

Anwendungsbegleitende Datenerhebung zu Etranacogen Dezaparvovec: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0)

2. Addendum zum Projekt A22-83

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: A23-99

Version: 1.0

Stand: 24.11.2023

IQWiG-Berichte – Nr. 1678

DOI: 10.60584/A23-99

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Anwendungsbegleitende Datenerhebung zu Etranacogen Dezaparovec: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0) – 2. Addendum zum Projekt A22-83

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

13.10.2023

Interne Projektnummer

A23-99

DOI-URL

<https://dx.doi.org/10.60584/A23-99>

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 Etranacogen Dezaparvovec: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0); 2. Addendum zum Projekt A22-83 [online]. 2023 [Zugriff: TT.MM.JJJJ]. URL: <https://dx.doi.org/10.60584/A23-99>.

Schlagwörter

Etranacogen Dezaparvovec, Hämophilie B, Nutzenbewertung, Register, Forschungsdesign, Peer Review – Forschung

Keywords

Etranacogene Dezaparvovec, Hemophilia B, Benefit Assessment, Registries, Research Design, Peer Review – Research

An dem Addendum beteiligte Mitarbeiterinnen und Mitarbeiter des IQWiG

- Christina Keksel
- Tatjana Hermanns
- Katharina Hirsch
- Katherine Rascher
- Ulrike Seay
- Volker Vervölgyi

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abkürzungsverzeichnis.....	vi
1 Hintergrund.....	1
2 Prüfung der Unterlagen zur Planung der AbD von Etranacogen Dezaparvovec	2
2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen	2
2.2 Spezifische Anmerkungen zum Studienprotokoll und SAP	2
2.2.1 Fragestellung gemäß PICO	2
2.2.1.1 Population.....	2
2.2.1.2 Intervention und Kontrolle	3
2.2.1.3 Endpunkte (Outcomes).....	3
2.2.2 Datenquelle	8
2.2.3 Rekrutierungs- und Beobachtungszeitraum, Zuordnung zu den Behandlungsgruppen und Indexdatum.....	9
2.2.4 Fallzahlplanung.....	11
2.2.5 Identifikation und Erhebung von Confoundern, Adjustierung für Confounder... 12	12
2.2.5.1 Identifikation und Erhebung wichtiger Confounder.....	12
2.2.5.2 Berücksichtigung wichtiger Confounder bei der Analyse.....	13
2.2.6 Anmerkungen zu den geplanten Auswertungen für die Endpunkte der AbD	15
2.2.7 Untersuchung auf potenzielle Effektmodifikatoren und Subgruppenanalysen ..	17
2.2.8 Qualität und Vollständigkeit der Datenerhebung.....	18
2.2.8.1 Maßnahmen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung.....	18
2.2.8.2 Umgang mit fehlenden Daten	18
2.2.9 Weitere Anpassungen der AbD	19
2.2.10 Interpretation der Ergebnisse der AbD	20
2.2.11 Zeitplan für die AbD.....	20
2.2.12 Sonstige Anmerkungen	21
2.3 Zusammenfassung und Fazit	22
3 Literatur	23
Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern ..	26

Tabellenverzeichnis

	Seite
Tabelle 1: Übersicht über die Zeitpunkte und Zeiträume der geforderten und geplanten Statusberichte und Zwischenanalysen	21

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AbD	Anwendungsbegleitende Datenerhebung
ABR	annualisierte Blutungsrate
AAV5	Adeno-assoziiertes Virus vom Serotyp 5
BPI-SF	Brief Pain Inventory – Short Form
CRO	Contract Research Organization (Auftragsforschungsinstitut)
DHR	Deutsches Hämophileregister
EPAR	European Public Assessment Report
FSW	Fine Stratification Weights
G-BA	Gemeinsamer Bundesausschuss
GLM	generalisiertes lineares Modell
Haemo-QoL-A	Hemophila-specific Quality of Life Questionnaire for Adults
HJHS	Haemophilia Joint Health Score
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IPTW	Inverse Probability of Treatment Weighting
ITT	Intention to treat
MedDRA	Medical Dictionary for Regulatory Activities (Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung)
MMRM	Mixed Model repeated Measures (gemischtes Modell mit Messwiederholungen)
PICO	Population, Intervention, Comparison, Outcome
PRO	Patient-reported Outcome (patientenberichteter Endpunkt)
PS	Propensity Score
PT	Preferred Term (bevorzugter Begriff)
pU	pharmazeutischer Unternehmer
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SAP	statistischer Analyseplan
SGB	Sozialgesetzbuch
SMD	standardisierte Mittelwertdifferenz
SMQ	Standardized MedDRA Query (standardisierte MedDRA-Abfrage)
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 13.10.2023 mit der Prüfung des Studienprotokolls (SP) und des statistischen Analyseplans (SAP) zur anwendungsbegleitenden Datenerhebung (AbD) zu Etranacogen Dezaparovec beauftragt.

In seiner Sitzung am 12.05.2023 hat der G-BA beschlossen, eine AbD und Auswertungen nach § 35a Absatz 3b Satz 1 SGB V für den Wirkstoff Etranacogen Dezaparovec in der Indikation „Behandlung von Erwachsenen mit schwerer und mittelschwerer Hämophilie B (angeborener Faktor-IX-Mangel) ohne Faktor-IX-Inhibitoren in der Vorgeschichte“ zu fordern [1,2]. Dem Beschluss liegt unter anderem das vom IQWiG erstellte Konzept für eine AbD zu Etranacogen Dezaparovec zugrunde (AbD-Konzept A22-83 vom 10.03.2023) [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 [4-6] und mit der Prüfung dieser Unterlagen beauftragt. Neben dem G-BA-Beschluss zu Etranacogen Dezaparovec sollen die Inhalte der diesbezüglichen Beratung des pU zur Studienplanung der AbD berücksichtigt werden (Beratungsanforderung 2023-B-164 [7]).

2 Prüfung der Unterlagen zur Planung der AbD von Etranacogen Dezaparovec

2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen

Der pU hat, wie im G-BA-Beschluss zur AbD gefordert [1,2], ein SP und einen SAP zur Prüfung vorgelegt, beide Dokumente in der Version vom 09.10.2023 [4,5]. Zudem hat der pU einen Bericht zur Methodik der Confounderidentifikation vorgelegt [6].

Gemäß Beschluss des G-BA ist die AbD vorzugsweise als vergleichende Registerstudie oder, sofern eine vergleichende Registerstudie nicht umsetzbar ist, als vergleichende Studie unter Nutzung einer spezifisch für die vorliegende anwendungsbegleitende Datenerhebung 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

Der pU beschreibt, dass er sich derzeit im Austausch mit dem DHR befindet, um die Umsetzung der von ihm geplanten Anpassungen der Datenfelder sowie der Source Data Verification zu diskutieren. Zwar gibt der pU an, dass die Anpassungen im DHR vor dem Beginn der Datenerhebung erfolgen sollen. Der pU plant jedoch, in den vom G-BA geforderten Statusberichten 6 und 18 Monate nach Beginn der AbD jeweils Angaben zum Stand der technischen Umsetzung der Änderungen im DHR sowie zur Umsetzung der Source Data Verification vorzulegen. Dies ist nicht sachgerecht. Alle erforderlichen Anpassungen müssen vor Beginn der AbD erfolgen und in SP und SAP hinterlegt sein.

2.2.1 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, Comparison, Outcome). Die folgenden Abschnitte beurteilen die Umsetzung des PICO im SP.

2.2.1.1 Population

Gemäß G-BA-Beschluss soll die AbD für folgende Population durchgeführt werden: Erwachsene mit schwerer und mittelschwerer Hämophilie B (angeborener Faktor-IX-Mangel) ohne Faktor-IX-Inhibitoren in der Vorgeschichte.

Die Einschlusskriterien der geplanten AbD entsprechen laut SP den Vorgaben des G-BA. Der pU plant den Einschluss erwachsener Patienten mit angeborener mittelschwerer oder schwerer Hämophilie B (Faktor-IX-Mangel), definiert durch eine endogene Faktor-IX-Aktivität von < 5 %. Mit den Einschlusskriterien grenzt der pU die Population auf Patienten ein, die mit Faktor-IX-Präparaten vorbehandelt sind. Die Einschlusskriterien sind grundsätzlich nachvollziehbar, wobei das Kriterium Vortherapie mit Faktor-IX-Präparaten keine Voraussetzung für die Behandlung mit Etranacogen Dezaparvovec und damit kein notwendiges Einschlusskriterium ist. Ungeachtet dessen wird jedoch davon ausgegangen, dass dieses Kriterium keine Auswirkungen auf die Rekrutierung der Patienten hat, da mit hoher Wahrscheinlichkeit annähernd alle Patienten im Kindesalter diagnostiziert werden und damit zu Studieneinschluss bereits Vortherapien mit Faktor-IX-Präparaten erhalten haben. Die vom pU benannten Ausschlusskriterien sind angemessen.

Der pU beschreibt, dass nur der Einschluss der Patienten im DHR dokumentiert wird, da einige der Ein- und Ausschlusskriterien auch zukünftig nicht im DHR dokumentiert werden können. Dies ist nicht sachgerecht. Aus den Tragenden Gründen zum Beschluss des G-BA [2] geht hervor, dass alle Ein- und Ausschlusskriterien verpflichtend im DHR zu erfassen sind.

2.2.1.2 Intervention und Kontrolle

Die Benennung der zu vergleichenden Therapieoptionen entspricht den Vorgaben des G-BA (Intervention: Etranacogen Dezaparvovec; Kontrolle: rekombinante oder aus humanem Plasma gewonnene Blutgerinnungsfaktor-IX-Präparate).

2.2.1.3 Endpunkte (Outcomes)

Allgemeine Anmerkungen

Der pU beschreibt, dass einige Datenfelder zur Erhebung von Endpunkten und Confoundern im DHR bereits vorhanden sind, aber nicht oder nicht vollständig verpflichtend auszufüllen sind. Der pU plant die Vollständigkeit der Dokumentation durch finanzielle Anreize zu erhöhen (siehe Abschnitt 2.2.2). Das ist nicht ausreichend. In den Tragenden Gründen zum Beschluss des G-BA [2] wird ausgeführt, dass die Anpassungen am DHR die Einführung von verpflichtend zu dokumentierenden Datenfeldern zu relevanten Endpunkten umfassen. Alle erforderlichen Anpassungen sind vor Beginn der AbD durchzuführen und in SP und SAP zu hinterlegen. Außerdem gibt es in den Studienunterlagen unterschiedliche Angaben darüber, ob lediglich die teilnehmenden Zentren oder ob auch die eingeschlossenen Patienten incentiviert werden sollen, um beispielsweise das Ausfüllen der Instrumente zu patientenberichteten Endpunkten (PROs) zu fördern. Letzteres ist vor dem Hintergrund des zwangsweise offenen Designs der Studie nicht sachgerecht. In den Studienunterlagen ist klarzustellen, wie die Incentivierung aussehen soll.

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

Mortalität

Todesfälle

Der pU plant die Erfassung eines Todesfalls über das Datenfeld „Grund des Ausscheidens“ zu operationalisieren. Dieses kann die Ausprägungen verstorben, Zentrenwechsel sowie Widerruf der Einwilligung annehmen. Der Todeszeitpunkt soll über das Datenfeld „Datum des Ausscheidens“ bestimmt werden, wobei ein exaktes Datum oder unbekannt angegeben werden kann.

Die vom pU vorgeschlagene Operationalisierung ist angemessen.

Morbidität

Blutungen

Der G-BA fordert in seinem Beschluss [1] die Erhebung der folgenden Blutungsendpunkte: schwere Blutungen, lebensbedrohliche Blutungen, Gelenkblutungen sowie behandelte Blutungen.

Der pU plant behandelte Blutungen, schwere Blutungen, lebensbedrohliche Blutungen und Gelenkblutungen jeweils als annualisierte Blutungsrate (ABR) auszuwerten. Zur Operationalisierung geht aus dem SP hervor, dass alle Blutungen erfasst werden, die eine Therapie auslösen. Hierzu plant er, die Angaben zu Therapiestart und -ende, Anlass (z. B. Verdacht auf Blutung, spontane Blutung, etc.), Lokalisierung (z. B. Gelenk, Muskel, gastrointestinal, etc.) und Schweregrad (leicht, schwer, lebensbedrohlich, unbekannt) zugrunde zu legen.

Für die Operationalisierung von schweren und lebensbedrohlichen Blutungen gibt der pU an, die vom DHR verwendeten Definitionen zu verwenden, die denen des Pediatric Network on haemophilia management entsprechen, jedoch ohne Angabe einer Quelle. Schwere Blutungen sind demnach definiert als Blutungen, 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 können. Die Operationalisierung der schweren und lebensbedrohlichen Blutungen erscheint sachgerecht. Der pU gibt zudem an, dass die Definitionen bei der Eingabe in das DHR ersichtlich sind. Dies ist allerdings nicht abschließend zu prüfen, da die vom DHR genutzten Definitionen weder aus dem Gesamtdatensatz noch aus dem Handbuch hervorgehen.

Die vorgesehenen Auswahlmöglichkeiten für die Erfassung des Therapieanlasses (Verdacht auf Blutung, spontane Blutung, traumatische Blutung sowie Blutung, Ursache unbekannt) ermöglichen eine hinreichend sichere Erfassung der behandelten Blutungen. Im Falle von schweren und lebensbedrohlichen Blutungen wird der Anlass der Bedarfsbehandlung über die

Optionen spontane Blutung und traumatische Blutung konkreter erhoben. Gelenkblutungen werden anhand des DHR hinreichend sicher erfasst.

Insgesamt sind die Operationalisierungen der Blutungsendpunkte angemessen.

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 [8]. Der pU plant aus der Skala Schmerzintensität das Item 5 „durchschnittlicher Schmerz“ 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 (insbesondere durch Hämarthrose) und chronischem Schmerz (insbesondere durch hämophile Arthropathie) essenziell [9,10]. In einer solchen Situation erscheint eine Operationalisierung, die die Entwicklung im Studienverlauf mitberücksichtigt, sinnvoll. Für eine solche Operationalisierung ist eine häufigere Erhebung als 2-mal jährlich notwendig.

Gelenkfunktion, gemessen mit einem validierten Instrument

Für die Gelenkfunktion plant der pU das Instrument Haemophilia Joint Health Score (HJHS) zu verwenden. Die Erhebung des Gesamtscores, Global Gait Scores sowie der Scores der einzelnen Gelenke soll 2-mal jährlich erfolgen. Bei dem HJHS handelt es sich um ein validiertes und etabliertes Instrument [11,12], 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 Hemophila-specific Quality of Life Questionnaire for Adults (Haemo-QoL-A) 2-mal jährlich zu erheben. Dabei sollen der Gesamtscore sowie die Scores der 6 Skalen (körperliche Funktionsfähigkeit, Rollenfunktion, Sorgen, Konsequenzen von Blutungen, emotionale Auswirkungen und Bedenken über die Behandlung) erfasst werden. Der Haemo-QoL-A ist ein validiertes und etabliertes Instrument zur Erhebung der gesundheitsbezogenen Lebensqualität bei Patienten mit Hämophilie [13,14]. Die Verwendung des Haemo-QoL-A in der AbD ist sachgerecht.

Allgemeine Anmerkungen zu den PROs und der Gelenkfunktion

Responderanalysen

Für die Endpunkte Schmerz, Gelenkfunktion und gesundheitsbezogene Lebensqualität beschreibt der pU, dass diese als Responderanalysen ausgewertet werden sollen. Der pU plant, bei den PROs jeweils die Verbesserung und Verschlechterung, bei der Gelenkfunktion die Verschlechterung zu untersuchen. Gemäß SP soll das Responsekriterium jeweils $\geq 15\%$ der Skalenspannweite betragen. Das ist jeweils sachgerecht. Als Responder gelten gemäß SP alle Patienten, die mindestens 2-mal eine Veränderung im Score um $\geq 15\%$ im Vergleich zur Baseline aufweisen. Diese Operationalisierung ist nicht sinnvoll interpretierbar. Die Bedeutung einer 2-maligen Veränderung um $\geq 15\%$ zu unterschiedlichen Zeitpunkten während des Studienverlauf ist unklar. Für die Nutzenbewertung, und damit auch für die AbD zu bevorzugen, sind aussagekräftige Responderanalysen, z. B. zu Beobachtungsende [15]. Für diese Art der Auswertung sind Informationen zum Studienverlauf, bspw. in Form von Verlaufskurven für die Interpretation des Ergebnisses der Responderanalyse erforderlich.

Erhebungszeitpunkte

Für die PROs und die Gelenkfunktion geht aus den Studienunterlagen hervor, dass diese alle 6 Monate erhoben werden sollen. Der pU weist in diesem Zusammenhang allerdings darauf hin, dass Unsicherheit über die tatsächlichen Erhebungszeitpunkte bestehe und die Meldezeitpunkte nicht beeinflussbar seien. Dies kann dazu führen, dass sich die Erhebungszeitpunkte zwischen den Studienarmen systematisch unterscheiden, was zu einer relevanten Verzerrung der Ergebnisse führen kann. Für die Durchführung der AbD und die Generierung aussagekräftiger Daten zum Krankheitsverlauf unter Therapie ist eine Vereinheitlichung der Erhebungs- und Meldezeitpunkte erforderlich. Dabei sollte der pU angemessene Toleranzfenster wählen und durch geeignete Maßnahmen sicherstellen, dass fehlende Werte vermieden werden.

Sollten einheitliche Erhebungszeitpunkte innerhalb des DHR nicht möglich sein, sollte der pU für die Erhebung der PROs sowie der Gelenkfunktion eine Erhebung durch ein externes Zentrum prüfen. Dieses Vorgehen wird beispielsweise in der AbD zu Brexucabtagen Autoleucel beim Mantelzelllymphom umgesetzt [16]. Durch dieses Vorgehen kann die Erhebung der PROs kontrolliert werden.

Unerwünschte Ereignisse

Gemäß Forderungsbeschluss des G-BA sollen für die AbD folgende unerwünschte Ereignisse (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-IX-Hemmkörpern
 - thromboembolische Ereignisse
 - symptomatische Leberschäden
 - maligne Neubildungen

Im Studienprotokoll ist neben der Erhebung der vom G-BA geforderten UE-Endpunkte auch die Gesamtrate von UEs jeglicher Schweregrade vorgesehen. Die Ergebnisse auf Basis aller UEs sind in der Regel nicht interpretierbar und aus diesem Grund ist die Erhebung verzichtbar.

Der pU plant für die UE-Endpunkte neben der Auswertung der Gesamtraten auch die Auswertung nach Systemorganklasse (SOC) / bevorzugtem Begriff (PT). Dies ist mit einem erheblichen Aufwand verbunden, gleichzeitig ist eine Einordnung aller auftretenden UEs für die AbD nicht gefordert. Unter Abwägung von Aufwand und Aussagekraft für die AbD erscheint die Auswertung nach SOC / PT nicht sinnvoll.

Der pU gibt an, alle UE-Endpunkte von einem Auftragsforschungsinstitut (CRO) mittels medizinischem Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung (MedDRA) codieren zu lassen. Gemäß der Tragenden Gründe zum Beschluss des G-BA [2] ist dies für die spezifischen UEs ideal, die Codierung für alle SUEs (und UEs) ist jedoch verzichtbar.

Die Auswertung der UE-Endpunkte soll gemäß SP als Anteil der Patienten mit mindestens 1 Ereignis erfolgen. Der Vergleich der Anteile der Patienten mit mindestens 1 Ereignis ist angemessen.

Bezüglich der Operationalisierung und der geplanten Auswertungen der vom G-BA geforderten UE-Endpunkte ergeben sich folgende Kritikpunkte:

Es ist unklar, auf Basis welcher Informationen die spezifischen UEs ausgewertet werden sollen. Aus dem SP geht hervor, dass es im DHR ein Auswahlfeld für das jeweilige spezifische UE geben soll und anschließend (mit Ausnahme der spezifischen UEs Bildung von Faktor-IX-Hemmkörpern [UEs und SUEs] und thromboembolische Ereignisse [SUEs]) das UE über einen Freitext weiter spezifiziert wird. Die im Freitext dokumentierten Ereignisse sollen von einem CRO mittels MedDRA codiert werden (siehe oben). Wie diese Daten in das DHR eingegeben werden, geht aus dem SP nicht hervor. Sofern die Auswertung ausschließlich auf der Auswahl der spezifischen UEs im DHR basiert, kann auf die Codierung mittels MedDRA verzichtet

werden, vorausgesetzt die Definition der spezifischen UEs ist sachgerecht. Sofern die Auswertung jedoch auf Basis der MedDRA-Codierungen erfolgen soll, ist sicherzustellen, dass die MedDRA-Codierungen im DHR verpflichtend dokumentiert werden. Zudem sind in diesem Fall für jedes spezifische UE entsprechende MedDRA-Codierungen (z. B. PT-Liste oder standardisierte MedDRA-Abfrage [SMQ]) in den Studienunterlagen zu ergänzen, die für das jeweilige spezifische UE relevant sind.

Ergänzende Informationen zu Anzahl der verbrauchten Faktorkonzentrate und Zeitpunkt der Wiederaufnahme einer prophylaktischen Behandlung

Der pU plant die Anzahl der verbrauchten Faktorkonzentrate, getrennt nach Bedarfs- und Prophylaxebehandlung sowie den Zeitpunkt der Wiederaufnahme einer prophylaktischen Behandlung zu erheben. Die jeweilige Operationalisierung ist sachgerecht.

2.2.2 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 (AbD-Konzept A22-83 [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 gibt an, dass er sich derzeit im Austausch mit dem DHR befindet, um die Umsetzung der von ihm geplanten Anpassungen der Datenfelder sowie der Source Data Verification zu diskutieren. Aus den Angaben im SP und SAP lässt sich jedoch weiterhin Anpassungsbedarf bei folgenden Kriterien ableiten:

- Erhöhung der Anzahl der Patienten in Einzelmeldung: Im SP ist beschrieben, dass der pU davon ausgeht, dass finanzielle Anreize die Anzahl der Patienten in Einzelmeldung erhöhen. In den Studienunterlagen gibt es unterschiedliche Angaben darüber, ob lediglich die teilnehmenden Zentren oder ob auch die eingeschlossenen Patienten incentiviert werden sollen (siehe dazu auch Abschnitt 2.2.1.3). Es ist klarzustellen, wie die Incentivierung aussehen soll.
- Verpflichtend zu dokumentierende Angaben im Register: Aus den Angaben im SP geht hervor, dass lediglich der Einschluss in die Studie im DHR dokumentiert werden soll, nicht jedoch die einzelnen Ein- und Ausschlusskriterien (siehe Abschnitt 2.2.1.1). Zudem beschreibt der pU, dass für einige Datenfelder zu relevanten Endpunkten und zu Confoundern keine verpflichtende Meldung im DHR besteht (siehe Abschnitt 2.2.1.3). Für relevante Daten, die im Zuge der AbD erhoben werden sollen, sind jedoch spätestens mit Beginn der AbD alle relevanten Datenfelder verpflichtend zu dokumentieren.

- **Patientencharakteristika:** Aus den Studienunterlagen des pU geht nicht hervor, welche Patientencharakteristika zur Beschreibung der Population in der AbD erhoben werden sollen. Diese sind in den Studienunterlagen zu ergänzen, einschließlich der für die verpflichtende Erhebung vorgesehenen Datenfelder im DHR.
- **Einheitliche Erhebungs- und Meldezeitpunkte:** Wie im Konzept beschrieben, besteht für das DHR die 1-mal jährliche Meldeverpflichtung. Im SP sind Erhebungsintervalle von 6 Monaten ohne die Angabe von Toleranzfenstern beschrieben. Der pU gibt an, dass Unsicherheiten über die tatsächlichen Erhebungs- und Meldungszeitpunkte bestehen, die nicht beeinflussbar seien. Der pU sollte angemessene Toleranzfenster wählen und durch geeignete Maßnahmen sicherstellen, dass fehlende Werte vermieden werden. Die Dokumentation der erhobenen Daten sollte möglichst für alle Patienten einheitlich direkt im Anschluss an die jeweilige Erhebung erfolgen, um Meldeverzögerungen zu vermeiden.
- **Sicherstellung der Richtigkeit der Daten mittels Source Data Verification:** Der pU plant die Source Data Verification durch eine CRO durchführen zu lassen. Hierfür ist ein Zugang zu einer Monitoringumgebung im DHR erforderlich. Der pU gibt an, dass die technische Umsetzbarkeit zwar bestätigt wurde, die Entscheidung des Lenkungsausschusses jedoch noch aussteht. Diese und alle weiteren Anpassungen am DHR müssen vor Beginn der AbD abgeschlossen sein.

Für den Fall, dass die Umsetzung der erforderlichen Änderungen im DHR nicht möglich ist, plant der pU mit dem G-BA in Kontakt zu treten und das Aufsetzen einer alternativen Datenplattform zu diskutieren. Einen Zeitrahmen dafür legt der pU nicht fest, gibt aber an, in den Statusberichten 6 und 18 Monate nach Beginn der AbD über den Stand der technischen Umsetzung der Änderungen im DHR zu berichten (siehe dazu auch Abschnitt 2.2.11). Dies ist nicht sachgerecht. Vor Beginn der AbD muss klar sein, welche Datenquelle genutzt wird. Bei Nutzung des DHR ist sicherzustellen, dass alle notwendigen Anpassungen vor Beginn der AbD erfolgen. Dies ist im SP und SAP zu hinterlegen.

2.2.3 Rekrutierungs- und Beobachtungszeitraum, Zuordnung zu den Behandlungsgruppen und Indexdatum

Rekrutierungs- und Beobachtungszeitraum

Als Rekrutierungszeitraum für die AbD legt der pU den Zeitraum von 05/2024 bis 01.01.2026 fest. Die Datenerhebung soll bis zum 31.12.2028 erfolgen (siehe dazu auch Abschnitt 2.2.11). Für die Nutzenbewertung sind die Auswertungen der mit der AbD erhobenen Daten spätestens bis zum 02.11.2029 vorzulegen. Eine Datenerhebung bis Ende 04/2029 erscheint daher angemessen. Da die Beobachtungszeit für die AbD gemäß Beschluss des G-BA 3 Jahre betragen soll, erscheint eine Verlängerung des Rekrutierungszeitraums bis Ende 04/2026 möglich, sofern die benötigte Fallzahl bis 01/2026 nicht erreicht werden kann.

Zuordnung zu den Behandlungsgruppen

Gemäß Abbildung 1 in SP und SAP und der zugehörigen Beschreibung werden alle Patienten, die während des Rekrutierungszeitraums eine Behandlung mit Etranacogen Dezaparovec bekommen, dem Interventionsarm zugeordnet. Patienten, die im Verlauf der Beobachtung von der Vergleichstherapie auf Etranacogen Dezaparovec wechseln, werden unterschiedlich berücksichtigt, abhängig davon, wie lange sie bereits seit Einschluss in die Studie unter der Vergleichstherapie beobachtet wurden. Diejenigen Patienten, die länger als 2 Jahre unter der Behandlung mit Faktor-IX-Präparaten beobachtet wurden und dann erst auf Etranacogen Dezaparovec wechseln, sollen im Vergleichsarm ausgewertet und bis zum Studienende weiter beobachtet werden. Patienten, die nach ≤ 2 Jahren auf eine Behandlung mit Etranacogen Dezaparovec wechseln, sollen hingegen im Interventionsarm ausgewertet werden. Für diese Patienten stellt der Zeitpunkt des Wechsels der Beobachtungsstart für die AbD dar, wobei eine neue Erhebung der Baselinedaten erfolgen soll. Die Beobachtungszeit unter Vergleichstherapie dieser Patienten wird für die AbD nicht berücksichtigt. Dieses Vorgehen ist sachgerecht.

An einigen Stellen (bspw. Abschnitt 8.2.6 im SAP) ist die Zuordnung zu den Behandlungsgruppen bei Wechsel im Verlauf der Beobachtung von der Vergleichstherapie auf Etranacogen Dezaparovec abweichend beschrieben. Dies sollte wie oben beschrieben angepasst werden.

Indexdatum

Der pU legt als Indexdatum (vom pU als Reference Date bezeichnet), d. h. das Datum für den Beobachtungsstart, das Datum der ersten Behandlung mit Etranacogen Dezaparovec oder der Vergleichstherapie nach dem Patienteneinschluss und der ersten Datenübermittlung an das DHR fest, mit Ausnahme von Therapiewechslern (siehe oben). Dieses Vorgehen ist 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 [17,18]. 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.

2.2.4 Fallzahlplanung

Im SP und SAP gibt der pU an, dass aufgrund des nicht interventionellen Designs der AbD nicht beeinflussbar sei, wie viele Patienten in die AbD eingeschlossen werden. Daher plant der pU alle Patienten, die die Einschlusskriterien erfüllen, in die AbD einzuschließen ohne die Benennung einer konkreten Fallzahl.

Der pU beschreibt 2 Ansätze zur Fallzahlplanung. Zum einen führt er Schätzungen beruhend auf einer verschobenen Nullhypothese von $RR = 0,5$ durch, wie sie im Beschluss des G-BA beschrieben ist. Zum anderen legt er Schätzungen vor, bei denen er mit eigenen Anpassungen die in den Allgemeinen Methoden des IQWiG [15] beschriebenen Kriterien für einen dramatischen Effekt zugrunde legt. Statt eines relativen Risikos von 10 (Kehrwert 0,1) verwendet der pU ein relatives Risiko von 0,5. Diese, vom pU als orientierende Fallzahlplanung benannten Schätzungen, führt er für verschiedene Blutungsendpunkte, jeweils operationalisiert als ABR durch. Zur Schätzung verwendet er ein Negativ-Binomial-Modell. Auf den Ansatz mittels eines dramatischen Effektes wird an dieser Stelle nicht weiter eingegangen.

Die allgemeinen Annahmen der Fallzahlplanung unter der verschobenen Nullhypothese sind, eine Power von 80 %, ein $\alpha = 0,05$ (2-seitig), eine verschobene Nullhypothese (Rate Ratio = 0,5), eine Dispersion von 1,5 und ein Behandlungsverhältnis von 1 : 5 (Intervention : Kontrolle). Diese Annahmen sind sachgerecht.

Je nach Blutungsendpunkt trifft der pU verschiedene Annahmen zur ABR in den Behandlungsarmen. Hierfür nutzt der pU Ergebnisse aus der Studie HOPE-B, der Zulassungsstudie zu Etranacogen Dezaparovec und zitiert dabei den European Public Assessment Report (EPAR) [19]. Die Angaben im EPAR stimmen jedoch nicht mit den Angaben im SAP und SP überein (z. B. ABR der mit Faktor-IX behandelten Blutungen für die Kontrollgruppe: 3,45 im SP / SAP und 3,56 im EPAR). Dies ist vom pU anzupassen.

Der pU begründet mit Hilfe der dargestellten Szenarien die Wahl der behandlungsbedürftigen Blutungen als primären Endpunkt, da für diesen Endpunkt die erforderliche Fallzahl für einen Wirksamkeitsnachweis gering sei und mit einer guten Datenqualität im DHR erfassbar sei. Der pU benennt keine konkrete geplante Fallzahl. Er gibt jedoch an, dass er im Rahmen der 2. Interimsanalyse eine Rekalkulation der Fallzahlen vornehmen will (zum Zeitplan siehe Abschnitt 2.2.11). Dieses Vorgehen ist nicht sachgerecht. Der pU sollte sich für ein Szenario und damit für eine (vorläufige) Fallzahl entscheiden und diese auch in die Machbarkeitsprüfung einbeziehen. Im Rahmen der Interimsanalyse können diese begründet angepasst werden.

2.2.5 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 [20-22],

- 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.

2.2.5.1 Identifikation und Erhebung wichtiger Confounder

Zur Identifikation der für die vorliegende Fragestellung relevanten Confounder legt der pU einen in sich abgeschlossenen Bericht über eine systematische Literatursuche zur Identifikation von Confoundern vor (Annex A1 des Studienprotokolls [6]). Die Beurteilung der technischen Durchführung der Informationsbeschaffung des pU zur Identifikation der Confounder findet sich in Anhang A.

Durch das Vorgehen des pU ist nicht sichergestellt, dass eine ausreichend vollständige Liste an potenziell relevanten Confoundern, einschließlich der zugehörigen statistischen Evidenz, identifiziert wurde. Zwar resultiert aus der Literatursuche eine umfangreiche Liste potenzieller Confounder, allerdings wurden diese Informationen aus Sekundärquellen extrahiert, die keine geeigneten Primärstudien in ausreichendem Umfang enthalten und auch für die Bewertung der Confounder notwendige statistische Evidenz nicht berichten.

Ungeachtet der Mängel bei der Identifikation der potenziellen Confounder ist auch das Vorgehen zur Einschätzung der Confounder als wichtig, weniger wichtig oder nicht wichtig nicht sachgerecht. Von den 34 identifizierten potenziellen Confoundern wurden 4 als sehr wichtig eingestuft: Faktor-IX-Restaktivität, Alter, Dosierung (Intensität der Prophylaxe) 12 Monate vor Studieneinschluss und Gelenkstatus. Keiner der potenziellen Confounder wurde als weniger wichtig eingestuft. Für den Ausschluss der 30 als nicht wichtig eingestuften potenziellen Confounder liegt nicht immer eine Begründung vor, z. B. wurde Adhärenz ohne Begründung ausgeschlossen. Inhaltlich inkonsistent ist das Vorgehen bezüglich des potenziellen Confounders Blutungsrate: Diese wurde von den klinischen Experten zunächst als nicht wichtig eingestuft. Im Folgenden wurde jedoch die ABR_12 (ABR 12 Monate vor Studieneinschluss) von den klinischen Experten als zusätzlicher relevanter Confounder benannt.

Darüber hinaus bestehen Unterschiede zwischen den vom DHR und vom pU 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]. Beispielsweise wurde Adipositas zwar vom pU als potenzieller Confounder identifiziert, von den klinischen Experten allerdings als nicht wichtig eingestuft.

Zusätzlich benennt der pU 4 Interaktionen der Confounder: Dosierung der Faktor-IX-Prophylaxe*ABR_12, Alter*Gelenkstatus, Gelenkstatus*Faktor-IX-Restaktivität sowie ABR_12*Faktor-IX-Restaktivität. Im Rahmen der Modellbeschreibung des Propensity Scores geht der pU jedoch nicht auf die genannten Interaktionen ein. Es bleibt unklar, ob diese in der Modellierung berücksichtigt werden oder nicht. Dies ist vom pU zu spezifizieren.

Insgesamt kann aufgrund der unzureichenden Informationsbeschaffung zu Confoundern und in Verbindung mit der nachfolgenden Experteneinschätzung auf Basis einer ungeeigneten Identifikation und Aufbereitung der statistischen Evidenz nicht von einer sachgerechten Confounderliste für die AbD ausgegangen werden. Es erscheint daher erforderlich, das Dokument zur Methodik der Confounderidentifikation grundlegend unter Berücksichtigung der oben genannten Punkte zu überarbeiten. Darüber hinaus sollten die Datenfelder zu den Confoundern spätestens mit Beginn der AbD verpflichtend im DHR zu dokumentieren sein.

2.2.5.2 Berücksichtigung wichtiger Confounder bei der Analyse

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

Der pU beschreibt, dass eine erste Prüfung der Balanciertheit der Behandlungsarme ohne vorherige Adjustierung der Daten erfolgen soll. Hierfür berechnet er die standardisierte Mittelwertdifferenz (SMD) der oben angegebenen Confounder. Wenn alle SMDs ihrem Betrage nach kleiner als 0,1 sind, ist aus Sicht des pU eine Confounderadjustierung und damit die Verwendung des Propensity Scores (PS) nicht notwendig. Dieses Vorgehen ist nicht sachgerecht. Ein PS-Verfahren ist dennoch vorzuziehen, um verbleibende Imbalancen auszugleichen.

Darüber hinaus beschreibt der pU, ein logistisches Regressionsmodell unter Einbezug der identifizierten Confounder zur Bildung des PS zu verwenden, wenn für mindestens einen Confounder der Betrag der SMD größer als 0,1 ist. Hierbei listet der pU auf, mit welchem Skalenniveau und welchem Trennwert welcher Confounder in das Modell aufgenommen werden soll. Anmerkungen zur Berücksichtigung möglicher Interaktionen nimmt der pU nicht vor, obwohl er verschiedene Interaktionen benennt (siehe Abschnitt 2.2.5.1). Dieses Vorgehen ist nicht sachgerecht. Sind mögliche Interaktionen identifiziert, müssen diese in das PS-Modell aufgenommen werden.

Sollte das PS-Modell nicht konvergieren, soll eine Logistische-Ridge-Regression verwendet werden. Sofern das PS-Modell weiterhin nicht konvergiert plant der pU die Durchführung von naiven Vergleichen. Dies ist sachgerecht, jedoch geht der pU nicht darauf ein, welche Auswirkungen dies auf die Interpretation der Ergebnisse hat, dies ist vom pU zu ergänzen.

Sofern ein PS verwendet wird, wird die Dichteverteilung der PS bezüglich ihrer Überlappung geprüft und Patienten aus der Analysepopulation entfernt, deren PS in Bereichen liegt die sich nicht überlappen (Trimming). Weder begründet der pU sein Vorgehen noch gibt er entsprechende Literaturverweise an. Stürmer et al. [23] verglichen unterschiedliche Trimming- und PS-Verfahren, darunter auch das vom pU beschriebene Verfahren. Aus den Ergebnissen ihrer Simulationsstudie geht hervor, dass andere Trimming-Verfahren zu weniger verzerrten Effektschätzungen führen. Der pU sollte entweder sein aktuelles Vorgehen begründen oder ein alternatives Verfahren wählen. In beiden Fällen sollte der pU erläutern, warum aus seiner Sicht das gewählte Verfahren für die AbD geeignet ist, z. B. anhand geeigneter Literatur.

Ein Kriterium, wann die Untersuchungen zu einer ausreichenden Überlappung führen, gibt der pU nicht an. Stattdessen beschreibt er, dass die so erhaltene Analysepopulation stark von der Zielpopulation des G-BA abweichen kann. Dieses Vorgehen ist nicht sachgerecht. Auch wenn es zurzeit noch keine etablierten Kriterien zur Einordnung der Überlappung gibt, sollte in den Studienunterlagen festgelegt werden zu wie viel Prozent eine Überlappung mindestens gewährleistet sein muss.

Auswahl des Verfahrens für die Confounderadjustierung

Der pU beschreibt, dass er zur Confounderadjustierung gewichtungsbasierte PS-Verfahren verwenden will. Hierbei benennt er das Inverse Probability of Treatment Weighting (IPTW) sowie das Verfahren der Fine Stratification Weights (FSW). Für die FSW wählt der pU eine feste Anzahl von 5 annähernd gleichgroßen Strata. Das beschriebene Vorgehen ist sachgerecht.

Den Umgang mit extremen Gewichten beim IPTW-Verfahren plant der pU mittels Winsorizing des untersten und obersten 1 % der IPTW-Gewichte zu berücksichtigen.

Um zu entscheiden, ob und wenn ja, welches PS-Verfahren zur weiteren Anwendung kommt, will der pU die Balanciertheit bewerten. Diese wird, wie bereits beschrieben, mittels SMD für die einzelnen Confounder betrachtet. Hierbei verwendet der pU das Verfahren, bei dem die Summe der Beträge der SMD am geringsten ist.

Zusätzlich betrachtet der pU die Balanciertheit, um zu prüfen, ob das PS-Modell aus seiner Sicht erfolgreich war. Dabei wird von einer ausreichenden Balanciertheit ausgegangen, wenn bei jedem Confounder 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 [24]. 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 [25]. Sollte bei einem oder mehreren Confoundern die SMD größer als 0,25 sein, wird der Ansatz eines PS-Verfahrens verworfen und der pU nutzt einen naiven Vergleich für die Auswertungen. Dieses Vorgehen ist sachgerecht.

Der pU stellt sein Vorgehen in einem Flussdiagramm dar, wobei dieses nicht vollständig deckungsgleich mit den Angaben im Text ist. Zum einen wird das PS-Verfahren SMRW (Standardized Mortality Ratio Weighting) aufgeführt, welches jedoch nicht vom pU als PS-Verfahren vorgesehen ist. Zum anderen ist das Vorgehen bei der Wahl des PS-Verfahrens, über die Betrachtung der Balanciertheit, nicht deckungsgleich mit dem im Text beschriebenen. Dies ist vom pU zu korrigieren.

Beschreibung der Zielpopulation nach PS-Verfahren

Ein zentraler Aspekt bei nicht randomisierten Vergleichen mit Confounderadjustierung ist die detaillierte Beschreibung der Teilpopulation, die sich aus der Anwendung eines PS-Verfahrens ergibt, im weiteren Analysepopulation genannt. Es muss nach erfolgreicher Anwendung eines PS-Verfahrens entweder sorgfältig begründet werden, dass die resultierende Analysepopulation der ursprünglichen Zielpopulation in ausreichendem Maße entspricht, oder, wenn dies nicht der Fall ist, zu beschreiben, auf welche Teilpopulation der Zielpopulation sich die Ergebnisse beziehen.

Um diesem Anspruch zu genügen, gibt der pU an, eine deskriptive Beschreibung der Baselinecharakteristika pro Behandlungsarm der ausgeschlossenen bzw. in die Analysepopulation eingeschlossenen Patienten durchzuführen. Dieses Vorgehen ist nicht vollständig. Der pU sollte auch die Baselinecharakteristika für alle in die AbD eingeschlossenen Patienten gegenüberstellen. Entsprechende Änderungen sind in den Studienunterlagen vorzunehmen. Zudem ist wie in Abschnitt 2.2.1.1 beschrieben zu ergänzen, welche Patientencharakteristika erhoben werden sollen.

2.2.6 Anmerkungen zu den geplanten Auswertungen für die Endpunkte der AbD

Im SP und SAP sind zu den verschiedenen Operationalisierungen des primären Endpunkts sowie der sekundären einschließlich der UE-Endpunkte jeweils Ausführungen zu den geplanten Auswertungen enthalten. Diese werden nachfolgend adressiert.

Insgesamt fehlen genauere Angaben zur geplanten Teststatistik. Diese sind vom pU zu ergänzen.

Bei einigen Endpunkten, z. B. lebensbedrohlichen Blutungen, Gesamtüberleben und spezifischen UEs, ist davon auszugehen, dass nur wenige Ereignisse auftreten. Die vom pU beschriebenen Methoden können in dieser Datensituation zu verzerrten Effektschätzungen führen. Daher sollte der pU im SAP auf diese Datensituation eingehen und adäquate Analyseverfahren hierfür benennen.

Die Auswertungen der Hauptanalysen beruhen auf der Safety Population. Diese besteht aus allen eingeschlossenen Patienten, die die Intervention oder Vergleichstherapie erhalten haben sowie zu denen Einzelmeldungen im DHR vorliegen. Dabei plant der pU, die Zuteilung zu Interventions- und Kontrollarm wie in Abschnitt 2.2.3 beschrieben vorzunehmen.

Der pU plant zur Auswertung von Zähldaten (Blutungsendpunkte), operationalisiert über die ABR, ein Negativ-Binomial-Modell mit Log-Link-Funktion und Pearson- χ^2 -Skalierung für Standardfehler zu verwenden. Die Behandlung geht dabei als Kovariable und sofern möglich die PS-Gewichte als GewichtungsvARIABLE ein. Die Beobachtungszeit geht als Offset ein.

Zur Betrachtung von Ereigniszeitdaten plant der pU Ereigniszeitanalysen unter Verwendung des Cox Proportional Hazards Modell zur Schätzung des Effektmaßes Hazard Ratio. In das Modell geht die Behandlung als Kovariable und wenn möglich die PS-Gewichte als GewichtungsvARIABLE ein. Zur grafischen Darstellung sollen Kaplan-Meier-Kurven verwendet werden.

Der pU gibt an, bei der Betrachtung von binären Daten im Rahmen von Responderanalysen ein generalisiertes lineares Modell (GLM) beruhend auf einer Binomialverteilung, mit der Variable Behandlung, sofern möglich den PS-Gewichten als GewichtungsvARIABLE und je nach Effektmaß einer geeigneten Link-Funktion verwenden zu wollen. Der pU plant 3 Effektmaße darzustellen, relatives Risiko (Log-Link), Odds Ratio (Logit-Link) und Risikodifferenz (Identity-Link), jeweils mit 95 %-Konfidenzintervall.

Für stetige Daten, wie beispielsweise bei PROs gibt der pU an, diese mittels Responsekriterien (siehe Abschnitt 2.2.1.3) im Rahmen von Responderanalysen auszuwerten (siehe binäre Daten).

Alle beschriebenen, vom pU geplanten Analysemethoden sind sachgerecht.

Sensitivitätsanalysen

Der pU plant verschiedene Sensitivitätsanalysen, für den primären, die sekundären als auch die UE-Endpunkte durchzuführen.

Dabei beruhen 3 Sensitivitätsanalysen auf der Verwendung verschiedener Patientenspopulationen:

- Patienten, die Etranacogen Dezaparovec oder aus humanem Plasma gewonnene Faktor-IX-Präparate erhalten haben (im SAP als Plasma-derived FIX Analysis Set bezeichnet)
- Patienten, die Etranacogen Dezaparovec oder rekombinante Faktor-IX-Präparate erhalten haben (im SAP als Recombinant FIX Analysis Set bezeichnet)
- Patienten, die mindestens 3 Jahre unter Beobachtung für die AbD standen (im SAP als 3-year Follow-up Analysis Set bezeichnet)

Eine weitere Sensitivitätsanalyse besteht darin Beobachtungen für Patienten zu zensieren, wenn diese mindestens 2 Jahre unter Beobachtung mit der Kontrollbehandlung waren und danach auf die Intervention wechselten. Das Vorgehen des pU ist sachgerecht.

Für stetige Daten plant der pU neben der Auswertung über Responderanalysen auch Analysen mittels eines gemischten Modells mit Messwiederholungen (MMRM) durchzuführen. Hierbei beschreibt der pU nicht vollständig die Kovariablen die er ins Modell einbinden will. Zusätzlich fehlt eine klare Definition für Hedges' g.

Darüber hinaus plant der pU im Rahmen von Sensitivitätsanalysen Vorher-nachher-Vergleiche. Hierbei ist zu beachten, dass bei einem Vorher-nachher-Vergleich u. a. aufgrund methodischer Limitationen Unsicherheiten darüber bestehen, ob die beobachteten Effekte nach Wechsel der Therapie tatsächlich auf die Intervention oder auf andere patientenindividuelle Faktoren zurückzuführen sind. Daher lassen sich Ergebnisse eines Vorher-nachher-Vergleichs nicht valide interpretieren. In der Gesamtschau erscheinen Vorher-nachher-Vergleiche für die Nutzenbewertung nicht sinnvoll.

2.2.7 Untersuchung auf potenzielle Effektmodifikatoren und Subgruppenanalysen

Der pU plant Subgruppenanalysen für die Faktoren Alter, Geschlecht, Dosierung (Intensität der Prophylaxe) 12 Monate vor Studieneinschluss, Gelenkstatus, ABR 12 Monate vor Studieneinschluss, Faktor-IX-Restaktivität bei Studieneinschluss und Adeno-assoziiertes-Virus-vom-Serotyp-5(AAV5)-Antikörpertiter bei Studieneinschluss. Sowohl für den Gelenkstatus als auch für die ABR 12 Monate vor Studieneinschluss gibt der pU als Trennwert den Median zu Baseline der eingeschlossenen Patientenpopulation an. Dieses Vorgehen ist nicht sachgerecht. Es ist jeweils ein inhaltlich begründeter Trennwert a priori festzulegen, der nicht von den Ergebnissen der Studie abhängt.

Der pU beschreibt, dass er für jeden Endpunkt das unter Abschnitt 2.2.6 beschriebene jeweilige Modell um einen Interaktionsterm (Behandlung*Subgruppenmerkmal) erweitern will. Als Teststatistik gibt er beispielhaft den Wald-Test für Analysen von Zählraten und Responderanalysen sowie den Likelihood-Ratio-Test im Rahmen von Ereigniszeitanalysen an. Dabei werden nur dann Subgruppenanalysen vorgenommen, wenn mindestens 10 Patienten

in jeder Subgruppe (Zählraten), beziehungsweise 10 Ereignisse in einer Subgruppe (binäre Daten, Überlebenszeiten) vorliegen. Das vom pU geplante Vorgehen ist sachgerecht.

2.2.8 Qualität und Vollständigkeit der Datenerhebung

2.2.8.1 Maßnahmen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung

Der pU beschreibt, dass die Source Data Verification durch eine CRO durchgeführt werden soll. In jedem teilnehmenden Zentrum soll eine Erstvisite sowie eine Abschlussvisite durchgeführt werden. Die Häufigkeit des regelmäßigen Monitorings im Verlauf der Studie wird gemäß SP in Abhängigkeit von der Rekrutierungsrate sowie der Dokumentation des jeweiligen Zentrums festgelegt. Ein Monitoringplan liegt nicht vor und wird gemäß Angaben des pU nach Beauftragung einer CRO erstellt. Wie die Source Data Verification für die AbD im Detail geplant ist, kann daher nicht geprüft werden.

Eine Source Data Verification soll gemäß SP für 100 % der Patienten für die Ein- und Ausschlusskriterien, Confounder zu Baseline sowie den primären Endpunkt erfolgen und für 10 % der Patienten für alle sekundären Endpunkte. Gemäß Beschluss des G-BA sollte eine Source Data Verification für alle Endpunkte mit Ausnahme des primären Endpunkts für mindestens 10 % zufällig ausgewählter Patienten je Erhebungszentrum durchgeführt werden. Die Source Data Verification für die sekundären Endpunkte ist im SP entsprechend zu spezifizieren, sodass je Erhebungszentrum mindestens 10 % zufällig ausgewählte Patienten umfasst sind.

2.2.8.2 Umgang mit fehlenden Daten

Gemäß G-BA-Beschluss 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].

Als Maßnahmen zur Vermeidung von fehlenden Werten gibt der pU finanzielle Anreize an, um die Erhebungsqualität und den Anteil an Einzelmeldungen zu erhöhen.

Zum Umgang und zur Vermeidung unplausibler Werte gibt der pU an, dass diese durch die Anwendung der Source Data Verification sowie Schulungen minimiert werden. Unter Berücksichtigung der vom DHR durchgeführten IT-gestützten Plausibilitätsprüfungen ist dies sachgerecht.

Der pU plant fehlende Datumsangaben wie folgt zu ersetzen: Fehlt bei Nutzen- und UE-Endpunkten der Tag, wird der 15. des jeweiligen Monats angenommen. Der pU geht nicht darauf ein, wie mit fehlenden Monatsangaben umgegangen werden soll. Die geplante Ersetzung des fehlenden Tages ist sachgerecht. Es ist unklar, ob für Monate keine Ersetzung

geplant ist und, welche Auswirkungen dieses Vorgehen hat. Grundsätzlich sind Anstrengungen zur Minimierung des Anteils fehlender Werte bei Datumsangaben vorzunehmen sowie klar zu definieren, wie fehlende Monatsangaben ersetzt werden sollen.

Der pU gibt an, für die Confounder-Variablen multiple Imputation durchführen zu wollen, sofern mehr als 5 % fehlenden Werten vorliegen. Dabei will er multiple Imputation mittels Chained Equations verwenden. Die beschriebene Methode ist prinzipiell geeignet, jedoch ergibt sich aus der Darstellung nicht, wie die multiple Imputation 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. Das genaue Vorgehen ist vom pU zu spezifizieren. In der Literatur sind verschiedene Ansätze für die Kombination von PS-Methoden mit multipler Imputation beschrieben [26,27].

Für die Endpunkte der Endpunktkategorien Mortalität, Morbidität und gesundheitsbezogene Lebensqualität plant der pU unterschiedliche Strategien, je nach Skalenniveau der Daten. Für Ereigniszeitdaten plant er keine Ersetzung, sondern die Zensierung zum letzten Zeitpunkt unter Beobachtung. Bei binären Daten und Zähldaten werden fehlende Werte eines Patienten durch entsprechende Werte eines Patienten der Kontrollgruppe ersetzt (wenn möglich unter Berücksichtigung des PS der Patienten). Dieses Vorgehen ist sachgerecht.

Für UE-Endpunkte sieht der pU keine Ersetzungsstrategie vor. Es soll lediglich eine zusammenfassende Darstellung fehlender Werte erfolgen. Dieses Vorgehen ist nicht sachgerecht. Da in einem GLM Patienten mit fehlenden Werten aus der Analyse ausgeschlossen werden (sofern es sich nicht um rekurrente Ereignisse handelt), ist nicht auszuschließen, dass ein relevanter Anteil an Patienten in den Analysen des pU fehlen werden. Dies hat Auswirkungen auf die Interpretierbarkeit der Ergebnisse. Auch für die UE-Endpunkte sind, ähnlich wie bei den Endpunkten der Endpunktkategorien Mortalität, Morbidität und gesundheitsbezogene Lebensqualität, Ersetzungsstrategien zu benennen.

Aus den Ausführungen des pU ist nicht ersichtlich, ab welchem Anteil an fehlenden Werten eine Ersetzung als nicht mehr sinnvoll angesehen wird sowie die daraus resultierenden Konsequenzen. Dies ist nicht sachgerecht und vom pU zu ergänzen.

2.2.9 Weitere Anpassungen der AbD

Estimand

Der primäre Estimand einer AbD entspricht der Treatment Policy Strategy und beinhaltet die Auswertung nach dem Intention-to-Treat(ITT)-Prinzip für alle patientenrelevanten Endpunkte. Grundsätzlich ist das Vorgehen des pU sachgerecht, sofern die Zuordnung zu den Behandlungsgruppen dem in Abschnitt 2.2.3 beschriebenen Schema entspricht.

Abbruchkriterien

Gemäß G-BA-Beschluss sind im SP und SAP Angaben zu Abbruchkriterien wegen Vergeblichkeit darzulegen [1]. Zwar gibt der pU im SAP und SP an, dass er die Vergeblichkeitsprüfung im Rahmen der 2. und 3. Zwischenanalyse (18 und 36 Monate nach Beginn der AbD) durchführen und die Entscheidung zu einem notwendigen Studienabbruch in Abstimmung mit dem G-BA treffen will, jedoch finden sich keine Angaben zu spezifischen Abbruchkriterien in den Unterlagen. Diese sind vom pU zu ergänzen.

2.2.10 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 [25].

Davon abweichend gibt der pU an, für die Interpretation der Ergebnisse der AbD keine verschobene Hypothesentestung durchzuführen. Ebenso will er den dramatischen Effekt zur Einordnung seiner Ergebnisse heranziehen. Diesen definiert er als eine Effektschätzung $< 0,5$ (relativ) und ein Signifikanzniveau von 1 %. Die vom pU gewählte Definition weicht von der des IQWiG ab, welches als Orientierung die von Glasziou et al. genannten Grenzen eines Signifikanzniveaus von 1 % sowie eines relativen Risikos von 10 heranzieht [15,28]. Unabhängig vom Vorgehen des pU werden im Rahmen der erneuten Nutzenbewertung die vom pU vorgelegten Ergebnisse unter Berücksichtigung einer verschobenen Nullhypothese interpretiert.

2.2.11 Zeitplan für die AbD

Der G-BA fordert Statusberichte nach 6, 18, 36 und 54 Monaten nach Beginn der AbD sowie Zwischenanalysen nach 18, 36 und 54 Monaten. Auf Basis der Zwischenanalyse nach 18 Monaten soll gemäß Beschluss des G-BA anhand der dann möglichen, genaueren Effektannahmen eine endgültige Fallzahlschätzung vorgenommen werden. Zudem ist bei jeder Zwischenanalyse eine Prüfung auf Vergeblichkeit vorzunehmen.

Der Zeitplan des pU weicht von der Forderung des G-BA ab, da sich der pU an den aktuellen Fristen des DHR orientiert. Das DHR stellt die Daten eines Jahres jeweils in der Mitte des Folgejahres zur Verfügung. Entsprechend gibt der pU für seine Berichte zwar angemessene Datenschnitte an, allerdings decken die dabei eingehenden Daten nicht die vom G-BA geforderten Zeiträume ab, mit großem Zeitverzug (siehe Tabelle 1). Der pU gibt an, dass sowohl die Fallzahlschätzung als auch die Vergeblichkeitsprüfung erst in der

2. Zwischenanalyse (nach 36 Monaten) erfolgen kann, da diese aufgrund der verzögerten Datenverfügbarkeit zum Zeitpunkt der 1. Zwischenanalyse als nicht durchführbar erachtet werden. Dies ist nicht sachgerecht. Es liegt in der Verantwortung des pU, Daten vorzulegen, die den vom G-BA geforderten Zeitraum abdecken. Dazu sollten jeweils Daten bis zu 4 Monate vor der jeweiligen Zwischenanalyse umfasst sein.

Tabelle 1: Übersicht über die Zeitpunkte und Zeiträume der geforderten und geplanten Statusberichte und Zwischenanalysen

	Vom G-BA geforderter Zeitpunkt	Geplanter Datenschnitt ^a	Eingehende Daten	Zeitverzug ^b
1. Statusbericht	6 Monate nach Beginn der AbD	–	–	6 Monate
2. Statusbericht & 1. Zwischenanalyse	18 Monate nach Beginn der AbD	07/2025	8 Monate (bis 31.12.2024)	10 Monate
3. Statusbericht & 2. Zwischenanalyse	36 Monate nach Beginn der AbD	01/2027	20 Monate (bis 31.12.2025)	16 Monate
4. Statusbericht & 3. Zwischenanalyse	54 Monate nach Beginn der AbD	07/2028	44 Monate (bis 31.12.2027)	10 Monate
a. Der geplante Datenschnitt liegt jeweils 4 Monate vor der Einreichung beim G-BA. b. bezogen auf die geplante Einreichung beim G-BA				

Überdies gibt der pU an, in den vom G-BA geforderten Statusberichten 6 und 18 Monate nach Beginn der AbD jeweils Angaben zum Stand der technischen Umsetzung der Änderungen im DHR sowie zur Umsetzung der Source Data Verification vorzulegen (siehe Abschnitt 2.2). Dies ist nicht sachgerecht. Alle erforderlichen Anpassungen müssen vor Beginn der AbD erfolgen.

2.2.12 Sonstige Anmerkungen

Einschränkung der teilnehmenden Zentren

Gemäß SP sollen nur Zentren an der AbD teilnehmen, in denen mindestens 10 Patienten mit Hämophilie B behandelt werden.

Der pU selbst geht davon aus, dass aufgrund der Seltenheit der Erkrankung nur eine begrenzte Anzahl an Patienten für die AbD zur Verfügung steht. Zudem gibt der pU an, dass die Bereitschaft der Behandlungszentren zur Teilnahme an der AbD nicht vorhersehbar ist. Vor diesem Hintergrund sollte keine Einschränkung der teilnehmenden Behandlungszentren erfolgen. Vielmehr sollte der pU durch geeignete Maßnahmen sicherstellen, dass eine ausreichende Anzahl an Patienten in die Studie eingeschlossen wird.

2.3 Zusammenfassung und Fazit

Die Prüfung des SP (Version vom 09.10.2023), SAP (Version vom 09.10.2023) und des Berichts zur Methodik bei der Confounderidentifikation für die AbD zu Etranacogen Dezaparvovec 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-Anpassungen am DHR (siehe Abschnitt 2.2)
- Verpflichtende Erhebung aller relevanten Daten im DHR (siehe Abschnitte 2.2.1.1, 2.2.1.3 und 2.2.2)
- Angaben zu geplanten Patientencharakteristika (siehe Abschnitt 2.2.2)
- Einheitliche Erhebungs- und Meldezeitpunkte (siehe Abschnitte 2.2.1.3 und 2.2.2)
- konkrete Aussagen zur (vorläufigen) Fallzahl (siehe Abschnitt 2.2.4)
- Identifikation und Erhebung der relevanten Confounder (siehe Abschnitt 2.2.5.1) und Adjustierung für Confounder (siehe Abschnitt 2.2.5.2)
- genauere Angaben zur geplanten Teststatistik bei den geplanten Auswertungen für die Endpunkte der AbD (siehe Abschnitt 2.2.6)
- detaillierte Beschreibung der Maßnahmen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung (siehe Abschnitt 2.2.8.1)
- detaillierte Beschreibung des Umgangs mit fehlenden Werten (siehe Abschnitt 2.2.8.2)
- Beschreibung der Abbruchkriterien im Rahmen der Prüfung auf Vergeblichkeit (siehe Abschnitt 2.2.9)
- Interpretation der Ergebnisse der AbD (siehe Abschnitt 2.2.10)
- Zeitplan für die AbD (siehe Abschnitt 2.2.11)

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); Etranacogen Dezaparovec (Hämophilie B); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2023 [Zugriff: 25.10.2023]. URL: https://www.g-ba.de/downloads/39-261-6010/2023-05-12_AM-RL-XII_Etranacogen-Dezaparovec_2022-AbD-005_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); Etranacogen Dezaparovec (Hämophilie B); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2023 [Zugriff: 25.10.2023]. URL: https://www.g-ba.de/downloads/40-268-9494/2023-05-12_AM-RL-XII_Etranacogen-Dezaparovec_2022-AbD-005_Forderung_TrG.pdf.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Etranacogen Dezaparovec (Hämophilie B) – Bewertung gemäß § 35a SGB V; AbD-Konzept [online]. 2023 [Zugriff: 11.07.2023]. URL: https://www.iqwig.de/download/a22-83_etranacogen-dezaparovec-haemophilie-b_abd-konzept_v1-1.pdf.
4. CSL Behring. Routine Practice Data Collection and Evaluation of etranacogene dezaparovec (Hemgenix) and prophylactic F actor IX (FIX) replacement in severe and moderately severe haemophilia B without a history of FIX inhibitors; a prospective, non interventional study mandated by GBA; Study Protocol [unveröffentlicht]. 2023.
5. CSL Behring. Routine Practice Data Collection and Evaluation of etranacogene dezaparovec (Hemgenix) and prophylactic Factor IX (FIX) replacement in severe and moderately severe haemophilia B without a history of FIX inhibitors: a prospective, non interventional study mandated by GBA; Statistical Analysis Plan (SAP) [unveröffentlicht]. 2023.
6. CSL Behring. Methodology of Confounder Identification; Etranacogene dezaparovec [unveröffentlicht]. 2023.
7. Gemeinsamer Bundesausschuss. Niederschrift zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV; Beratungsanforderung 2023-B-164 [unveröffentlicht]. 2023.
8. 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.

9. 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>.
10. 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>.
11. 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>.
12. 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>.
13. 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>.
14. 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>.
15. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 21.09.2023]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf.
16. 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.
17. 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>.
18. 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>.

19. European Medicines Agency. Hemgenix; Assessment report [online]. 2023 [Zugriff: 06.11.2023]. URL: https://www.ema.europa.eu/documents/assessment-report/hemgenix-epar-public-assessment-report_en.pdf.
20. 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>.
21. 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>.
22. Pufulete M, Mahadevan K, Johnson TW et al. Confounders and co-interventions identified in non-randomized studies of interventions. *J Clin Epidemiol* 2022; 148: 115-123. <https://dx.doi.org/10.1016/j.jclinepi.2022.03.018>.
23. Stürmer T, Webster-Clark M, Lund JL et al. Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study. *Am J Epidemiol* 2021; 190(8): 1659-1670. <https://dx.doi.org/10.1093/aje/kwab041>.
24. 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>.
25. 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_versorgungснаhe-daten-zum-zwecke-der-nutzenbewertung_rapid-report_v1-1.pdf.
26. 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>.
27. 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>.
28. Glasziou P, Chalmers I, Rawlins M et al. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351. <https://dx.doi.org/10.1136/bmj.39070.527986.68>.

Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern

Informationsbeschaffung

Zur Identifizierung von relevanten Confoundern führt der pU eine Informationsbeschaffung nach systematischen Übersichten und Leitlinien zu Hämophilie B in folgenden Quellen durch:

- bibliografische Recherchen in MEDLINE und Handsuche auf ausgewählten Webseiten nach Leitlinien (letzte Suche am 16.05.2023)
- bibliografische Recherchen in MEDLINE und der Cochrane Database of Systematic Reviews nach systematischen Übersichten / Metaanalysen (letzte Suche am 16.05.2023)

Die Validierung der identifizierten Confounder erfolgte durch 3 klinische Experten im Rahmen eines Workshops. Hierfür wurden alle identifizierten Confounder diskutiert und in 3 Gruppen (wichtig, weniger wichtig und nicht wichtig) kategorisiert. Abschließend nutzt der pU die zuvor identifizierte Evidenz für die Bestimmung von Interaktionen zwischen den relevanten Confoundern.

Kommentar zur Informationsbeschaffung

Die Informationsbeschaffung des pU ist nicht geeignet, eine vollständige Liste an Confoundern zu identifizieren.

Dies hat insbesondere folgende Gründe:

Für die systematische Identifizierung von Confoundern muss gewährleistet sein, dass geeignete Primärstudien in ausreichendem Umfang identifiziert werden. Dazu gehören insbesondere epidemiologische Studien sowie Untersuchungen auf Effektmodifikationen in randomisierten kontrollierten Studien (RCTs). Ein Rückgriff auf Sekundärliteratur, wie im vorliegenden Fall, ist prinzipiell möglich. Allerdings muss zusätzlich gewährleistet sein, dass die Sekundärliteratur die geeigneten Primärstudien in ausreichendem Umfang enthält und auch für die Bewertung der Confounder notwendige statistische Evidenz berichtet wird. Dieser Prüfschritt fehlt beim Vorgehen des pU. Dass ein solcher Prüfschritt nicht erfolgte, wird auch dadurch deutlich, dass trotz der vom pU genannten Ein- und Ausschlusskriterien eine systematische Übersicht mit Hämophilie A Patienten eingeschlossen wird. Bei 2 weiteren vom pU als relevant eingeschlossenen Publikationen handelt es sich um unsystematische Übersichtsarbeiten auf Basis von Expertenmeinungen bzw. einer selektiven Auswahl von Studien. Es wäre stattdessen sinnvoll gewesen, eine Suche nach Primärstudien mit Stoppkriterien wie in Pufulete 2022 beschrieben anzuschließen [22].