann erst bei Vorlage des finalen SP und des angepassten Meldebogens für das DHR erfolgen.

2.2.10 Datenquelle: Erhebung von Baselinedaten

Auflage des G-BA

Es ist im SP festzulegen, dass alle Ein- und Ausschlusskriterien mit eindeutiger Operationalisierung zum Indexdatum erhoben werden. Für Patienten, die im Verlauf der Studie auf eine Behandlung mit Valoctocogen Roxaparvovec wechseln und entsprechend der untenstehenden Ausführungen zur Zuordnung der Behandlungsgruppen dem Valoctocogen

Roxaparvovec-Arm zugeordnet werden, ist festzulegen, dass die Baselinedaten zum Indexdatum (Datum der Applikation von Valoctocogen Roxaparvovec) erneut erhoben werden.

Im SP ist zu streichen, dass zur Erhebung der Baselinedaten nur auf Patienten zugegriffen wird, für die bereits seit mindestens 12 Monaten Daten im Register dokumentiert wurden.

Es ist im SP festzulegen, dass das Alter der Patienten und der AAV5 -Status in der Datenquelle eindeutig erfasst werden. Zudem ist sicherzustellen, dass der AAV5-Status, sofern erhoben, für die Patienten beider Studienarme erfasst wird. Dies ist im SP zu spezifizieren.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP und im SAP ist unverändert festgehalten, dass die Beschreibung der Patientencharakteristika entweder auf Basis von Daten die vor dem Indexdatum in das DHR überführt wurden oder zum Indexdatum erhoben werden erfolgen soll. Eine eindeutige Festlegung, die Patientencharakteristika zum Indexdatum zu erheben, findet sich nicht. Auch für Patienten, die im Verlauf der Studie auf eine Behandlung mit Valoctocogen Roxaparvovec wechseln, findet sich keine Festlegung, die Patientencharakteristika zum Indexdatum (Datum der Applikation von Valoctocogen Roxaparvovec) erneut zu erheben. Im SP und im SAP ist unverändert festgehalten, dass zur Erhebung der Patientencharakteristika nur auf Patienten zugegriffen wird, für die bereits seit mindestens 12 Monaten Daten im Register dokumentiert wurden. Der pU gibt an, dass vom DHR zurzeit geprüft werde, ob die Extraktion des Geburtsmonats der Patienten aus dem DHR möglich ist. Unter der Voraussetzung, dass die entsprechenden Anpassungen im DHR erfolgen, soll eine Erfassung für spezifische AAV-Vektoren in beiden Studienarmen möglich sein.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU bezgl. der Erhebung der Patientencharakteristika ist nicht sachgerecht. Der pU adressiert die Auflagen des G-BA weder hinsichtlich der Erhebung der Patientencharakteristika zum Indexdatum bzw. der ggf. erneuten Erhebung nach Therapiewechsel auf Valoctocogen Roxaparvovec noch in Bezug darauf, dass zu streichen ist, dass für Patienten bereits seit 12 Monaten Daten im DHR vorliegen müssen. Sowohl SP als auch SAP müssen dahin gehend angepasst werden. Es ist festzulegen, dass das Alter und der AAV5-Status der Patienten in beiden Studienarmen eindeutig erfasst werden.

2.2.11 Datenquelle: Definitionen und Operationalisierung

Auflage des G-BA

Im SP und SAP sind alle zu erhebenden Daten (u. a. Expositionen, klinische Ereignisse, Endpunkte, Confounder) mit ihrer entsprechenden Operationalisierung a priori festzulegen. Für die Operationalisierung sind soweit möglich Standardklassifikationen zu nutzen. Die

Erhebung der erforderlichen Daten muss vor Beginn der AbD in der gewählten Datenquelle sichergestellt sein.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU gibt an, dass alle für die Analyse heranzuziehenden Variablen a priori im SP und im SAP definiert werden. Für Komorbiditäten ist weiterhin nur Freitexteingabe vorgesehen. Die Operationalisierung der Blutungen wurde ergänzt (siehe 2.2.4). Zur Operationalisierung der UEs siehe 2.2.7 und zu Confoundern siehe 2.2.12. Die erforderlichen Anpassungen der Datenfelder im DHR sollen vor dem Start der AbD erfolgen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Es ist sachgerecht, dass der pU plant, alle Variablen a priori zu definieren. Allerdings setzt der pU sein geplantes Vorgehen nicht durchgängig um. Bspw. ist für die Komorbiditäten weiterhin nur eine Freitexteingabe geplant, obwohl der pU selbst anmerkt, dass dadurch eine sinnvolle Auswertung der Daten ggf. nicht möglich ist. Zur Einschätzung der Operationalisierungen der Blutungen, der UEs und der Confounder siehe Abschnitte 2.2.4, 2.2.7 und 2.2.12. Eine Einschätzung der Anpassungen im DHR ist zurzeit nicht möglich, da diese noch nicht umgesetzt wurden.

2.2.12 Datenquelle: Confounder

Auflage des G-BA

Es ist vor Beginn der AbD eine systematische Literaturrecherche nach potentiell relevanten Confoundern durchzuführen und unter Einbindung von Experten zu ergänzen.

Der Abschnitt zur Identifikation und Festlegung von Confoundern im SP ist grundlegend zu überarbeiten. Es muss sichergestellt sein, dass alle a priori identifizierten relevanten Confounder in der gewählten Datenquelle sachgerecht ab dem Beginn der AbD erhoben werden. Sollte eine Erhebung bestimmter Parameter nicht möglich sein, ist diese Unsicherheit im SP zu adressieren und die Berücksichtigung bei der Ergebnisinterpretation zu beschreiben.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU plant, vor dem Start der AbD eine Literaturrecherche zur Identifikation potenzieller Confounder durchzuführen. Falls die identifizierten Confounder nach Rücksprache mit Klinikern als für die Bildung des PS notwendig erachtet werden, plant der pU im Austausch mit dem DHR zu klären, inwieweit die Erhebung der Confounder im DHR machbar ist. Anhang 2 zum SP Version 1.0, in dem das Vorgehen zur Identifikation und Festlegung relevanter Confounder beschrieben ist, wurde nicht überarbeitet.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU ist nicht sachgerecht. Zum einen ist eine systematische Literaturrecherche zur Identifikation potenzieller Confounder erforderlich. Zum anderen ist zu gewährleisten, dass alle relevanten Confounder ab Studienbeginn im DHR erhoben werden. Sollte eine Erhebung bestimmter Parameter nicht möglich sein, ist diese Unsicherheit im SP zu adressieren und die Berücksichtigung bei der Ergebnisinterpretation zu beschreiben. Die Methodik zur Identifikation und Festlegung von Confoundern ist grundlegend zu überarbeiten.

2.2.13 Datenquelle: Meldezeitpunkte

Auflage des G-BA

Die Meldezeitpunkte in der gewählten Datenquelle müssen sicherstellen, dass die Daten der AbD für eine fristgerechte Einreichung der im Beschluss des G-BA vom 2. Februar 2023 [1] festgelegten Zwischenanalysen und des Dossiers für die erneute Nutzenbewertung zur Verfügung stehen. Diese Vorgabe muss in den Studienunterlagen hinterlegt werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die Studienzentren werden aufgefordert, die Daten für alle an der AbD teilnehmenden Patienten etwa 6 Monate vor der jeweiligen Zwischenanalyse zu aktualisieren.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Es liegt im Verantwortungsbereich des pU sicherzustellen, dass die Daten fristgerecht eingereicht werden. Dazu sind konkrete Maßnahmen zu beschreiben, bspw. die Entsendung von Monitoren in alle Studienzentren.

2.2.14 Datenquelle: Vollständigkeit der Daten

Auflage des G-BA

Es ist im SP zu beschreiben, welche Maßnahmen zur Schulung der behandelnden Ärztinnen und Ärzte für eine sachgerechte Datenerhebung ergriffen werden.

Zudem sind im SP Maßnahmen zu adressieren, die zur Sicherstellung der Vollständigkeit der Daten für jeden Patienten durchgeführt werden.

Im Rahmen der gewählten Datenquelle ist sicherzustellen, dass die relevanten Daten für die anwendungsgebegleitende Datenerhebung bei der Eingabe der Daten in die Datenquelle nicht nur optionale, sondern verpflichtende Datenfelder darstellen. Im SP ist entsprechend festzulegen, dass alle relevanten Datenfelder für die Durchführung der AbD obligat zu erheben sind.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Mit den überarbeiteten Studienunterlagen legt der pU einen vorläufigen Entwurf eines Clinical Operations Plan (COP) pU vor. Nur ausgewählte Studienzentren sollen gemäß des COP überwacht werden. Bei den weiteren beschriebenen Maßnahmen zur geplanten Schulung der behandelnden Ärzte und des beteiligten Personals durch vom pU qualifizierte Monitore bleibt jeweils unklar, ob sie ebenfalls nur in ausgewählten Studienzentren oder in allen teilnehmenden Studienzentren durchgeführt werden sollen. Die 1. Schulung soll vor dem Start der AbD stattfinden. Die Vollständigkeit der Daten im Verlauf der AbD soll durch 2-mal jährliche Besuche der Monitore in den Studienzentren sowie durch regelmäßigen Kontakt der Monitore zum Studienpersonal sichergestellt werden. Der pU geht davon aus, dass vor Beginn der AbD die Erhebung aller relevanten Datenfelder im DHR obligat sein wird.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die im SP beschriebenen Maßnahmen zur Schulung der Ärzte und des Studienpersonals sowie zur Sicherstellung der Vollständigkeit der Daten sind wegen der beschriebenen Unklarheiten nicht sachgerecht. Die im COP zu beschreibenden Maßnahmen müssen in allen Studienzentren umgesetzt werden. Die konkrete inhaltliche Ausgestaltung der Maßnahmen ist zum gegenwärtigen Zeitpunkt nicht bewertbar, da erst ein vorläufiger Entwurf des COP vorliegt. Die weiteren beschriebenen Maßnahmen zur Schulung der Ärzte sind ebenfalls in allen Studienzentren umzusetzen. Dies ist im SP klar zu stellen. Die obligate Erhebung aller Datenfelder im DHR muss vor Beginn der AbD umgesetzt sein.

2.2.15 Datenquelle: Source Data Verification

Auflage des G-BA

Im SP ist der technische Prozess der Source Data Verification eindeutig und nachvollziehbar zu beschreiben. Der Study Monitoring Plan ist an das SP anzuhängen oder separat für die erneute Überprüfung der Studienunterlagen einzureichen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die Source Data Verification soll für 100 % der Datenfelder für die Ein- und Ausschlusskriterien sowie den primären Endpunkt durchgeführt werden. Für weitere 10 % zufällig ausgewählte der an der AbD teilnehmenden Patienten soll eine Source Data Verification für alle Endpunkte erfolgen. In Abschnitt 11.8 des SP wird beschrieben, dass der Abgleich der in das DHR gemachten Einträge anhand der Quelldateien, wie z. B. Patientenakten, Laborberichten, Krankenhausberichten für SUEs erfolgen soll. Für weitere Details verweist der pU auf den vorläufigen COP. In Abschnitt 10.5 steht unverändert, dass die in das DHR eingegebenen Daten als Quelldatei dienen sollen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU geplante Source Data Verification anhand der Quelldateien wie z. B. der Patientenakte ist sachgerecht. Dies sollte im SP vereinheitlicht werden. Voraussetzung ist, dass die Source Data Verification an jedem der an der AbD teilnehmenden Zentren durchgeführt wird (siehe auch Abschnitt 2.2.14). Weitere Details zur Source Data Verification werden im separaten COP beschrieben.

2.2.16 Studiendesign: Estimand

Auflage des G-BA

Für die AbD ist als primärer Estimand die Treatment Policy Strategy zu hinterlegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Ein Estimand nach Treatment Policy Strategy wird nun beschrieben, allerdings nicht als primärer Estimand, sondern gleichberechtigt mit dem Hypothetical Estimand.

Es ist geplant, Patienten aus dem Vergleichsarm, die nach dem Rekrutierungszeitraum eine Behandlung mit Valoctocogen Roxaparvovec erhalten, von der Auswertung auszuschließen

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Gemäß der Auflage des G-BA ist der Treatment Policy Estimand als primärer Estimand festzulegen. Dies wurde nicht umgesetzt.

Die Umsetzung der Treatment Policy Strategy ist nicht sachgerecht, da geplant ist, dass Patienten aus dem Vergleichsarm, die nach dem Rekrutierungszeitraum eine Behandlung mit Valoctocogen Roxaparvovec erhalten, nicht in die Auswertungen eingehen. Entsprechend der Ausführungen in Abschnitt 2.2.18 sind diese Patienten in Abhängigkeit der Beobachtungszeit unter der Vergleichstherapie den Studienarmen zuzuordnen.

Darüber hinaus fehlt im SP bislang eine sachgerechte Erfassung des Indexdatums und der damit verbundenen Erhebung der Baselinedaten, insbesondere auch für Patienten, die im Verlauf der Studie auf eine Behandlung mit Valoctocogen Roxaparvovec wechseln (siehe Abschnitt 2.2.10).

2.2.17 Studiendesign: Rekrutierung

Auflage des G-BA

Im SP ist zu beschreiben, welche Maßnahmen ergriffen werden, um die Überführung eines höheren Anteils der Patienten aus der Sammelmeldung in die Einzelmeldung herbeizuführen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In den Kapiteln 6 und 7 des SPs ist ergänzt, dass Patienten mit der Zustimmung zur Teilnahme an der AbD auch der Einzelmeldung an das DHR zustimmen müssen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Der Hinweis, dass Patienten mit der Zustimmung zur Teilnahme an der AbD auch der Einzelmeldung an das DHR zustimmen müssen, findet sich bereits in Abschnitt 5.2 von Version 1.0 des SPs. Der pU beschreibt keine weiteren konkreten Maßnahmen um die Anzahl der Patienten in Einzelmeldung zu erhöhen.

2.2.18 Studiendesign: Zuordnung zu den Behandlungsgruppen

Auflage des G-BA

Im SP sind Angaben zu ergänzen, wie mit den Patienten im Vergleichsarm bei einem Wechsel auf die Behandlung mit Valoctocogen Roxaparvovec umgegangen wird. Diesbezüglich ist eine angemessene Beobachtungszeit zu definieren, ab welcher die Patienten dem Vergleichsarm oder dem Valoctocogen Roxaparvovec-Arm zugeordnet werden. Hierbei sind auch Patienten zu berücksichtigen, die kurz nach dem angesetzten Rekrutierungszeitraum auf eine Behandlung mit Valoctocogen Roxaparvovec wechseln.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In den Studienunterlagen ist ergänzt, dass Patienten, die nach dem Ende der Rekrutierungsphase vom Vergleichsarm auf die Behandlung mit Valoctocogen Roxaparvovec wechseln, nicht in die Analysen eingehen. Patienten im Valoctocogen Roxaparvovec-Arm, die wieder eine Prophylaxe mit Faktor-VIII-Präparaten erhalten, werden im Valoctocogen Roxaparvovec-Arm ausgewertet.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU bezüglich der Patienten, die vom Vergleichsarm auf eine Behandlung mit Valoctocogen Roxaparvovec wechseln, ist nicht sachgerecht. Gemäß den Vorgaben des G-BA ist in der vorliegenden Situation eine Strategie im Sinne einer Intention-to-treat(ITT)-Auswertung zu verfolgen, wobei Patienten, die im Verlauf der Beobachtung von der Vergleichstherapie auf Valoctocogen Roxaparvovec wechseln, in Abhängigkeit der Beobachtungszeit unter der Vergleichstherapie den Studienarmen zugeordnet werden sollen. Diejenigen Patienten, die bereits eine vom pU zu definierende, angemessen lange Zeit unter der Behandlung mit Faktor-VIII-Präparaten beobachtet wurden, sodass bereits aussagekräftige Daten für die Vergleichsgruppe vorliegen (z. B. 2 Jahre bei einer geplanten Beobachtungszeit von 3 Jahren) und dann erst auf Valoctocogen Roxaparvovec wechseln, sind im Vergleichsarm auszuwerten und bis zum Studienende weiter zu beobachten. Patienten, die nach einer kürzeren vom pU zu definierender Zeit auf eine Behandlung mit Valoctocogen

Roxaparovec wechseln und für die noch eine angemessene Beobachtungszeit unter der Genterapie zu erwarten ist, sind im Interventionsarm auszuwerten. Für diese Patienten stellt der Zeitpunkt des Wechsels den Beobachtungsstart für die AbD dar, wobei eine neue Erhebung der Baselinedaten erforderlich ist. Die Beobachtungszeit unter Vergleichstherapie dieser Patienten ist für die AbD nicht zu berücksichtigen.

Das geplante Vorgehen für Patienten im Valoctocogen Roxaparovec-Arm, die wieder eine Prophylaxe mit Faktor-VIII-Präparaten erhalten, ist sachgerecht.

2.2.19 Studiendesign: Fallzahlplanung

Auflage des G-BA

Im SP sind die Angaben zur Methodik der Fallzahlschätzung entsprechend den Ausführungen im SAP zu korrigieren.

Darüber hinaus sind die Maßnahmen zu beschreiben, die ergriffen werden, um eine ausreichend hohe Anzahl an Patienten unter einer Behandlung mit Faktor-VIII-Präparaten in die AbD zu rekrutieren.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU hat die Angaben zur Fallzahlplanung im SP den Angaben im SAP angeglichen. Der pU beschreibt weiterhin keine Maßnahmen, um eine ausreichend hohe Anzahl an Patienten unter einer Behandlung mit Faktor-VIII-Präparaten in die AbD zu rekrutieren.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Anpassung von SP und SAP bezüglich der Fallzahlplanung ist sachgerecht. Der pU hat weiterhin Maßnahmen zu beschreiben, um eine ausreichend hohe Anzahl an Patienten unter einer Behandlung mit Faktor-VIII-Präparaten in die AbD zu rekrutieren.

2.2.20 Studiendesign: Beendigung der Studienteilnahme

Auflage des G-BA

Die Beendigung der Studienteilnahme alleine aufgrund der klinischen Einschätzung der behandelnden Ärztin bzw. des behandelnden Arztes ist aus den Studienunterlagen zu streichen. Es sind konkrete Kriterien zu definieren, auf Basis welcher die Patienten aus der Studie ausscheiden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SP wurde gestrichen, dass die Studienteilnahme alleine aufgrund der klinischen Einschätzung der behandelnden Ärztin bzw. des behandelnden Arztes beendet werden kann. Als Kriterien für die Beendigung der Studienteilnahmen werden im SP unverändert Tod, Lost to Follow-up, Einnahme eines Prüfartzneimittels sowie Einschlusskriterien nicht erfüllt

genannt, mit dem Zusatz, dass diese Aufzählung der Gründe nicht abschließend sei. Im SAP wird als weiterer Grund für die Beendigung der Studienteilnahme genannt, dass Patienten aus dem Vergleichsarm nach Studienbeginn eine Behandlung mit Valoctocogen Roxaparovec erhalten.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die im SP genannten Gründe für die Beendigung der Studienteilnahme sind nur zum Teil sachgerecht. Zwar wurde gestrichen, dass die Studienteilnahme aufgrund der Einschätzung der behandelnden Ärztin bzw. des behandelnden Arztes beendet werden kann. Der Hinweis, dass es außer den aufgelisteten Gründen noch weitere geben könne, ist nicht sachgerecht. Die Liste der Gründe zur Beendigung der Studienteilnahme muss vor Studienbeginn vollständig festgelegt sein. Zum sachgerechten Umgang mit Patienten aus dem Vergleichsarm, die nach Studienbeginn eine Behandlung mit Valoctocogen Roxaparovec erhalten, siehe Abschnitt 2.2.18.

2.2.21 Studiendesign: Abbruchkriterien

Auflage des G-BA

Im SP und SAP sind Abbruchkriterien aufgrund von Vergeblichkeit zu ergänzen. Änderungen an AbD müssen in Abstimmung mit dem G-BA erfolgen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Für die Zwischenauswertung zum Zeitpunkt 18 Monate hat der pU sowohl im SP als auch im SAP Abbruchkriterien für Vergeblichkeit basierend auf der Anzahl rekrutierter Patientinnen und Patienten ergänzt. Des Weiteren soll die Anzahl der für die Auswertung verfügbaren Variablen berücksichtigt werden. Ebenfalls ergänzt wurde, dass Änderungen an der AbD in Abstimmung mit dem G-BA erfolgen sollen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Abbruchkriterien für Vergeblichkeit basierend auf der Anzahl rekrutierter Patientinnen und Patienten für die Zwischenauswertung zum Zeitpunkt 18 Monate sind sachgerecht. Unklar bleibt hingegen, was mit der Anzahl der für die Auswertung verfügbaren Variablen gemeint ist. Hier fehlt eine Präzisierung, wann eine Variable als für die Auswertung verfügbar gilt. Es fehlt auch ein Abbruchkriterium in Zusammenhang mit der Anzahl verfügbarer Variablen.

Die Ergänzung, dass Änderungen an der AbD in Abstimmung mit dem G-BA erfolgen sollen, ist sachgerecht.

2.2.22 Studiendesign: Zwischenanalysen

Auflage des G-BA

Die geplante Zwischenanalyse 6 Monate nach Beginn der anwendungsbegleitenden Datenerhebung ist zu streichen. Die Zwischenanalysen sind entsprechend der im Beschluss vom 2. Februar 2023 hinterlegten Zeitpunkte 18 Monate, 36 Monate und 54 Monate nach dem Beginn der anwendungsbegleitenden Datenerhebung durchzuführen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die geplante Zwischenanalyse nach 6 Monaten ist im SP gestrichen. Im SAP wurde darüber hinaus auch der nach 6 Monaten geplante Statusbericht der AbD (Angaben zum Verlauf der Datenerhebung) gestrichen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Es ist sachgerecht, dass die geplante Zwischenanalyse nach 6 Monaten gestrichen wurde. Der Statusbericht an den G-BA 6 Monate nach Beginn der AbD ist weiterhin erforderlich.

2.2.23 Auswertung der Daten: verschobene Hypothesengrenze

Auflage des G-BA

Im SP und SAP ist unter Berücksichtigung des nicht randomisierten Studiendesigns festzulegen, dass für die Auswertung und Ergebnisinterpretation der Daten eine verschobene Hypothesengrenze von 0,2 bis 0,5 in Abhängigkeit der Qualität der Datenerhebung und -auswertung herangezogen wird.

Zudem ist im SP und SAP ein Abschnitt zu ergänzen, welcher die Ergebnisinterpretation der Daten unter Berücksichtigung des nicht randomisierten Studiendesigns und unter Verwendung einer entsprechenden verschobenen Hypothesengrenze (im Bereich zwischen 0,2 bis 0,5) thematisiert.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt in Abschnitt 7.6 des SAP die Fallzahlberechnung für den Endpunkt annualisierte Blutungsrate (ABR) mit dem relativen Risiko (RR) als Effektmaß. Er setzt dabei aufgrund des nicht randomisierten Designs eine verschobene Hypothesengrenze von 0,5 an. Der pU argumentiert weiter, dass er keine geringere Hypothesengrenze anwenden kann (z. B. 0,2), da nach Aussagen von Vertretern der Gesellschaft für Thrombose und Hämostaseforschung e. V. und der Deutschen Gesellschaft für Hämatologie und Onkologie e. V. die zur Verfügung stehende Population geringer sei als 700, wovon nur 50 % in Einzelmeldung seien. Des Weiteren führt er aus, dass zusätzlich zur verschobenen Hypothesengrenze eine PS-Methodik angewendet werde, um die Unsicherheit des nicht randomisierten Vergleiches zu berücksichtigen.

In Abschnitt 7.9 des SAP beschreibt der pU die statistischen Verfahren und die Darstellung für die deskriptiven Analysen der einzelnen Therapiearme in Abhängigkeit der Daten. Zum Vergleich von Valoctocogen Roxaparvovec mit der Kontrolle werden die statistischen Verfahren beschrieben. Die Darstellung der Ergebnisse umfasst nach Abschnitt 7.9 und Tabelle 6 im SAP Regressionskoeffizienten mit 95 %-Konfidenzintervallen und p-Werte, wobei p-Werte $< 0,05$ als „statistisch signifikant“ bezeichnet werden. Zudem beschreibt der pU, dass die Analysen auf beobachteten Daten und nicht auf „projected“ Daten beruhen sollen. Alle p-Werte zu sekundären Endpunkten seien nominal.

Ein entsprechender Abschnitt zur Ergebnisinterpretation unter Berücksichtigung des nicht randomisierten Studiendesigns fehlt sowohl im Studienprotokoll als auch im SAP.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Hierzu ist anzumerken, dass sich der als Ergebnis der AbD anzunehmende Effekt aus dem wahren Unterschied zwischen Valoctocogen Roxaparvovec und der Vergleichstherapie sowie der durch das nicht randomisierte Studiendesign bedingten Verzerrung zusammensetzt. Wie im Konzept zur AbD [3] beschrieben und in den Beratungen verdeutlicht, kann aus einer nicht randomisierten Studie aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten daher erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Die konkrete Schwelle ergibt sich durch die Qualität der Daten, u. a. durch das Wissen über und die Berücksichtigung der relevanten Confounder, beispielsweise durch die Anwendung einer PS-Methodik. Der pU begründet die Grenze von 0,5 jedoch ausschließlich über die geschätzte zu erwartende Patientenzahl ohne Berücksichtigung einer potenziellen Verzerrung, der Qualität der Daten und der PS-Methodik. Davon unabhängig liegt es letztlich im Verantwortungsbereich des pU, eine ausreichend große Fallzahl zu rekrutieren, um einen bestehenden wahren Unterschied zwischen Valoctocogen Roxaparvovec und der Vergleichstherapie durch die AbD mit ausreichender Sicherheit nachzuweisen. Insofern ist die Interpretation der Ergebnisse nicht durch die im SP verwendete Hypothesengrenze für die Fallzahlberechnung des pU präjudiziert (hier 0,5), sondern ergibt sich letztlich aus der abschließenden Qualität der Datenerhebung und -auswertung (in der Regel durch eine verschobene Hypothesengrenze im Bereich 0,2 bis 0,5 [12]).

Die in Abschnitt 7.9 des SAP beschriebene Darstellung der Ergebnisse für den Vergleich von Valoctocogen Roxaparvovec zur Kontrolle ist nicht ausreichend. Für die AbD ist es nicht relevant, ob ein p-Wert für eine Effektschätzung eine vorgegebene Irrtumswahrscheinlichkeit unterschreitet, maßgeblich ist die obere 95 %-Konfidenzintervallgrenze des RR bezüglich des Rate Ratios, bzw. deren Inverses, im Vergleich zu einer verschobenen Nullhypothese. Dementsprechend sind Ergebnisse darzustellen. Da die Analysen sowohl eine PS-Methodik beinhalten als auch die Multiple Imputation zur Berücksichtigung fehlender Werte, sollte der

pU klarstellen, was darunter zu verstehen ist, wenn Analysen auf beobachteten Daten beruhen.

Ein entsprechender Abschnitt zur Ergebnisinterpretation unter Berücksichtigung des nicht randomisierten Studiendesigns ist nachzureichen (siehe auch Abschnitt 2.2.27).

2.2.24 Auswertung der Daten: Auswertungspopulation

Auflage des G-BA

Im SAP ist zu streichen, dass in die Auswertung nur Patienten einbezogen werden, zu denen vollständige Daten zu allen Variablen vorliegen, die in das PS-Verfahren einfließen sollen. Es sind Maßnahmen zu etablieren, um die Vollständigkeit der Daten sicherzustellen sowie klare Kriterien festzulegen, wie mit fehlenden Werten in der Analyse umgegangen wird.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In Abschnitt 7.8.1 des SAP hat der pU bei der Beschreibung der Schätzung der PS keine Änderung vorgenommen. In Abschnitt 7.9.5.1 des SAP beschreibt der pU zum Umgang mit fehlenden Werten ein Verfahren der Multiplen Imputation (siehe Abschnitt 2.2.32). Zur Sicherstellung der Vollständigkeit der Daten beschreibt der pU, dass in den teilnehmenden Behandlungszentren Schulungen der behandelnden Ärzte und des beteiligten Personals durch vom pU qualifizierte Monitore stattfinden sollen (siehe 2.2.14).

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die fehlende Umsetzung bei der Beschreibung der Schätzung der PS ist nicht sachgerecht. Ebenfalls ist die Beschreibung des Ersetzungsverfahrens nicht ausreichend (siehe Abschnitt 2.2.32). Der pU sollte sicherstellen, dass bei der Schätzung der PSs ein geeignetes Verfahren zur Berücksichtigung fehlender Werte adäquat angewendet wird, damit möglichst viele Patienten in die Analysen eingehen. Das entsprechende Verfahren ist zu spezifizieren. Zu Maßnahmen zur Sicherstellung der Vollständigkeit der Daten siehe ebenso Abschnitt 2.2.32.

Die im SP beschriebenen Maßnahmen zur Schulung der Ärzte und des Studienpersonals sowie zur Sicherstellung der Vollständigkeit der Daten sind wegen der beschriebenen Unklarheiten nicht sachgerecht (siehe 2.2.14).

2.2.25 Auswertung der Daten: Sensitivitätsanalysen

Auflage des G-BA

Für die Endpunkte der Kategorie Mortalität, Morbidität und gesundheitsbezogene Lebensqualität sind Sensitivitätsanalysen festzulegen, in denen die Patienten, die im Vergleichsarm auf eine Behandlung mit Valoctocogen Roxaparvovec gewechselt haben und weiterhin dem Vergleichsarm zugeordnet werden, zum Zeitpunkt des Wechsels zensiert werden.

Zudem sind Sensitivitätsanalysen anhand von Verfahren festzulegen, die angewendet werden können, wenn nicht in beiden Behandlungsgruppen zum Beobachtungsstart mit einer neuen Therapie begonnen wird (z. B. Prevalent new User Design).

Weiterhin sind Sensitivitätsanalysen für die getrennte Auswertung der Daten zu Valoctocogen Roxaparovec gegenüber den Daten zu Faktor-XIII-Präparaten und zu Emicizumab zu präspezifizieren.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt keine Sensitivitätsanalyse, bei der Patienten, die im Vergleichsarm auf eine Behandlung mit Valoctocogen Roxaparovec gewechselt haben und weiterhin dem Vergleichsarm zugeordnet werden, zum Zeitpunkt des Wechsels zensiert werden. Des Weiteren sind keine Sensitivitätsanalysen anhand von Verfahren festgelegt worden, die angewendet werden können, wenn nicht in beiden Behandlungsgruppen zum Beobachtungsstart mit einer neuen Therapie begonnen wird (z. B. Prevalent new User Design).

Der pU beschreibt im SAP Analysen zu den verschiedenen Therapien des Vergleichsarms, die er teilweise als Subgruppenanalyse und teilweise als Sensitivitätsanalyse bezeichnet.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die fehlenden Sensitivitätsanalysen sind vom pU zu ergänzen.

Für die Sensitivitätsanalyse zu den verschiedenen Therapien des Vergleichsarms ist zu präzisieren, dass es sich hierbei um die getrennte Auswertung der Daten zu Valoctocogen Roxaparovec gegenüber den Daten zu Faktor-XIII-Präparaten und zu Emicizumab handelt (siehe auch Abschnitt 2.2.18).

2.2.26 Auswertung der Daten: Subgruppenanalysen

Auflage des G-BA

Im SAP sind die statistischen Tests für Subgruppenanalysen zu beschreiben. Entsprechend der Modulvorlagen für die Nutzenbewertung ist die Durchführungen weiterer Subgruppenanalysen unter anderem zur Krankheitsschwere und dem Alter zu prüfen.

Inkonsistente Angaben bezüglich der Betrachtung des Merkmals AAV5-Status zwischen SP und SAP sind zu vereinheitlichen (Subgruppenanalysen vs. Sensitivitätsanalysen).

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt in Abschnitt 7.9.5 des SAP verschiedene Sensitivitäts- und Subgruppenanalysen. Es ist konsistent in SP und SAP festgehalten, dass für den AAV5-Status

Subgruppenanalysen geplant sind. Darüber hinaus plant der pU Subgruppenanalysen für die Merkmale Alter und Wirkstoffklasse der Faktor-XIII-Prophylaxebehandlungen durchzuführen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU beschriebenen Anpassungen der statistischen Methoden sind nicht sachgerecht. Der Likelihood-Ratio-Test ist prinzipiell geeignet. Allerdings ist bei der Analyse unklar, ob die verschiedenen Faktor-XIII-Prophylaxebehandlungen gegeneinander verglichen werden sollen, oder ob die Effekte der Prophylaxebehandlungen gegenüber Valoctocogen Roxaparvovec miteinander verglichen werden sollen. Dies ist zu präzisieren. Bei den geplanten Subgruppenanalysen ist das Merkmal Krankheitsschwere zu ergänzen. Die Anpassungen bezgl. der Merkmale AAV5-Status und Alter sind sachgerecht, allerdings sollte der Likelihood-Ratio-Test im Abschnitt zum AAV5-Status ergänzt werden.

Der pU beschreibt nicht, unter welchen Bedingungen Subgruppenanalysen durchgeführt werden sollen. Dieses ist zu ergänzen.

2.2.27 Auswertung der Daten: Propensity-Score-Verfahren

Auflage des G-BA

Im SAP ist zu hinterlegen, dass alle zuvor identifizierten relevanten Confounder im Regressionsmodell zur Schätzung des PS berücksichtigt werden. Es ist eine angemessene alternative Strategie für den Fall zu hinterlegen, dass das Regressionsmodell nicht konvergiert. Es ist festzulegen, dass eine Diskussion der Ergebnisse insbesondere auch unter Berücksichtigung der Balanciertheit der nicht im Modell berücksichtigten Variablen erfolgt.

Die Hauptanalyse zur Confounderadjustierung anhand des Standardized Mortality Ratio Weighting (SMRW-Gewichtung) ist zu streichen und eine geeignete Analysemethode, welche sich auf den Average Treatment Effect (ATE) bezieht (z. B. Inverse Probability of Treatment Weighting [IPTW]) zu hinterlegen. Der Umgang mit extremen Gewichten ist im SAP eindeutig zu spezifizieren.

Es ist festzulegen, wann bei den vorgesehenen Untersuchungen zur Überlappung (visuelle Untersuchung der Dichtekurve und Histogramme der PS, c-Statistik) von einer ausreichenden Überlappung ausgegangen wird.

Für den Umgang mit Imbalancen ist die zugrundeliegende Methodik nachvollziehbar und transparent darzulegen. Die Gründe für die Eignung der gewählten Methodik sind im SAP zu beschreiben, sowie die Kriterien für die Modellauswahl für die finale Analyse darzulegen.

Im SAP ist eine klare Hierarchie der infrage kommenden PS-Verfahren und der Prüfkriterien zur Auswahl der robustesten Methode festzulegen. Hierbei sind konkrete Angaben zur Überprüfung einer ausreichenden Überlappung und Balanciertheit zu machen. Es ist zu

ergänzen, welche Konsequenzen sich ergeben, sofern kein PS-Verfahren identifiziert wird, mit dem eine ausreichende Überlappung und Balanciertheit erreicht wird.

Im SAP ist klarzustellen, dass die Schätzung des PS die Wahrscheinlichkeit darstellt, mit der ein Patient eine Intervention erhält, und diese unabhängig vom betrachteten Endpunkt erfolgt.

Es ist im SAP festzulegen, dass Sensitivitätsanalysen mit verschiedenen PS-Verfahren für alle relevanten Endpunkte und nicht nur für den primären Endpunkt durchgeführt werden.

Ausführungen zur Erforderlichkeit einer detaillierten Beschreibung der Patientenpopulation, welche sich aus der Anwendung des jeweiligen PS-Verfahrens ergibt, einschließlich der Erforderlichkeit eines Abgleichs dieser Patientenpopulation mit der ursprünglichen Zielpopulation der anwendungsbegleitenden Datenerhebung sind zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im SAP ist beschrieben, dass alle relevanten Confounder im PS-Modell berücksichtigt werden sollen.

Als primäres Verfahren wählt der pU die Schätzung des ATE mittels IPTW, PS-Matching (PSM) wird als Sensitivitätsanalyse dargestellt. PSM kann zum primären Verfahren werden, dazu beschreibt der pU einen Algorithmus unter Berücksichtigung von extremen Gewichten und der Überlappung.

Für den Umgang mit extremen Gewichten und der daraus möglicherweise resultierenden Verzerrung im IPTW-Verfahren beschreibt der pU stabilisierte Gewichte. Zusätzlich sollen Gewichte, die größer als 5 sind, auf 5 gesetzt werden (Truncation). Des Weiteren beschreibt der pU ein Trimming von nichtüberlappenden Regionen von PSs, sofern extreme stabilisierte Gewichte weiterhin existieren. Wenn ein Trimming für das IPTW-Verfahren erforderlich ist, betrachtet der pU ein PSM-Verfahren, sofern dieses ein 1:1 oder 1:2 – Matching erlaubt. In diesem Fall soll das PSM als primäre Analyse verwendet werden. Bei der Beschreibung vom PSM-Verfahren beschreibt der pU sein Vorgehen bez. des Verteilungsverhältnisses. Er definiert ein hierarchisches Vorgehen, wobei er bei einer 1:4 Verteilung startet. Dieses wird für die Analysen herangezogen, wenn mindestens 90 % der Patienten unter der Intervention gematcht werden können. Gelingt dies nicht, so wird mit dem gleichen Kriterium ein 1:3 Verhältnis geprüft. Nachfolgend werden die Verteilungen 1:2 und 1:1 geprüft. Eine Einschränkung auf den primären Endpunkt für PSM findet sich im SAP nicht.

Der pU beschreibt Untersuchungen zur Überlappung (visuelle Untersuchung der Dichtekurve und Histogramme der PS, c-Statistik). Kriterien für eine ausreichende Überlappungen werden nicht angegeben.

Zur Bewertung der Balanciertheit soll die standardisierte Mittelwertdifferenz (SMD) für die einzelnen Confounder herangezogen werden. Dabei wird von einer ausreichenden Balanciertheit ausgegangen, wenn bei allen Confoundern ein Unterschied von höchstens 0,25 SMD festgestellt wird. Wird für einen Confounder trotz Modellanpassungen keine Balanciertheit erreicht, so plant der pU, diesen Confounder nicht zur Schätzung der PSs heranzuziehen. Stattdessen soll der Confounder in der entsprechenden Regressionsmethode als Adjustierungsvariable berücksichtigt werden.

Falls IPTW und PSM keine ausreichenden Überlappungen und Modellgüte erreichen, plant der pU eine multivariable Regressionsanalyse mit Adjustierung mit Confoundern. Die Ergebnisse dieser Analysen seien mit dem G-BA zu diskutieren.

Im SAP wurde gestrichen, die Schätzung der PS in Abhängigkeit von den betrachteten Endpunkten zu diskutieren.

Die Forderung des G-BA, Ausführungen zur Erforderlichkeit einer detaillierten Beschreibung der Patientenpopulation, welche sich aus der Anwendung des jeweiligen PS-Verfahrens ergibt, einschließlich der Erforderlichkeit eines Abgleichs dieser Patientenpopulation mit der ursprünglichen Zielpopulation der anwendungsbegleitenden Datenerhebung zu ergänzen, wird vom pU nicht adressiert.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Berücksichtigung aller relevanten Confounder bei den PS-Verfahren ist sachgerecht.

Die vom pU vorgeschlagene Grenze von 0,25 SMD zur Bewertung der Balanciertheit ist zwar sachgerecht, stellt allerdings das schwächste Kriterium für eine ausreichende Balanciertheit dar. Des Weiteren ist die Berücksichtigung von Confoundern, für die keine Balanciertheit erreicht werden kann, als Adjustierungsvariablen sachgerecht. Beide Punkte sind jedoch bei der Interpretation der Ergebnisse auf Basis einer verschobenen Hypothesengrenze(siehe Abschnitt 2.2.23) entsprechend zu berücksichtigen [12].

Die Beschreibung des PS-Verfahrens ist nicht sachgerecht. Es bleibt unklar, ob der pU stabilisierte Gewichte benutzen möchte oder ob die Benutzung optional ist. Der pU hat die Verwendung der stabilisierten Gewichte entsprechend zu präzisieren. Als Kriterium für die Truncation wählt der pU ohne weitere Begründung den festen Wert 5. In der vom pU zitierten Literatur [13,14] sind jedoch zweiseitige, relative Kriterien, z. B. anhand von Perzentilen, diskutiert. Der pU sollte darlegen, warum die Wahl für die Truncation in der vorliegenden Situation angemessen ist. Die Trimming ist nicht konkret spezifiziert und es liegt keine Definition von extremen stabilisierten Gewichten vor. Zur Bewertung der Überlappung findet sich lediglich die Angabe, dass die Histogramme und Dichtekurven der PS visuell untersucht

sowie die c-Statistik berichtet werden soll. Insgesamt fehlen weiterhin Kriterien, wann die Überlappung als ausreichend angesehen wird.

Der Algorithmus zur Wahl des PS-Verfahrens ist nicht sachgerecht. Es kann nicht ausgeschlossen werden, dass qua Konstruktion aufgrund der oben beschriebenen Mängel die PSM als primäres Verfahren verwendet wird. Damit handelt es sich nicht um ein hierarchisches Vorgehen. Des Weiteren ist nicht verständlich, warum bei der Auswahl des Verfahrens schon ein 1:1 oder 1:2 Verteilungsverhältnis zwischen Intervention und Kontrolle genügt, um PSM als primäres Verfahren gegenüber dem IPTW zu bevorzugen. Der pU benennt ein Verhältnis von 1:4 als erstrebenswert und geht nur auf kleinere Verhältnisse, wenn weniger als 90 % der Interventionsgruppe berücksichtigt wird.

Es ist sachgerecht, dass das PSM nicht auf den primären Endpunkt eingeschränkt ist.

Die grafischen Darstellungen der PSs sind sachgerecht. Allerdings gibt der pU keine Kriterien für eine ausreichende Überlappung an. Diese Kriterien sind zu spezifizieren.

Für den Fall, dass IPTW und PSM keine ausreichenden Überlappungen und Modellgüte erreichen, ist die geplante multivariable Regressionsanalyse mit Adjustierung mit Confoundern nicht sachgerecht, da hierbei mit Extrapolationen gearbeitet wird, die zu nicht interpretierbaren Ergebnissen führen [15,16]. Ein naiver Vergleich ist vom pU nicht spezifiziert. Dazu ist folgendes anzumerken:

Aus einer nicht randomisierten Studie kann aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Eine (positive oder negative) Aussage zum Nutzen oder Schaden ergibt sich dann, wenn das Konfidenzintervall für den beobachteten Effekt ober- bzw. unterhalb einer zu definierenden Schwelle liegt (Test auf verschobene Nullhypothese). Die konkrete Schwelle ergibt sich durch die Qualität der Daten im Einzelfall, u. a. durch das Wissen über wichtige Confounder [12]. In der vorliegenden AbD wird dies beispielsweise durch die Wahl des schwächsten möglichen Kriteriums für die Balanciertheit der Daten dahin gehend beeinflusst, dass dies potenziell zu einer höheren Schwelle bei der Ergebnisinterpretation führt.

Ein entsprechender Abschnitt zur Ergebnisinterpretation unter Berücksichtigung des nicht randomisierten Studiendesigns fehlt sowohl im Studienprotokoll als auch im SAP und sollte ergänzt werden. Unabhängig davon wird auf Basis der üblichen Präsentation der Ergebnisse in Form von Effektschätzungen mit Konfidenzintervall in der späteren Nutzenbewertung bei der Interpretation der Ergebnisse dieser Ansatz zur Anwendung kommen

Der pU soll Ausführungen zur Erforderlichkeit einer detaillierten Beschreibung der Patientenpopulation, welche sich aus der Anwendung des jeweiligen PS-Verfahrens ergeben,

einschließlich der Erforderlichkeit eines Abgleichs dieser Patientenpopulation mit der ursprünglichen Zielpopulation der anwendungsbegleitenden Datenerhebung, ergänzen.

2.2.28 Auswertung der Daten: binäre Endpunkte

Auflage des G-BA

Für binäre Endpunkte mit vergleichender Beobachtungsdauer ist das relative Risiko als Effektschätzer festzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU hat für alle binären Endpunkte als Effektmaß das relative Risiko aus einem gewichteten log-binomial-Modell festgelegt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Umsetzung ist sachgerecht. Allerdings ist die Auflage nicht überall umgesetzt, bspw. nicht in dem einführenden Absatz zu Abschnitt 7.9 des SAP sowie in der Überschrift für Table 8 des SAP. Die Auflage des G-BA ist konsistent in SP und SAP umzusetzen.

2.2.29 Auswertung der Daten: Blutungen

Auflage des G-BA

Die Angaben zum negativ-binomial-Modell für die Auswertung der Blutungen sind im SAP unter Berücksichtigung aller Ereignisse, die zwischen Indexdatum und Studienende beobachtet werden, sowie der Spezifizierung eines Offsets zu vervollständigen. Zudem ist die Anwendung des Zero-inflated negativ-binomial-Modells im SAP eindeutig zu spezifizieren.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU spezifiziert im SAP, dass alle während des Follow-up-Zeitraumes beobachteten Blutungsereignisse in den entsprechenden Analysen berücksichtigt werden.

Als Offset für das Negativ-binomial-Modell spezifiziert der pU die log-transformierte individuelle Follow-up-Zeit.

Des Weiteren spezifiziert der pU die Anwendung des Zero-inflated negativ-binomial-Modell. Dieses wird angewendet, wenn das Negativ-binomial-Modell nicht konvergiert oder wenn der Vergleich zwischen dem Negativ-binomial-Modell und dem Zero-inflated negativ-binomial-Modell anhand des Akaike-Informations-Kriteriums (AIC) eine bessere Modellanpassung mit dem Zero-inflated-Modell aufzeigt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind sachgerecht.

2.2.30 Auswertung der Daten: patientenberichtete Endpunkte (PROs)

Auflage des G-BA

Die für die PROs vorgesehenen generalisierte gemischte Modelle für Messwiederholungen sind im SAP detailliert zu beschreiben. Als Responsekriterium ist ein Wert von mindestens 15 % der Skalenspannweite zu definieren. Zudem ist im SAP die Auswertung in Form von Verlaufskurven für die Interpretation der Responderanalysen festzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In Abschnitt 7.9.4.4 des SAP beschreibt der pU die geplanten Auswertungen des Instruments Hemophilia Joint Health Score (HJHS) für die Erfassung Gelenkfunktion sowie der PRO-Instrumente Hemophila-specific Quality of Life Questionnaire for Adults (Haemo-QoL-A) und Brief Pain Inventory – short form (BPI-SF).

Der pU beschreibt eine Analyse der Differenz zu Studienbeginn mittels eines generalisierten gemischten Modells für Messwiederholungen (MMRM) mit unstrukturierter Kovarianzmatrix. Es sollen alle Patienten mit 1 Wert zu Studienbeginn und mindestens 1 Follow-Up-Wert berücksichtigt werden. Die Analyse soll auf dem letzten beobachteten Follow-up-Wert basieren. Der pU plant, sowohl ein ungewichtetes als auch ein mittels IPTW gewichtetes MMRM für das Effektmaß SMD durchzuführen. Diese Analysen sind für alle 3 Instrumente vorgesehen. Die Ersetzung fehlender Werte ist nicht vorgesehen.

Des Weiteren beschreibt der pU Responderanalysen für das Instrument HJHS. Der pU definiert ein Ereignis als eine Verbesserung um einen präspezifizierten Schwellenwert für „one of the main outcomes at a certain time point“. Für den Schwellenwert verweist der pU auf die vom IQWiG bevorzugte Schwelle von 15 % der Skalenbreite des jeweiligen Instruments. Einen konkreten Wert gibt der pU für das Instrument HJHS nicht an.

Die Veränderung der Responderanteile über die Zeit soll für alle 3 Instrumente grafisch dargestellt werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU beschriebene Auswertung mittels MMRM ist nicht sachgerecht. Durch die Verwendung des letzten beobachteten Wertes sind die Ergebnisse potenziell hoch verzerrt und nicht interpretierbar, da sich aufgrund unterschiedlicher Beobachtungszeiten die Zeitpunkte zwischen den Patienten unterscheiden werden. Angemessen wäre entweder eine Analyse zu einer Differenz zu einem festen Zeitpunkt, wie beispielsweise die Differenz in der Änderung von Studienbeginn zu Monat 36, oder eine Analyse zur mittleren Differenz in der Änderung im Vergleich zu Studienbeginn über den gesamten Studienzeitraum.

Es ist nicht plausibel, dass der pU für die Auswertung eine ungewichtete und eine mittels IPTW gewichtete Methode durchzuführen plant. Die Analysen für die 3 Instrumente sollte nach dem PS-Verfahren (bzw. Multiple Regressionsanalyse) durchgeführt werden, das durch das hierarchische Auswahlverfahren hervorgegangen ist (siehe Abschnitt 2.2.27). Des Weiteren ist für die SMD keine Operationalisierung angegeben. Es bleibt unklar, ob die SMD aus den beobachteten Werten oder aus den Schätzungen des MMRM ergeben soll. Es ist korrekt, dass in einem MMRM keine Ersetzung fehlender Werte für die Endpunkte durchgeführt werden muss. Unter der Annahme Missing at Random führt dies zu unverzerrten Schätzern. Dies gilt allerdings nicht für fehlende Werte für Confounder. Das in Abschnitt 7.9.5.1 des SAP vorgesehene und noch zu präzisierende Verfahren der multiplen Imputation ist an dieser Stelle zu ergänzen.

Die für den HJHS vorgesehenen Responderanalysen sind nicht sachgerecht. Die Definition eines Ereignisses ist unklar, da nicht ausreichend definiert, auf welche Skalen (Einzelitems und / oder Gesamtscore) sich die Definition bezieht. Eine Verbesserung / Verschlechterung definiert als 1 Ereignis in mindestens 1 der Gelenke ist nicht sachgerecht.

Die Darstellung der Responderanteile in Verlaufskurven ist sachgerecht. Eine Darstellung der mittleren beobachteten Werte als Verlaufskurven sollte vom pU ergänzt werden.

2.2.31 Auswertung der Daten: unerwünschte Ereignisse (UE)

Auflage des G-BA

Für die UE ist eine Auswertung unabhängig von der Anzahl der aufgetretenen Ereignisse festzulegen. Das statistische Verfahren für die Auswertung der UE ist im SAP vollständig zu beschreiben. Für die UE ist eine Analyse festzulegen, in der die Patienten, die im Vergleichsarm auf eine Behandlung mit Valocotocogen Roxaparvovec gewechselt haben und weiterhin dem Vergleichsarm zugeordnet werden, zum Zeitpunkt des Wechsels nicht zensiert werden. Zudem ist für diese Patienten eine weitere Analyse mit Zensierung zum Zeitpunkt des Wechsels festzulegen. Für Patienten, die bei einer erneuten Faktor-VIII-Therapie weiterhin dem Interventionsarm zugeordnet werden, soll für die Analyse keine Zensierung erfolgen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt die Auswertung der UEs sowohl für den Treatment Policy Estimand als auch für Hypothetical Estimand. Im Folgenden wird – im Sinne einer Auswertung nach dem ITT-Prinzip – nur der Treatment Policy Estimand berücksichtigt.

Der pU beschreibt, dass Auswertungen zu den Nebenwirkungen je nach Endpunkt mit einem Negativ-binomial-Modell für das Effektmaß Rate Ratio oder einem log-binomial-Modell für das Effektmaß relatives Risiko durchgeführt werden. Als einzige Methode zur Confounderadjustierung wird IPTW genannt, wobei Confounder, deren Balanciertheit nicht

erreicht wurde, ins Modell aufgenommen werden sollen. Für die Ergebnisdarstellung sollen die Punktschätzer, 95 %-Konfidenzintervalle sowie p-Werte zu allen Modellvariablen dargestellt werden. Im Falle von weniger als 10 Ereignissen plant der pU als Annäherung eine logistische Regression mit Firth-Korrektur.

Der pU macht keine Angaben zur Zensierung von Patienten, die die Behandlung wechseln, vorgesehen ist die Berücksichtigung aller Ereignisse während des Follow-ups (siehe auch Abschnitt 2.2.18).

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die statistischen Auswertungen sind nicht präzise genug beschrieben. Der pU erwähnt ausschließlich IPTW zur Confounderadjustierung und nicht das beschriebene hierarchische Verfahren. Des Weiteren bleibt die multiple Imputation zur Ersetzung fehlender Werte in Confoundern unerwähnt. Zum Umgang mit Patienten, die die Behandlung wechseln, macht der pU keine Angaben. Eine Sensitivitätsanalyse bez. der Patienten, die vom Kontrollarm auf den Interventionsarm wechseln, ist nicht beschrieben (siehe auch Abschnitt 2.2.25). Die Sensitivitätsanalyse ist zu ergänzen.

Das methodische Vorgehen bei wenigen Ereignissen ist sachgerecht.

Die Berücksichtigung aller Ereignisse, die im Follow-up beobachtet werden ist sachgerecht.

2.2.32 Auswertung der Daten: Umgang mit fehlenden Werten

Auflage des G-BA

Die Angaben zum Umgang mit fehlenden Werten sind in den Studienunterlagen klarzustellen und zu vereinheitlichen. Es sind sinnvolle Ersetzungsstrategien für fehlende Werte darzulegen und die entsprechende Methodik zu präzisieren.

Die geplante Ersetzung des Monats führt potentiell zu erheblichen Verzerrungen und ist nicht sachgerecht. Diese Festlegung ist daher zu streichen. Vom pU ist stattdessen zu ergänzen, welche Anstrengungen unternommen werden, um die Rate fehlender Werte bei der Datumsangabe zu minimieren.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der Umgang des pU mit fehlenden Daten ist in Abschnitt 7.9.5.1 des SAP beschrieben. Der pU führt an, dass er die Source Data Verification benutzen möchte, um den Anteil fehlender Werte oder nicht vollständig erhobener Variablen zu reduzieren. Des Weiteren soll bei den Site Monitoring Visits auf das Problem fehlender oder unvollständiger Dateneingabe und Query Status eingegangen werden. Dateneingabefelder zu demografischen Informationen, Vorgeschichte, Therapieadministration und klinischen Endpunkten sollen für Studienteilnehmende im DHR verpflichtend gemacht werden.

Der pU beschreibt die deskriptive Darstellung der fehlenden Werte. Weiterhin beschreibt der pU, dass eine Ersetzung fehlender Werte in Abhängigkeit der Muster der fehlenden Daten und der Auswirkung der fehlenden Daten auf die Interpretierbarkeit der Ergebnisse erwogen wird. Der pU konkretisiert hierbei lediglich, dass Variablen mit mehr als 40 % fehlender Werte nicht im Modell für die PS-Schätzung aufgenommen werden sollen. Für Confounder mit weniger als 40 % fehlender Werte beschreibt der pU die Methode der multiplen Imputation (MI) mittels der Methode Fully Conditional Specification (FCS) / Chained Equations (MICE) zur Ersetzung fehlender Werte.

Zu Datumsangaben macht der pU die Angabe, bei fehlenden Angaben zum Monat des Geburtsdatums, diese zu ersetzen. Ein konkretes Vorgehen wird nicht beschrieben. Weitere fehlende Datumsfelder sollen nicht imputiert werden, zum Umgang macht der pU keine Angaben.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Source Data Verification ist kein geeignetes Instrument zur Vermeidung fehlender Werte, sie dient in erster Linie der Sicherstellung der Richtigkeit der Registerdaten. Die beschriebenen Site Monitoring Visits und die Einführung von verpflichtenden Feldern in der Datenbank werden als geeignet und zur Erfüllung der Auflage des G-BA als sachgerecht erachtet.

Die beschriebene Methode der MI mittels FCS / MICE ist prinzipiell geeignet. Jedoch ergibt sich aus der Darstellung nicht, wie die MI konkret mit der Schätzung der PS und der nachfolgenden Effektschätzung für die Endpunkte zusammengebracht werden soll. Dies betrifft die Einschätzung der Balanciertheit und der Überlappung sowie die Modellauswahl zum PS-Verfahren (IPTW oder Matching). Des Weiteren wird die Anzahl an Imputationen nicht angegeben. Das genaue Vorgehen ist vom pU zu spezifizieren. In der Literatur sind verschiedene Ansätze für die Kombination von PS-Methoden mit MI beschrieben [17-20].

Fehlende Confounderdaten sollen laut pU mittels MI ersetzt werden, sofern die Menge der fehlenden Daten höchstens 40 % beträgt. Für den Fall, dass mehr Daten fehlen, soll der entsprechende Confounder aus der Adjustierung herausgenommen werden. Einen relevanten Confounder mit mehr als 40 % fehlenden Daten bei der Adjustierung nicht zu beachten ist nicht sachgerecht. Vielmehr ist vom pU zu beschreiben, auf welche Weise dann mit dem beträchtlichen Verlust von Informationen in den Auswertungen umzugehen ist, und unter welchen Voraussetzungen überhaupt noch der Versuch einer Adjustierung für Confounder sinnvoll ist. Ein hoher Anteil fehlender Daten für relevante Confounder führt potenziell zu einer unzureichenden Adjustierung und in der Folge zu nicht validen Ergebnissen. Es sind daher besondere Anstrengungen im Rahmen der AbD zu unternehmen, dass die Erhebung der Confounder vollständig ist. Sollten aus Sicht des pU und der Registerbetreibenden hierfür

Maßnahmen erforderlich sein, die über die adressierten Maßnahmen hinausgehen, sollten diese in der AbD und damit auch im SP verankert werden.

Der Umgang mit fehlenden Angaben zu Endpunkten ist nicht beschrieben. Das ist nicht sachgerecht, da dies zu Informationsverlusten und zu erheblichen Verzerrungen der Ergebnisse führen kann und ist daher vom pU zu ergänzen. Ausnahme sind Analysen mittels MMRM, bei denen aufgrund der statistischen Methodik keine Ersetzung fehlender Werte für die Endpunkte notwendig ist.

3 Literatur

1. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: (AM-RL) Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V); Valoctocogen Roxaparvovec (schwere Hämophilie A); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2023 [Zugriff: 02.08.2023]. URL: https://www.g-ba.de/downloads/39-261-5866/2023-02-02_AM-RL-XII_Valoctocogen_Roxaparvovec_AbD-002_Forderung_BAnz.pdf.
2. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V); Valoctocogen Roxaparvovec (schwere Hämophilie A); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2023 [Zugriff: 02.08.2023]. URL: https://www.g-ba.de/downloads/40-268-9250/2023-02-02_AM-RL-XII_Valoctocogen_Roxaparvovec_AbD-002_Forderung_TrG.pdf.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzept für eine anwendungsbegleitende Datenerhebung – Valoctocogen Roxaparvovec; Rapid Report [online]. 2023 [Zugriff: 07.02.2023]. URL: https://www.iqwig.de/download/a22-20_anwendungsbegleitende-datenerhebung-valoctocogen-roxaparvovec_rapid-report_v1-0.pdf.
4. Biomarin. Comparative Effectiveness of Roctavian to Standard of Care Hemostatic Therapies in Germany Among People with Severe Hemophilia A: A Prospective Non-Interventional Study Utilizing Data Collected in the German Hemophilia Register [unveröffentlicht]. 2023.
5. IQVIA. Comparative Effectiveness of Roctavian to Standard of Care Hemostatic Therapies in Germany Among People with Severe Hemophilia A: A Prospective Non-Interventional Study Utilizing Data Collected in the German Hemophilia Register; Statistical Analysis Plan (SAP) [unveröffentlicht]. 2023.
6. Gemeinsamer Bundesausschuss. Niederschrift zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV; Beratungsanforderung 2023-B-065 [unveröffentlicht]. 2023.
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Anwendungsbegleitende Datenerhebung zu Valoctocogen Roxaparvovec: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0); 2. Addendum zum Projekt A22-20 [online]. 2023 [Zugriff: 27.11.2023]. URL: https://www.iqwig.de/download/a23-64_valoctocogen-roxaparvovec_addendum-zum-projekt-a22-20_v1-0.pdf.

8. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Feststellung im Verfahren der anwendungsbegleitenden Datenerhebung und von Auswertungen nach § 35a Absatz 3b des Fünften Buches Sozialgesetzbuch (SGB V):

Valoctocogen Roxaparvovec (schwere Hämophilie A) – Vorlage von Studienprotokoll und Statistischem Analyseplan [online]. 2023 [Zugriff: 27.11.2023]. URL: https://www.g-ba.de/downloads/39-261-6179/2023-09-21_AM-RL-XII_Valoctocogen_2020-AbD-002_Feststellung.pdf.

9. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Feststellung im Verfahren der anwendungsbegleitenden Datenerhebung und von Auswertungen nach § 35a Absatz 3b des Fünften Buches Sozialgesetzbuch (SGB V): Valoctocogen Roxaparvovec (schwere Hämophilie A) – Vorlage von Studienprotokoll und Statistischem Analyseplan [online]. 2023 [Zugriff: 27.11.2023]. URL: https://www.g-ba.de/downloads/40-268-9788/2023-09-21_AM-RL-XII_Valoctocogen_2020-AbD-002_Feststellung_TrG.pdf.

10. Deutsches Hämophileregister. Gesamtdatensatz dhr 2.0 [online]. 2023 [Zugriff: 02.08.2023]. URL: <https://www.pei.de/SharedDocs/Downloads/DE/regulation/meldung/dhr-deutsches-haemophileregister/dhr-20-datensatz.pdf?blob=publicationFile&v=10>.

11. Deutsches Hämophileregister. Handbuch, Version 2.3 [online]. 2023 [Zugriff: 07.08.2023]. URL: <https://www.pei.de/SharedDocs/Downloads/DE/regulation/meldung/dhr-deutsches-haemophileregister/dhr-20-handbuch.pdf?blob=publicationFile&v=9>.

12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V; Rapid Report [online]. 2020 [Zugriff: 26.07.2022]. URL: https://www.iqwig.de/download/a19-43_versorgungsnaher-daten-zum-zwecke-der-nutzenbewertung_rapid-report_v1-1.pdf.

13. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015; 34(28): 3661-3679. <https://dx.doi.org/10.1002/sim.6607>.

14. Thoemmes F, Ong AD. A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. Emerging Adulthood 2016; 4(1): 40-59. <https://dx.doi.org/10.1177/2167696815621645>.

15. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res 2011; 46(3): 399-424. <https://dx.doi.org/10.1080/00273171.2011.568786>.

16. Williamson E, Morley R, Lucas A et al. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res* 2012; 21(3): 273-293.

<https://dx.doi.org/10.1177/0962280210394483>.

17. Eiset AH, Frydenberg M. Considerations for Using Multiple Imputation in Propensity Score-Weighted Analysis - A Tutorial with Applied Example. *Clin Epidemiol* 2022; 14: 835-847. <https://dx.doi.org/10.2147/CLEP.S354733>.

18. Granger E, Sergeant JC, Lunt M. Avoiding pitfalls when combining multiple imputation and propensity scores. *Stat Med* 2019; 38(26): 5120-5132.

<https://dx.doi.org/10.1002/sim.8355>.

19. Ling A, Montez-Rath M, Mathur M et al. How to Apply Multiple Imputation in Propensity Score Matching with Partially Observed Confounders: A Simulation Study and Practical Recommendations. *Journal of Modern Applied Statistical Methods* 2020; 19(1).

<https://dx.doi.org/10.22237/jmasm/1608552120>.

20. Segalas C, Leyrat C, Carpenter JR et al. Propensity score matching after multiple imputation when a confounder has missing data. *Stat Med* 2023; 42(7): 1082-1095.

<https://dx.doi.org/10.1002/sim.9658>.