

Anwendungsbegleitende Datenerhebung zu Brexucabtagen Autoleucel: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0)

2. Addendum zum Projekt A21-130

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 11th, 12th, and 13th segments from the left.

ADDENDUM

Projekt: A22-136

Version: 1.0

Stand: 08.02.2023

IQWiG-Berichte – Nr. 1511

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Anwendungsbegleitende Datenerhebung zu Brexucabtagen Autoleucel: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0) – 2. Addendum zum Projekt A21-130

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

22.12.2022

Interne Projektnummer

A22-136

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
- Ralf Bender
- Thomas Kaiser
- Guido Skipka
- Siw Waffenschmidt

Schlagwörter

Brexucabtagen Autoleucel, Lymphom – Mantelzell, Register, Nutzenbewertung, Forschungsdesign, Peer Review – Forschung

Keywords

Brexucabtagene Autoleucel, Lymphoma – Mantle Cell, Registries, Benefit Assessment, Research Design, Peer Review – Research

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis.....	v
1 Hintergrund.....	1
2 Prüfung der Unterlagen zur Planung der AbD von Brexucabtagen Autoleucel	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	6
2.2.3 Studienpopulation der Registerstudie	8
2.2.3.1 Fallzahlplanung	8
2.2.3.2 Rekrutierung der Studienpopulation.....	9
2.2.3.3 Auswertungspopulationen	9
2.2.4 Identifikation und Erhebung von Confoundern, Adjustierung für Confounder... 10	10
2.2.4.1 Identifikation und Erhebung wichtiger Confounder.....	10
2.2.4.2 Berücksichtigung wichtiger Confounder bei der Analyse.....	11
2.2.5 Anmerkungen zu den geplanten Auswertungen für die Endpunkte der AbD	13
2.2.5.1 Auswertung des Endpunkts Gesamtmortalität	13
2.2.5.2 Auswertung der PRO-Endpunkte (Symptome und gesundheitsbezogene Lebensqualität)	13
2.2.5.3 Auswertung der UE-Endpunkte	14
2.2.6 Untersuchung auf potenzielle Effektmodifikatoren und Subgruppenanalysen ..	14
2.2.7 Qualität und Vollständigkeit der Datenerhebung.....	14
2.2.7.1 Maßnahmen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung.....	14
2.2.7.2 Umgang mit fehlenden Daten	15
2.2.8 Anpassung der AbD	16
2.2.9 Interpretation der Ergebnisse der AbD	17
2.2.10 Zeitplan für die AbD.....	17
2.2.11 Sonstige Anmerkungen	17
2.3 Zusammenfassung und Fazit	18

3 Literatur 20
Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern .. 25

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AbD	Anwendungsbegleitende Datenerhebung
AT	As treated
BTK	Bruton-Tyrosin-Kinase
DRM	Data Review Meeting
CTCAE	Common Terminology Criteria for Adverse Events
ECOG	Eastern Co-operative Oncology Group
EMCL-R	European Mantel-Cell-Lymphoma Registry (Europäisches Mantelzelllymphom-Register)
EORTC	European Organisation For Research And Treatment Of Cancer
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
MCL	Mantle Cell Lymphoma (Mantelzell-Lymphom)
MedDRA	Medical Dictionary for Regulatory Activities
OS	Overall Survival (Gesamtüberleben)
PICO	Patient, Intervention, Comparator, Outcome
PS	Propensity Score
PT	Preferred Term (bevorzugter Begriff)
pU	pharmazeutischer Unternehmer
SAP	statistischer Analyseplan
SOC	System Organ Class (Systemorganklasse)
SGB	Sozialgesetzbuch
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 22.12.2022 mit der Prüfung des Studienprotokolls (SP) und des statistischen Analyseplans (SAP) zur anwendungsbegleitenden Datenerhebung (AbD) zu Brexucabtagen Autoleucel beauftragt.

In seiner Sitzung am 21.07.2022 hat der G-BA beschlossen, eine AbD und Auswertungen nach § 35a Absatz 3b Satz 1 SGB V für den Wirkstoff Brexucabtagen Autoleucel in der Indikation „rezidiertes oder refraktäres Mantelzelllymphom“ zu fordern [1,2]. Dem Beschluss liegt unter anderem das vom IQWiG erstellte Konzept für eine AbD zu Brexucabtagen Autoleucel zugrunde (Rapid Report A21-130 vom 31.03.2022 [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 Brexucabtagen Autoleucel sollen die Inhalte der diesbezüglichen Beratung des pU zur Studienplanung der AbD berücksichtigt werden (Beratungsanforderung 2022-B-219 [6]).

2 Prüfung der Unterlagen zur Planung der AbD von Brexucabtagen Autoleucel

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 1.0 vom 21.12.2022 [4,5]. Teil des Studienprotokolls ist ein in sich abgeschlossener Bericht über eine systematische Literatursuche zur Identifikation von Confoundern (Version 1 vom 08.12.2022, Anhang 2 des Studienprotokolls).

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 Europäische Mantelzelllymphom-Register (EMCL-R) als primäre Datenquelle genutzt werden, sofern die im G-BA-Beschluss aufgeführten Qualitätskriterien erfüllt sind. Der pU wählt das EMCL-R als Datenquelle für die AbD.

In den nachfolgenden Abschnitten werden die vom pU vorgelegten Dokumente dahingehend geprüft, ob durch sie die Durchführung der AbD unter Verwendung des EMCL-R 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 Fragestellung gemäß PICO

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

2.2.1.1 Population

Gemäß G-BA-Beschluss soll die AbD für folgende Population durchgeführt werden: Erwachsene Patientinnen und Patienten mit rezidiviertem oder refraktärem Mantelzell-Lymphom (MCL) nach 2 oder mehr systemischen Therapien, die einen Bruton-Tyrosinkinase-(BTK)-Inhibitor einschließen. Dabei sind gemäß G-BA-Beschluss zudem die Kriterien für die Eignung einer Behandlung mit Brexucabtagen Autoleucel anzuwenden, das bedeutet: Es sollen nur solche Patientinnen und Patienten eingeschlossen werden, für die prinzipiell eine Therapie mit Brexucabtagen Autoleucel in Frage kommen. Patientinnen und Patienten, für die eine Therapie mit Brexucabtagen Autoleucel z. B. aufgrund des Allgemeinzustands nicht in

Frage kommt, sollen nicht in die AbD eingeschlossen werden, auch nicht in die Vergleichsgruppe.

Zwar stimmen die Einschlusskriterien der geplanten AbD gemäß Studienprotokoll prinzipiell mit den Vorgaben des G-BA überein. Allerdings bleibt in der vorliegenden Version der Dokumente unklar, ob die Anforderung „Anwendung der Kriterien für die Eignung einer Behandlung mit Brexucabtagen Autoleucel“ inhaltlich sachgerecht umgesetzt wird. Denn aus dem Studienprotokoll, Abschnitt 4.1 „Inclusion Criteria“ geht hervor, dass die konkrete Umsetzung dieses Punktes noch gar nicht feststeht („currently under discussion (feasibility of implementation by EMCL-R not clear)“). Da vor Start der Datenerhebung feststehen muss, wie das Eignungskriterium umgesetzt wird und diese Umsetzung sachgerecht sein muss, stellt dies einen wesentlichen Mangel im Studienprotokoll dar.

In diesem Zusammenhang fällt auf, dass in Abbildung 1 des Studienprotokolls zum Patientenfluss eine Gruppe abgegrenzt wird, für die nach Einschätzung des Tumorboards eine Therapie mit Brexucabtagen Autoleucel nicht in Frage kommt, die aber der Vergleichsgruppe zugeordnet wird. Dies ist aus den zuvor genannten Gründen nicht sachgerecht.

Schließlich ist offen, ob die alleinige Angabe „prinzipielle Eignung für Brexucabtagen Autoleucel: ja“ ausreichend ist, um das mit dem Eignungskriterium verbundene Ziel der Positivität (grundsätzliche Eignung der Patientin bzw. des Patienten für Brexucabtagen Autoleucel und eine der vom G-BA genannten Vergleichstherapien) zu erreichen. Es sollten soweit möglich konkrete Ausschlusskriterien benannt werden, aufgrund derer eine Therapie mit Brexucabtagen Autoleucel auszuschließen ist. Dazu gehören mindestens die Kontraindikationen von Cyclophosphamid und Fludarabin, da eine Lymphodepletion mit diesen Wirkstoffen im Vorfeld einer Therapie mit Brexucabtagen Autoleucel erforderlich ist [7].

2.2.1.2 Intervention und Kontrolle

Die Benennung der zu vergleichenden Therapieoptionen stimmt mit den Vorgaben des G-BA überein (Intervention: Brexucabtagen Autoleucel; Kontrolle: Patientenindividuelle Therapie, mögliche Wirkstoffe und Wirkstoffkombinationen konkretisiert in [1]).

2.2.1.3 Endpunkte (Outcomes)

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

Gesamtüberleben

Dieser Endpunkt ist Gegenstand der AbD und wird gemäß Studienprotokoll zur AbD sachgerecht erhoben.

Patientenberichtete Endpunkte (PROs): Symptome und gesundheitsbezogene Lebensqualität

Die Erhebung der PROs zur Symptomatik und zur gesundheitsbezogenen Lebensqualität stellt eine wesentliche Stärke der geplanten AbD dar.

Zwar werden diese Endpunkte im EMCL-R bislang nicht erhoben. Zum Zwecke der AbD soll jedoch eine separate Erhebung durch eine dritte Partei erfolgen, verbunden mit einer Erinnerung im Falle fehlenden Rücklaufs. Für die PRO-Erhebung werden standardisierte Erhebungszeitpunkte vorgegeben, die unter Abwägung von Aufwand, Machbarkeit und Aussagekraft sachgerecht erscheinen (Erhebung jeweils 1, 3, 6, 12, 24 und 36 Monate nach Indexdatum). In diesem Zusammenhang ist besonders positiv hervorzuheben, dass die Erhebung bis Studienende (Tod oder Beobachtungsende) erfolgen soll und nicht vorzeitig, z. B. bei Progression der Erkrankung, beendet wird. Dadurch wird eine Beschreibung der Symptome und der gesundheitsbezogenen Lebensqualität über den gesamten Studienverlauf ermöglicht. Es sollte allerdings erwogen werden, dies durch entsprechende Auswertungen zu gewährleisten (siehe Abschnitt 2.2.5.2).

Zur PRO-Erhebung werden von der European Organisation For Research And Treatment Of Cancer (EORTC) entwickelte Fragebogen herangezogen, und zwar der indikationsübergreifende Fragebogen QLQ-C30 sowie der Lymphom-spezifische Fragebogen QLQ-NHL-HG29. Zu letzterem liegt zwar noch keine Publikation zur abschließenden Validierungsstudie der Phase IV vor. Auf Nachfrage hat die hierfür verantwortliche Gruppe am 18.01.2023 jedoch bestätigt, dass die Validierungsstudie abgeschlossen wurde und sich daraus keine Veränderungen am bislang verfügbaren Fragebogen ergeben haben. Die Verwendung der beiden genannten Fragebogen für die geplante AbD ist daher sachgerecht.

Ebenfalls sachgerecht ist die im Studienprotokoll dargestellte Zuordnung der einzelnen Skalen der beiden Fragebogen zum Endpunkt „Symptome“ bzw. „gesundheitsbezogene Lebensqualität“. Das gilt zwar prinzipiell gemäß Manual des Fragebogens QLQ-C30 auch für die Skala „Financial difficulties“ zur Kategorie „Symptome“. Allerdings bildet diese Skala kein Symptom im eigentlichen Sinne ab, weshalb sie regelhaft nicht bei der Nutzenbewertung herangezogen wird (siehe z. B. [8-10]). Sie sollte daher auch nicht Gegenstand der geplanten AbD sein.

Offen bleibt aus den Beschreibungen im Studienprotokoll, ob die behandelnden Zentren Informationen über die Ergebnisse der jeweiligen PRO-Erhebung erhalten. Für einzelne onkologische Entitäten wurde nachgewiesen, dass eine medizinische Betreuung, die PRO-Informationen im Verlauf heranzieht, einer Betreuung ohne solche Informationen überlegen ist [11-13]. Es wäre daher grundsätzlich sinnvoll, diese Informationen an die behandelnden Zentren zeitnah weiterzugeben. Für die AbD ist davon unabhängig sicherzustellen und im

Studienprotokoll entsprechend zu beschreiben, dass das diesbezügliche Vorgehen für beide Behandlungsgruppen konsistent ist, d. h. die Informationen regelhaft nicht oder für beide Gruppen vollständig zur Verfügung gestellt werden.

Unerwünschte Ereignisse

Der pU plant, im Zusammenhang mit der Erhebung unerwünschter Ereignisse (UEs) folgende Endpunkte zu erheben:

- UEs, die zu einer Hospitalisierung führen oder eine bestehende Hospitalisierung verlängern
- UEs, die zu einer Hospitalisierung führen oder eine bestehende Hospitalisierung verlängern, und bei denen nach Einschätzung der Studienärztin bzw. des Studienarztes das jeweilige UE auf die Behandlung zurückzuführen ist
- UEs, die zum Tod führen
- definierte spezifische UEs, die zu einer Hospitalisierung führen oder eine bestehende Hospitalisierung verlängern

Dabei sollen die einzelnen UEs nach dem Medical Dictionary for Regulatory Activities (MedDRA) kodiert werden unter Verwendung der System Organ Class (SOC) und der Preferred Terms (PT). Der pU führt im Studienprotokoll aus, dass damit der Komplex UEs ausreichend abgedeckt sei, auch wenn der G-BA in seinen Anforderungen neben spezifischen UEs die Erhebung von SUEs, schweren UEs und Abbrüchen wegen UEs gefordert habe.

Schwerwiegende UEs (SUEs)

Mit der Erhebung von UEs, die zur Hospitalisierung führen bzw. eine bestehende Hospitalisierung verlängern, und solchen, die zum Tod führen, sei eine ausreichende Annäherung an den Endpunkt schwerwiegende UEs (SUEs) gewährleistet. Dies ist nachvollziehbar, allerdings sollte daher auch eine gemeinsame Auswertung dieser beiden Komponenten geplant werden.

Schwere UEs

Im Studienprotokoll wird zur Begründung für den Verzicht auf die Erhebung schwerer UEs zum einen ausgeführt, dass die Erhebung schwerer UEs unter Verwendung der Common Terminology Criteria for Adverse Events (CTCAE) mit Grad 3 nicht Bestandteil der Versorgungspraxis sei. Dies gilt allerdings auch für die Erhebung von Symptomen und der gesundheitsbezogenen Lebensqualität mit validierten Skalen, die daher spezifisch für die AbD zusätzlich erhoben werden. Es wäre daher prinzipiell möglich, auch schwere UEs zum Zwecke der AbD auf eine ähnliche Art und Weise separat zu erheben.

Des Weiteren gibt der pU an, dass neben SUEs auch schwere UEs durch die Kombination der aus UEs, die zur Hospitalisierung führen bzw. eine bestehende Hospitalisierung verlängern, und solchen, die zum Tod führen, ausreichend abgebildet seien. Im Gegensatz zur allgemeinen Definition der SUE sind jedoch schwere UEs dadurch gekennzeichnet, dass auch Ereignisse, die die Aktivitäten des täglichen Lebens deutlich beeinträchtigen, aber nicht zur Hospitalisierung führen, davon umfasst sind. Entsprechend ist in der Studie ZUMA-2 (Zulassungsstudie für Brexucabtagen Autoleucel) die Rate schwerer UEs insgesamt und für verschiedene spezifische UEs deutlich höher als die SUE-Rate [14].

Unter Abwägung von Aufwand und Relevanz für die vorliegende AbD sollte zusammenfassend erwogen werden, an Stelle einer vollständigen Erhebung schwerer UEs lediglich für die im Studienprotokoll genannten spezifischen UEs das jeweilige in der CTCAE-Klassifikation genannte Kriterium für einen CTCAE-Grad 3 oder höher oder das allgemeine Kriterium „deutliche Beeinträchtigung der Aktivitäten des täglichen Lebens“ zusätzlich zu erheben und diese Ereignisse entsprechend auszuwerten.

Abbruch wegen UEs

Der Verzicht auf die Erhebung des Endpunkts „Abbruch wegen UE“ ist unter Abwägung von Aufwand und Relevanz für die vorliegende AbD nachvollziehbar, zumal der Endpunkt aufgrund der unterschiedlichen Therapieformen (einmalige Therapie mit Brexucabtagen Autoleucel gegenüber länger andauernden Vergleichstherapien) kaum sinnvoll zu interpretieren ist.

Spezifische UEs

Die im Studienprotokoll genannte Liste spezifischer UEs bildet die wichtigsten spezifischen UEs für beide Vergleichsarme ausreichend ab und ist daher sachgerecht.

Weitere Anmerkungen

Die geplante Abgrenzung von UEs, bei denen nach Einschätzung der Studienärztin bzw. des Studienarztes das jeweilige UE auf die Behandlung zurückzuführen ist, ist nicht erforderlich. Diese Art der UE-Auswertung ist für die Nutzenbewertung prinzipiell nicht relevant, und damit auch nicht für die geplante AbD.

2.2.2 Datenquelle

Der pU wählt das EMCL-R als Datenquelle für die AbD. Dieses Register wurde auch im Konzept für die AbD als potenziell geeignetes Indikationsregister identifiziert (Rapid Report A21-130 [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.

In Abschnitt 5.1 des Studienprotokolls findet sich zu den vom G-BA benannten Qualitätskriterien sowie weiteren Kriterien, die sich aus den Tragenden Gründen zum AbD-Beschluss ableiten lassen, jeweils eine Einschätzung, ob diese aus Sicht des pU erfüllt sind. Aus den Ausführungen im Studienprotokoll wird insgesamt sichtbar, dass umfangreiche Anpassungen am EMCL-R vorgenommen wurden oder werden sollen, mit denen ein Großteil der Qualitätskriterien adressiert wird. Zu folgenden Kriterien gibt es jedoch weiterhin Änderungsbedarf:

- Kriterium „Exakte Definitionen bzw. Operationalisierung von Expositionen (Art und Dauer der medikamentösen Therapie und anderer Begleittherapien), klinischen Ereignissen, Endpunkten und Confoundern“: Der Erfüllungsgrad wird im Studienprotokoll als „project-specific“ bezeichnet, wobei unklar bleibt, was dies im Zusammenhang mit diesem Kriterium bedeutet. Darüber hinaus finden sich zwar an verschiedenen Stellen im Studienprotokoll und SAP Definitionen und Operationalisierungen zu Baseline-Daten oder Ereignissen im Studienverlauf. Allerdings liegt eine entsprechende zusammenfassende Variablenliste nur für die Baseline-Daten vor (Tabelle 4 im Studienprotokoll bzw. Tabelle 2 im SAP), nicht jedoch für die Verlaufsdaten (z. B. Endpunkte). Eine eindeutige Variablenliste der Verlaufsdaten für die AbD ist zwingend zu ergänzen. Zudem ist die Variablenliste für Baseline-Daten zu finalisieren (es fehlt die Operationalisierung der PRO-Erhebung zu Baseline).
- Kriterium „Verwendung exakter Datumsangaben zur Patientin bzw. zum Patienten, zur Erkrankung, zu wichtigen Untersuchungen und zu Behandlungen / Interventionen“: Der Erfüllungsgrad wird im Studienprotokoll als „limited“ bezeichnet, da für „assessments“ exakte Datumsangaben fehlen. Sofern sich diese Einschränkung allein auf anamnestische Angaben beschränkt, erscheint dies akzeptabel. Aus Tabelle 4 im Studienprotokoll ist z. B. ersichtlich, dass im Falle einer zuvor durchgeführten Stammzelltherapie nur eine kategorielle Einstufung im EMCL-R dokumentiert wird (mehr oder höchstens 12 Monate Abstand zwischen der Stammzelltherapie und dem Indexdatum der Registerstudie). Aufgrund der fehlenden Variablenliste der Verlaufsdaten bleibt die Relevanz der nur eingeschränkten Erfüllung des Kriteriums für die AbD jedoch insgesamt unklar.
- Kriterium „Strategien zur Vermeidung ungewollter Selektionen beim Patienteneinschluss, um eine Repräsentativität zu erreichen“: Der Erfüllungsgrad wird im Studienprotokoll als „N/A“ (nicht zutreffend) bezeichnet. Dies wird damit begründet, dass alle Patientinnen und Patienten, die die Einschlusskriterien erfüllen, im EMCL-R dokumentiert werden können. Dies allein vermeidet jedoch weder Selektion, noch gewährleistet es Repräsentativität der Studienpopulation für die Registerstudie. Im Fall der vorliegenden AbD werden ausweislich der Angaben in Abschnitt 3.2 des Studienprotokolls (Screening Procedure) offenbar Maßnahmen zur Intensivierung der Rekrutierung eingesetzt, die sich zum einen allein auf Zentren beziehen, die eine

Therapie mit Brexucabtagen Autoleucel durchführen können („CAR-T qualified centers will be contacted weekly“). Zum anderen informiert der pU die Registerbetreiber aktiv, falls ein Zentrum eine Therapie mit Brexucabtagen Autoleucel anfordert. Es sollte erwogen werden, die erst genannte Maßnahme auf andere Studienzentren zu erweitern und, sofern möglich, die 2. genannte Maßnahme auf Vergleichstherapien. Dies wäre generell sinnvoll im Hinblick auf eine verstärkte Rekrutierung für die Vergleichsgruppe.

- Kriterium „Gewährleistung wissenschaftlicher Unabhängigkeit und Transparenz“: Der Erfüllungsgrad wird im Studienprotokoll mit „yes“ angegeben. Dem widerspricht jedoch der Hinweis in Abschnitt 11 des Studienprotokolls, dass alle Abstracts, Poster und Publikationen auch vom pU freigegeben werden müssen.
- Ausführungen aus den Tragenden Gründen zum Beschluss zur Erweiterung des Datensatzes um bislang nicht dokumentierte Confounder: Der Erfüllungsgrad wird im Studienprotokoll mit „yes“ angegeben. Da jedoch die Auswahl der Confounder nicht sachgerecht ist (siehe Abschnitt 2.2.4) ist unklar, ob dieser Erfüllungsgrad auch für eine adäquate Confounderliste angenommen werden kann. Dies wäre auf Basis der erweiterten Confounderliste zu gewährleisten.

2.2.3 Studienpopulation der Registerstudie

2.2.3.1 Fallzahlplanung

Der pU beschreibt eine Fallzahlberechnung für den Endpunkt Mortalität mit dem Hazard Ratio (HR) als Effektmaß. Er setzt dabei aufgrund des nicht randomisierten Designs richtigerweise eine verschobene Hypothesengrenze an, allerdings ist diese mit einer Konfidenzintervallgrenze von 0,85 außerhalb des Bereichs, der für eine sachgerechte Interpretation der Daten erforderlich ist (in der Regel 0,2 bis 0,5 je nach Qualität der Datenerhebung und -auswertung). Der pU argumentiert, dass er nicht von einem Effekt in einer solchen Größenordnung ausgeht. Hierzu ist anzumerken, dass sich der als Ergebnis der AbD anzunehmende Effekt aus dem wahren Unterschied zwischen Brexucabtagen Autoleucel und der Vergleichstherapie sowie der durch das nicht randomisierte Studiendesign bedingten Verzerrung zusammensetzt. Wie im Konzept zur AbD beschrieben und in den Beratungen verdeutlicht, kann aus einer nicht randomisierten Studie aufgrund potenziell unbekannter Confounder aus den in der Studie beobachteten Effekten daher erst ab einer bestimmten Effektstärke eine Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Die konkrete Schwelle ergibt sich durch die Qualität der Daten, u. a. durch das Wissen über und die Berücksichtigung der relevanten Confounder. Der pU begründet die Abschwächung der Grenze jedoch ausschließlich über die zu erwartende wahre Effektstärke ohne Berücksichtigung einer potenziellen Verzerrung.

Davon unabhängig liegt es letztlich im Verantwortungsbereich des pU, eine ausreichend große Fallzahl zu rekrutieren, um einen bestehenden wahren Unterschied zwischen Brexucabtagen

Autoleucel und der Vergleichstherapie durch die AbD mit ausreichender Sicherheit nachzuweisen. Insofern ist die Interpretation der Ergebnisse nicht durch die im Studienprotokoll verwendete Hypothesengrenze für die Fallzahlberechnung des pU präjudiziert (hier 0,85), sondern ergibt sich letztlich aus der abschließenden Qualität der Datenerhebung und -auswertung (in der Regel durch eine verschobene Hypothesengrenze im Bereich 0,2 bis 0,5 [15], siehe auch Abschnitt 2.2.9).

2.2.3.2 Rekrutierung der Studienpopulation

Das EMCL-R ist als europäisches Netzwerk ausgelegt und bindet daher auch Zentren außerhalb Deutschlands ein. Im Studienprotokoll finden sich zwar vereinzelt Hinweise auf die mögliche Einbindung von Zentren außerhalb Deutschlands in die AbD. Allerdings ist der entsprechende Prozess offenbar noch nicht abgeschlossen und die Einbindung weiterer europäischer Zentren unklar:

- In Abschnitt 3.2 des Studienprotokolls heißt es: „The selection of further countries remains to be determined.“
- In Abschnitt 3.6 des Studienprotokolls heißt es: „The possibility/feasibility to recruit patients from other European EMCL-R sites and collect data as required by the present protocol is being studied.“

Die europäische Ausrichtung des EMCL-R war auch Gegenstand des Fachaustauschs zum AbD-Konzept [16]. Dabei wurde festgestellt, dass die Einbindung der europäischen Zentren außerhalb Deutschlands voraussichtlich für die Durchführung der AbD erforderlich ist, um die notwendige Patientenzahl insbesondere für die Vergleichsgruppe zu rekrutieren. Es wäre daher sinnvoll, wenn die Einbindung der Länder bzw. Zentren außerhalb Deutschlands vor Beginn der eigentlichen Datenerhebung geklärt und dies im Studienprotokoll beschrieben wird.

2.2.3.3 Auswertungspopulationen

Der pU beabsichtigt, 2 Auswertungspopulationen zu verwenden. Dies ist zum einen eine Intent-to-treat(ITT)-Population, die alle Personen umfasst, für die eine der zu vergleichenden Therapien geplant ist. Dies soll die primäre Analyse sein, was der Anforderung des G-BA entspricht. Zum anderen soll zusätzlich eine As-treated(AT)-Population ausgewertet werden, die nur die Personen enthält, die die geplante Therapie auch erhalten haben. In den Hauptanalysen sollen mögliche Therapiewechsel (Treatment Switches) unbeachtet bleiben. Lediglich im Rahmen von Sensitivitätsanalysen sollen mögliche Therapiewechsler zum Zeitpunkt des Wechsels zensiert werden.

Das Vorgehen ist insgesamt sachgerecht.

2.2.4 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 [17-19],

- 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.

2.2.4.1 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 (Version 1 vom 08.12.2022, Anhang 2 des Studienprotokolls [4]). Eine 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 dieser Suche eine umfangreiche Liste potenzieller Confounder, allerdings ist die mangelnde Aufbereitung der zugehörigen statistischen Evidenz für die Zwecke der AbD unzureichend. Dies kann auch nicht durch die darauf folgende Experteneinschätzung ersetzt werden, da bei dieser die statistische Evidenz berücksichtigt werden muss.

Davon unabhängig ist das Vorgehen der im Studienprotokoll vorgenommenen Einstufung der einzelnen Confounder in „substantial impact“ (d. h. für die AbD relevant) und „no substantial impact“ (d. h. für die AbD nicht relevant) insbesondere aus folgenden Gründen nicht sachgerecht:

- Für die Confounder Komorbiditäten und ECOG (Eastern Co-operative Oncology Group)-Stadium wird als Begründung für die Einstufung „non substantial“ angegeben, dass diese Teil der Einschlusskriterien seien und daher per se das Positivitätskriterium erfüllt sei. Dies ist nicht nachvollziehbar: Weder werden Patientinnen und Patienten mit Komorbiditäten aus der Studie ausgeschlossen, noch werden nur Patientinnen und Patienten mit einem spezifischen ECOG-Stadium (z. B. 1) eingeschlossen, sodass die Verteilung sowohl der Komorbiditäten als auch des ECOG-Stadiums zwischen den

Gruppen unterschiedlich sein kann. Komorbiditäten und das ECOG-Stadium sind relevante Confounder für die vorliegende Fragestellung.

- Für die Confounder Alter und Tumor-Stadium (nach Ann Arbor) wird als Begründung angegeben, dass diese in Subgruppenanalysen untersucht werden. Dies ist jedoch keine adäquate Begründung für deren Nichtberücksichtigung. Richtig ist hingegen, dass im Falle der Durchführung einer Subgruppenanalyse das untersuchte Merkmal bei der konkreten Subgruppenanalyse nicht als Confounder berücksichtigt werden muss (also z. B. das Alter nicht bei den Subgruppenanalysen zu Altersgruppen). Alter und Tumor-Stadium sind relevante Confounder für die vorliegende Fragestellung.
- Mehrere der vom pU für das Nutzenbewertungsverfahren von Brexucabtagen Autoleucel identifizierten und auch in der Beratungsanfrage für die AbD von ihm als „wichtig“ bezeichneten Confounder werden vom pU jetzt nicht mehr berücksichtigt. Dazu gehören beispielsweise die Dauer und das Ansprechen auf die vorherige BTK-Inhibitor-Therapie, eine extranodale Erkrankung, der Befall des Knochenmarks und das Geschlecht.
- Obwohl der pU in Abschnitt 1 des Studienprotokolls die Leukozytenzahl neben dem LDH-Wert, dem Alter und dem ECOG-Stadium als unabhängigen prognostischen Faktor bezeichnet, berücksichtigt er diesen nicht in seiner Confounder-Liste.

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, den Abschnitt zur Identifikation und Festlegung der Confounder im Studienprotokoll grundlegend unter Berücksichtigung der oben genannten Punkte zu überarbeiten.

2.2.4.2 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

Der pU plant die Confounderadjustierung mittels Propensity Score (PS) Matchings. Es soll ein optimales PS-Matching im Verhältnis 2:1 ohne Zurücklegen mit einem Caliper-Abstand von 0,25 der gepoolten Standardabweichung des Logit der PS durchgeführt werden. Das Verfahren des PS-Matchings ist ausreichend genau beschrieben und sachgerecht.

Zur Bewertung der Überlappung findet sich lediglich die Angabe, dass die PS-Histogramme visuell untersucht werden. Es fehlen Kriterien, wann diese visuelle Untersuchung zu einer ausreichenden Überlappung führt und wann nicht.

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 [20]. Die vom pU vorgeschlagene Grenze von 0,25 ist somit zwar sachgerecht, stellt aber das schwächste Kriterium für eine ausreichende Balanciertheit dar. Dies ist bei der Interpretation der Ergebnisse auf Basis einer verschobenen Hypothesengrenze entsprechend zu berücksichtigen [15].

Sofern keine ausreichende Überlappung und Balanciertheit erreicht werden kann, sollen naive Vergleiche oder andere PS-basierte Verfahren (wie z. B. Inverse Probability of Treatment Weighting [IPTW] [21]) durchgeführt werden. Weiterführende Angaben, welche Methoden in diesen Fall zur Anwendung kommen sollen, und welche der Methoden unter welchen Bedingungen jeweils gewählt wird, gibt es allerdings nicht. Diese Vorgehensweise ist nicht sachgerecht.

Die übliche Vorgehensweise wäre, zunächst alternative Verfahren wie z. B. Gewichtungsmethoden zur Anwendung des Propensity Score zu wählen, sodass eine ausreichende Überlappung und eine ausreichende Balanciertheit der zu vergleichenden Gruppen erreicht wird. Der Entscheidungsalgorithmus zur Anpassung der Propensity-Score-Analyse bei fehlender Überlappung und Balanciertheit nach Anwendung des ersten Verfahrens [22] ist im SAP entsprechend zu ergänzen. Ebenso ist zu ergänzen, welche Konsequenz gezogen wird, sofern kein Propensity-Score-Verfahren gefunden werden kann, mit dem eine ausreichende Überlappung und eine ausreichende Balanciertheit der zu vergleichenden Gruppen erreicht werden kann. In einem solchen Fall ist der Versuch einer Effektschätzung weder mithilfe von Propensity Scores noch mithilfe von anderen Adjustierungsverfahren ohne Propensity Scores sinnvoll, und die untersuchte Fragestellung sollte überdacht werden [22,23]. Im Falle eines dann durchgeführten naiven Vergleichs ohne Adjustierung sind die Konsequenzen bei der Ergebnisinterpretation zu beachten und zu beschreiben (Nachweis eines dramatischen Effekts) [15,24].

Beschreibung der Zielpopulation nach PS-Matching

Ein zentraler Aspekt bei nicht randomisierten Vergleichen mit Confounderadjustierung ist die detaillierte Beschreibung der Zielpopulation, die sich aus der Anwendung eines PS-Verfahrens ergibt. Insbesondere beim PS-Matching werden erwartungsgemäß Patienten, die zunächst den Einschlusskriterien entsprechen, aus der Analyse ausgeschlossen. Es muss daher nach erfolgreicher Anwendung eines PS-Verfahrens entweder sorgfältig begründet werden, dass die resultierende AbD-Population der ursprünglichen Zielpopulation in ausreichendem Maß entspricht, oder, wenn dies nicht der Fall ist, zu beschreiben, auf welche Teilpopulation der

ursprünglichen Zielpopulation sich die resultierende Analyse bezieht. Dies fehlt im vorliegenden SAP.

2.2.5 Anmerkungen zu den geplanten Auswertungen für die Endpunkte der AbD

Im Studienprotokoll und SAP sind zu den verschiedenen Endpunkten jeweils endpunktspezifische Ausführungen zu den geplanten Auswertungen enthalten. Diese werden nachfolgend adressiert.

2.2.5.1 Auswertung des Endpunkts Gesamtmortalität

Für den Endpunkt Gesamtmortalität sind Auswertungen mittels Cox-Regression (Hazard Ratio, Schätzer und Konfidenzintervall) und Log-Rank-Test geplant. Zusätzlich sollen Kaplan-Meier-Kurven, medianes Überleben und Überlebensraten in 6-Monatszeitspannen inklusive Konfidenzintervallen dargestellt werden. Dieses Vorgehen ist sachgerecht.

2.2.5.2 Auswertung der PRO-Endpunkte (Symptome und gesundheitsbezogene Lebensqualität)

Als primäre Analyse der PRO-Endpunkte ist für jede Skala eine Auswertung der Zeit bis zur erstmaligen Verschlechterung vorgesehen, und zwar mit jeweils 2 verschiedenen Responsekriterien 10 bzw. 15 Punkte (entsprechend 10 % bzw. 15 % der Skalenspannweite). Im SAP sind darüber hinaus verschiedene Sensitivitätsanalysen beschrieben. Zur Auswertung der PRO-Endpunkte ist folgendes anzumerken:

- Zwar ist ein Responsekriterium von 15 % der Skalenspannweite das Standardkriterium für Nutzenbewertungen nach § 35a SGB V [24,25], allerdings gilt dies nicht für die EORTC-Fragebogen. Für diese sieht der G-BA allein ein Responsekriterium von 10 Punkten vor [26]. Auch bei der geplanten AbD sind daher nur Responderanalysen mit dem 10-Punkte-Responsekriterium erforderlich.
- Die geplante Auswertung der Zeit bis zur erstmaligen Verschlechterung ist für die AbD grundsätzlich geeignet, auch wenn solche Auswertungen derzeit international diskutiert werden [27]. Analysen zur Zeit bis zur erstmaligen Verbesserung oder Verschlechterung sind bei Nutzenbewertungen im onkologischen Bereich derzeit oft die einzig sinnvoll interpretierbare Responderanalyse, da sich in den zugrunde liegenden Studien die Beobachtungsdauer zumeist zwischen den Behandlungsarmen relevant unterscheidet. Ursache ist das vorzeitige Beobachtungsende im Falle einer Progression oder eines Behandlungsabbruchs [28]. Gemäß vorliegendem Studienprotokoll ist bei der geplanten AbD jedoch gerade kein vorzeitiges Beobachtungsende geplant. Es sollte daher erwogen werden, als primäre Auswertung der PRO-Endpunkte eine andere Operationalisierung zu wählen, die die Folgerhebungen auch nach der erstmaligen Verschlechterung berücksichtigt, z. B. die Zeit bis zur 1- oder 2-malig bestätigten Verschlechterung.

- Als Sensitivitätsanalyse sind u. a. für jeden Erhebungszeitpunkt Vergleiche der mittleren Veränderung gegenüber Baseline vorgesehen. Diese sollten ergänzt werden um Responderanalysen (Verschlechterung um mindestens 10 Punkte) für jeden Erhebungszeitpunkt.

2.2.5.3 Auswertung der UE-Endpunkte

Es ist geplant, für UE-Endpunkte jeweils die Zeit bis zum Erstereignis auszuwerten, und es sollen Statistiken analog zu den Statistiken für das Gesamtüberleben (Cox-Regression, Kaplan-Meier-Analysen) sowie die absoluten und relativen Häufigkeiten der Personen mit Ereignis dargestellt werden. Zusätzlich sollen die Analysen der Zeiten bis zum Erstereignis nach System Organ Class (SOC) und Preferred Terms (PT) kategorisiert werden. Dieses Vorgehen ist sachgerecht.

2.2.6 Untersuchung auf potenzielle Effektmodifikatoren und Subgruppenanalysen

Der pU plant Subgruppenanalysen für die Faktoren Alter, Geschlecht, Krankheitsstadium (Ann Arbor I, II, III, IV) und Land für die ITT-Population. Sensitivitätsanalysen sind zu den Subgruppenanalysen nicht geplant. Mögliche Effektmodifikationen sollen mittels Interaktionstests geprüft werden. Die Subgruppenanalysen sollen nur durchgeführt werden, falls mindestens 10 Personen in jeder Subgruppe enthalten und bei binären Endpunkten (inklusive Zeit bis Erstereignis) mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind. Dieses Vorgehen ist sachgerecht.

2.2.7 Qualität und Vollständigkeit der Datenerhebung

2.2.7.1 Maßnahmen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung

Zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung sollen gemäß Studienprotokoll verschiedene Maßnahmen eingesetzt werden.

Dazu gehören die vom G-BA angeforderte Source Data Verification, ein Monitoring der teilnehmenden Zentren sowie ein Zentralmonitoring, verbunden mit Nachfragen bei fehlenden oder unplausiblen Werten. Besondere Maßnahmen werden zudem bei der Erhebung der PRO-Endpunkte eingesetzt (siehe Abschnitt 2.2.1.3). Diese Maßnahmen sind zusammenfassend prinzipiell geeignet, die Rate fehlender Werte zu verringern und die Qualität der Datenerhebung zu verbessern.

Auffällig ist jedoch, dass im Studienprotokoll an verschiedenen Stellen die Vollständigkeit der Datenerhebung vom pU selbst zum Teil in Frage gestellt wird, und zwar bezüglich einer Datenerhebung außerhalb der Zentren, die eine Therapie mit Brexucabtagen Autoleucel durchführen können. Konkret betrifft dies nach Angaben im Studienprotokoll potenziell beide Behandlungsgruppen. Zum einen kann die Betreuung nach Therapie mit Brexucabtagen

Autoleucel auch außerhalb der Zentren fortgesetzt werden, in denen diese Therapie angewendet wurde. Zum anderen werden Patientinnen und Patienten, die die Vergleichstherapie erhalten, potenziell während der gesamten AbD außerhalb solcher Zentren betreut.

Dazu ist zunächst anzumerken, dass in Deutschland diejenigen Zentren, die Patientinnen und Patienten mit Brexucabtagen Autoleucel behandelt haben, für die Vollständigkeit der Datenerhebung zu diesen Patientinnen und Patienten über den gesamten Zeitraum der AbD verantwortlich sind. Denn der G-BA hat die Befugnis zur Versorgung mit Brexucabtagen Autoleucel auf diejenigen Zentren beschränkt, die an der AbD teilnehmen [29], und damit verbunden, dass diese Zentren „auf eine möglichst vollständige Datenübermittlung hinzuwirken“ haben [30]. In der mündlichen Anhörung zu diesem Beschluss des G-BA wurde in diesem Zusammenhang bestätigt, dass die Datenübermittlung „in der Langzeitversorgung ... etabliert [ist] und ... von daher als sichergestellt betrachtet werden“ kann [31], ohne dass diese Feststellung auf deutsche Zentren beschränkt wurde. Insofern ist die oben beschriebene, an verschiedenen Stellen des Studienprotokolls enthaltene Relativierung der Vollständigkeit der Datenerhebung nicht nachvollziehbar. Sie ist davon unabhängig auch nicht sachgerecht, da dadurch das Ziel der AbD gefährdet wäre. Sollten daher weitere Maßnahmen zur Sicherstellung der Vollständigkeit und Qualität der Datenerhebung außerhalb der Zentren, die eine Therapie mit Brexucabtagen Autoleucel durchführen können, notwendig sein, sollten diese in der AbD und damit auch im Studienprotokoll verankert werden.

2.2.7.2 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].

Während die Maßnahmen zur Vermeidung fehlender Daten im Studienprotokoll im Detail beschrieben sind, werden im SAP lediglich rudimentäre Angaben zur Ersetzung von fehlenden Confounderdaten zur Adjustierung und zum Umgang mit unvollständigen Datumsangaben gemacht.

Fehlende Confounderdaten sollen laut pU mittels multipler Imputation (ohne weitere Spezifizierung) ersetzt werden, sofern die Menge der fehlenden Daten höchstens 30% beträgt. Für den Fall, dass mehr Daten fehlen, soll der entsprechende Confounder aus der Adjustierung herausgenommen werden. Einen relevanten Confounder mit mehr als 30% fehlenden Daten bei der Adjustierung nicht zu beachten ist nicht sachgerecht. Vielmehr ist vom pU zu beschreiben, auf welche Weise dann mit dem beträchtlichen Verlust von Informationen in den Auswertungen umzugehen ist, und unter welchen Voraussetzungen überhaupt noch der Versuch einer Adjustierung für Confounder sinnvoll ist. Ein hoher Anteil fehlender Daten für relevante Confounder führt potenziell zu einer unzureichenden Adjustierung und in Folge zu

nicht validen Ergebnissen. Es sind daher besondere Anstrengungen im Rahmen der AbD zu unternehmen, dass die Erhebung der Confounder vollständig ist. Sollten aus Sicht des pU und der Registerbetreibenden hierfür Maßnahmen erforderlich sein, die über die in Abschnitt 2.2.7.1 des vorliegenden Addendums adressierten Maßnahmen hinausgehen, sollten diese in der AbD und damit auch im Studienprotokoll verankert werden.

Fehlende Datumsangaben sollen wie folgt ersetzt werden: Sofern nur der Tag fehlt, wird der 1. des jeweiligen Monats oder, im Falle eines Enddatums für ein Ereignis (z. B. Ende eines UE), der letzte Tag (z. B. 31.01.) des jeweiligen Monats angenommen. Sofern auch der Monat fehlt, soll entsprechend der erste Kalendermonat (Januar) bzw. der letzte Kalendermonat (Dezember) angenommen werden. 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 die Rate fehlender Werte bei Datumsangaben zu minimieren.

Zu fehlenden Daten für Endpunkte wird lediglich beschrieben, dass fehlende Daten nicht ersetzt werden, da es sich um Routinedaten ohne verpflichtendes Follow-up handelt. Das ist nicht sachgerecht, da dies zu Informationsverlusten und zu erheblichen Verzerrungen der Ergebnisse führen kann. Der pU ist zum einen aufgefordert, sinnvolle Ersetzungsstrategien für fehlende Daten zu Endpunkten darzulegen. Zum anderen sind auch für die Erhebung der Endpunkte besondere Anstrengungen im Rahmen der AbD zu unternehmen, um eine möglichst vollständige Datenerhebung zu gewährleisten. Dies ist für PRO-Endpunkte der Fall, und es ist kein Grund ersichtlich, nicht auch für andere relevante Endpunkte geeignete Maßnahmen durchzuführen (siehe auch Abschnitt 2.2.7.1).

2.2.8 Anpassung der AbD

Gemäß G-BA-Beschluss zur AbD sind im Studienprotokoll und SAP Angaben zu Abbruchkriterien wegen Vergeblichkeit darzulegen [1]. Darüber hinaus soll auf Basis der ersten Zwischenanalyse ggf. die Fallzahlschätzung aktualisiert werden. Dazu soll der Endpunkt Mortalität herangezogen sowie die verschobene Hypothesengrenze in Anlehnung an das Vorgehen im Konzept des IQWiG [3] berücksichtigt werden.

Zu beiden Punkten finden sich keine Angaben in den vom pU vorgelegten Dokumenten. Dies sollte 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 Onasemnogen-Abeparvovec sowie die Ausführungen im Addendum zur Überprüfung dieses Studienprotokolls [32,33]).

2.2.9 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 relevante Confounder [15]. 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 2.2.4.2).

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

2.2.10 Zeitplan für die AbD

Die im Studienprotokoll benannten Zeitpunkte für die Vorlage von Statusberichten, Zwischenanalysen und für die finale Analyse entsprechen denen, die der G-BA im Beschluss zur AbD vorgegeben hat. Sie sind daher sachgerecht.

2.2.11 Sonstige Anmerkungen

Der pU gibt im Studienprotokoll und im SAP an, dass vor Datenbankschluss ein Data Review Meeting (DRM) durchgeführt werden soll. Dies soll zum einen noch bestehende Inkonsistenzen in den Daten auflösen, was sachgerecht ist. Zum anderen wird jedoch angegeben, dass die Entscheidungen im DRM potenziell die im SAP beschriebenen Methoden ergänzen oder sogar überschreiben können. Dies ist nicht sachgerecht: Das Verfahren der AbD sieht vor, dass der pU für die AbD ein Studienprotokoll und ein SAP zu erstellen und mit dem G-BA abzustimmen hat. Die AbD ist auf Basis der abgestimmten Dokumente durchzuführen. Änderungen im Nachgang, die nicht mit dem G-BA abgestimmt werden, gefährden das Ziel und die Verwendbarkeit der Datenerhebung. Es ist also mindestens zu ergänzen, dass Änderungsvorschläge, die sich aus dem DRM ergeben, mit dem G-BA abzustimmen sind, bevor sie bei der Auswertung umgesetzt werden.

2.3 Zusammenfassung und Fazit

Die Prüfung des SP (Version 1.0 vom 21.12.2022) und SAP (Version 1.0 vom 21.12.2022) für die AbD zu Brexucabtagen Autoleucel 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.

Positiv hervorzuheben sind insbesondere die Erweiterung hinsichtlich der Erhebung der PRO-Daten zu Symptomen und zur gesundheitsbezogenen Lebensqualität sowie die Erweiterung der Maßnahmen zur Verbesserung der Datenqualität.

Eine Anpassung ist hingegen insbesondere in folgenden Bereichen erforderlich:

- eindeutige Definition des Eignungskriteriums für eine Therapie mit Brexucabtagen Autoleucel (siehe Abschnitt 2.2.1.1)
- für die im Studienprotokoll genannten spezifischen UEs sollte das jeweilige in der CTCAE-Klassifikation genannte Kriterium für einen CTCAE-Grad 3 oder höher oder das allgemeine Kriterium „deutliche Beeinträchtigung der Aktivitäten des täglichen Lebens“ zusätzlich erhoben und diese Ereignisse entsprechend ausgewertet werden (siehe Abschnitt 2.2.1.3)
- Ergänzung einer eindeutigen Variablenliste der Verlaufsdaten für die AbD; Finalisierung der Variablenliste für Baseline-Daten (siehe Abschnitt 2.2.2)
- Erweiterung der aktiven Rekrutierungsschritte im Hinblick auf die gezielte Rekrutierung von Patientinnen und Patienten auch für die Vergleichsgruppe, sowohl auf nationaler wie auf internationaler Ebene (siehe Abschnitte 2.2.2 und 2.2.3.2)
- Adjustierung für Confounder: Festlegung und Erhebung der relevanten Confounder (siehe Abschnitt 2.2.4.1) und detaillierte Beschreibung des konkreten Vorgehens bei der Adjustierung für Confounder (siehe Abschnitt 2.2.4.2)
- Ergänzung, dass eine detaillierte Beschreibung der AbD-Population erforderlich ist, die sich aus der Anwendung eines PS-Verfahrens ergibt, einschließlich eines Abgleichs dieser AbD-Population mit der ursprünglichen Zielpopulation der AbD (siehe Abschnitt 2.2.4.2)
- Beschränkung der Responderanalysen für PRO-Endpunkte auf solche mit einer Responseschwelle von 10 Punkten (siehe Abschnitt 2.2.5.2)
- ggf. Definition und Beschreibung weiterer Maßnahmen zur Sicherstellung der Vollständigkeit und Qualität der Datenerhebung, insbesondere außerhalb der Zentren, die eine Therapie mit Brexucabtagen Autoleucel durchführen können (siehe Abschnitte 2.2.7.1 und 2.2.7.2)
- detaillierte Beschreibung des Umgangs mit fehlenden Werten (siehe Abschnitt 2.2.7.2)

- Ergänzung der Prüfung auf Vergeblichkeit und des Vorgehens zur möglichen Anpassung der AbD im Rahmen der Zwischenanalysen in Abstimmung mit dem G-BA (siehe Abschnitte 2.2.8 und 2.2.11)
- Ergänzung eines Abschnitt zur Ergebnisinterpretation unter Berücksichtigung des nicht randomisierten Studiendesigns und unter Verwendung einer entsprechenden verschobenen Hypothesengrenze (siehe Abschnitt 2.2.9)

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 SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiertes oder refraktäres Mantelzell-Lymphom); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/39-261-5540/2022-07-21_AM-RL-XII_Autologe-Anti-CD19-transduzierte-CD3-positive-Zellen_AbD-008_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 SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiertes oder refraktäres Mantelzell-Lymphom); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/40-268-8672/2022-07-21_AM-RL-XII_Autologe-Anti-CD19-transduzierte-CD3-positive-Zellen_AbD-008_Forderung_TrG.pdf.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzept für eine anwendungsbegleitende Datenerhebung – Brexucabtagen autoleucel; Rapid Report [online]. 2022 [Zugriff: 31.08.2022]. URL: https://www.iqwig.de/download/a21-130_anwendungsbegleitende-datenerhebung-brexucabtagen-autoleucel_rapid-report_v1-0.pdf.
4. University Medical Center of the Johannes Gutenberg-University, Kite Pharma. Real world effectiveness and safety of brexucabtagene autoleucel versus patient-individual therapy in relapsed/refractory mantle cell lymphoma: A European Mantle Cell Lymphoma Network (EMCL) registry study mandated by the G-BA; Study Project Plan Version 1.0 [unveröffentlicht]. 2022.
5. University Medical Center of the Johannes Gutenberg-University, Kite Pharma. Statistical Analysis Plan V1.0 [unveröffentlicht]. 2022.
6. Gemeinsamer Bundesausschuss. Niederschrift zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV Beratungsanforderung 2022-B-219 [unveröffentlicht]. 2022.
7. Kite, Gilead Sciences. Fachinformation Tecartus [online]. 2022 [Zugriff: 01.02.2023]. URL: <https://www.fachinfo.de/suche/fi/023222>.

8. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V); Enfortumab Vedotin (Urothelkarzinom, vorbehandelt mit Platin-basierter Chemotherapie und PD-(L)1-Inhibitor) [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/39-261-5748/2022-12-01_AM-RL-XII_Enfortumab-Vedotin_D-790.pdf.
9. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V); Polatuzumab Vedotin (neues Anwendungsgebiet: Diffus großzelliges B-Zell-Lymphom, Kombination mit Rituximab, Cyclophosphamid, Doxorubicin und Prednison (R-CHP)) [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/39-261-5750/2022-12-01_AM-RL-XII_Polatuzumab-Vedotin_D-827_BAnz.pdf.
10. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V); Nivolumab (Neues Anwendungsgebiet: Urothelkarzinom, PD-L1-Expression $\geq 1\%$, adjuvante Therapie) [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/39-261-5661/2022-10-20_AM-RL-XII_Nivolumab_D-821_BAnz.pdf.
11. Lindberg-Scharf P, Steinger B, Koller M et al. Long-term improvement of quality of life in patients with breast cancer: supporting patient-physician communication by an electronic tool for inpatient and outpatient care. Support Care Cancer 2021; 29(12): 7865-7875. <https://dx.doi.org/10.1007/s00520-021-06270-1>.
12. Klinkhammer-Schalke M, Steinger B, Koller M et al. Diagnosing deficits in quality of life and providing tailored therapeutic options: Results of a randomised trial in 220 patients with colorectal cancer. Eur J Cancer 2020; 130: 102-113. <https://dx.doi.org/10.1016/j.ejca.2020.01.025>.
13. Basch E, Deal AM, Dueck AC et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017; 318(2): 197-198. <https://dx.doi.org/10.1001/jama.2017.7156>.
14. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiviertes oder refraktäres Mantelzell-Lymphom) [online]. 2021 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/39-261-4964/2021-08-05_AM-RL-XII_Autologe-Anti-CD19-transduzierte-CD3-positive-Zellen_D-633_BAnz.pdf.

15. 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.
16. Gemeinsamer Bundesausschuss. Zusammenfassende Dokumentation über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiviertes oder refraktäres Mantelzell-Lymphom); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/40-268-8673/2022-07-21_AM-RL-XII_Autologe-Anti-CD19-transduzierte-CD3-positive-Zellen_AbD-008_Forderung_ZD.pdf.
17. 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>.
18. 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>.
19. 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>.
20. 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>.
21. Kuss O, Blettner M, Borgermann J. Propensity Score: an Alternative Method of Analyzing Treatment Effects. Dtsch Arztebl Int 2016; 113(35-36): 597-603. <https://dx.doi.org/10.3238/arztebl.2016.0597>.
22. 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>.

23. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011; 46(3): 399-424. <https://dx.doi.org/10.1080/00273171.2011.568786>.
24. Institute for Quality and Efficiency in Health Care. General Methods Version 6.1 (German version) [online]. 2022 [Zugriff: 27.01.2022]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
25. Gemeinsamer Bundesausschuss. Dossier zur Nutzenbewertung gemäß § 35a SGB V; Modul 4; Dokumentenvorlage, Version vom 16.12.2021 [online]. 2021 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/17-98-4825/2021-12-16_An12_6_Modul4.pdf.
26. Gemeinsamer Bundesausschuss. Antworten auf häufig gestellte Fragen zum Verfahren der Nutzenbewertung; Wie soll, vor dem Hintergrund der Veröffentlichung des Methodenpapiers 6.0 des IQWiG am 5. November 2020, derzeit in der Dossiererstellung mit der Bestimmung von klinischen Relevanzschwellen bei komplexen Skalen umgegangen werden? [online]. [Zugriff: 05.08.2021]. URL: <https://www.g-ba.de/themen/arzneimittel/arzneimittel-richtlinie-anlagen/nutzenbewertung-35a/faqs>.
27. Fiero MH, Roydhouse JK, Bhatnagar V et al. Time to deterioration of symptoms or function using patient-reported outcomes in cancer trials. *Lancet Oncol* 2022; 23(5): e229-e234. [https://dx.doi.org/10.1016/S1470-2045\(22\)00021-3](https://dx.doi.org/10.1016/S1470-2045(22)00021-3).
28. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Abemaciclib (Mammakarzinom; Kombination mit Fulvestrant) – Nutzenbewertung gemäß § 35a SGB V (Ablauf Befristung); Dossierbewertung [online]. 2022 [Zugriff: 01.03.2022]. URL: https://www.iqwig.de/download/a21-153_abemaciclib_nutzenbewertung-35a-sgb-v_v1-0.pdf.
29. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiertes oder refraktäres Mantelzell-Lymphom); Beschränkung der Versorgungsbefugnis [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/39-261-5541/2022-07-21_AM-RL-XII_BrexCel_2021-AbD-008_Versorgungsbeschraenkung_BAnz.pdf.
30. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiertes oder refraktäres Mantelzell-Lymphom); Beschränkung der Versorgungsbefugnis [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/40-268-8674/2022-07-21_AM-RL-XII_BrexCel_2021-AbD-008_Versorgungsbeschraenkung_TrG.pdf.

31. Gemeinsamer Bundesausschuss. Zusammenfassende Dokumentation über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Autologe Anti-CD19-transduzierte CD3-positive Zellen (rezidiertes oder refraktäres Mantelzell-Lymphom); Beschränkung der Versorgungsbefugnis [online]. 2022 [Zugriff: 01.02.2023]. URL: https://www.g-ba.de/downloads/40-268-8675/2022-07-21_AM-RL-XII_BrexCel_2021-AbD-008_Versorgungsbeschraenkung_ZD.pdf.

32. 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.

33. 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 führt der pU eine Informationsbeschaffung nach systematischen Übersichten und Leitlinien zum MCL in folgenden Quellen durch:

- bibliografische Recherchen in MEDLINE und der Cochrane Database of Systematic Reviews nach Leitlinien und systematischen Übersichten (letzte Suche am 15.11.2022)
- Handsuche auf ausgewählten Webseiten (Konferenzen, German Guideline Program in Oncology, Google, Google Scholar, PubMed)

Zudem führt der pU an, zusätzlich eine cursorische Handsuche nach Studien durchgeführt zu haben, um die identifizierten Confounder zu validieren, da durch die vorherige systematische Recherche kaum relevante statistische Evidenz identifiziert wurde.

Kommentar zur Informationsbeschaffung

Die vom pU durchgeführte Recherche des pU ist zwar prinzipiell geeignet, systematische Übersichten und Leitlinien als Basis für die Identifizierung von Confoundern einschließlich der zugehörigen statistischen Evidenz zu finden. Allerdings ist die weitere Vorgehensweise des pU nicht sachgerecht. Dies hat insbesondere folgende Gründe:

Für die systematische Identifizierung von Confoundern muss gewährleistet sein, dass hierfür geeignete Primärstudien in ausreichendem Umfang identifiziert werden. Dazu gehören insbesondere epidemiologische Studien und Untersuchungen auf Effektmodifikationen in randomisiert kontrollierten Studien (RCTs) [19]. Ein Rückgriff auf Sekundärliteratur, wie im vorliegendem Fall, ist prinzipiell möglich, sofern gewährleistet ist, dass die Sekundärliteratur die geeigneten Primärstudien mit ausreichender statistischer Evidenz in ausreichendem Umfang enthält und die für die Bewertung der Confounder notwendigen Informationen in ausreichendem Umfang berichtet. Dieser Prüfschritt fehlt bei der Informationsbeschaffung des pU. Davon unabhängig wird aus der Auflistung in Appendix B des Berichts deutlich, dass keine der identifizierten Übersichten geeignete Primärstudien in ausreichendem Umfang identifiziert hat. Auch der pU selbst stellt fest, dass durch seine systematische Suche nach Sekundärliteratur kaum ausreichende statistische Evidenz identifiziert wurde. Allerdings zieht er daraus die falsche Konsequenz einer cursorischen Handsuche, woraus sich letztlich ein nicht systematisches Vorgehen ergibt. Es wäre statt dessen sinnvoll gewesen, eine Suche nach Primärstudien mit Stopp-Kriterien wie in Pufulete 2022 beschrieben anzuschließen [19].

Aus den beschriebenen methodischen Gründen ist insgesamt nicht sichergestellt, dass eine ausreichend vollständige Liste an potenziell relevanten Confoundern einschließlich der zugehörigen statistischen Evidenz identifiziert wurde.