

Anwendungsbegleitende Datenerhebung zu Valoctocogen Roxaparvovec: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0)

2. Addendum zum Projekt A22-20

A horizontal bar composed of several colored segments. The word 'ADDENDUM' is written in white capital letters on a dark blue segment. The segments to the left are light blue, and the segments to the right are dark blue.

ADDENDUM

Projekt: A23-64

Version: 1.0

Stand: 14.08.2023

IQWiG-Berichte – Nr. 1615

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Anwendungsbegleitende Datenerhebung zu Valoctocogen Roxaparvovec: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0) –

2. Addendum zum Projekt A22-20

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

04.07.2023

Interne Projektnummer

A23-64

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

An dem Addendum beteiligte Mitarbeiterinnen und Mitarbeiter des IQWiG

- Anna-Katharina Barnert
- Lars Beckmann
- Katharina Hirsch
- Ulrike Seay
- Siw Waffenschmidt
- Volker Vervölgyi

Schlagwörter

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

Keywords

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

2.3 Zusammenfassung und Fazit	24
3 Literatur	25
Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern ..	29

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AbD	Anwendungsbegleitende Datenerhebung
ABR	annualisierte Blutungsrate
AIC	Akaike Information Criterion
ATE	Average Treatment Effect
ATT	Average Treatment Effect in Treated
G-BA	Gemeinsamer Bundesausschuss
IPTW	Inverse Probability of Treatment Weighting
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
MedDRA	Medical Dictionary for Regulatory Activities
PICO	Patient, Intervention, Comparator, Outcome
PS	Propensity Score
PT	Preferred Term (bevorzugter Begriff)
pU	pharmazeutischer Unternehmer
SAP	statistischer Analyseplan
SGB	Sozialgesetzbuch
SMD	standardisierte Mittelwertdifferenz
SMRW	Standardized Mortality Ratio Weighting
SOC	System Organ Class (Systemorganklasse)
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) 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 Studienprotokoll 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]).

2 Prüfung der Unterlagen zur Planung der AbD von Valoctocogen Roxaparovec

2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen

Der pU hat, wie im G-BA-Beschluss zur AbD gefordert [1,2], ein Studienprotokoll und einen SAP zur Prüfung vorgelegt, beide Dokumente in der Version vom 29.06.2023 [4,5].

Gemäß Beschluss des G-BA ist die AbD als vergleichende Registerstudie oder, sofern eine vergleichende Registerstudie nicht umsetzbar ist, als vergleichende Studie unter Nutzung einer spezifisch für die AbD aufzusetzenden Datenplattform (studienindividuelle Datenerhebung) umzusetzen [1]. Zudem soll im Falle einer vergleichenden Registerstudie das Deutsche Hämophileregister (DHR) als primäre Datenquelle genutzt werden, sofern die im G-BA-Beschluss aufgeführten Qualitätskriterien erfüllt sind. Der pU wählt das DHR als Datenquelle für die AbD.

In den nachfolgenden Abschnitten werden die vom pU vorgelegten Dokumente dahin gehend geprüft, ob durch sie die Durchführung der AbD unter Verwendung des DHR einschließlich der zugehörigen Auswertungen ausreichend genau beschrieben wird, ob die Planung der AbD einschließlich der Auswertung der Daten inhaltlich und methodisch sachgerecht ist, und ob die im G-BA-Beschluss genannten Anforderungen adäquat umgesetzt wurden.

2.2 Spezifische Anmerkungen zum Studienprotokoll und SAP

2.2.1 A priori Festlegung aller Definitionen, Operationalisierungen und Analysen

Im SP und SAP beschreibt der pU an verschiedenen Stellen, dass er Definitionen, Operationalisierungen (bspw. von Endpunkten oder Confoundern) oder Auswertungen erst anhand der beobachteten Daten oder im Verlauf der AbD festlegen will. Dieses Vorgehen ist nicht sachgerecht. Alle relevanten Daten müssen vor Beginn der AbD festgelegt und im SP und SAP beschrieben werden.

2.2.2 Fragestellung gemäß PICO

Die Fragestellung des G-BA für die AbD und die anschließende Auswertung der Daten sind im Beschluss des G-BA mithilfe des PICO-Schemas niedergelegt (PICO: Population, Intervention, Kontrolle, Outcome). Die folgenden Abschnitte beurteilen die Umsetzung des PICO im SP und SAP.

2.2.2.1 Population

Gemäß G-BA-Beschluss soll die AbD für folgende Population durchgeführt werden: Erwachsene mit schwerer Hämophilie A (angeborener Faktor-VIII-Mangel) ohne Faktor-VIII-Hemmkörper in der Vorgeschichte.

Die Einschlusskriterien der geplanten AbD stimmen gemäß SP prinzipiell mit den Vorgaben des G-BA überein. Der pU schränkt die Patientenpopulation auf männliche Patienten ein. Da die Hämophilie A nahezu ausschließlich Männer betrifft, ist dies nachvollziehbar. Weder aus dem G-BA-Beschluss noch aus den Angaben der Fachinformation geht jedoch hervor, dass sich die Anwendung von Valoctocogen Roxaparvovec und somit auch die Patientenpopulation der AbD auf Männer beschränkt. Darüber hinaus müssen die Ausschlusskriterien entsprechend den Vorgaben der Fachinformation von Valoctocogen Roxaparvovec noch um das Vorliegen aktiver Infektionen, entweder akut oder unkontrolliert chronisch und Patienten mit bekannter signifikanter Leberfibrose oder -zirrhose ergänzt werden [7].

Der pU beschränkt durch die Einschlusskriterien die Patientenpopulation auf Patienten, die bereits mindestens 12 Monate mit einer prophylaktischen Therapie behandelt werden. Warum er diese Einschränkung wählt, geht aus den Angaben im SP und SAP nicht hervor. Da jedoch davon ausgegangen wird, dass alle Patienten mit schwerer Hämophilie A eine Prophylaxebehandlung erhalten und durch die gesetzliche Meldepflicht im DHR entsprechend für alle Patienten Daten für die letzten Monate vorliegen, erscheint dies zunächst keine Konsequenz zu haben. Der pU beschreibt jedoch auch, dass er zur Erhebung der Baseline-Daten nur auf Patienten zugreifen möchte, für die bereits seit mindestens 12 Monaten Daten im DHR dokumentiert sind. Aus dieser Einschränkung resultiert, dass der pU nur auf Patienten für die AbD zurückgreifen kann, die bereits mindestens 12 Monate vor Einschluss in die AbD ihre Zustimmung zur Einzelmeldung im DHR gegeben haben, da mit einer Sammelmeldung nur aggregierte Daten vorliegen [3]. Nach Angaben des DHR waren im Jahr 2019 insgesamt 2637 Patienten mit schwerer Hämophilie A erfasst. Von diesen 2637 Patienten haben 1064 Patienten einer Einzelmeldung ihrer Daten zugestimmt, von denen 715 Patienten zum damaligen Zeitpunkt keine Hemmkörper gegen den Faktor VIII aufwiesen. Wie viele dieser 715 Patienten die Einschlusskriterien der AbD erfüllen, ist nicht absehbar [3]. Entsprechend der orientierenden Fallzahlschätzung auf Basis der annualisierten Blutungsrate, die auch der pU im SAP verwendet, wird die benötigte Stichprobengröße auf circa 397 Patienten geschätzt. Der pU berücksichtigt dabei allerdings nicht, dass laut Beschluss des G-BA sichergestellt werden soll, dass eine ausreichende Anzahl an Patienten unter einer Faktor-VIII-Therapie rekrutiert wird (siehe Abschnitt 2.2.5.1). Die benötigte Fallzahl sollte trotzdem prinzipiell über die bereits in Einzelmeldung vorliegende Patientenzahl im DHR erreichbar sein. Gegebenenfalls kann es dennoch notwendig sein, Patienten aus der Sammelmeldung in die Einzelmeldung zu überführen. Dies wurde auch im Fachaustausch zu Valoctocogen Roxaparvovec diskutiert [8]. Hierfür sollte der pU geeignete Maßnahmen darlegen und die Rekrutierung der Patientenpopulation nicht von vornherein nur auf Patienten beschränken, die bereits ihre Zustimmung zur Einzelmeldung gegeben haben.

Der pU beschreibt die Ein- und Ausschlusskriterien anhand der Patientencharakteristika zu prüfen, die bei der Aufnahme in das DHR dokumentiert wurden. Dieses Vorgehen ist nicht

sachgerecht. Wie oben ausgeführt, kann er auf diese Weise nur Patienten für die AbD identifizieren, für die bereits ein Einverständnis zur Einzelmeldung im DHR vorliegt. Darüber ist so nicht sichergestellt, dass die herangezogenen Patientencharakteristika noch aktuell sind und sich keine Änderungen ergeben haben, die einer Teilnahme an der AbD entgegenstehen bzw. zum Heranziehen veralteter Patientencharakteristika führen. Dies betrifft insbesondere die Angaben zum Leberstatus und dem Vorliegen von Hemmkörpern gegen Faktor VIII, die sich zwischen dem Zeitpunkt der Aufnahme in das DHR und dem Einschluss in die AbD verändert haben können.

Der pU gibt an das Alter in 5-Jahres-Kategorien darzustellen. Dieses Vorgehen ist nicht nachvollziehbar. Für jeden Patienten sollte das Alter eindeutig aus den Daten hervorgehen.

Alle Ein- und Ausschlusskriterien müssen verpflichtend im DHR und mit eindeutiger Operationalisierung erfasst werden. Die Patientencharakteristika müssen erneut zum Indexdatum erhoben werden, damit zu Baseline aktuelle Daten vorliegen. Dies betrifft auch Patienten, die von einer Behandlung mit einer Kontrolltherapie auf eine Behandlung mit Valoctocogen Roxaparvovec wechseln und dem Interventionsarm zugeordnet werden (siehe Abschnitt 2.2.4).

2.2.2.2 Intervention und Kontrolle

Die Benennung der zu vergleichenden Therapieoptionen stimmt mit den Vorgaben des G-BA überein (Intervention: Valoctocogen Roxaparvovec; Kontrolle: Therapie nach ärztlicher Maßgabe unter Berücksichtigung von rekombinanten oder aus humanem Plasma gewonnenen Blutgerinnungsfaktor-VIII-Präparaten und Emicizumab).

In den Tragenden Gründen weist der G-BA darauf hin, dass Emicizumab für die AbD zwar als geeigneter Komparator angesehen wird, aber noch nicht feststeht, ob Emicizumab für die Nutzenbewertung Teil der zweckmäßigen Vergleichstherapie sein wird [2]. Aus diesen Gründen soll der pU geeignete Maßnahmen etablieren und im SP und SAP darlegen, über die sichergestellt werden kann, dass ausreichend Patienten, die mit Blutgerinnungsfaktor-VIII-Präparaten behandelt werden, in die AbD eingeschlossen werden. Dies ist insbesondere bei der Fallzahlschätzung zu berücksichtigen (siehe Abschnitt 2.2.5.1) Der pU geht auf diesen Punkt weder im SP noch im SAP genauer ein.

2.2.2.3 Endpunkte (Outcomes)

Der pU berücksichtigt die vom G-BA im Beschluss zur AbD geforderten Endpunkte wie folgt:

Mortalität

Der pU plant den Endpunkt Gesamtmortalität im Rahmen der unerwünschten Ereignisse zu erheben. Das ist sachgerecht. Darüber hinaus sollen die Hämophilie assoziierten Todesfälle

dargestellt werden. Diese Operationalisierung wird im Beschluss des G-BA nicht gefordert und wird für die Nutzenbewertung nicht herangezogen. Relevant für die Nutzenbewertung sind Todesfälle jeglicher Ursache. Im SP wird die Erhebung der Gesamtmortalität nicht in der Tabelle 8, in der die Erhebungen zu den Sicherheitsendpunkten gelistet sind, aufgeführt. Die Angaben zur Gesamtmortalität müssen ergänzt werden.

Morbidität

Blutungen

Aus dem Forderungsbeschluss des G-BA [1] geht hervor, dass für die AbD schwere Blutungen, lebensbedrohliche Blutungen sowie Gelenkblutungen erhoben und getrennt dargestellt werden sollen.

Der pU plant behandelte Blutungen, schwere Blutungen, lebensbedrohliche Blutungen und Gelenkblutungen zu erheben und als annualisierte Blutungsraten (ABR) auszuwerten, sowie den Anteil der Patienten mit Blutungsfreiheit. Der pU beschreibt, dass der Anlass einer Bedarfsbehandlung anhand der Datenfelder im DHR wie folgt spezifiziert werden soll: Verdacht auf Blutung, spontane Blutung sowie unbekannter Grund. Bei den schweren und lebensbedrohlichen Blutungen erscheint es jedoch nicht plausibel als Grund den Verdacht einer Blutung oder einen unbekanntem Grund anzugeben. Dies sollte angepasst werden. Im Protokoll wird der kombinierte Endpunkt Major Bleeds benannt, in den sowohl die schweren Blutungen als auch die lebensbedrohlichen Blutungen eingehen, separate Auswertungen sind nur für die lebensbedrohlichen Blutungen vorgesehen. Dieses Vorgehen ist nachvollziehbar. Der pU legt jedoch nicht dar, wie schwere Blutungen und lebensbedrohliche Blutungen für die AbD operationalisiert werden sollen. Wichtig für eine vollständige und möglichst unverzerrte Dokumentation von Blutungsereignissen ist eine einheitliche, eindeutige und möglichst objektive Definition dieser Ereignisse. Laut Gesamtdatensatz kann im DHR derzeit unter anderem der Schweregrad einer Blutung als leicht, schwer, lebensbedrohlich oder unbekannt angegeben werden [9]. Es finden sich jedoch weder im Gesamtdatensatz noch im Handbuch des DHR [10] genaue Kriterien, wann eine Blutung als schwer einzustufen ist. Es besteht deshalb das Risiko, dass in den einzelnen Zentren generell oder für individuelle Patienten unterschiedliche Definitionen angewendet werden. Wenn die bestehenden Datenfelder für die Blutungsereignisse genutzt werden sollen, ist es erforderlich, diese mit klaren Definitionen zu versehen und ggf. die behandelnden Ärztinnen und Ärzte entsprechend dieser Erhebung zu schulen. Zudem wird laut Gesamtdatensatz für jede Blutung die Lokalisierung angegeben, die unter anderem Gelenk oder Zielgelenk sein kann. Es wird daher davon ausgegangen, dass Gelenkblutungen anhand des DHR hinreichend sicher erfasst werden können. Für die AbD ist es erforderlich im Studienprotokoll die Kriterien für schwere Blutungen und lebensbedrohliche Blutungen präzise zu beschreiben. Die geplanten Operationalisierungen der Endpunkte zu Blutungen als ABR bzw. als Anteil der Patienten mit Blutungsfreiheit sind sachgerecht.

Beginn Erhebung von Blutungen im Interventionsarm

Der pU plant die Erhebung der Endpunkte zu den Blutungen im Valoctocogen-Arm ab 5 Wochen nach Verabreichung der Intervention, 3 Tage nach der letzten Routine FVIII Prophylaxe oder 27 Wochen nach der letzten Emicizumab Prophylaxe zu beginnen, je nachdem, welches dieser Ereignisse als Letztes auftritt. Für den Vergleichsarm soll die Erhebung mit dem Indexdatum beginnen. Dies ist nicht sachgerecht. Der erste Zeitraum nach Gabe von Valoctocogen Roxaparvovec ist Teil der Behandlungsstrategie, insofern sollte die Erhebung der Blutungsereignisse ebenfalls ab dem Indexdatum erfolgen. Das vom pU beschriebene Vorgehen könnte beispielsweise zu einem Immortal Time Bias führen [11]. Erforderlich ist die Erhebung aller Endpunkte für alle Patienten ab dem Indexdatum (siehe Abschnitt 2.2.4).

Schmerz, gemessen mit einem validierten Instrument

Der pU plant den Endpunkt Schmerz mit der Brief Pain Inventory – Short Form (BPI-SF) zu erheben.

Der BPI-SF ist ein generisches Instrument zur Messung der Schmerzintensität und der Beeinträchtigung durch Schmerzen, das auch in der Indikation Hämophilie herangezogen werden kann. Die im Studienprotokoll vorgesehene Auswertung der Skalen Schmerzintensität und Beeinträchtigung durch Schmerzen entspricht den Vorgaben des Handbuchs [12]. Der pU plant aus der Skala Schmerzintensität das Item 5 „durchschnittlicher Schmerz“ und aus der Skala Beeinträchtigung durch Schmerzen der Durchschnitt der Items 9A – G auszuwerten. Die Verwendung des BPI-SF für die geplante AbD ist sachgerecht.

Häufigkeit der Erhebung von Schmerzen

Die Erhebung des BPI-SF soll 2-mal jährlich erfolgen. Bei Patienten mit Hämophilie ist die Unterscheidung von akutem (insb. durch Hämarthrose) und chronischem Schmerz (insb. durch hämophile Arthropathie) essenziell [13,14]. In einer solchen Situation erscheint eine Operationalisierung, die die Entwicklung im Studienverlauf mitberücksichtigt, sinnvoll. Für eine solche Operationalisierung ist eine häufigere Erhebung als 1- bis 2-mal jährlich notwendig.

Gelenkfunktion, gemessen mit einem validierten Instrument

Für die Gelenkfunktion plant der pU das Instrument Haemophilia Joint Health Score (HJHS) einzusetzen. Die Häufigkeit der Erhebung wird im SP nicht konsistent beschrieben und wird mit 1-mal bzw. 2-mal jährlich angegeben. Dies muss angeglichen werden, wobei eine 2-mal jährliche Erhebung zu bevorzugen ist. Der pU beschreibt, dass er den Gesamtscore und den Global Gait Score zu Baseline sowie zu jedem Follow-up-Termin bestimmen möchte. Bei dem HJHS handelt es sich um ein validiertes und etabliertes Instrument [15,16], das bereits im DHR erfasst wird. Die Erfassung der Gelenkfunktion mit dem HJHS ist sachgerecht.

Gesundheitsbezogene Lebensqualität

Der pU plant die gesundheitsbezogene Lebensqualität mit dem Haemo-QoL-A 2-mal jährlich zu erheben. Der Haemo-QoL-A ist ein validiertes und etabliertes Instrument zur Erhebung der gesundheitsbezogenen Lebensqualität bei Patienten mit Hämophilie. Die im SAP vorgesehenen Berechnungen der einzelnen Skalen und des Gesamtscores entsprechen den Vorgaben der Validierungspublikationen [17,18]. Die Verwendung des Haemo-QoL-A in der AbD ist sachgerecht.

Allgemeine Anmerkung zur Erhebung der PROs sowie der Gelenkfunktion

Der pU beschreibt, dass er davon ausgeht, dass die Patienten alle 6 Monate zur Kontrolle kommen und plant die PROs jeweils an diesen Visiten zu erheben. Er beschreibt, daher Erhebungen zu Baseline und zu den Monaten 6, 12, 18, 24, 30 und 36 mit einem Toleranzfenster von ± 3 Monaten für jeden Erhebungszeitpunkt nach Baseline. Dieses Vorgehen ist nicht sachgerecht, da die Erhebungszeitpunkte durch die Wahl der Toleranzfenster aneinandergrenzen. Dies hätte beispielsweise bei den Auswertungen über ein MMRM Konsequenzen, bei dem die Werte ohne Berücksichtigung des eigentlichen Erhebungsdatums den geplanten Zeitpunkten zugeordnet werden, obwohl diese ggf. direkt aneinandergrenzen. Der pU sollte daher angemessene Toleranzfenster wählen und über geeignete Maßnahmen sicherstellen, dass fehlende Werte vermieden werden.

Für die Erhebung der PROs sowie der Gelenkfunktion könnte der pU auch eine Erhebung durch ein externes Zentrum prüfen. Dieses Vorgehen wird beispielsweise in der AbD zu Brexucabtagen Autoleucel beim Mantelzelllymphom umgesetzt [19]. Durch diese Vorgehensweise wäre die Erhebung der PROs nicht an die Zahl der Visiten gebunden.

Unerwünschte Ereignisse

Gemäß Forderungsbeschluss des G-BA sollen für die AbD folgende UEs erfasst werden:

- schwerwiegende unerwünschte Ereignisse (SUEs) (operationalisiert als Ereignisse, die zur Hospitalisierung oder zum Tod führen; Gesamtrate
- spezifische unerwünschte Ereignisse (sofern angezeigt mit Angabe des jeweiligen Schweregrads)
 - Bildung von Faktor-VIII-Hemmkörpern
 - thromboembolische Ereignisse
 - symptomatische Leberschäden
 - maligne Neubildungen

Im Studienprotokoll ist die Erhebung der vom G-BA geforderten UE Endpunkte vorgesehen, bezüglich der Operationalisierung und der geplanten Auswertungen ergeben sich allerdings Kritikpunkte:

Laut Studienprotokoll ist geplant SUEs operationalisiert als Ereignisse, die zur Hospitalisierung oder zum Tod führen, zu erheben. Allerdings sollen nur diejenigen Ereignisse, die im Zusammenhang mit der Behandlung der Hämophilie auftreten, in die Analysen eingehen. Dies ist nicht sachgerecht. Gemäß Anforderung des G-BA soll die Gesamtrate aller Ereignisse, die zur Hospitalisierung oder zum Tod führen ausgewertet werden.

Für die spezifischen UE ist die Angabe des Schweregrades nicht vorgesehen. Auch für die spezifischen UEs sollte neben der Erfassung der jeweiligen Gesamtraten die Rate an Ereignissen erfasst werden, die zur Hospitalisierung oder zum Tod führen.

Auch für die thromboembolischen Ereignisse sollen nur diejenigen Ereignisse, die im Zusammenhang mit der Behandlung der Hämophilie auftreten, in die Analysen eingehen. Dies ist nicht sachgerecht, erforderlich ist die Auswertung aller thromboembolischer Ereignisse.

Die Bildung von Faktor-VIII-Hemmkörpern soll auf Basis positiver Testergebnisse ggf. in Verbindung mit der Verabreichung einer Immuntoleranztherapie erhoben werden. Dieses Vorgehen ist sachgerecht. Im Studienprotokoll ist geplant, das spezifische UE schwere Leberschäden, operationalisiert als Leberversagen oder Zirrhose, zu erheben. Im statistischen Analyseplan wird die Operationalisierung dahin gehend präzisiert, dass Leberversagen, Leberfibrose und Leberzirrhose anhand der Child-Pugh-Kriterien A, B und C erhoben werden sollen. Laut Beschluss des G-BA sollen symptomatische Leberschäden erfasst werden. Der pU sollte daher sicherstellen, dass durch die Operationalisierung Leberschäden erfasst werden, die nicht allein auf beispielsweise Veränderungen von Laborwerten beruhen. Maligne Neubildungen sollen erhoben werden und ohne Einschränkungen in die Analysen eingehen. Die Erfassung maligner Neubildungen kann im DHR derzeit nur über Freitext dokumentiert werden (unter dem Datenfeld sonstige relevantes Ereignis) [9]. Wie der pU selbst im SP beschreibt, können Unterschiede in der Erhebungspraxis zwischen den Zentren entstehen. Um die Ereignisse eindeutig zu erfassen, muss im Studienprotokoll spezifiziert werden, wie Ereignisse zu malignen Neubildungen im DHR einheitlich dokumentiert werden.

Der pU plant Patienten nach einem Behandlungswechsel zu zensieren. Im Sinne einer ITT-Auswertung sind alle Nebenwirkungen bis zum Studienende zu erheben und entsprechend auszuwerten

Ergänzende Informationen zu:***Anzahl der verbrauchten Faktorkonzentrate sowie Emicizumab, getrennt nach Bedarfs- und Prophylaxebehandlung******Zeitpunkt der Wiederaufnahme einer prophylaktischen Behandlung***

Es ist vorgesehen die Menge der Faktorkonzentrate sowie die Behandlung mit Emicizumab, getrennt nach Bedarfs- und Prophylaxebehandlung zu erheben. Dies entspricht den Vorgaben des G-BA. Der Endpunkt Zeitpunkt der Wiederaufnahme einer prophylaktischen Behandlung im Valoctocogen-Arm ist operationalisiert als 4 Dosen von Emicizumab oder 18 Dosen eines Faktorpräparats mit verlängerter Halbwertszeit oder 24 Dosen eines Faktorpräparats mit normaler Halbwertszeit. Dies ist nicht sachgerecht. Die Wiederaufnahme jeglicher prophylaktischen Behandlung sollte dokumentiert und in der Nutzenbewertung dargestellt werden.

2.2.3 Datenquelle

Der pU wählt das DHR als Datenquelle für die AbD. Dieses Register wurde auch im Konzept für die AbD als potenziell geeignetes Indikationsregister identifiziert (Rapid Report A22-20 [3]). Im AbD-Konzept wurde dargelegt, welche Anpassungen am Register notwendig sind, um die international einheitlichen und im G-BA-Beschluss zur AbD benannten Qualitätskriterien zu erfüllen.

Der pU führt weder im SP noch im SAP spezifisch aus, welche Anpassungen am DHR mit welcher Frist geplant bzw. bereits vorgenommen wurden. Aus den Angaben im SP und SAP lässt sich jedoch weiterhin Anpassungsbedarf bei folgenden Kriterien ableiten:

- Erhöhung des Anteils Zahl der Patienten mit Einzelmeldung: Wie in Abschnitt 2.2.2.1 beschrieben, schränkt der pU die Patientenpopulation auf Patienten ein, für die bereits eine Zustimmung zur Einzelmeldung vorliegt. Der pU sollte beschreiben, welche Maßnahmen er ergreift, um bei Bedarf weitere Patienten aus der Sammelmeldung in die Einzelmeldung überführen zu können.
- Exakte Definition und Operationalisierung der Expositionen, klinischen Ereignisse, Endpunkte und Confounder: Der pU beschreibt im SAP über welche Datenfelder im DHR er die für die AbD relevanten Daten erheben möchte. Teilweise sind diese Datenfelder jedoch noch nicht im DHR etabliert. So bestehen beispielsweise bei der Dokumentation der Erhebungen mit dem BPI-SF oder dem Haemo-QoL-A noch Unsicherheiten, was im DHR dokumentiert werden soll. Darüber hinaus beschreibt der pU im SP, dass die Definitionen der Variablen anhand der Rohdaten verfeinert werden sollen. Dieses Vorgehen ist nicht sachgerecht. Es müssen alle zu erhebenden Daten mit ihrer entsprechenden Operationalisierung prädefiniert und im SP bzw. SAP beschrieben werden sowie deren Umsetzung im DHR vor dem Start der AbD gewährleistet sein. Wo

möglich, sollten darüber hinaus Standard-Klassifikationen genutzt werden, um eine einheitliche Codierung der Daten zu gewährleisten. Der pU beschreibt, dass z. B. Angaben zu Komorbiditäten als Freitext erfasst werden sollen und merkt selbst an, dass dies dazu führen kann, dass die Daten zu Komorbiditäten nicht sinnvoll ausgewertet werden können. Der pU muss durch die Wahl der Operationalisierungen und eindeutigen Definition der entsprechenden Datenfelder gewährleisten, dass zu allen relevanten Patientencharakteristika bzw. Confoundern sowie Endpunkten verwertbare Daten vorliegen.

- Einheitliche Meldezeitpunkte: Wie im Konzept beschrieben, besteht für das DHR die 1-mal jährliche Meldeverpflichtung. Der pU koppelt die Erhebungen an die regulär anstehenden Visiten der Patienten und macht im SP die Angabe, dass die Dokumentation im DHR mindestens 1-mal jährlich erfolgen soll, jedoch häufiger erfolgen kann, wenn dies im Anschluss an die jeweilige Visite erfolgt. Die Dokumentation der erhobenen Daten sollte grundsätzlich möglichst für alle Patienten einheitlich direkt im Anschluss an die jeweilige Visite erfolgen, um Meldeverzögerungen zu vermeiden.
- Verpflichtend zu dokumentierende Angaben im Register: Aus den Angaben im SP geht hervor, dass für einige Datenfelder zu relevanten Endpunkten, beispielsweise zur Erhebung von Schmerzen oder der gesundheitsbezogenen Lebensqualität keine verpflichtende Meldung im DHR besteht. Für relevante Daten, die im Zuge der AbD erhoben werden sollen, müssen jedoch alle relevanten Datenfelder spätestens mit Start der AbD verpflichtend zu dokumentieren sein.
- Datenfelder zu Confoundern: Der pU beschreibt im SP und SAP nicht konkret, wie er die Confounder im DHR erheben möchte. Der pU möchte erst vor der Zwischenanalyse nach 18 Monaten eine systematische Recherche nach Confoundern durchführen und abgleichen, ob diese im DHR erhoben werden. Das Vorgehen ist nicht sachgerecht. Es ist erforderlich, dass vorab alle wichtigen Confounder (inklusive wichtiger Wechselwirkungen) mithilfe eines systematischen Ansatzes identifiziert werden und sichergestellt wird, dass diese spätestens mit Start der AbD im DHR verpflichtend erhoben werden.
- Sicherstellung der Richtigkeit der Daten mittels Source Data Verification: Der pU beschreibt, dass er durch Abgleich der Daten, die in den Zentren ins DHR eingegeben wurden mit denen vom DHR übermittelten aggregierten Daten eine Source Data Verification für geschätzt 100 % der Datenfelder für die Ein- und Ausschlusskriterien und den primären Endpunkt durchführen möchte. Eine Source Data Verification umfasst den Abgleich zwischen der Quelldatei (in der Regel der Patientenakte) und den in das Register gemachten Einträgen. Vor Ort sollte entsprechend ein Abgleich zwischen den ans DHR gemeldeten Daten und den entsprechenden Angaben der jeweiligen Patientenakte erfolgen. Der pU verweist für weitere Angaben auf den Study Monitoring

Plan (SMP), der nicht vorliegt. Wie die Source Data Verification für die AbD im Detail geplant ist, ist somit nicht zu überprüfen. Entsprechend des G-BA Beschlusses sollte eine Source Data Verification für 100 % der Patienten je Erhebungszentrum für den primären Endpunkt und für mindestens 10 % zufällig ausgewählter Patienten je Erhebungszentrum für alle weiteren Endpunkte über den Zeitraum seit Beginn der Datensammlung durchgeführt werden [1].

2.2.4 Beobachtungsstart und Indexdatum

Rekrutierungszeitraum

Als Rekrutierungszeitraum (vom pU als „Identification Window“ bezeichnet) für die AbD legt der pU das 1. Quartal 2024 bis Ende des 3. Quartals 2025 fest. Da die Beobachtungszeit für die AbD gemäß Beschluss des G-BA 3 Jahre betragen soll [1] und der finale Studienbericht für Februar 2029 erwartet wird, erscheint dies angemessen.

Zuordnung zu den Behandlungsgruppen

Im Studienprotokoll werden alle Patienten, die während des Rekrutierungszeitraums eine Behandlung mit Valoctocogen Roxaparovec bekommen, dem Interventionsarm zugeordnet. Dieses Vorgehen ist sachgerecht. In den Fällen, in denen die Patienten im Rekrutierungszeitraum zunächst dem Vergleichsarm zugeordnet wurden, stellt sich die Frage, wie mit den bis dahin erhobenen Daten im Vergleichsarm umgegangen werden soll. Der pU macht dazu weder im SP noch im SAP Angaben. In der vorliegenden Situation sollte eine Strategie im Sinne einer ITT-Auswertung verfolgt werden, wobei Patienten, die im Verlauf der Beobachtung von der Vergleichstherapie auf Valoctocogen Roxaparovec wechseln, unterschiedlich berücksichtigt werden, abhängig davon, wie lange sie bereits unter der Vergleichstherapie beobachtet wurden. Diejenigen Patienten, die bereits eine 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 Roxaparovec wechseln, sollten im Vergleichsarm ausgewertet und bis zum Studienende weiter beobachtet werden. Patienten, die nach einer kürzeren Zeit auf eine Behandlung mit Valoctocogen Roxaparovec wechseln und für die noch eine angemessene Beobachtungszeit unter der Genterapie zu erwarten ist, sollten hingegen im Interventionsarm ausgewertet werden. 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 wird für die AbD nicht berücksichtigt.

Für dieses Vorgehen ist im SP/SAP festzulegen, wie die oben genannte angemessene Beobachtungszeit definiert ist. Gemäß Beschluss des G-BA sollen die Patienten mindestens 3 Jahre beobachtet werden [1]. Zu beachten ist dabei auch, wie mit den Patienten umgegangen wird, die kurze Zeit nach dem angesetzten Rekrutierungszeitraum auf die Genterapie

wechseln, da auch für diese Patienten noch eine angemessene Beobachtungszeit unter der Gentherapie erwartbar ist.

Um den Einfluss der Gentherapie in der Vergleichsgruppe zu untersuchen, sollten Sensitivitätsanalysen durchgeführt werden, in denen Patienten, die im Vergleichsarm auf Valoctocogen Roxaparvovec gewechselt haben und in der Primäranalyse nicht der Interventionsgruppe zugeordnet werden, zum Zeitpunkt des Wechsels zensiert werden.

Indexdatum

Der pU legt das Indexdatum, d. h. das Datum für den Beobachtungsstart, für den Interventions- und den Vergleichsarm unterschiedlich fest. Im Vergleichsarm wird als Indexdatum für alle Endpunkte das Datum der schriftlichen Einwilligung zugrunde gelegt. Für den Interventionsarm wird das Datum der Applikation von Valoctocogen Roxaparvovec als Indexdatum festgelegt. Dieses Vorgehen ist sachgerecht. Wie beschrieben, werden Patienten, die im Rekrutierungszeitraum zunächst mit der Vergleichstherapie behandelt wurden und in diesem Zeitraum auf eine Therapie mit Valoctocogen Roxaparvovec wechseln, dem Interventionsarm zugeordnet. In diesen Fällen müssen die Basischarakteristika erneut zum Indexdatum (dann Datum der Applikation von Valoctocogen Roxaparvovec) erhoben werden. Eine Übertragung der Basischarakteristika vom Datum der schriftlichen Einwilligung ist nicht sachgerecht.

In der Literatur sind Verfahren beschrieben, wie mit der Situation umgegangen werden kann, wenn nicht in beiden Behandlungsgruppen zum Beobachtungsstart mit einer neuen Therapie begonnen wird, wie beispielsweise das Prevalent new User Design [20,21]. Ziel dieser Verfahren ist es, mögliche Verzerrungen durch eine falsche Wahl des Beobachtungsstarts zu reduzieren. Statistische Methoden zur Auswertung von Daten aus diesem Studiendesign berücksichtigen Daten, die vor dem Index-Datum liegen.

Methoden für die statistische Auswertung sind in der Literatur beschrieben, sodass die Auswertungen als Sensitivitätsanalysen gerechnet werden sollten. Dies ist entsprechend im SP und SAP zu beschreiben. Dabei ist aber zu beachten, dass fortlaufend während der Studie potenzielle zeitabhängige Confounder erhoben werden müssen.

Der pU wählt darüber hinaus endpunktabhängig unterschiedliche Zeitpunkte für den Beobachtungsstart. Für den Interventionsarm unterscheidet der pU zwischen der Erhebung der Blutungen sowie weiteren Endpunkten. Während die Erhebung der Blutungen im Interventionsarm erst 5 Wochen nach der Applikation von Valoctocogen Roxaparvovec, 3 Tage nach der letzten prophylaktischen Gabe eines Faktor-VIII-Präparats oder 27 Wochen nach der letzten prophylaktischen Gabe von Emicizumab beginnt, je nachdem was später auftrat, erfolgt die Erhebung für die weiteren Endpunkte dagegen ab dem Indexdatum. Der pU begründet den späteren Start der Erhebung von Blutungen im Interventionsarm damit, dass

nach Applikation der Gentherapie eine Art Brückentherapie bis zur ausreichenden endogenen Produktion von Faktor VIII verabreicht werden muss. Diese Brückentherapie ist jedoch als Teil der Intervention zu betrachten. Insofern ist das geplante Vorgehen des pU nicht sachgerecht. Im Vergleichsarm fällt der Beobachtungsbeginn für alle Endpunkte dagegen einheitlich auf das Indexdatum. Die Erhebung aller Endpunkte sollte für beide Arme einheitlich ab dem Indexdatum erfolgen.

2.2.5 Studienpopulation der Registerstudie

2.2.5.1 Fallzahlplanung

Der pU beschreibt im SAP entsprechend des Beschlusses [1] eine Fallzahlberechnung für den Morbiditätsendpunkt Blutungen, operationalisiert als ABR für behandelte Blutungen. Die Beschreibung des pU zur Fallzahlplanung ist im SAP nachvollziehbar und ausreichend beschrieben. Bei der Fallzahlplanung muss jedoch berücksichtigt werden, dass Emicizumab für die AbD zwar als geeigneter Komparator angesehen wird, jedoch noch nicht absehbar ist, ob Emicizumab auch Bestandteil der zweckmäßigen Vergleichstherapie sein wird. Es muss somit sichergestellt sein, dass für die AbD eine ausreichende Anzahl an Patienten unter einer Faktor-VIII-Therapie rekrutiert werden. Die geschätzte Fallzahl muss sich in diesem Fall auf die Patienten unter einer Faktor-VIII-Therapie beziehen.

Die Angaben zur Fallzahlplanung sind zwischen SP und SAP inkonsistent. Entgegen der Angaben im SAP beschreibt der pU im SP, dass der zugrunde liegende Test zur Fallzahlschätzung einem t-Test entspricht. Da für den primären Endpunkt eine Abhängigkeitsstruktur der rekurrent auftretenden Ereignisse angenommen wird, sind die Angaben zur Methodik der Fallzahlschätzung im SP nicht korrekt. Der Abschnitt im SP ist vom pU gemäß den Angaben im SAP zu korrigieren.

Weiterhin beschreibt der pU, dass die zur Testung von primären und sekundären Endpunkten verwendete Hypothesengrenze nicht verschoben wird. Aus einer nicht randomisierten Studie kann jedoch 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 [22]. Die konkrete Schwelle ergibt sich durch die Qualität der Daten, u. a. durch das Wissen über und die Berücksichtigung der wichtigen Confounder. Für die Auswertung und Ergebnisinterpretation der Daten ist daher eine verschobene Hypothesengrenze von 0,2 bis 0,5 in Abhängigkeit von der Qualität der Datenerhebung und -auswertung heranzuziehen.

2.2.5.2 Auswertungspopulationen

Der pU beschreibt im SAP, dass die Auswertungspopulation gegebenenfalls von der gesamten Studienpopulation abweichen kann, da er nur Patienten in die Analysen einbeziehen möchte, zu denen vollständige Daten zu allen Variablen vorliegen, die in den PS einfließen. Dieses

Vorgehen ist nicht sachgerecht. Insbesondere vor dem Hintergrund, dass der pU die Confounder, für die im PS adjustiert werden soll, erst zur Zwischenanalyse nach 18 Monaten festlegen möchte und keine verpflichtende Meldung zu relevanten Datenfeldern vorgibt, besteht ein hohes Risiko, dass die Daten vieler Patienten aus der Gesamtpopulation nicht in der Auswertungspopulation berücksichtigt werden. Der pU sollte einerseits Maßnahmen etablieren, um die Vollständigkeit der Daten für jeden Patienten sicherzustellen und andererseits klare Kriterien festlegen, wie er mit fehlenden Werten in den Analysen umgeht (siehe Abschnitt 2.2.9.2).

Darüber hinaus merkt der pU an, dass die behandelnde Ärztin oder der behandelnde Arzt auf Basis ihrer klinischen Einschätzung die Studienteilnahme von Patienten beenden können. Die betreffenden Patienten würden damit aus der weiteren Beobachtung ausgeschlossen werden, was zu einer deutlichen Verzerrung führen kann. Dieses Vorgehen ist nicht nachvollziehbar. Anhand welcher Kriterien die Patienten aus der Studie und somit aus der Beobachtung entfernt werden, sollte klar definiert und a priori festgelegt werden.

2.2.6 Identifikation und Erhebung von Confoundern, Adjustierung für Confounder

Einen zentralen Aspekt bei vergleichenden Studien ohne Randomisierung stellt die adäquate Adjustierung für Confounder dar, um interpretierbare Schätzungen des interessierenden Effekts zu erhalten. Für eine adäquate Kontrolle für Confounder ist es daher unabhängig von der verwendeten Methodik zur Adjustierung notwendig [23,24],

- vorab alle wichtigen Confounder (inklusive wichtiger Wechselwirkungen) zu identifizieren und diese wichtigen Confounder in der Registerstudie vollständig zu erheben sowie
- die wichtigen Confounder in angemessener Form bei der Auswertung im Modell zu berücksichtigen.

2.2.6.1 Identifikation und Erhebung wichtiger Confounder

Der pU beschreibt in Anhang 2 zum SP, dass er relevante Confounder auf Basis der Patientencharakteristika der 1-armigen Beobachtungsstudie 270-902 sowie der 1-armigen Studie 270-301 identifiziert. Für die Auswahl potenziell relevanter Confounder bindet er klinische Sachverständige ein. Eine Beurteilung der Durchführung der Informationsbeschaffung des pU zur Identifikation der Confounder findet sich in Anhang A.

Die Durchführung einer systematischen Literaturrecherche plant der pU erst zur Zwischenanalyse nach 18 Monaten. Die dann identifizierten relevanten Confounder sollen mit den erhobenen Variablen im DHR abgeglichen werden. Darüber hinaus plant der pU sich bei der Literaturrecherche vor allem auf in Deutschland erhobene Daten zu fokussieren.

Durch das Vorgehen des pU ist nicht sichergestellt, dass bis zu Studienbeginn in Q1/2024 eine ausreichend vollständige Liste an potenziell relevanten Confoundern identifiziert wird. In der Liste potenzieller Confounder ist beispielsweise die Region enthalten, die in der vorliegenden AbD keine Relevanz besitzt, da die AbD das DHR als primäre Datenquelle nutzt und somit keine unterschiedlichen Regionen vorliegen. Auch eine Fokussierung bei der Literaturrecherche auf in Deutschland erhobene Daten ist nicht sachgerecht. Darüber hinaus bestehen bereits Unterschiede in den vom DHR und vom pU in Anhang 2 des SP identifizierten Confoundern. Das DHR hat im Rahmen der Konzepterstellung als potenzielle Confounder Alter, Körpergewicht, Blutungsdauer und -häufigkeit, Therapieanlass, zugrunde liegende Mutation, Komorbiditäten, Familienhistorie sowie medizinisch relevante Ereignisse benannt [3]. Der pU benennt Alter, BMI, Region, Zahl der betroffenen Gelenke, die jährliche Blutungsrate zu Baseline, den Faktor-VIII-Verbrauch zu Baseline sowie die Art der verwendeten Faktor-VIII-Präparate. Zusätzlich merkt der pU im SP unter dem Abschnitt der Limitationen der Beobachtungsstudie an, dass individuelle Lebensführungen der Patienten, beispielsweise die körperliche Aktivität, nicht abgebildet werden können und dies zu Verzerrungen führen kann. Der pU muss sicherstellen, dass er alle relevanten Faktoren erhebt. Sollte eine Erhebung bestimmter Parameter nicht möglich sein, ist diese Unsicherheit in der Interpretation der Ergebnisse zu berücksichtigen.

Durch das Vorgehen des pU die systematische Literaturrecherche erst zur Zwischenanalyse nach 18 Monaten durchzuführen und dann abzugleichen, ob das DHR Daten zu den dann identifizierten Confoundern erhebt, besteht die Gefahr, dass Daten zu wichtigen Confoundern nicht im DHR erhoben werden und somit nicht angemessen in der Analyse berücksichtigt werden können.

Der pU plant zudem, aufgrund der zu erwartenden geringen Anzahl von Patienten im Interventionsarm, nur 6 Confounder bei der Schätzung des PS zu berücksichtigen. Dies ist nicht sachgerecht. In der Regel müssen alle wichtigen Confounder im Regressionsmodell zur Schätzung der PS berücksichtigt werden, da die Balanciertheit für alle wichtigen Confounder erreicht werden soll. Für den Fall, dass das Regressionsmodell nicht konvergiert, soll der pU eine angemessene Strategie festlegen. Die Ergebnisse sind diesbezüglich zu diskutieren und ggf. bei der Interpretation der Ergebnisse zu berücksichtigen, dies gilt insbesondere auch für die Balanciertheit der nicht im Modell berücksichtigten Variablen.

Insgesamt kann aufgrund der unzureichenden Informationsbeschaffung zu Confoundern nicht von einer sachgerechten Confounderliste für die AbD ausgegangen werden. Es erscheint daher erforderlich, den Abschnitt zur Identifikation und Festlegung der Confounder im Studienprotokoll grundlegend unter Berücksichtigung der oben genannten Punkte zu überarbeiten. Darüber hinaus sollten die Datenfelder zu den Confoundern spätestens mit Start der AbD verpflichtend im DHR zu dokumentieren sein.

2.2.6.2 Berücksichtigung wichtiger Confounder bei der Analyse

Eine hinreichende Prä-Spezifikation des Vorgehens zur Adjustierung für Confounder ist für die Auswertung von vergleichenden Studien ohne Randomisierung von besonderer Bedeutung.

Auswahl des Verfahrens für die Confounderadjustierung

Als Hauptanalyse wählt der pU zur Confounderadjustierung das Gewichtungsverfahren des Standardized Mortality Ratio Weighting (SMRW-Gewichtung), wobei Patienten, welche mit Valoctocogen Roxaparvovec behandelt wurden, jeweils ein Gewicht von 1 erhalten. Patienten im Vergleichsarm werden so gewichtet, dass ihre Confounderverteilung der Population im Interventionsarm entspricht. Ausreichende Angaben zum zu verwendenden Varianzschätzer macht der pU nicht. Das Verfahren der SMR-Gewichtung ist für eine AbD nicht angemessen. Der entsprechende Schätzer ist der Average Treatment Effect in Treated [ATT], welcher explizit den Effekt in der Population beschreibt, aus der die Interventionsgruppe stammt [25]. Bei der Planung einer nicht randomisierten Studie sollte die Zielpopulation entsprechend einer Target Trial Emulation nach klaren, für beide Behandlungsgruppen geltenden Ein- und Ausschlusskriterien rekrutiert werden, sodass für die Analysen eine möglichst repräsentative Patientenpopulation (=Auswertepopulation) herangezogen wird [26]. Der entsprechende Effekt in der Zielpopulation entspricht dem Average Treatment Effect (ATE). Der ATE kann bspw. mittels Inverse Probability of Treatment Weighting (IPTW) geschätzt werden. Der SP und der SAP sind entsprechend anzupassen.

Der Umgang mit extremen Gewichten ist vom pU nicht eindeutig beschrieben. Zur Identifizierung von Extremwerten verweist er auf sein Vorgehen zur Identifizierung von Extremwerten bei beobachteten Baselinecharakteristika und Endpunkten im SAP. Falls extreme Gewichte beobachtet werden, so sollen die Charakteristika der Patienten, die in die PS-Schätzung eingehen, auf Plausibilität untersucht werden.

Des Weiteren beschreibt der pU ein Verfahren zum Umgang mit extremen Gewichten in der SMRW-Analyse („Winsorization“ bzw. „Trimming“). Hierbei werden jedoch nur extreme Gewichte im Kontrollarm berücksichtigt und Werte größer als das 95 %-Perzentil der Verteilung bei den Kontrollen auf eben dieses 95 %-Perzentil gesetzt werden. Die Gewichte im Interventionsarm werden für alle Patienten auf 1 gesetzt. Das vom pU beschriebene Verfahren führt jedoch dazu, dass unabhängig von der Beobachtung von extremen Werten ein Trimming durchgeführt wird.

Als Sensitivitätsanalyse plant der pU die Confounderadjustierung mittels PS-Matchings. Es soll ein „greedy“ PS-Matching mit zufälliger Zuordnung zunächst im Verhältnis 1:1 (je nach beobachteten Daten kann das Verhältnis angepasst werden bis 1:5) ohne Zurücklegen mit einem Caliper-Abstand von 0,2 der gepoolten Standardabweichung des Logit des PS durchgeführt werden. Als Sensitivitätsanalyse ist eine Analyse mit einem Caliper-Abstand von

0,01 geplant. Das Verfahren des PS-Matching ist ausreichend beschrieben. Allerdings ist sowohl im SP als auch im SAP unklar, welches die ausschlaggebende Sensitivitätsanalyse ist, da verschiedene Verhältnisse des Matchings (SP: 1:1 bis zu 1:4, SAP: bis zu 1:5) angegeben sind. Die primäre Sensitivitätsanalyse sollte möglichst viele Patienten berücksichtigen. Es ist jedoch fraglich, ob die primäre Verwendung eines 1:1, wie im SP beschrieben, bei einem erwarteten Verhältnis zwischen Intervention und Kontrolle von 1:5 [1] adäquat ist, da dadurch der Stichprobenumfang stark eingeschränkt wird.

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. Es fehlen Kriterien, wann diese Untersuchungen zu einer ausreichenden Überlappung führen.

Zur Bewertung der Balanciertheit wird die standardisierte Mittelwertdifferenz (SMD) für die einzelnen Confounder herangezogen. Dabei wird von einer ausreichenden Balanciertheit ausgegangen, wenn bei allen Confoundern ein Unterschied von höchstens 0,25 SMD festgestellt wird. Die in der Literatur vorgeschlagenen akzeptablen Grenzen für eine ausreichende Balanciertheit liegen im Bereich zwischen 0,1 und 0,25 [27]. Die vom pU vorgeschlagene Grenze von 0,25 ist zwar sachgerecht, stellt allerdings das schwächste Kriterium für eine ausreichende Balanciertheit dar. Dies ist bei der Interpretation der Ergebnisse auf Basis einer verschobenen Hypothesengrenze entsprechend zu berücksichtigen [22].

Das PS-Matching ist prinzipiell als Sensitivitätsanalyse geeignet. Patienten, die nicht gematcht werden, werden in der Analyse nicht berücksichtigt. Die Effektschätzung ist daher eine Aussage über eine Teilpopulation der Gesamtpopulation und kann daher mit der ATE-Schätzung verglichen werden.

Weiterhin beschreibt der pU, dass das PS-Matching dann als Hauptanalyse herangezogen wird, wenn in Folge der SMRW-Gewichtung Gewichte getrimmt werden müssen (siehe oben) und beim PS-Matching mindestens 95 % der Interventionsgruppe jeweils mindestens 3 Kontrollpatienten zugeteilt bekommen. Bei einer erwarteten Verteilung von 1:5 Patienten (Intervention : Kontrolle) [1] ist es daher wahrscheinlich, dass das vom pU geplante Vorgehen dazu führt, das PS-Matching zur primäre Analyse wird. Sofern der pU ein Gewichtungsverfahren als primäre Analyse verwenden will, ist das von ihm beschriebene Verfahren nicht geeignet und muss angepasst werden.

Der pU gibt im SAP ein mehrschrittiges Verfahren an, um mit Imbalance umzugehen. Dazu gehört die Transformationen der Ausgangsvariablen und das Hinzufügen von Interaktionstermen. Bei bleibender Imbalance soll diese vermutlich mittels Double-Matching [28,29] berücksichtigt werden. Hierbei werden mögliche Confounder, bei denen auch nach der Gewichtung Imbalancen zwischen den Behandlungsgruppen vorliegen, mit in das

Regressionsmodell zur Effektschätzung berücksichtigt. Der pU beschreibt jedoch weder die Methodik noch zitiert er geeignete Literatur. Auch diskutiert er nicht, warum mit diesem Ansatz in der vorliegenden Situation angemessen für Imbalance korrigiert werden kann. Des Weiteren spezifiziert der pU nicht, nach welchen Kriterien die Modellauswahl für die finalen Analysen vorgenommen wird.

Gemäß Beschluss des G-BA soll der pU eine klare Hierarchie der infrage kommenden PS-Verfahren und die Prüfkriterien zur Auswahl der robustesten Methoden festlegen [1]. Dabei muss die Darstellung der Prüfkriterien konkrete Angaben zur Überprüfung einer ausreichenden Überlappung und Balanciertheit enthalten. Die vorgelegten Angaben genügen dieser Anforderung nicht. Der pU beschreibt weder eine Hierarchie der PS-Verfahren, noch liegen ausreichende Angaben zur Überlappungs- und Balanciertheitsprüfung vor. Die Angabe einer Hauptanalyse mittels Gewichtung und einer Sensitivitätsanalyse mit Matching entspricht nicht einem hierarchischen Vorgehen. Der pU plant naive Analysen darzustellen und diese im Weiteren mit den confounderadjustierten Analysen (mittels PS) zu vergleichen. Hierbei bleibt der Entscheidungsalgorithmus des pU unklar. Konsequenzen für die Ergebnisinterpretation eines naiven Vergleiches beschreibt der pU ebenfalls nicht.

Weiterhin ist zu ergänzen, welche Konsequenz gezogen wird, wenn kein PS-Verfahren gefunden werden kann, mit dem eine ausreichende Überlappung und Balanciertheit der zu vergleichenden Gruppen erreicht werden kann.

Der pU beschreibt im SAP, dass er den PS für den primären Endpunkt, jedoch nicht für andere sekundäre Endpunkte schätzt. Dieser Abschnitt ist nicht verständlich, da der PS die Wahrscheinlichkeit darstellt, mit der ein Patient die Intervention erhält, bedingt auf die Verteilung unter den Confoundern. Somit ist die Schätzung der PS unabhängig von den betrachteten Endpunkten.

Sensitivitätsanalysen

Wie im Vorfeld beschrieben, gibt der pU an, dass er Sensitivitätsanalysen für den primären Endpunkt durchführen wird. Sensitivitätsanalysen für weitere Endpunkte werden nur in Abhängigkeit von dem Unterschied in den Analysen des primären Endpunkts durchgeführt. Das Vorgehen des pU ist nicht vollständig, da Sensitivitätsanalysen nicht auf den primären Endpunkt eingeschränkt werden sollten, sondern die Ergebnisse aller patientenrelevanten Endpunkte bezüglich ihrer Robustheit geprüft werden sollten.

Beschreibung der Zielpopulation nach PS-Verfahren

Ein zentraler Aspekt bei nicht randomisierten Vergleichen mit Confounderadjustierung ist die detaillierte Beschreibung der Zielpopulation, die sich aus der Anwendung eines PS-Verfahrens ergibt. Die durch das PS-Verfahren gebildete Patientenpopulation (= Analysepopulation) muss repräsentativ für die jeweilige Fragestellung sein. Sollten relevante Unterschiede zwischen der

Analysepopulation und der ursprünglichen Gesamtpopulation erkennbar sein, ist anzugeben auf welche Teilpopulation sich die Ergebnisse der durchgeführten Analysen beziehen. Eine detaillierte Beschreibung des pU bezüglich der Prüfung der Analysepopulation und Zielpopulation fehlen in den vorgelegten Dokumenten. Dies ist vom pU zu überarbeiten.

2.2.7 Anmerkungen zu den geplanten Auswertungen für die Endpunkte der AbD

Im SP und SAP sind zu den verschiedenen Endpunkten jeweils endpunktspezifische Ausführungen zu den geplanten Auswertungen enthalten. Insgesamt ist anzumerken, dass für binäre Endpunkte in Nutzenbewertungen bei vergleichbaren Beobachtungsdauern das relative Risiko als Effektmaß vorzuziehen ist. Dies sollte im SAP und SP vom pU berücksichtigt werden. Weitere Aspekte der geplanten Auswertungen werden nachfolgend adressiert.

2.2.7.1 Auswertung des Endpunkts Gesamtmortalität

Der pU plant den Endpunkt Gesamtmortalität im Rahmen der UEs zu erheben und auszuwerten. Dieses Vorgehen des pU ist sachgerecht. Eine entsprechende Kommentierung bezüglich der geplanten Analysen ist dem Abschnitt 2.2.7.4 zu entnehmen.

2.2.7.2 Auswertung des Endpunkts Blutungen

Der pU plant, den Endpunkt Blutungen, operationalisiert über die ABR, mittels eines gewichteten Negativ-binomial-Modelles auszuwerten. Die Angaben des pU sind nicht vollständig. Die Auswertung soll alle Ereignisse die zwischen Indexdatum und Studienende beobachtet werden, berücksichtigen. Aufgrund individueller Beobachtungszeiten ist die Spezifizierung eines Offsets im Modell notwendig.

Der pU beschreibt weiter ein Zero-inflated Negativ-binomial-Modell. Allerdings bleibt die Anwendung dieses Modells im SAP unklar. Einerseits wird beschrieben, dass das Modell zur Anwendung kommen soll, wenn es „zu viele“ Patienten ohne Blutungsereignisse gibt. Ab welcher Grenze der pU von zu vielen Patienten ausgeht, spezifiziert er nicht. An einer anderen Stelle beschreibt der pU, dass dieses Modell zum Einsatz kommt, wenn die primäre Analyse (gewichtetes Negativ-binomial-Modell) nicht konvergiert. Anhand eines Vergleichs der Akaike Information Criterion (AIC) der 2 Modelle soll entschieden werden, welches Ergebnis herangezogen wird. Dieses Vorgehen ist unverständlich, da aus dem primären Modell keine validen Ergebnisse vorliegen. Die Anwendung des Zero-inflated Negativ-binomial-Modells ist eindeutig zu spezifizieren.

Responderanalysen

Für die Auswertung der Endpunkte zu Blutungen als Anteil der Patienten mit Blutungsfreiheit beschreibt der pU eine Auswertung über den vollen Beobachtungszeitraum. Für die statistische Auswertung gibt es unterschiedliche Angaben im SP und SAP. Im SP beschreibt der pU als Effektmaß die absolute Risikodifferenz, während im SAP das Effektmaß Odds Ratio (OR)

dargestellt ist. Beide Effektmaße sind nicht sachgerecht, da für Nutzenbewertungen das RR heranzuziehen ist. Das Ergebnis ist unter Berücksichtigung der unterschiedlichen individuellen Beobachtungszeiten und den Zensierungsgründen zu diskutieren.

2.2.7.3 Auswertung der PRO-Endpunkte (Schmerzen und gesundheitsbezogene Lebensqualität)

Der pU plant, den Haemo-QoL-A und den BPI-SF als stetige Daten auszuwerten. Dies plant er mittels generalisierter gemischter Modelle für Messwiederholungen. Diese Modelle sind jedoch detailliert zu beschreiben, insbesondere das Effektmaß. Darüber hinaus gibt der pU an, Responderanalysen für die Verbesserung bzw. Verschlechterung in Betracht zu ziehen. Gemäß dem Methodenpapier des IQWiG werden für patientenberichtete Endpunkte Responderanalysen bevorzugt. Wie in den Allgemeinen Methoden des Instituts [30,31] erläutert, sollte ein Responsekriterium, damit es hinreichend sicher eine für die Patienten spürbare Veränderung abbildet, prädefiniert mindestens 15 % der Skalenspannweite eines Instruments entsprechen. Für diese Art der Auswertung sind seltenere Erhebungen (wie die vom pU geplante 2-mal jährliche Erhebung) zunächst ausreichend. Allerdings sind Informationen zum Studienverlauf, bspw. in Form von Verlaufskurven für die Interpretation des Ergebnisses der Responderanalyse erforderlich.

2.2.7.4 Auswertung der UE-Endpunkte

Es ist für alle UE-Endpunkte geplant, die jeweiligen Endpunkte nur dann vergleichend zu analysieren, wenn mindestens 10 Ereignisse in beiden Studienarmen auftreten. Dies ist nicht sachgerecht. Auswertungen aller patientenrelevanten Endpunkte müssen für die Nutzenbewertung gemäß den Vorgaben in den Dossievorlagen unabhängig von der Anzahl der aufgetretenen Ereignisse vorgelegt werden (siehe auch Abschnitt 2.2.2.3). Welche statistischen Verfahren er für den Vergleich der beiden Arme verwenden will, gibt der pU nicht an. Die vom pU geplanten Auswertungen sind nicht vollständig beschrieben. Um geringere und höhere Schäden abzubilden ist ein Vergleich der Ergebnisse beider Behandlungsarme, z. B. anhand des relativen Risikos, notwendig. Für den Fall, dass nur sehr wenige oder in einem Behandlungsarm keine Ereignisse beobachtet werden, sind entsprechende statistische Verfahren wie die Firth-Korrektur und Profile-Likelihood-Konfidenzintervalle zu verwenden. Im Sinne einer ITT-Auswertung sind alle Nebenwirkungen bis zum Studienende zu erheben und entsprechend auszuwerten. Die Zensierung von Ereignissen nach einem Therapiewechsel ist daher nicht sachgerecht.

2.2.8 Untersuchung auf potenzielle Effektmodifikatoren und Subgruppenanalysen

Der G-BA fordert im Beschluss, dass im SP und SAP Angaben zur Untersuchung potenzieller Effektmodifikatoren gemacht werden [1]. Der pU beschreibt 2 Subgruppenmerkmale, zum

einen die Prophylaxe-Therapie und zum anderen den AAV5 Status (AAV5 Status bekannt vs. unbekannt) der Patienten.

Im Beschluss wird ausgeführt, dass Emicizumab für die AbD zwar als geeigneter Komparator angesehen wird, aber noch nicht feststeht, ob Emicizumab für die Nutzenbewertung Teil der zweckmäßigen Vergleichstherapie sein wird [6]. Aus diesen Gründen sollten geeignete Maßnahmen etabliert und im SP dargelegt werden, über die sichergestellt werden kann, dass ausreichend Patienten, die mit Blutgerinnungsfaktor-VIII-Präparaten behandelt werden, in die AbD eingeschlossen werden. In diesem Zuge sollten auch Sensitivitätsanalysen für die getrennte Auswertung der Daten zu Valoctocogen Roxaparvovec gegenüber den Daten zu rekombinanten oder aus humanem Plasma gewonnenen Blutgerinnungsfaktor-VIII-Präparaten als auch den Daten zu Emicizumab im SP und SAP präspezifiziert werden. Der pU beschreibt jedoch lediglich, dass er die verschiedenen Therapien innerhalb des Vergleichsarms mit Valoctocogen Roxaparvovec vergleichen will. Eine gemeinsame Analyse der Faktor-Präparate gegen Valoctocogen Roxaparvovec wird nicht beschrieben und sollte ergänzt werden.

SP und SAP sollten entsprechend angepasst werden. Weiterhin sollten die unterschiedlichen Angaben bezüglich der Betrachtung des Merkmals AAV5 Status zwischen SAP (Subgruppenanalyse) und SP (Sensitivitätsanalyse) vereinheitlicht werden. Für die Sensitivitätsanalyse zum AAV5-Status ist darüber hinaus erforderlich die Operationalisierung im DHR anzupassen (Spezifikation des Virustyps, gegen den Antikörper vorliegen) und den AAV5-Status, wenn erhoben, für Patienten beider Arme zu erfassen.

Der pU beschreibt nicht, ob und wie statistische Tests auf Subgruppenunterschiede durchgeführt werden.

Für die Nutzenbewertung gemäß der Vorgaben in den Dossiervorlagen sollte darüber hinaus geprüft werden, ob Subgruppenanalysen für weitere relevante Merkmale durchgeführt werden sollten. Dies betrifft insbesondere Analysen zu Merkmalen für die Krankheitsschwere sowie das Alter.

2.2.9 Qualität und Vollständigkeit der Datenerhebung

2.2.9.1 Maßnahmen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung

Zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung soll gemäß Studienprotokoll eine Source Data Verification für geschätzt 100 % der Datenfelder für die Ein- und Ausschlusskriterien und den primären Endpunkt durchgeführt werden (siehe Abschnitt 2.2.3). Für weitere Angaben verweist er auf den Study Monitoring Plan (SMP), der jedoch nicht vorliegt. Für die AbD wäre eine Source Data Verification anhand einer Stichprobe

von z. B. 5 % oder 10 % der Datensätze sinnvoll [1]. Die Überprüfung kann sich mit Blick auf die AbD auf die für die AbD relevanten Datenfelder beschränken. Neben der Source Data Verification plant der pU automatisierte Prüfungen sowie manuelle Prüfungen durch das Prüfpersonal beim routinemäßigen Monitoring der Prüfzentren.

Spezifische Maßnahmen zur Sicherstellung der Vollständigkeit der Datenerhebungen benennt der pU nicht. Wie bereits beschrieben, sollte der pU durch die Vorgabe verpflichtender Datenfelder die Vollständigkeit der Datenerhebungen erhöhen (siehe auch Abschnitt 2.2.3).

2.2.9.2 Umgang mit fehlenden Daten

Gemäß des G-BA-Beschlusses zur AbD sind Angaben zum erwarteten Umfang und den Gründen für fehlende Daten, sowie Maßnahmen zur Vermeidung fehlender Daten und Auswertungsstrategien zum Umgang mit fehlenden Daten darzulegen [1].

Zur möglichen Ersetzung fehlender Werte macht der pU widersprüchliche Angaben im SAP im Abschnitt zur PS-Schätzung (Abschnitt 7.8.1) und dem Abschnitt zu fehlenden Werten (Abschnitt 7.9.5.4). Der pU sollte zum einen sinnvolle Ersetzungsstrategien für fehlende Daten darlegen und die entsprechende Methodik präspezifizieren, zum anderen soll er spezifische Maßnahmen ergreifen, durch die eine möglichst vollständige Datenerhebung gewährleistet wird.

Fehlende Datumsangaben sollen wie folgt ersetzt werden: Sofern nur der Tag fehlt, wird der 15. des jeweiligen Monats angenommen. Sofern auch der Monat fehlt, soll der 1. Juli angenommen werden. Fehlen alle Angaben, so wird kein Datum ersetzt. Während die geplante Ersetzung des fehlenden Tages sachgerecht ist, führt die geplante Ersetzung des Monats potenziell zu erheblichen Verzerrungen und ist nicht sachgerecht. Es sind vielmehr entsprechende Anstrengungen erforderlich, um den Anteil fehlender Werte bei Datumsangaben zu minimieren.

Insbesondere problematisch ist, dass der pU keine finale Confounder-Selektion vorlegt (siehe Abschnitt 2.2.6) und für die Auswertungspopulation nur Patienten mit vollständigen Angaben zu den im PS-Verfahren berücksichtigten Variablen berücksichtigen will (siehe Abschnitt 2.2.5.2). Durch dieses Vorgehen besteht das Risiko die Zahl der Patienten in der Auswertungspopulation drastisch zu reduzieren.

2.2.10 Anpassung der AbD

Estimand

Der pU bestimmt als primären Estimand der Studie den Hypothetical Estimand, ohne den Estimand ausreichend zu spezifizieren [32]. Im Rahmen einer AbD sind jedoch Estimands nach dem Hypothetical – Ansatz nicht angemessen. Der primäre Estimand einer AbD entspricht der

Treatment Policy Strategy und beinhaltet insbesondere die Auswertung nach dem ITT-Prinzip für alle patientenrelevanten Endpunkte. Der pU sollte daher sowohl SP als auch SAP entsprechend anpassen, dies betrifft insbesondere die Punkte zum Indexdatum und Start des Follow-up als auch den Umgang mit Behandlungswechslern.

Abbruchkriterien

Gemäß G-BA-Beschluss zur AbD sind im Studienprotokoll und SAP Angaben zu Abbruchkriterien wegen Vergeblichkeit darzulegen [1]. Dazu finden sich keine ausreichenden Angaben in den vom pU vorgelegten Dokumenten. Die Prüfung auf Vergeblichkeit sollte in der Gesamtschau aller Daten erfolgen. Weder die isolierte Betrachtung der zu einem Zeitpunkt erreichten Fallzahl, noch die alleinige Betrachtung der ABR ist dabei sachgerecht. Die genauen Abbruchkriterien sollten ergänzt werden. Dabei ist darauf zu achten, dass eine etwaige Entscheidung für einen Abbruch der AbD sowie für eine Änderung der Fallzahlschätzung in Abstimmung mit dem G-BA erfolgen sollte (siehe dazu auch das Studienprotokoll zur AbD zu Onasemogen-Abeparvovec sowie die Ausführungen im Addendum zur Überprüfung dieses Studienprotokolls [33,34]).

2.2.11 Interpretation der Ergebnisse der AbD

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 [22]. 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 (siehe Abschnitt 2.2.5.2).

2.2.12 Zeitplan für die AbD

Der G-BA fordert Auswertungen zu Zwischenanalysen nach 18, 36 und 54 Monaten nach Beschlussdatum. Der pU plant zusätzlich noch eine Zwischenanalyse nach 6 Monaten. Zu jeder der Zwischenanalyse plant der pU auch eine Prüfung auf Abbruch wegen Vergeblichkeit durchzuführen. Die Rationale der Durchführung einer zusätzlichen Zwischenanalyse nach 6 Monaten und der damit verbundenen Prüfung auf Vergeblichkeit erschließt sich aus den Angaben im SP und SAP nicht. Es ist fraglich, ob nach 6 Monaten bereits ausreichend Informationen vorliegen, um eine Prüfung auf Vergeblichkeit durchzuführen.

2.3 Zusammenfassung und Fazit

Die Prüfung des SP (Version vom 29.06.2023) und SAP (Version vom 29.06.2023) für die AbD zu Valoctocogen Roxaparvovec hat ergeben, dass die Anforderungen des G-BA zur Durchführung der AbD einschließlich der zugehörigen Auswertung der Daten zum Teil, aber nicht vollständig umgesetzt wurden.

Eine Anpassung ist insbesondere in folgenden Bereichen erforderlich:

- A-priori-Festlegung aller Definitionen, Endpunkte, Operationalisierungen und Endpunkte (siehe Abschnitt 2.2.1), auch in Bezug auf die Datenfelder im DHR (siehe Abschnitt 2.2.3)
- Maßnahmen, über die sichergestellt werden kann, dass ausreichend Patienten mit einer Faktor-VIII-Prophylaxe in die AbD eingeschlossen werden (siehe Abschnitt 2.2.2.1 und 2.2.5.1)
- Erhebung aller Endpunkte ab Indexdatum im Interventions- und Vergleichsarm (siehe Abschnitt 2.2.4)
- Angaben zum Umgang mit Behandlungswechslern vom Vergleichsarm in den Interventionsarm während und kurz nach Beendigung des Rekrutierungszeitraums (siehe Abschnitt 2.2.4)
- Verpflichtende Erhebung aller relevanten Daten im DHR (siehe Abschnitt 2.2.3)
- Auslegung der Fallzahlplanung auf den Vergleich von Valoctocogen Roxaparvovec mit Faktor-VIII-Präparaten (siehe Abschnitt 2.2.5.1)
- Adjustierung für Confounder: Festlegung und Erhebung der relevanten Confounder (siehe Abschnitt 2.2.6.1) und detaillierte Beschreibung des konkreten Vorgehens bei der Adjustierung für Confounder (siehe Abschnitt 2.2.6.2)
- Berücksichtigung aller UE-Ereignisse in den Analysen (siehe Abschnitt 2.2.7.4)
- Definition und Beschreibung der Maßnahmen zur Sicherstellung der Vollständigkeit und Qualität der Datenerhebung, insbesondere in Bezug auf die Source Data Verification (siehe Abschnitt 2.2.9.1)
- detaillierte Beschreibung des Umgangs mit fehlenden Werten (siehe Abschnitt 2.2.9.2)
- Beschreibung der Abbruchkriterien im Rahmen der Prüfung auf Vergeblichkeit (siehe Abschnitt 2.3)

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. Biomarin. ROCTAVIAN 2 × 10¹³ Vektorgenome/ml Infusionslösung [online]. [Zugriff: 02.08.2023]. URL: <https://www.fachinfo.de/suche/fi/023888>.

8. Gemeinsamer Bundesausschuss. Zusammenfassende Dokumentation über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buch Sozialgesetzbuch (SGB V); Valoctocogen Roxaparvovec (schwere Hämophilie A); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2023 [Zugriff: 09.08.2023]. URL: https://www.g-ba.de/downloads/40-268-9251/2023-02-02_AM-RL-XII_Valoctocogen_Roxaparvovec_AbD-002_Forderung_ZD.pdf.
9. Deutsches Hämophilierregister. Gesamtdatensatz dhr 2.0 [online]. 2023 [Zugriff: 02.08.2023]. URL: https://www.pei.de/SharedDocs/Downloads/DE/regulation/meldung/dhr-deutsches-haemophilierregister/dhr-20-datensatz.pdf?__blob=publicationFile&v=10.
10. Deutsches Hämophilierregister. Handbuch, Version 2.3 [online]. 2023 [Zugriff: 07.08.2023]. URL: https://www.pei.de/SharedDocs/Downloads/DE/regulation/meldung/dhr-deutsches-haemophilierregister/dhr-20-handbuch.pdf?__blob=publicationFile&v=9.
11. Hernan MA, Sauer BC, Hernandez-Diaz S et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016; 79: 70-75. <https://dx.doi.org/10.1016/j.jclinepi.2016.04.014>.
12. Cleeland CS. The Brief Pain Inventory User Guide [online]. 2009 [Zugriff: 18.04.2023]. URL: https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf.
13. Humphries TJ, Kessler CM. The challenge of pain evaluation in haemophilia: can pain evaluation and quantification be improved by using pain instruments from other clinical situations? *Haemophilia* 2013; 19(2): 181-187. <https://dx.doi.org/10.1111/hae.12023>.
14. Santoro C, Di Minno MND, Corcione A et al. Improving assessment and management of pain in hemophilia: an Italian Delphi consensus statement. *Blood Rev* 2022; 51: 100885. <https://dx.doi.org/10.1016/j.blre.2021.100885>.
15. Feldman BM, Funk SM, Bergstrom BM et al. Validation of a new pediatric joint scoring system from the International Hemophilia Prophylaxis Study Group: validity of the hemophilia joint health score. *Arthritis Care Res (Hoboken)* 2011; 63(2): 223-230. <https://dx.doi.org/10.1002/acr.20353>.
16. St-Louis J, Abad A, Funk S et al. The Hemophilia Joint Health Score version 2.1 Validation in Adult Patients Study: A multicenter international study. *Res Pract Thromb Haemost* 2022; 6(2): e12690. <https://dx.doi.org/10.1002/rth2.12690>.
17. Rentz A, Flood E, Altisent C et al. Cross-cultural development and psychometric evaluation of a patient-reported health-related quality of life questionnaire for adults with haemophilia. *Haemophilia* 2008; 14(5): 1023-1034. <https://dx.doi.org/10.1111/j.1365-2516.2008.01812.x>.

18. Limperg PF, Terwee CB, Young NL et al. Health-related quality of life questionnaires in individuals with haemophilia: a systematic review of their measurement properties. *Haemophilia* 2017; 23(4): 497-510. <https://dx.doi.org/10.1111/hae.13197>.
19. Gilead Sciences. Anwendungsbegleitende Datenerhebung und Beschränkung der Versorgungsbefugnis: Brexucabtagen-Autoleucel/Autologe Anti-CD19-transduzierte CD3+ Zellen – Mantelzelllymphom: Studienunterlagen [online]. 2023 [Zugriff: 09.08.2023]. URL: https://www.g-ba.de/downloads/40-268-9661/2023-07-20_AM-RL-XII_BrexCel_2021-AbD-008_Ueberpruefung-SP-SAP-Start-AbD_Studienunterlagen.pdf.
20. Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017; 26(4): 459-468. <https://dx.doi.org/10.1002/pds.4107>.
21. Webster-Clark M, Ross RK, Lund JL. Initiator Types and the Causal Question of the Prevalent New-User Design: A Simulation Study. *Am J Epidemiol* 2021; 190(7): 1341-1348. <https://dx.doi.org/10.1093/aje/kwaa283>.
22. 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.
23. Berger ML, Mamdani M, Atkins D et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value Health* 2009; 12(8): 1044-1052. <https://dx.doi.org/10.1111/j.1524-4733.2009.00600.x>.
24. Cox E, Martin BC, Van Staa T et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value Health* 2009; 12(8): 1053-1061. <https://dx.doi.org/10.1111/j.1524-4733.2009.00601.x>.
25. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ* 2019; 367: l5657. <https://dx.doi.org/10.1136/bmj.l5657>.
26. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* 2016; 183(8): 758-764. <https://dx.doi.org/10.1093/aje/kwv254>.

27. Stuart EA, Lee BK, Leacy FP. Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *J Clin Epidemiol* 2013; 66(8 (Suppl)): S84-S90.e1. <https://dx.doi.org/10.1016/j.jclinepi.2013.01.013>.
28. Nguyen TL, Collins GS, Spence J et al. Comparison of the ability of double-robust estimators to correct bias in propensity score matching analysis. A Monte Carlo simulation study. *Pharmacoepidemiol Drug Saf* 2017; 26(12): 1513-1519. <https://dx.doi.org/10.1002/pds.4325>.
29. Austin PC. Double propensity-score adjustment: A solution to design bias or bias due to incomplete matching. *Stat Methods Med Res* 2017; 26(1): 201-222. <https://dx.doi.org/10.1177/0962280214543508>.
30. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 6.1 [online]. 2022 [Zugriff: 24.01.2022]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
31. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden 6.0 [online]. 2020 [Zugriff: 29.06.2022]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_dwa-entwurf-fuer-version-6-0_v1-0.pdf.
32. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials [online]. 2020 [Zugriff: 03.08.2023]. URL: <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline>.
33. Novartis Gene Therapies. Routine data collection and evaluations of onasemnogene abeparvovec in Germany; Study Protocol; Version 2.02 [online]. 2021 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/17-98-5254/2022-01-20_AM-RL_Onasemnogen-Abeparvovec_abD_Studienprotokoll.pdf.
34. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Anwendungsbegleitende Datenerhebung zu Onasemnogen-Abeparvovec: Prüfung des Studienprotokolls und des statistischen Analyseplans; 2. Addendum zum Auftrag A20-61 [online]. 2022 [Zugriff: 11.04.2022]. URL: https://www.iqwig.de/download/a21-149_abd-onasemnogen-abeparvovec-pruefung-protokoll-sap_addendum-zum-auftrag-a20-61_v1-0.pdf.

Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern

Informationsbeschaffung

Zur Identifizierung von relevanten Confoundern listet der pU 14 Baselinecharakteristika der 1-armigen Studie 270-902 (prophylactic FVIII replacement with CFCs) sowie der laufenden Folgestudien 270-301 (Valoctocogen Roxaparvovec) auf (zu finden im Appendix 2 zum SP). Nach Prüfung eines statistischen Zusammenhangs zwischen den potenziellen Confounder und der ABR, basierend auf behandelten Blutungen, und der Beratung durch 3 externe Studienberaterinnen oder Studienberater wurde eine Liste mit 6 Confoundern identifiziert.

Darüber hinaus gibt der pU an, vor der Zwischenanalyse nach 18 Monaten ein Literaturreview zu erstellen, das weitere potenzielle Confounder identifizieren soll. Diese sollen dann mit den Variablen verglichen werden, die derzeit im Deutschen Hämophileregister gesammelt werden. Zum Zwischenbericht soll dann die Liste der Confounder finalisiert werden.

Kommentar zur Informationsbeschaffung

Durch das Vorgehen des pU ist nicht sichergestellt, dass eine ausreichend vollständige Liste an potenziell relevanten Confoundern identifiziert wurde. Dies kann auch nicht durch die darauffolgende Experteneinschätzung ausgeglichen werden.

Es ist nicht bekannt, wie die 14 potenziellen Confounder für die beiden Studien 270-902 und 270-301 identifiziert wurden und ob diese auf einer systematischen Recherche unter Einbindung von Fach-Experten beruhen.

Der pU selektiert aus der Liste von 14 potenziellen Confounder erstmal solche, die in der Studienpopulation der Studie 270-902 einen statistisch signifikanten Zusammenhang mit der log-transformierten ABR zeigen. Dieses Vorgehen ist für eine Confounderadjustierung nicht angemessen, da hierbei für prognostische Faktoren unter der Kontrolle als auch unter der Intervention, sowie für Effektmodifikatoren adjustiert werden muss.

Auch der Hinweis auf ein Literaturreview, das vor der Zwischenanalyse nach 18 Monaten erstellt wird, ist nicht sachgerecht, da damit eine Präspezifikation von Confoundern bei der Studienplanung nicht gewährleistet ist.