

**Wissenschaftliche Ausarbeitung  
ausgewählter Aspekte zur Generierung  
versorgungsnaher Daten und deren  
Auswertung zum Zwecke der  
Nutzenbewertung von Arzneimitteln nach  
§ 35a SGB V**

**DOKUMENTATION DER ANHÖRUNG ZUM VORLÄUFIGEN  
RAPID REPORT**



# Impressum

## Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

## Thema

Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V

## Auftraggeber

Gemeinsamer Bundesausschuss

## Datum des Auftrags

28.01.2025

## Interne Projektnummer

A25-13

## Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Siegburger Str. 237

50679 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: [berichte@iqwig.de](mailto:berichte@iqwig.de)

Internet: [www.iqwig.de](http://www.iqwig.de)

# Inhaltsverzeichnis

	Seite
<b>Abkürzungsverzeichnis.....</b>	<b>iii</b>
<b>1 Dokumentation der Anhörung .....</b>	<b>1</b>
<b>2 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll .....</b>	<b>2</b>
<b>2.1 Teilnehmerliste der wissenschaftlichen Erörterung .....</b>	<b>2</b>
<b>2.2 Liste der Stellungnahmen, zu denen keine Vertreterin bzw. kein Vertreter an     der wissenschaftlichen Erörterung teilgenommen hat.....</b>	<b>3</b>
<b>2.3 Tagesordnung der wissenschaftlichen Erörterung .....</b>	<b>4</b>
<b>2.4 Protokoll der wissenschaftlichen Erörterung.....</b>	<b>4</b>
<b>2.4.1 Begrüßung und Einleitung .....</b>	<b>4</b>
<b>2.4.2 Tagesordnungspunkt 1: Confounderidentifikation und -auswahl .....</b>	<b>5</b>
<b>2.4.3 Tagesordnungspunkt 2: Schätzung notwendiger Fallzahlen vor dem         Hintergrund unzureichender Vorabinformationen.....</b>	<b>15</b>
<b>2.4.4 Tagesordnungspunkt 3: Schwellenwert für die verschobene Nullhypothese .....</b>	<b>28</b>
<b>2.4.5 Tagesordnungspunkt 4: Alternative Methoden zu Propensity-Score-         Verfahren bei kleinen Fallzahlen und vielen Confoundern.....</b>	<b>33</b>
<b>2.4.6 Tagesordnungspunkt 5: Hürden bei Nutzung materieller Anreize als         Instrument zur Sicherstellung der Vollständigkeit der Daten.....</b>	<b>37</b>
<b>2.4.7 Tagesordnungspunkt 6: Verschiedenes .....</b>	<b>39</b>
<b>Anhang A Dokumentation der Stellungnahmen .....</b>	<b>41</b>

# Abkürzungsverzeichnis

<b>Abkürzung</b>	<b>Bedeutung</b>
AbD	anwendungsbegleitende Datenerhebung
AMNOG	Arzneimittelmarktneuordnungsgesetz
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
BMS	Bristol-Myers Squibb GmbH & Co. KGaA
BPI	Bundesverband der Pharmazeutischen Industrie e. V.
BSC	Best supportive Care
DAG	Directed acyclic Graph (gerichteter azyklischer Graph)
DEGAM	Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin
DGFN	Deutsche Gesellschaft für Nephrologie e. V.
DGIM	Deutsche Gesellschaft für Innere Medizin e. V.
DGKJ	Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V.
DGOU	Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V.
DGP	Gesellschaft für Neuropädiatrie e. V.
DKG	Deutsche Krankenhausgesellschaft e. V.
DRFZ	Deutsches Rheuma-Forschungszentrum Berlin
EU	Europäische Union
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KöR	Körperschaft des öffentlichen Rechts
MD Bund	Medizinischer Dienst Bund
PRO	Patient-reported Outcome (patientenberichteter Endpunkt)
PS	Propensity Score
pU	pharmazeutischer Unternehmer
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SGB	Sozialgesetzbuch
TOP	Tagesordnungspunkt
vfa	Verband Forschender Arzneimittelhersteller e. V.

## 1 Dokumentation der Anhörung

Am 30.07.2025 wurde der vorläufige Rapid Report in der Version 1.0 vom 28.07.2025 veröffentlicht und zur Anhörung gestellt. Bis zum 19.08.2025 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 27 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang A abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 16.09.2025 im IQWiG diskutiert. Das Wortprotokoll der wissenschaftlichen Erörterung befindet sich in Kapitel 2.

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des vorläufigen Rapid Reports überprüft. Eine Würdigung der in der Anhörung vorgebrachten wesentlichen Aspekte befindet sich im Anhang des finalen Rapid Reports. Im finalen Rapid Report sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der finale Rapid Report ist auf der Website des IQWiG unter [www.iqwig.de](http://www.iqwig.de) veröffentlicht.

## 2 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

### 2.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Acikgöz, Özlem	Deutsche Krankenhausgesellschaft e. V. (DKG)
Annacker, Vincent	Pfizer Pharma GmbH
Ahrens, Lucas	Ecker + Ecker GmbH
Beckmann, Lars	IQWiG
Berkemeier, Fabian	IGES Institut GmbH
Beuermann, Janik	Ecker + Ecker GmbH
Bluhmki, Tobias	Bristol-Myers Squibb GmbH & Co. KGaA (BMS)
Böhm, Doris	Bristol-Myers Squibb GmbH & Co. KGaA (BMS)
Bogner, Kathrin	AMS Advanced Medical Services GmbH
Bussilliat, Paul	Verband Forschender Arzneimittelhersteller e. V. (vfa)
Dintsiros, Charalabos-Markos	Bayer Vital GmbH
Eberle, Sonja	MSD Sharp & Dohme GmbH
Eisele, Lewin	Janssen-Cilag GmbH
Giesl, Nicole	Merck Healthcare Germany GmbH
Göbel, Jana	IQWiG
Haug, Nina	iOMEDICO AG
Hofmann-Xu, Liping	Bristol-Myers Squibb GmbH & Co. KGaA (BMS)
Johner, Simone	IQWiG
Jülich, Fabian	CSL Behring GmbH
Kabilka, Anna	SmartStep Consulting GmbH
Kaiser, Thomas	IQWiG (Moderation)
Kauffmann, Stephanie	Pfizer Pharma GmbH
Keksel, Christina	IQWiG
Knoerzer, Dietrich	Roche Pharma AG
Kucka, Annett	Servier Forschung und Pharma-Entwicklung GmbH
Kürschner, Niclas	Pfizer Pharma GmbH
Lange, Stefan	Medizinischer Dienst Bund (KÖR)
Masoudi, Ehsan	Lilly Deutschland GmbH
Mehnert, Angelika	Janssen-Cilag GmbH
Meisel, Roland	Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ)
Mikulić, Ulrike	IQWiG
Musiolik, Katharina	AbbVie Deutschland GmbH & Co. KG
Nink, Katrin	IQWiG
Petri, Anne-Kathrin	IQWiG
Petry, Dennis	Vertex Pharmaceuticals (Germany) GmbH

Name	Organisation / Institution / Firma / privat
Pfitzer, Lisa	MSD Sharp & Dohme GmbH
Rasch, Andrej	Verband Forschender Arzneimittelhersteller e. V. (vfa)
Regierer, Anne	Deutsches Rheuma-Forschungszentrum Berlin (DRFZ)
Reiter, Bennet	AMS Advanced Medical Services GmbH
Richter, Swantje	AMS Advanced Medical Services GmbH
Rörtgen, Thilo	Protokollant
Sauerbruch, Tilman	Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)
Schmitt, Susanne	Vertex Pharmaceuticals (Germany) GmbH
Seay, Ulrike	IQWiG
Sindern, Jörn	Janssen-Cilag GmbH
Skipka, Guido	IQWiG
Sternberg, Kati	Abbvie Deutschland GmbH & Co. KG
Struck, Karolin	SmartStep Consulting GmbH
Sussmann, Stephanie	IGES Institut GmbH
Vaitsiakhovich, Tatsiana	Sanofi-Aventis Deutschland GmbH
Vervölgyi, Volker	IQWiG
Volmer, Timm	SmartStep Consulting GmbH
Weinmann-Menke, Julia	Deutsche Gesellschaft für Nephrologie e. V. (DGfN)
Wöhl, Caroline	IQWiG
Wöhling, Heike	Vertex Pharmaceuticals (Germany) GmbH
Zietze, Henny A.	Sanofi-Aventis Deutschland GmbH

## 2.2 Liste der Stellungnahmen, zu denen keine Vertreterin bzw. kein Vertreter an der wissenschaftlichen Erörterung teilgenommen hat

In der folgenden Tabelle werden Stellungnahmen genannt, zu denen trotz Einladung kein Stellungnehmender oder Vertreter zur wissenschaftlichen Erörterung erschienen ist.

Organisation / Institution / Firma / Privatperson
Bundesverband der Pharmazeutischen Industrie e. V. (BPI)
Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM)
Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU)

## 2.3 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
<b>TOP 1</b>	Confounderidentifikation und -auswahl
<b>TOP 2</b>	Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen
<b>TOP 3</b>	Schwellenwert für die verschobene Nullhypothese
<b>TOP 4</b>	Alternative Methoden zu Propensity-Score-Verfahren bei kleinen Fallzahlen und vielen Confoundern
<b>TOP 5</b>	Hürden bei Nutzung materieller Anreize als Instrument zur Sicherstellung der Vollständigkeit der Daten
<b>TOP 6</b>	Verschiedenes

## 2.4 Protokoll der wissenschaftlichen Erörterung

Datum: 16.09.2025, 13:33 bis 15:32 Uhr

Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),  
Siegburger Str. 237, 50679 Köln

Moderation: Thomas Kaiser

### 2.4.1 Begrüßung und Einleitung

**Moderator Thomas Kaiser:** Ich begrüße Sie ganz herzlich zu der wissenschaftlichen Erörterung der Stellungnahmen zum vorläufigen Rapid Report mit dem wunderbaren Titel „Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“. Alle, mindestens aber alle Informierten, die heute teilnehmen, wissen, dass das als eine Art Nachfolgeprojekt zum Rapid Report A19-43 gesehen werden kann, der einen ähnlichen Titel hatte. Aber hier geht es jetzt um ein paar Spezifika, die sich im Verlauf der Jahre, der Bearbeitung, insbesondere der anwendungsbegleitenden Datenerhebung, ergeben haben. Das ist der Auftrag des Gemeinsamen Bundesausschusses. Wir möchten heute Ihre Stellungnahmen respektive bestimmte Punkte, die für uns für die Bearbeitung wichtig sind und wo es einige Unklarheiten gegeben hat, mit Ihnen diskutieren.

Bevor ich zu ein paar Rahmenbedingungen komme, weil ich gerade bei den Stellungnahmen bin: Das bedeutet auch, Sie müssen, wenn Sie einen Redebeitrag haben, nicht Ihre Stellungnahmen wiederholen. Wir haben alle Stellungnahmen gelesen, bearbeiten die. Uns geht es heute darum, uns darauf zu konzentrieren - wir haben eine recht sportliche Tagesordnung -, was an wesentlichen Punkten in den Stellungnahmen genannt worden ist und wozu wir eben gerne einen Austausch wünschen, um unseren Rapid Report an der einen oder

anderen Stelle noch besser, anders, wie auch immer, zu machen, wenn es sinnvolle Argumente gegeben hat.

Kurz zu den administrativen Punkten. Ich gehe davon aus, dass Sie Ihre Zustimmung zur Erstellung des Wortprotokolls aufrechterhalten. Das haben Sie nämlich mit der Bekanntgabe Ihrer Teilnahme heute und Ihrem Anmelden gemacht. Sollten Sie das nicht mehr wollen, dann müssten Sie den virtuellen Raum verlassen, aber ich gehe davon aus, dass das für keinen von Ihnen der Fall ist.

Es ist wichtig, dass Sie sich jedes Mal, wenn Sie einen Beitrag machen wollen, durch die Handhebefunktion - die ist ganz unten in der Leiste - melden. Es sind so viele Teilnehmende dabei, dass wir nicht auf physische Handhebung reagieren können, sondern bitte nutzen Sie die Handhebefunktion. Das hat auch den Vorteil, dass wir dann die Reihenfolge erkennen können. Und sagen Sie bitte jedes Mal wieder Ihren Namen, auch wenn Sie wiederholte Beiträge haben, damit wir auf jeden Fall Ihren Beitrag Ihrer Person im Wortprotokoll zuordnen können.

Gut. Wenn es ansonsten von Ihrer Seite keine Fragen gibt, dann starten wir in die Tagesordnung, mit

#### **2.4.2 Tagesordnungspunkt 1: Confounderidentifikation und -auswahl**

**Jana Göbel:** In Abschnitt 5.2 des vorläufigen Rapid Reports schlagen wir ein 3-schrittiges Verfahren zur systematischen Confounderidentifikation vor. In verschiedenen Stellungnahmen von Ihnen wurde der große Aufwand angemerkt, eine Einschätzung, die wir teilen, und das geplante Vorhaben begrüßt, weitere Schritte für eine mögliche Aufwandsreduktion zu prüfen.

In den Stellungnahmen wurden auch verschiedene Anregungen genannt, die wir in das Folgeprojekt einfließen lassen werden, aber zu denen wir heute die Gelegenheit nutzen möchten, uns hier an dieser Stelle dazu auszutauschen.

Zum einen möchten wir uns zur Relevanzbewertung austauschen. Hier wurde von der DKG angeregt, die Qualität der Quelle zu berücksichtigen, aus denen ein Confounder extrahiert wurde. Uns ist an der Stelle unklar, welche Kriterien für die Bewertung der Qualität der Quelle gelten, insbesondere vor dem Hintergrund des Ziels, Confounder zu identifizieren. Das möchten wir gerne mit Ihnen diskutieren, was unter der Qualität der Quelle zu verstehen ist und auch wie diese berücksichtigt werden kann.

In der Stellungnahme vom MD Bund wurde angeregt, Variablen, die lediglich in einer Quelle als Patientencharakteristikum genannt werden, nur dann zu berücksichtigen, wenn das inhaltlich plausibel erscheint, also als potenziellen Confounder zu berücksichtigen. Wenn wir das richtig verstanden haben, würden wir den Variablen, für die in Beobachtungsstudien

adjustiert wurde, somit eine höhere Relevanz beimessen als Variablen, welche lediglich als Patientencharakteristikum in einer RCT genannt werden.

Somit möchten wir uns zu dem Aspekt der Relevanzbewertung zu zwei Punkten austauschen, zum einen, welche Aspekte der Qualität der Quellen zu berücksichtigen sind, und zum anderen, was ein sinnvoller Schwellenwert für die Häufigkeit der Nennung eines Confounders wäre.

Darüber hinaus möchten wir auf einen weiteren Aspekt zu sprechen kommen. In einer Stellungnahme vom MD Bund wurde angemerkt, die Korrelation von potenziellen Confoundern bei der Auswahl der zu erhebenden Variablen stärker zu berücksichtigen. Hierzu würden wir uns gerne austauschen, welche Schritte Sie vorschlagen, die Korrelation auf Ebene der Confounderidentifikation zu berücksichtigen, und mit welcher Begründung ein Ausschluss einer Variable gerechtfertigt werden kann.

**Moderator Thomas Kaiser:** Vielen Dank. - Ich fasse kurz zusammen. Wir haben also drei Themen, und vielleicht arbeiten wir die nach und nach ab und fangen mal an mit der Frage Relevanzbewertung der identifizierten Quellen zur potenziellen Identifikation von Confoundern. Was sind mögliche Instrumente, und was sind aber auch mögliche Schwellenwerte, um eine Quelle zu berücksichtigen oder nicht zu berücksichtigen? Melden Sie sich gerne über die Handhebefunktion. - Frau Acikgöz.

**Özlem Acikgöz:** Sie haben uns ja direkt angesprochen. Was wir sehr begrüßen, ist, dass es ein Verfahren sein soll, wo die klinische Relevanz über klinische Experten einbezogen werden soll. Dafür haben Sie natürlich absolut Unterstützung von uns. Wir haben uns natürlich bei der Fülle der Confounder, die jetzt bei der exemplarisch durchgeführten Analyse herausgekommen sind, gefragt, ob man sich die Evidenz dahinter, so wie das ja sonst auch üblich ist, mal anguckt, wo man diese Confounder extrahiert hat, insbesondere auch vor dem Hintergrund, dass wir es ja hier häufig mit neuartigen Therapien und Wirkstoffen zu tun haben und wo die klinische Beurteilung vielleicht nicht ganz so einfach ausfällt als wenn man da als klinischer Experte schon langjährige Erfahrung hat. Das wäre unsere Anregung. Sie haben ja selber gesagt, dass Sie Ihr Arbeitspapier an der Stelle noch mal überarbeiten wollen. Aber die Quelle der Confounder, was man herausgefunden hat, auf welcher Evidenzgrundlage sozusagen die zugrunde liegende Studie stattgefunden hat, das wäre unsere Anmerkung an der Stelle.

**Moderator Thomas Kaiser:** Vielen Dank. - Gibt es weitere Anmerkungen? - Stefan Lange.

**Stefan Lange:** Ich bin mir jetzt nicht ganz sicher. Ich würde jetzt zu dieser Frage der Qualitätsbewertung nicht so viel sagen wollen. Ich vermute, dass es jetzt auch darum gehen soll, unsere Idee ... Frau Göbel, Sie hatten das zu Beginn gesagt, wenn zum Beispiel ein möglicher

Confounder nur in einer Studie irgendwie extrahiert werden konnte. Darum soll es doch jetzt auch gehen, nicht?

**Moderator Thomas Kaiser:** Ist richtig. Also, wenn es jetzt zu der Relevanzbewertung keine weiteren ..., und es gab keine weiteren, dann gerne jetzt zu dem Thema der Quellen - Patientencharakteristika.

**Stefan Lange:** Danke. - Nach dem Schwellenwert ist natürlich ... Das ist einerseits berechtigt, andererseits noch nicht so weit ausgereift. Ich glaube, das kann man entwickeln. Ich stelle mir jetzt mal so vor, man hat irgendwie für eine Fragestellung - keine Ahnung - 30 Beobachtungsstudien gefunden in der Recherche und eben nur in einer oder von mir aus auch zwei findet man eben ein solches Merkmal, für das gegebenenfalls adjustiert worden ist, oder in entsprechenden RCTs als Baseline-Charakteristikum. Man wird einen solchen Schwellenwert natürlich anders anlegen müssen, wenn man nur vier zugrunde liegende Studien hat. Dann ist N gleich 1 eine ganze Menge, 25 %. Oder wenn man 30 hat, dann ist N gleich 1 eben nur noch 3 %. Also, die Idee, die dahintersteckt, ist ja, dass damit eine Wertung verbunden ist, dass das vielleicht nicht so bedeutsame Confounder sein könnten, wenn sie nur selten genannt werden, potenzielle Confounder.

Konkret zur Schwelle haben wir uns tatsächlich bis jetzt keine Gedanken gemacht. Ich weiß gar nicht, ob es dann dazu wieder Empirie gibt zum Zusammenhang zwischen der Häufigkeit von Nennung und der - in Anführungszeichen - Relevanz. Aber, ich meine, wir sprechen ja über Möglichkeiten, den Aufwand zu reduzieren. Wenn ich das richtig gesehen habe, waren ja ursprünglich in dem Beispielprojekt zunächst mal eine Auswahl von über 100 potenziellen Confoundern. Damit muss man irgendwie umgehen können. Die sind dann zwar weiter sukzessive reduziert worden, aber war immer noch bei einer sehr beachtlichen Größenordnung. Also, das wäre jetzt ein Gedanke, aber noch mal konkret auf die Frage, da würde ich mich jetzt ehrlich gesagt nicht aus dem Fenster lehnen wollen, da eine konkrete Schwelle im Augenblick zu benennen. Da müsste man vielleicht tatsächlich mal ein bisschen nach Empirie gucken.

Vielleicht noch ein Aspekt dazu. Wir haben das, glaube ich, nicht angemerkt. Im Grunde genommen steht es auch im Methodenpapier des IQWiG und auch in den Rapid Reports, aber, ich finde, nicht ganz so prominent, wie es vielleicht sein könnte. Man kann sich der Frage ja wirklich auch noch mal leicht besser nähern, indem man versucht, dieses Prinzip der direkten azyklischen Graphen noch weiter auszuarbeiten und dem einen größeren Stellenwert einzuräumen. Da wird man ja möglicherweise auch Anhaltspunkte für die Relevanz oder auch Nichtrelevanz von möglichen Confoundern finden.

Und last but not least müsste man sich auch noch mal darüber Gedanken machen - das geht jetzt über diese Schwellenwertfrage hinaus -, was eigentlich als Confounder tatsächlich

berücksichtigt werden soll. Reden wir von Effektmodifikatoren, reden wir von übrigens instrumentellen Variablen? Da müsste man wahrscheinlich auch noch mal ein bisschen gucken, dass man da nicht eventuell auch noch Eigentore schießt.

So weit erst mal.

**Moderator Thomas Kaiser:** Vielen Dank. Vielen Dank auch noch mal für den Hinweis, der in Richtung Korrelation geht. Also, da zu überlegen, wie man gegebenenfalls an der Stelle mit einer weiteren Bearbeitung die Menge der abschließend für die Datenerhebung relevanten Confounder reduziert, das ist ja die Zielsetzung an der Stelle.

Dann nur eine Anmerkung von mir - ich habe dann Volker Vervölgyi hier intern und dann Herrn Meisel auf der Rednerliste -, dass wir, wenn wir über Aufwand sprechen, zwei übergeordnete Punkte haben. Der eine Punkt ist Aufwand für die Identifikation, den man theoretisch größer machen kann, um auch zu reduzieren, denn auch die Reduktion ist ein gewisser Aufwand, und dann aber Aufwand für die tatsächliche Datenerhebung und die Durchführung der Studie. Ich glaube, dieses Verhältnis muss man sich immer anschauen, wenn man über Aufwand spricht. Es kann sich also lohnen, einen etwas größeren Aufwand in die Identifikation und Reduktion zu stecken, um dann für die gesamte Durchführung des Forschungsvorhabens deutlich weniger Aufwand zu haben. Das einfach nur so als ein Grundgedanke.

Ich habe dann eben jetzt Volker Vervölgyi, dann Herrn Meisel, dann noch mal Stefan.

**Volker Vervölgyi:** Ich war eben nicht schnell genug, mich zu melden. Ich habe noch eine Rückfrage an Frau Acikgöz zu ihrem Beitrag. Wenn ich Sie richtig verstanden habe, ist Qualität der Quelle bezogen auf eher in Richtung Nähe zur Fragestellung. Also, wie sehr bildet das eigentlich das ab, was ich auch tatsächlich nachher untersuchen möchte, als im Sinne von Qualität, Verzerrung oder Güte, wie gut oder schlecht eine Studie ist? Vielleicht können Sie dazu noch kurz was sagen.

Eine Rückfrage habe ich tatsächlich auch noch. Sie hatten eben von neuen Therapiearten gesprochen, die tatsächlich in der AbD ja gar nicht so selten sind. In der Regel ist es dann natürlich auch so - das hatten Sie ja gerade auch schon gesagt -, dass Studien gegebenenfalls das gar nicht abbilden können. Die Frage ist halt tatsächlich: Wie groß ist dann auch die Erfahrung der klinischen Experten auf dem Gebiet? Also, sind die klinischen Experten genau in diesem Gebiet, wo es um neuartige Therapieformen geht, besser geeignet als publizierte Literatur?

**Moderator Thomas Kaiser:** Das war jetzt eine Frage direkt an Frau Acikgöz. Herr Meisel, wenn Sie erlauben, dann würde ich zunächst Frau Acikgöz die Gelegenheit geben, darauf kurz zu antworten.

**Özlem Acikgöz:** Also, ich hatte das ja vorhin angesprochen genau bei diesen Therapien oder Wirkstoffen, die ja in den AbDs auch Eingang finden, dass dann natürlich die klinische Expertise in Bezug auf die Patientengruppen, bestimmte Risikofaktoren bei bekannten Therapien, natürlich einfließen müssen und sollen, aber halt vor dem Hintergrund der begrenzten Erfahrungen mit dem gegenständlichen Wirkstoff, der vielleicht nur bis zu einem gewissen Grad nützlich ist. Das war unsere Anmerkung an der Stelle.

Zur Qualität. Eine Studie - das wissen Sie viel besser als ich - hat ja mehrere Qualitätskriterien. Natürlich spielt dabei auch die Evidenz eine Rolle, ob das jetzt ein Confounder ist, der auf Grundlage von ganz wenigen Patienten in einer Ministudie irgendwie kreiert worden ist, oder ob es eine größere Fallzahl ist in der Studie. Solche Dinge sollten ein bisschen im Grading Berücksichtigung finden. Darum geht es uns. Letztendlich geht es ja darum, die Confounder in einer Hierarchie oder in einer Abstufung für die Relevanz zu identifizieren, wo wir nicht mehr so eine Fülle von Confoundern haben, sondern sozusagen bestmögliche - „bestverfügbar“ will ich nicht sagen - oder am besten infrage kommende Confounder. Deshalb kann man jetzt nicht sagen, das ist die Qualität, die es auf jeden Fall braucht oder nicht, sondern in der Abstufung sich anzusehen, welche Confounder eigentlich die solideste Grundlage haben, so solide wie möglich natürlich, insbesondere bei den neuen Medikamenten.

**Moderator Thomas Kaiser:** Vielen Dank. - Dann bleiben wir eben noch mal kurz im Dialog. Volker.

**Volker Vervölgyi:** Noch mal eine kurze Rückfrage dazu. Dann, glaube ich, habe ich Sie richtig verstanden. Sie haben recht, es gibt natürlich ganz viele Arten von Qualitätsdimensionen in einer Studie. In der Tat ist es so, dass die Confounderidentifikation - das sehen wir auch immer ganz häufig - in nicht randomisierten Studien eher stiefmütterlich behandelt wird. Natürlich kann man dann auch - so verstehe ich Sie jetzt - schauen, wie man eigentlich in der einzelnen Beobachtungsstudie vorgegangen ist, die Confounder zu identifizieren, und das als Qualitätskriterium quasi zu berücksichtigen. Dann habe ich Sie, glaube ich, richtig verstanden. Danke schön.

**Moderator Thomas Kaiser:** Gut. Dann zunächst eine Anmerkung: Nichts gegen Stiefmütter, ich habe da nur gute Erfahrungen gemacht. Aber dann gehen wir zu Herrn Meisel. Bitte sehr.

**Roland Meisel:** Auch wenn wir uns in unserer Stellungnahme zu dem Thema der Confounder-identifikation nicht geäußert haben, möchte ich kurz einen Diskussionsbeitrag leisten basierend auf dem, was bisher gesagt wird. Ich denke, es ist ja sehr klar, dass die Confounder dazu dienen, Verzerrungen in einem späteren Vergleich zwischen dem zu prüfenden neuen Verfahren und der Standardtherapie zu vermindern. Deswegen, glaube ich, ist ganz wichtig, diese Suche - und da stimme ich dem Moderator sehr zu - zunächst breit zu gestalten, weil eben das Vergessen von Confoundern, auch wenn deren Evidenz vielleicht gering ist, natürlich

dazu führt, dass wir Verzerrungen einführen, die am Ende die primäre Frage nicht bewerten lassen.

Das führt mich zur Frage: Gibt es eine Anzahl von Referenzen, die da sein müssen? Ich glaube, das kann man auf keinen Fall tun, weil die Qualität der Referenz natürlich völlig unterschiedlich sein kann. Daten aus einer klinischen Studie, wo also eine prospektive, systematische Erfassung eines potenziellen Confounders gemacht wurde, müssen doch ganz anders eingeteilt werden als prospektive oder gar retrospektive Registerstudien. Deswegen, glaube ich, ist eine Zahl deutlich weniger relevant als die Qualität und die Art der Datenerhebung in der entsprechenden Referenz.

Und dann komme ich zu dem Punkt „Experteneinbezug“. Da finde ich sehr klar, dass schon in dieser frühen Phase der Identifikation potenzieller Confounder eine unbedingte Expertenbewertung erforderlich ist, weil ja immer in Beziehung zu setzen ist, was die Evidenz ist, die wir extrahieren können, bezogen auf das, was überhaupt über die Erkrankung, die eine seltene sein kann, bekannt ist. Also, vielleicht muss ja eine geringerwertige Evidenz bei einer Erkrankung, die sehr selten ist und die deswegen aufgrund von Fallzahlen keine hohe Evidenz hat, ganz anders bewertet werden als bei einer Erkrankung mit einer ganz hohen Prävalenz. Das führt mich dazu, dass ich und auch die DGKJ es unbedingt sinnvoll finden, eben schon in diesem frühen Prozess der Confounderidentifikation vor deren Bewertung und vor der Frage der Feasibility für eine Erhebung die Experten mit einzubeziehen.

**Moderator Thomas Kaiser:** Vielen Dank für die Anmerkung. - Kurze Info dazu: Wir haben das Thema auch in unserem Arbeitspapier zur Confounderidentifikation bearbeitet. Da ist natürlich die Frage, an welcher Stelle man das macht, eine völlig berechtigte, die allerdings auch mit einer Machbarkeit und Aufwand insgesamt für die Expertinnen und Experten zu tun hat. Aber das ist tatsächlich ein Thema, das man berücksichtigen, den richtigen Ort, den richtigen Zeitpunkt berücksichtigen und dann in einer Durchführbarkeit auch adressieren muss.

Ich habe jetzt Stefan Lange, dann Herrn Sauerbruch und dann bei uns intern noch mal Volker Vervölgyi. - Stefan, bitte.

**Stefan Lange:** Erst mal danke für den Hinweis. Das fand ich sehr gut mit dem Differenzieren von Aufwand im Vorfeld und Aufwand dann in der Durchführung. Ich glaube, Letzteres ist fast noch wichtiger. Wir reden ja im Zusammenhang mit der AbD häufig über seltene, wenn nicht gar sehr seltene Erkrankungen mit sehr, sehr niedrigen Fallzahlen zum Teil, die dann in klinische Studien eingeschlossen werden können oder auch in solche Beobachtungsstudien wie der AbD. Und dann muss ich natürlich überlegen, um überhaupt noch irgendwas sinnvoll schätzen zu können, dass ich das irgendwie reduzieren muss. Ich meine, das ist ja auch Ausdruck im Fazit des Berichts, wo ja relativ ernüchternd am Ende festgestellt wird, bei

seltenen Studien wissen wir doch nicht so ganz genau, ob das dann mit den PS-Methoden so wirklich hinhaut. Und das mag dann eben auch tatsächlich mit der Zahl der zu berücksichtigenden Confounder zusammenhängen.

Jetzt ist mir bei den beiden vorigen Redebeiträgen noch was eingefallen, was wir nicht adressiert haben, aber was ich mir erlaube, trotzdem hier zu sagen, Stichwort „Qualitätskriterien“. Ich bin nicht ganz sicher, ob es jetzt schon angesprochen war, aber es gibt natürlich Qualitätskriterien für Beobachtungsstudien.

Übrigens, im Gegensatz zu meinem Vorredner, würde ich tatsächlich Beobachtungsstudien hier den Vorzug geben. Denn in klinischen Studien ist es ja so, dass der Stellenwert der berühmten - das sagt man immer - Table 1 oder diese Darstellungen von Patientencharakteristika ja weniger der Frage dient, ob da irgendein Confounding ist. Das haben wir in klinischen Studien nicht. Das liegt in der Natur der Sache. Bei der Randomisierung brauche ich das alles gar nicht. Vielmehr dient sie der Charakterisierung der Studienpopulation. Dass nun immer wieder diese Unart gemacht wird, das irgendwie zu vergleichen und dann einen p-Wert anzugeben oder sonst etwas, das ändert nichts daran, dass nicht die stringentesten Kriterien zur Auswahl dieser Baselinecharakteristiken herangezogen wurden. Deswegen, glaube ich, würde ich eher a priori sagen, in Beobachtungsstudien hat man sich da unter Umständen sehr viel mehr Mühe gegeben. Dazu gibt es ja dann auch Instrumente, um dieses Vorgehen zu bewerten, könnte man dann einsetzen.

Aber mir fällt noch ein ganz wichtiges Qualitätskriterium diesbezüglich ein, nämlich ob ein Design einer Beobachtungsstudie vorab publiziert worden ist. Das ist nicht so häufig, das weiß ich wohl. Aber das kann natürlich auch schon ein gutes Kriterium sein, zu sagen, Mensch, da haben sich Leute echt viele Gedanken vorher schon gemacht über die Fragen, was denn hier irgendwie kausal miteinander verknüpft ist.

Last but not least - du hast es angesprochen - Korrelation ist, glaube ich, jetzt nicht Thema, aber darauf würde ich gerne noch mal trotzdem hinweisen. Wir finden das ja total sinnvoll, wie es beschrieben ist, Merkmale, die inhaltlich irgendwie einem gemeinsamen Konstrukt angehören, zu reduzieren. Aber man kann eben wahrscheinlich noch darüber hinausgehen. Auch wenn sie augenscheinlich nicht dasselbe Konstrukt messen, können sie ja trotzdem miteinander korreliert sein und damit natürlich auch in der Analyse entweder überflüssig sein oder gar Neues erzeugen, was ja kontraproduktiv wäre.

Okay, so weit.

**Moderator Thomas Kaiser:** Vielen Dank dafür, auch für die Hinweise zu dieser Unterscheidung zwischen Konstrukt und Korrelation zwischen verschiedenen Konstrukten, um da aber auch noch mal gegebenenfalls für die Durchführung des Forschungsvorhabens eine Reduktion des

Aufwands hinzubekommen. - Dann habe ich jetzt Herrn Sauerbruch, dann intern noch mal Volker Vervölgyi und dann noch mal Herrn Meisel.

**Tilman Sauerbruch:** Ich möchte auch noch mal eine Lanze brechen für die frühe Einbindung von Experten, vor allem, wenn sie auf dem Gebiet zum Beispiel seltener Erkrankungen forschen, weil sie hinsichtlich der Confounder auch die wissenschaftliche Plausibilität nicht nur hinsichtlich Statistik, sondern auch des Pathomechanismus ein Urteil abgeben können, ob das überhaupt sinnvoll ist und wie das eingreift in die Pathophysiologie der jeweiligen Erkrankung. Das wäre dann auch wichtig.

**Moderator Thomas Kaiser:** Vielen Dank für den Hinweis. - Dann habe ich Volker hier intern.

**Volker Vervölgyi:** Ich habe eine Frage, vor allem an Stefan Lange, oder auf den Punkt von Stefan Lange würde ich gerne eingehen. Tatsächlich ist es ja so, dass wir uns, als wir uns mit dem Arbeitspapier auseinandergesetzt haben, genau die gleichen Fragen gestellt haben. Also, hat eine RCT den gleichen Stellenwert wie eine Beobachtungsstudie? Wir haben gesehen, dass einzelne von den Faktoren, die wir gesehen haben, tatsächlich nur in einer RCT in der Table 1 standen. Da ist natürlich die Frage - das hattest du eben auch gesagt -, da ist es schwierig, einen Schwellenwert festzulegen. Das hängt sicherlich davon ab, wie viel Evidenz ich eigentlich insgesamt habe. Trotzdem ist es meines Erachtens schwierig, woran man das insgesamt festmacht. Die Frage ist, ob es dazu gute Empirie gibt. Du hast eben die DAGs angesprochen. Auch darüber haben wir natürlich diskutiert. Aber ein DAG ist ja eigentlich erst mal nur eine visuelle Darstellung der Annahmen, die ich habe. Da wäre halt die Frage, auf welcher Informationsbasis man eigentlich einen DAG erstellt. Also, macht man das erst mal möglichst literaturgestützt, oder macht man das halt unter Einbindung von Experten, die natürlich daran beteiligt werden sollten?

Eine weitere Frage, die ich noch habe, wäre zum Thema „Korrelation“. Korrelation ist ja nicht 0 und 1, sondern ist zwischen minus 1 und 1, das wissen wir alle. Die Frage ist auch da natürlich: Bräuchte man da nicht einen Schwellenwert, um zu sagen, die sind so hinreichend korreliert, dass man den einen oder den anderen Faktor halt nicht braucht?

Das wären ein paar Fragen, die ich habe.

**Moderator Thomas Kaiser:** Vielen Dank. - Dann würde ich - Herr Meisel, wenn Sie erlauben - Herrn Lange eben mal kurz die Möglichkeit geben, darauf zu antworten. Sie sind nicht vergessen.

**Stefan Lange:** Zwei Sachen dazu. Man dreht sich natürlich ein bisschen im Kreis. Was ist zuerst, DAG oder Grundlage für DAG? Ich denke mal, die Leute, die versuchen, so einen gerichteten azyklischen Graphen aufzumalen, haben sich natürlich vorher Gedanken gemacht. Auf welcher Basis auch immer, das muss man natürlich im Einzelfall gucken. Das kann eine

systematische Identifikation sein, das kann Expertise sein, das kann beides sein. Trotzdem hilft es, glaube ich, sehr, sich eben die verschiedenen Zusammenhänge, besonders wenn es eine Reihe sind, noch mal klarzumachen und auch eventuell da schon Anhaltspunkte für Korrelationen zu erkennen unter Umständen.

Noch mal, ich glaube, dass es im Sinne eines Qualitätskriteriums auch gelten kann, wenn ich eine Beobachtungsstudie habe, wo ich weiß, da ist viel Hirnschmalz reingesteckt worden im Vorfeld der Studie, um genau diese wesentlichen Faktoren zu identifizieren.

Also, Schwellenwert hatte ich schon gesagt. Für die Zahl könnte ich jetzt ad hoc nichts sagen, aber es ist immer so mit Schwellenwerten. Wenn man einen nennt, dann heißt es: Ja, wie kommst du denn da drauf? Und wenn man keinen nennt, dann wird man gefragt: Welches ist denn der Schwellenwert? Also, da muss man sich halt irgendwann mal festlegen. Man kann überlegen, ob man wiederum irgendeine sinnvolle Empirie dazu findet oder irgendeine sinnvolle Simulation oder sonst etwas, wo man den Eindruck hat, wenn ich die und die Größenordnung in der Korrelation habe, dann kann das nicht mehr so wahnsinnig viele Auswirkungen haben, bzw. das erzeugt wie gesagt sogar vielleicht mehr Rauschen als dass es tatsächlich Erkenntnisgewinn bringt. Da würde ich mich jetzt außer Stande setzen, zu sagen, da setzen wir jetzt mal 0,7, 0,5, 0,8 oder so etwas ein. Da bitte ich um Verständnis. Und am Ende, wie gesagt, würde es auch nicht helfen, weil dann würde man mich fragen: Warum?

**Moderator Thomas Kaiser:** Vielen Dank. Deswegen haben wir ja auch noch TOP 3 heute mit dem Schwellenwert für die Nullhypothese, aber dazu werden wir ja dann später kommen. - Ich habe jetzt Herrn Meisel.

**Roland Meisel:** Ich wollte nur kurz aus der weiteren Diskussion dieses Themas „Confounderreduktion“ einmal bezüglich des Einbezugs von Experten einordnen. Also, ich denke schon, und die DGKJ denkt, dass eben diese beiden Schritte, die Identifikation und Bewertung potenzieller Confounder und dann aber auch später das Assessment, ob diese realistisch im Setting einer AbD erhebbar sind, ganz getrennte Dinge sind. Die sollten getrennt voneinander betrachtet werden. In beiden ist Expertenwissen erforderlich. Denn es kann ja durchaus sein, dass für eine Erkrankung und eine Fragestellung bezüglich eines neuen Therapieverfahrens Confounder klar identifiziert werden, die aufgrund der zur Verfügung stehenden Patientenmenge dann am Ende eine AbD statistisch schwierig herausfordernd oder unmöglich machen. Und dann ist es doch wichtig zu bewerten, welchen Stellenwert diese Confounder haben. Also, ist es möglich, die niedriger Gerankten zu entfernen, um zu einem Studienkonzept zu kommen, was auch eine statistische Power erlangt, oder muss man auch auf der Basis von Expertenwissen sagen, diese Confounder sind so wichtig, damit ist eine AbD unfeasible? Deswegen plädieren wir dafür, dass eben die Bewertung möglicher Confounder in der Identifikationsphase, aber auch in der Frage: „Kann man sie überhaupt erheben, und sind

sie damit quasi für eine AbD wirksame Erhebungsparameter?“ ... unbedingt voneinander getrennt betrachten sollte.

**Moderator Thomas Kaiser:** Vielen Dank auch noch mal für das Statement, das auch noch mal so ein bisschen die Kaskade der gesamten Aufgabe Confounderidentifikation für ein Forschungsvorhaben beschreibt.

Da wir keine Meldungen mehr haben, würde ich mal so ein bisschen zusammenfassen. Tatsächlich, denke ich, ist es wichtig, zu unterscheiden zwischen Aufwand für diesen ersten Schritt vor dem Forschungsvorhaben und Aufwand fürs Forschungsvorhaben. Ich weiß, dass bezüglich des Aufwands für die Confounderidentifikation durchaus - Jana Göbel hat das ja am Anfang beschrieben - gesehen wird: Oh, das ist aber ein relativ großer Aufwand. - Ich glaube aber tatsächlich, es lohnt sich, und zwar nicht mit der Zielsetzung, so viele Confounder wie möglich zu identifizieren, sondern so sinnvoll wie möglich diejenigen zu identifizieren, die wirklich für das Forschungsvorhaben notwendig sind, und gegebenenfalls auch zu reduzieren dann, da durchaus auch Aufwand zu investieren, weil der Gesamtaufwand für das Forschungsvorhaben dann vermutlich deutlich geringer werden wird. Ob und inwieweit das, denn das ist unsere Erfahrung, durch eine mehrfache Einbindung auch externer Expertise gelingt, das muss man sich vielleicht im jeweiligen Feld dann anschauen. Nachvollziehbar, Herr Meisel, ist das, dass an verschiedenen Stellen da auch noch mal ein Input hilfreich ist. Und da kommen dann ja sowohl methodische Expertise als auch medizinische Fachexpertise zusammen, um das dann gemeinsam zu überlegen.

Das wäre für mich so die Schlussfolgerung, dieser gemeinsame Blick auf den Gesamtaufwand und durchaus am Anfang ein bisschen mehr investieren, um insgesamt dann das Forschungsvorhaben machbar zu machen.

Gut. Vielen Dank. Wir werden auf jeden Fall an dem Thema weiterarbeiten, denn wir haben ja auch noch das Folgeprojekt zu unserem Arbeitspapier, wo wir auch noch mal schauen, ob wir in dem gesamten Verlauf da auch noch mal Aufwand reduzieren können, ohne dass wir zu viel Informationsverlust haben.

Jetzt habe ich aber trotzdem noch mal ... Meine Zusammenfassung hat zwei Meldungen provoziert, zunächst Stefan Lange, dann Frau Weinmann-Menke.

**Stefan Lange:** Nein, das stimmt gar nicht. Das war nicht die wunderbare Zusammenfassung, sondern vielmehr ist mir noch ein Gedanke gekommen. Ich weiß nicht, ob der bisher schon eine Rolle gespielt hat, vielleicht ist er auch irrelevant, aber vielleicht könnte man noch mal darüber nachdenken.

Man müsste ja eigentlich auch noch differenzieren zwischen Confoundern im Hinblick auf das Outcome, was mich interessiert, aber mindestens genauso wichtig ist ja die Frage: Was

triggert denn die Therapieentscheidung, und sind das unter Umständen nicht auch unterschiedliche? Das ist mir jetzt gerade so, wie gesagt, in den Kopf gekommen. Vielleicht spielt es auch keine Rolle, aber vielleicht kann man das ja auch noch mal mitnehmen.

**Moderator Thomas Kaiser:** Vielen Dank auch noch mal für den Hinweis. - Dann habe ich Frau Weinmann-Menke.

**Julia Weinmann-Menke:** Wir hatten ja auch eine Stellungnahme eingereicht. Ich glaube ehrlich gesagt, es passt nicht wirklich zu den Confoundern, nur ich wollte noch mal darauf hinweisen, weil primär kann man sich natürlich Confounder angucken für eine Studie, nur in der Nephrologie haben wir das Problem, dass wir unsere Endpunkte gar nicht - in Anführungszeichen - nach den klassischen ..., die zu bewerten am Ende anstehen, bei einer Bewertung des Einsatzes eines neuen Medikaments erreicht werden können, weil im Endeffekt die Beobachtungszeit dann so lange wird. Also, wir sprechen dann nicht mehr von drei Jahren, sondern wir reden plötzlich von 10, 20, 30 Jahren, sodass hier unsere Anmerkung war, dass man von den klassischen Endpunkten abweichen muss und andere Endpunkte bzw. Surrogatparameter mit einbezieht, damit man überhaupt adäquate Studienergebnisse erreichen kann.

**Moderator Thomas Kaiser:** Vielen Dank für den Hinweis. Der ist ja - ich sage jetzt einfach mal - ein allgemeingültiger, der auch nicht nur bei nicht randomisierten Studien eine Rolle spielt, sondern ganz generell eine Rolle spielt, wobei innerhalb der Indikationen ja verschiedene Bewertungen schon durchgeführt worden sind und in den Studien auch in einer absehbaren Frist durchaus relevante Ergebnisse erhoben worden sind. Aber die Frage von Relevanz von Surrogaten und auch tatsächlich adäquater Surrogatvalidierung ist eine wichtige, aber tatsächlich kein Spezialthema jetzt für diesen Bericht. Trotzdem noch mal danke für den Hinweis.

Gut. Dann wären wir mit TOP 1 durch und kommen jetzt zu

#### **2.4.3 Tagesordnungspunkt 2: Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen**

Bitte sehr.

**Caroline Wöhl:** Unter diesem Tagesordnungspunkt möchten wir heute gerne zwei Punkte zur Diskussion stellen. Das ist zum einen die Abschätzung der Patientenzahlen und zum Zweiten die finale Fallzahlabschätzung zum Zeitpunkt von Zwischenanalysen.

Starten wir mit Punkt 1. Ziel der orientierenden Betrachtung von Fallzahlszenarien ist ja grundsätzlich die Beantwortung der Frage, welche Effekte bei gegebener Patientenzahl unter Berücksichtigung der verschobenen Nullhypothese und gegebenenfalls eines Ankers in einer AbD aufgedeckt werden können. Also, als Ausgangspunkt dient die für eine AbD potenziell zur Verfügung stehende Patientenzahl im zu bewertenden Anwendungsgebiet. In den Stellungnahmen von AbbVie und Pfizer wurde angemerkt, dass die tatsächlich in einer AbD rekrutierbare Patientenzahl nicht dieser theoretisch, also der potenziell zur Verfügung stehenden Patientenzahl entspreche und wir im Rahmen der Konzepterstellung realistische Annahmen treffen sollten.

Darüber hinaus hatte CSL Behring angeregt, diesen anteiligen Verlust von Patientinnen und Patienten, die zum Beispiel aufgrund von Matching oder Trimming in den Auswertungen nicht berücksichtigt werden können, in der Machbarkeitsanalyse aber wiederum über Szenarien einfließen zu lassen.

An diesem Punkt möchten wir gerne noch mal klarstellen - das haben wir auch im Rapid Report bereits getan -, dass es grundsätzlich in der Verantwortung des betroffenen pharmazeutischen Unternehmers liegt, hier eine ausreichende Patientenzahl zu rekrutieren, um darüber dann entsprechend sicherzustellen, dass mit einer AbD überhaupt aussagekräftige Ergebnisse generiert werden können.

Aber nun zu unseren beiden Fragen zu diesem Punkt.

Erstens: Auf Basis welcher Informationen würden Sie denn abschätzen, wie viele Patientinnen und Patienten schlussendlich für eine AbD zur Verfügung stehen? Und zweitens wäre die Frage, ob Sie uns Literatur nennen können, aus denen Richtwerte für die Anteile von Patientinnen und Patienten hervorgehen, die je nach gewähltem Propensity-Score-Verfahren nicht in der Auswertung berücksichtigt werden können.

**Moderator Thomas Kaiser:** Vielen Dank für die Einführung mit zwei klaren Fragen. The Floor is open. Zwei Firmen wurden direkt angesprochen, die das angesprochen haben in ihren Stellungnahmen. Ich sehe noch nichts. Vielleicht gibt es auch keine Literatur oder Überlegungen. - Gut, wenn das nicht der Fall ist, dann scheint das so zu sein. Dann können wir leider diese Fragen nicht weiter erörtern, müssen dann gucken, wie wir dann in der Bearbeitung des ... Da haben wir jemanden. Herr Annacker, bitte.

**Vincent Annacker:** Mit Literatur können wir auch nicht dienen, aber wir haben ja in unserer Stellungnahme das grundsätzliche Fazit gezogen, dass wir zu dem Schluss kommen, dass die orientierende Betrachtung von Fallzahlszenarien grundsätzlich nicht dazu geeignet ist, die Realisierbarkeit einer AbD einzuschätzen. Dieses Problem, dass man diese Zahlen zum Beispiel aus G-BA-Verfahren, - Nutzenbewertungsverfahren nicht heranziehen kann, ist sozusagen ein

Bauteil, der letztlich zu diesem Fazit führt. Also, es ist aus unserer Sicht dann auch ... Ja, da gibt es jetzt einfach keine Literatur, weil es unseres Erachtens nicht möglich ist, das sinnvoll heranzuziehen. Das zu Ihrer Frage zu entsprechender Literatur. Also, damit können wir nicht dienen, weil wir es einfach grundsätzlich nicht als sinnvoll erachten, sondern eben die orientierende Fallzahlschätzung weiterhin als einzigen sinnvollen Weg erachten.

**Moderator Thomas Kaiser:** Vielen Dank. Das nehmen wir so zur Kenntnis. - Dann habe ich Herrn Eisele.

**Lewin Eisele:** Bezuglich dieser Literatur. Wenn ich mich recht entsinne, haben wir für unsere AbD auf ein Paper von Austin et al. verwiesen. Der hatte Simulationsstudien gemacht, um dann für verschiedene Propensity-Score-Verfahren so einen Korrekturfaktor zu berechnen, den man dann benutzen kann, um zum Beispiel so eine orientierende Fallzahlschätzung noch anzupassen für das zu verwendende Propensity-Score-Verfahren. Also, Austin et al. Nein, ich glaube, Austin war der einzige Autor, und es war eine Simulationsstudie. Den Titel weiß ich nicht mehr genau, aber das wäre jetzt eine Referenz, die vielleicht zu Ihrer Frage passen würde.

**Moderator Thomas Kaiser:** Vielen Dank. Wenn ich da einmal nachfragen darf: Das heißt, das ist ein Zitat gewesen, eine Studie, die Sie im Rahmen einer AbD zitiert haben, aber jetzt nicht in der Stellungnahme. Welche AbD war das?

**Lewin Eisele:** Talquetamab.

**Moderator Thomas Kaiser:** Okay, das müssen wir uns dann einfach mal anschauen. Vielen Dank. - Dann habe ich Frau Sternberg.

**Kati Sternberg:** An der Stelle stimme ich Ihnen zu. Natürlich obliegt es in der Aufgabe des pU, entsprechende Patientenzahlen zu rekrutieren. Ich finde dennoch, dass es sehr, sehr wichtig ist, neben der theoretisch möglichen rekrutierbaren Patientenzahl noch eine Abstufung zu machen, was denn praktisch tatsächlich für einen pharmazeutischen Unternehmer möglich ist, zu rekrutieren. An der Stelle ist es natürlich schwierig, konkrete Literaturangaben, wie das zu bewerkstelligen ist, zu machen, weil es einfach sehr diffuse Gründe dafür gibt, warum diese Patienten nicht rekrutierbar sind. Als Beispiel nenne ich einfach mal, dass eine AbD in wenigen Zentren durchgeführt wird und es deshalb für die Patienten vermutlich komplexer ist, diese zu erreichen, oder Ähnliches. Da spielen ja ganz viele verschiedene Gründe rein, warum nicht alle Patienten in diese entsprechende AbD aufgenommen werden können. Da spielen auch solche Dinge rein, neue Produkte, die auf den Markt kommen werden in Zukunft, wie sich die entwickeln. Dazu gibt es natürlich kein Paper, aber natürlich gibt es theoretisch dazu schon Untersuchungen, und dann muss man entsprechend extrapoliieren, wie sich das in Zukunft entwickeln wird, was natürlich in einem sehr dynamischen Markt sehr schwierig ist.

Also, insgesamt sehe ich diese Aufgabe sehr, sehr schwierig, und ich würde an der Stelle, ähnlich wie das in mündlichen Anhörungen ja auch schon geschieht, auf die Expertenmeinung hoffen, die da schon hoffentlich ein zwar nicht quantifizierbares, aber vielleicht ganz gutes Gefühl haben, welche Patienten tatsächlich in der Praxis wirklich vorhanden sind.

**Moderator Thomas Kaiser:** Vielen Dank für die Anmerkung, Frau Sternberg. Das provoziert mich ein wenig zu der Rückmeldung, dass wir in den Anhörungen durchaus unterschiedliche Sichtweisen von Experten und Expertinnen haben. Die Registerbetreibenden sagen ganz oft, das klappt, und andere sagen, oh, das wird schwierig. Insofern ist das zwar einerseits eine nicht ganz unwichtige Information, die aber nicht immer zum Ziel führt. Aber trotzdem danke noch mal für den Hinweis.

Ich habe dann hier intern Volker Vervölgyi und dann Herrn Meisel.

**Volker Vervölgyi:** Ich hätte jetzt ein ähnliches Statement gesagt. Sie haben gerade selber gesagt, das ist alles sehr schwierig und es ist auch schwierig zu quantifizieren. Ich will mal kurz darauf hinweisen, in welcher Situation wir sind. Das ist ja auf Basis der Ebene des Konzeptes, das wir erstellen, und die Anhörung kommt ja erst danach. Also, Sie haben gerade gesagt, das ist natürlich auch fragestellungsabhängig, damit hängen ganz viele Faktoren zusammen. Deswegen wäre meine Rückfrage erst mal: Wie würden Sie eigentlich dann auf Ebene des AbD-Konzeptes für eine neue Therapie, die es noch nicht gibt, noch nicht zugelassen ist, die gerade im Zulassungsverfahren ist, gegebenenfalls auch ein neues Therapiekonzept hat usw., ... wie man das für eine ganz konkrete AbD berücksichtigen soll?

Ein anderer Punkt, der mir dabei eingefallen ist, als Sie eben ausgeführt haben, war: Man muss, glaube ich, auch trennen zwischen Faktoren, die vom pU beeinflussbar sind, also solche wie: „Kann ich zusätzliche Zentren rekrutieren? Kann ich die Erhebung der Daten im Verlauf vielleicht für die Patientinnen und Patienten möglichst einfach machen, indem ich vielleicht durch digitale Tools das alles regle und die nicht jedes Mal ins Zentrum kommen müssen?“, und solche, die man natürlich nicht unbedingt beeinflussen kann, wie das, was Frau Wöhl eben angesprochen hat, durch das Trimming zum Beispiel. Da lässt sich ja nichts machen. Wenn die Patientinnen und Patienten sind, wie sie sind, und die Methodik so ist, dann geht es einfach nicht anders. Aber es gibt sicherlich viele Faktoren, die auch beeinflussbar sind. Die müsste man natürlich da mitberücksichtigen auf Ebene des Konzeptes, um dann eine realistische Einschätzung machen zu können.

**Moderator Thomas Kaiser:** Möchten Sie dazu eine Rückmeldung geben? - Frau Sternberg, bitte.

**Kati Sternberg:** Da gebe ich Ihnen tatsächlich völlig recht. Das ist zu so einem frühen Zeitpunkt, zur Konzepterstellung, eine starke Herausforderung. Ich würde an der Stelle tatsächlich dafür

plädieren, frühzeitig Experten einzubinden. Weil, wie gesagt, die Datengrundlage fehlt uns zu diesem Zeitpunkt. Und was wir an der Stelle tun müssen, ist extrapolieren in die Zukunft. Das fällt uns bei dieser Patientenpopulation extrem schwierig, gerade wenn es um eine neue Anwendung geht.

**Moderator Thomas Kaiser:** Gut. Vielen Dank. - Dann habe ich Herrn Meisel, dann Frau Acikgöz und dann Herrn Sauerbruch.

**Roland Meisel:** Ich glaube, das wirklich Schwierige an dieser Frage ist natürlich die Quantifizierung. Ich habe weder Literatur noch eine Lösung dafür. Aber aus der Erfahrung mit einigen AbD-Konzepten, die zum Teil in eine Erhebung übergegangen sind, die zum Teil zu einer Einschätzung einer Nichtdurchführbarkeit geführt haben und zum Teil am Start stehen, möchte die DGKJ noch mal darauf hinweisen, dass es ganz wichtig ist, dass es wichtige soziale, kulturelle und auch sprachliche Barrieren für Patienteneinschlüsse gibt. Die verzerrten Ergebnisse und lassen zur Verfügung stehende Patientenzahlen, die sich vielleicht aus Registern oder aus epidemiologischen Abrechnungsdaten von Krankenkassen ergeben, einfach einordnungspflichtig werden. Das führt wieder dazu, dass für die Abschätzung verfügbarer Fallzahlen bezüglich der Feasibility - nicht der reinen Verfügbarkeit aufgrund einer Kalkulation, wir haben 6.000 Patienten in Deutschland, davon sind soundso viele in der Indikation, soundso viele im Alter - eben diese Faktoren zu berücksichtigen sind. Und wenn man dann eben an der Untergrenze einer Fallzahl ist, dann sollte das schon in der Konzeptbewertung mitberücksichtigt werden. Also, diese verfügbaren Fallzahlen müssen bezüglich dieser Barrieren mitbedacht werden.

Ein zweiter Punkt, der sehr wichtig ist, ist, bei komplexen AbDs, wo komplexe Erhebungen erforderlich sind, die eben bedeuten, dass es auch nur eine Auswahl von Gesundheitsdienstleistern gibt, die sich überhaupt für eine AbD qualifizieren werden, weil sie die Infrastruktur vorhalten, die durchzuführen, muss man auch berücksichtigen, dass eben nicht nur durch eine Patientenselektion oder eine Einschränkung der Patientenzahl, sondern auch durch eine Einschränkung der Anteil der Gesundheitsdienstleister, die überhaupt an so einer AbD sinnvoll teilnehmen können oder es werden wollen, sich eine Einschränkung ergeben kann. Das sind beides Dinge, die natürlich in den Anhörungen berücksichtigt werden und dann auch zu Wort gebracht werden, die man aber auch in der Planung und in dem ersten Design schon berücksichtigen sollte.

**Moderator Thomas Kaiser:** Vielen Dank für die Hinweise. - Dann habe ich Frau Acikgöz.

**Özlem Acikgöz:** Ich habe zwei Hinweise zu der Diskussion. Das eine ist, wir müssen unterscheiden zwischen der Rekrutierbarkeit und der Tatsache, ob ein Patient oder eine Patientin tatsächlich das Medikament erhält. Wir haben ja eine Beschränkung der Verordnungsbefugnis. Nur die Patienten und Patientinnen sollen das Medikament erhalten,

denen zumindest angeraten wird oder der Hinweis gegeben wird, du musst auch in eine AbD. Da liegt aus meiner Sicht nicht so das große Problem. Also, Patientinnen und Patienten werden sich, wenn sie das Medikament bekommen, wahrscheinlich auch rekrutieren lassen. Das Problem ist eher, und das wurde vorhin schon angesprochen, ob, wenn ich in Deutschland eine bestimmte Anzahl von Patienten und Patientinnen habe, die auch wirklich in den Genuss - in Anführungsstrichen, wenn es ein Genuss ist - der Therapie kommen. Da spielen tatsächlich Aspekte wie regionale Aspekte oder wo die Uniklinik die nächste ist, was es für Begleiterkrankungen und Ähnliches gibt, eine Rolle. Das ist das Erste, also der Appell, zu unterscheiden zwischen Rekrutierbarkeit und ob der Patient oder die Patienten dann wirklich das Medikament bekommt oder es bekommen kann.

Und das Zweite ist - das ist etwas, womit wir uns auch in dem Rahmen jetzt so ein bisschen beschäftigt haben -: Wenn man davon ausgeht, dass das eigentlich sehr schwierig ist in den allermeisten Fällen, wenn die Evidenzgrundlage nicht so gut ist für eine Fallzahlschätzung, würde das ja bedeuten, dass man eigentlich ganz oft auch mal Ethikkommissionen Studien vorlegen müsste, wo es keine Fallzahlplanungsschätzung oder Fallzahlplanung gibt. Also, das kann es ja an der Stelle auch nicht sein. Deshalb plädieren wir als DKG dafür, auch um die Angemessenheit und auch um den Aufwand und alles abschätzen zu können für eine AbD und Umsetzungsmöglichkeiten zu sehen, dass man das orientierend auf der Basis der verfügbaren Informationen macht. Man kann auch Experten ziemlich früh in solche Verfahren einbinden. Wir haben im Gemeinsamen Bundesausschuss das eine oder andere Verfahren, wo wir mit einer Expertenanhörung beginnen und sagen: Sagen Sie uns mal, was glauben Sie denn, wie viele Patienten tatsächlich infrage kommen werden für die Therapie? - Das ist jetzt nicht das AbD-Verfahren, sondern ein anderes Verfahren. Aber vielleicht wäre das noch mal ein Hinweis, wie man sich einer orientierenden Fallzahlschätzung nähern könnte, insbesondere natürlich auch mit dem Blick darauf, dass das immer noch anpassungsfähig im Verlauf der Studie, der AbD, wäre.

**Moderator Thomas Kaiser:** Vielen Dank für den Hinweis, der, wenn ich das richtig verstanden habe, so ein bisschen aber auch in das Gesamtverfahren mit dem Blick auf den G-BA noch mal gegangen ist, was man ja auch an der Stelle in unserem Rapid Report noch mal erwähnen kann: Macht es Sinn, im Vorfeld noch mal ein bisschen mehr zu gucken, zu investieren, bevor man das gesamte Verfahren startet, schon innerhalb des G-BA? Vielen Dank dafür. - Dann habe ich jetzt Herrn Sauerbruch und dann intern Herrn Vervölgyi.

**Tilman Sauerbruch:** Der Begriff „Register“ ist jetzt zweimal gefallen, unter anderem auch von Ihnen, Herr Kaiser. Da möchte ich noch mal etwas anregen. Ich meine, Sozialdaten müssen Sie hinnehmen, wie sie sind. Die Qualität und die Struktur von Registern können Sie aber formen. In dem Fall möchte ich anregen, dass vielleicht doch mal zwischen den wissenschaftlichen

Schwerpunktgesellschaften und dem IQWiG eine Diskussion hineinkommt, dass man im Grunde dann auch besser auf die Register zurückgreifen kann.

**Moderator Thomas Kaiser:** Vielen Dank. Dazu sind wir gerne bereit. Wir haben ja vielfältige Diskussionen auch nach dem Rapid Report A19-43 gemacht mit verschiedenen Fachgesellschaften, können das aber auch gerne noch erweitern. Wir werden mal sehen, ob mit dem noch zu erwartenden Registergesetz da auch noch mal was passiert. - Volker.

**Volker Vervölgyi:** Zu dem Thema „Rekrutierbarkeit“. Tatsächlich ist das Problem, und das ist die Frage, die ich eben schon hatte: Wie soll man das für eine konkrete Fragestellung antizipieren? Natürlich - Sie hatten das gesagt, Frau Acikgöz -, man könnte das ins Verfahren einbetten und dann halt frühzeitig Experten einbinden. Die Frage ist aber trotzdem: Wir versuchen, eine Studie durchzuführen in einer Situation, wozu es bisher keine vergleichenden Daten gibt. Der springende Punkt bei der Fallzahlschätzung, die Sie gerne hätten, ist ja nicht die Patientenzahl, die wir zur Verfügung haben - das ist sicherlich auch ein Punkt, über den man reden muss - und wie viele davon abgehen - aber der wichtigste Punkt an dieser Stelle ist, habe ich eigentlich ausreichend Informationen zum erwarteten Effekt? Das ist ja genau der Punkt, warum das in den Fällen, in denen wir keine orientierende Fallzahlschätzung gemacht haben, nicht ging, nämlich weil die Informationen, die wir sowohl auf Vergleichsseite oder auf Interventionsseite nicht ausreichend hatten oder die einfach nicht zusammenpassten, weil die komplett unterschiedliche Fragestellungen berührt haben ... Das ist der springende Punkt. Es ist ja nicht so, dass wir das nicht machen würden, wenn es ginge. Aber tatsächlich fängt man irgendwann an, sich methodisch zu verbiegen, wenn man versucht, das auf möglichst der verfügbaren Informationsbasis zu machen. Das ist ja genau der springende Punkt. Ich glaube, darin liegt die große Schwierigkeit.

**Moderator Thomas Kaiser:** Vielen Dank für den Hinweis noch mal, dass es also nicht alleine um Korrekturfaktoren geht, sondern bei dieser Überlegung, wie viele ... Oder anders gesagt: Ist aufgrund der zu erwartenden - ich sage jetzt mal nicht „Gesamtfälle in Deutschland“ - rekrutierbaren Personen in die AbD plus Abschlag gegebenenfalls aufgrund der Methodik dann aber noch eine sinnvolle AbD durchführbar? - Da sind wir bei Frau Sternberg und dann Stefan Lange.

**Kati Sternberg:** Ich stimme Ihnen zu, die Fallzahlplanung hat an der Stelle zwei Komponenten. Die eine ist, ob sich diese Fallzahl, die eine Fallzahlplanung berechnet, tatsächlich realisierbar ist, Stichwort „Population, Ein- und Ausschluss“. Und die andere Komponente sind natürlich die Effektschätzer.

Ich glaube, an der Stelle würde ich gerne dafür plädieren, etwas pragmatischer mit diesen Effektschätzern umzugehen. Ich glaube, in Ihrem Paper steht tatsächlich auch, dass wir uns mehr in den Zulassungsstudien befinden, dass wir da immer perfekte Fallzahlplanungen - ich

übertreibe ein wenig - haben, weil Evidenz vorliegt. Das ist nach meiner Erfahrung aus der Planung klinischer Studien in den ganz wenigsten Fällen so, weil wir uns auch für Phase-III-Studien tatsächlich sehr pragmatisch auf diese Evidenz stützen müssen für unsere klinischen Studien, die an der Stelle vorliegen. Das sind Phase-II-Studien in einer anderen Population mit ganz anderen Ein- und Ausschlusskriterien, teilweise einarmige Studien. Für Komparatoren nutzt man andere Studien mit anderen Ein- und Ausschlusskriterien, wo die Population dann nur so ähnlich ist. Also, ich glaube, an der Stelle können wir uns tatsächlich von der regulatorischen Fallzahlplanung etwas abschauen, die da deutlich freigiebiger ist in dem, dass sie alle Evidenz, die irgendwie verfügbar ist, nutzt an der Stelle. Und das müssen nicht mal klassische RCTs sein, auf denen die Fallzahlplanung beruht, sondern so wie wir das in der AbD bisher ja auch gesehen haben: Man macht eine Fallzahlplanung nach bestem Wissen und Gewissen, nutzt alle Evidenz, die man zur Verfügung hat, gegebenenfalls auch von niedrigerem Evidenzgrad als RCTs, und passt gegebenenfalls durch eine Interimsanalyse diese Fallzahlplanung noch mal an. - Diese Möglichkeit haben wir ja auch in der AbD, die ist ja durchaus gegeben.

Das heißt, mein Plädoyer wäre dafür, Fallzahlplanung unbedingt vorab, um grob abzuschätzen, ob diese Fallzahl überhaupt realisierbar ist. Und da kommt wieder die Realisierbarkeit ins Spiel. Nur dann ist die Durchführung ethisch. Aber an der zweiten Stelle wäre dann natürlich eine Interimsanalyse, um die Annahmen der Fallzahlschätzung, die vielleicht aus nicht perfekten Schätzungen von Nicht-RCTs kommen, noch einmal zu revidieren.

**Moderator Thomas Kaiser:** Vielen Dank auch noch mal für diesen Blickwinkel mit, wenn ich das richtig verstanden habe, dem Ausblick: Sind wir nicht tatsächlich in unseren Überlegungen strenger, als dass an anderen Stellen - sogar in klinischen Studien - gemacht wird? Aber danke noch mal für den Hinweis. Es hilft sicherlich, da noch mal eine Reflexion zu bekommen. - Stefan Lange.

**Stefan Lange:** Das geht in eine ähnliche Richtung, wobei das jetzt hier noch ein bisschen verkompliziert wird durch die verschobene Nullhypothese. Aber wenn wir uns davon kurz mal lösen, dann ist das meines Erachtens eine nicht ganz zutreffende Sichtweise, dass für die Fallzahlplanung leitend ist, das, was man meint, zeigen zu können. Was für die Fallzahlplanung leitend sein sollte aus meiner Sicht, ist: Was ist es denn wert, aufgedeckt zu werden? Und das sind unter Umständen zwei sehr unterschiedliche Paar Schuhe. Insofern könnte man sich auf den Standpunkt stellen, das juckt mich gar nicht, ob das realistisch ist. Da werden manche sagen: Oh, das ist aber mit der Ethik schwierig. - Aber, ich glaube, man sollte sich nicht zu sehr darauf fokussieren, zu sagen, ich muss jetzt irgendwie sehr präzise vorher wissen, was denn tatsächlich möglich ist mit dem neuen Arzneimittel. Ich glaube, da reichen unter Umständen ja doch eher grobe Abschätzungen. Also, bewege ich mich in einem Bereich einer

geringfügigen Verbesserung, oder bin ich doch in einem Bereich, wo unter Umständen Welten bewegt werden? Das vielleicht nur so als grundsätzlicher Gedanke.

Wie gesagt, durch die verschobene Nullhypothese kommt dann noch mal eine zusätzliche Komplexität hinein, die wahrscheinlich zu einem Großteil diese Auffassung, was es denn wert ist, aufgedeckt zu werden, schon ein Stück weit mit beinhaltet, wobei das wiederum verkompliziert wird, weil ja die 0,5, die jetzt so ein bisschen im Raum stehen als sozusagen niedrigster Schwellenwert ... Wie soll ich es sagen? Das ist ja nur ein reiner Versuch eines Korrekturfaktors für Bias. Das hat nichts damit zu tun, was es wert ist, aufgedeckt zu werden. Das muss man ja im Prinzip tatsächlich noch obendrauf setzen, wenn man das so versteht mit der Bias-Korrektur.

Ich weiß nicht genau - deswegen habe ich mich auch gemeldet -, ob wir noch über andere Aspekte im Zusammenhang mit der orientierenden Fallzahlschätzung ..., ob das nach der Tagesordnung geplant ist, darüber noch zu sprechen. Sonst hätte ich dann noch kurz was zu unserem Vorschlag ergänzt.

**Moderator Thomas Kaiser:** Gerne, wobei wir diesen Punkt mit der verschobenen Nullhypothese ja gleich noch haben. Aber wenn es noch andere Punkte gibt, gerne.

**Stefan Lange:** Ja, weniger die verschobene Nullhypothese als vielmehr die Überlegung .... Wir fanden zugegebenermaßen diese Darstellung in dem Papier, was sich auch teilweise in AbD-Konzepten gefunden hat, mit dieser orientierenden Fallzahlschätzung bedingt informativ und hilfreich. Ich würde deshalb auch noch mal anregen, ob man sich nicht alternative Darstellungsformate vorstellen kann. Eine Idee war eben, sich irgendwie auf Breite von Konfidenzintervallen zu kaprizieren. Also, bei gegebener Fallzahl, wie präzise kann ich denn überhaupt werden? Und dann kann man unter Umständen den Entscheidungsträgern ein bisschen anheimstellen, wie sie das sozusagen für sich mit der Sinnhaftigkeit einer AbD in Einklang bringen.

Man könnte genauso gut natürlich auch je nach Häufigkeit der Erkrankung - nehmen wir mal an, wir haben eine sehr, sehr seltene Erkrankung - eventuell noch mal überdenken, entweder andere Signifikanzniveaus anzusetzen oder aber eben auch - dazu kommen wir jetzt gleich - über andere Hypothesengrenzen bei der verschobenen Nullhypothese nachzudenken. Diese Darstellung, wie sie jetzt ist, fand ich persönlich, aber das ging jetzt meinen Kolleginnen und Kollegen im MD Bund ähnlich, nicht so wahnsinnig instruktiv.

**Moderator Thomas Kaiser:** Danke für den Hinweis. Den nehmen wir auch noch mal mit, wobei die letzten Darstellungen zumindest in so eine Richtung Szenariendarstellung gegangen sind, also sich zu überlegen, was denn potenziell aufdeckbare Unterschiede in einer Situation dieser oder jener tatsächlich rekrutierten Zahl sind. Wenn du da noch mal sagen kannst, was an der

Stelle besonders wenig informativ war oder welche Änderung besonders sinnvoll wäre aus eurer Sicht?

**Stefan Lange:** Ich habe das, glaube ich, schon verstanden, dass dort Effekte, die aufzudecken sind, und zwar unter den gegebenen Bedingungen, die da formuliert sind ... Also, ich habe einen bestimmten Schwellenwert für die verschobene Nullhypothese, und ich habe ein bestimmtes Signifikanzniveau. Jetzt könnte es ja sein, dass irgendwie dann doch im Gremium angesichts einer zum Beispiel maximal zur Verfügung stehenden Fallzahl von 120 oder 150 Menschen in diesem Land, wo ja die Probleme, glaube ich, am gravierendsten sind, wenn ich das richtig verstanden habe, und was sozusagen die Obergrenze ist dessen, was möglich ist, noch mal über Schwellen und Setzungen, die dahinterstecken, auch hinter dieser Abbildung, ... nachdenken können. Denn da kommt man dann auf Fallzahlen von 500, 700, 900. Na gut, dann stellt man nur fest, das klappt ja alles nicht. Insofern kann ich mir vorstellen, dass man das, wenn man das etwas plastischer macht, wie gesagt zum Beispiel mit der Breite von Konfidenzintervallen, also Unsicherheitsbereiche darstellt, ... könnte das unter Umständen informativer sein.

Jetzt habe ich aber auch noch eine Rückfrage; das ist ziemlich unverschämt. Mich würde noch interessieren - ich weiß nicht, ob das jemand beantworten kann -, was mir jetzt gerade aufgefallen ist, weil ich mir das alles noch mal angeschaut habe, inwieweit in diese Überlegung auch einbezogen ist, dass wir ja eine verschobene Nullhypothese mit einem Nichtzentralitätsparameter möglicherweise haben und diese Nichtzentralitätsparameter unter Umständen auch Einfluss haben auf zum Beispiel Überdeckungsraten und Breiten von Konfidenzintervallen. Ich weiß, das ist jetzt ein bisschen unverschämt, und das tut mir auch leid, aber es ist wie gesagt mir gerade erst eingefallen. Vielleicht kann man das noch mal irgendwie im Kopf bewegen. Das hat wahrscheinlich keinen allzu großen Einfluss, wobei bei der Verschiebung um 0,5 könnte es schon einen Einfluss haben. Guido Skipka sehe ich da, der kann was dazu sagen.

**Moderator Thomas Kaiser:** Genau, ist gar nicht unverschämt, sondern wir haben ja eine wunderbare Biometrieabteilung, und dazu wird Guido Skipka etwas sagen.

**Guido Skipka:** Hallo Stefan. Die Frage ist vielleicht ein bisschen theoretisch für mein Dafür-halten. Natürlich hat man verschobene Nullhypothese, ist die T-Verteilung halt nicht Zentralitätsparameter. Das ist natürlich richtig. Jetzt kann man die zentrale oder die nicht zentrale T-Verteilung sehr gut durch die zentrale approximieren. Natürlich begeht man einen gewissen Fehler. Je weiter man in dem Fall von der 1 weggeht, desto größer wird er. Ich glaube, es spielt tatsächlich in der Praxis eine untergeordnete Rolle, wenn man sich überlegt, was die Hauptstörgrößen hier bei Fallzahlszenarien angeht. Also, ich glaube, das ist mehr eine akademische Diskussion.

**Stefan Lange:** Danke.

**Moderator Thomas Kaiser:** Gut. Vielen Dank. Danke für die Einschätzung. Dann habe ich zunächst hier intern Volker Vervölgyi und dann noch mal Frau Sternberg.

**Volker Vervölgyi:** Ich bin eben noch mal hellhörig geworden, als du, Stefan, gesagt hastest: Was ist es denn wert, in der Studie aufgedeckt zu werden? Ich habe mich gefragt: Muss oder kann überhaupt im Vorhinein klar sein, dass da überhaupt ein Vorteil ist? Also, wir sind ja in der Situation, wir haben keine vergleichenden Daten, und dann kann natürlich erst mal alles herauskommen. Theoretisch kann es auch sein, dass es einfach keinen Vorteil gibt. Und daran anschließend: Kein Zusatznutzenbeleg ist ja auch eine Art von Ausmaß, also eine Quantifizierung. - Die Frage ist, was wollen wir eigentlich vermeiden? Also, wir wollen ja vermeiden, dass die Studie umsonst gemacht worden ist. Die Frage ist ja dann: Was ist eigentlich umsonst? Was wir auf jeden Fall vermeiden wollen, ist ja - das wäre zumindest mein Dafürhalten -, dass man am Ende, weil die Modelle nicht konvergieren, keine Effektschätzung machen kann, weil einfach aufgrund der Confounder und der Patienten, die eingehen, das nicht möglich ist. Alles darüber hinaus wäre erst mal, man hat tatsächlich einen Effekt und man hat eine Größenordnung eines Effekts, man hat eine Präzision, wie auch immer die ist, das hängt ja von der Anzahl der Patienten ab und so. Und das wäre erst mal ein Ergebnis. Also, wäre es nicht erst mal das primäre Ziel, erst mal ein Ergebnis zu erlangen und zu vermeiden halt genau den umgekehrten Fall, nämlich kein Ergebnis zu bekommen? Also, wäre das nicht eigentlich die Schwelle, über die man da reden müsste?

**Moderator Thomas Kaiser:** Stefan Lange hat genickt, kann sich vielleicht dem ein wenig anschließen oder auch nicht.

**Stefan Lange:** Ich muss gestehen, ich war ganz kurz abgelenkt, aber da gibt es ja auch eine Formulierung im Rapid Report, in dem es eben heißt, im Grunde genommen ist ja schon das Ergebnis an sich wertvoll, ob man nicht auch diesen Gedanken haben kann in dem Zusammenhang. Möglicherweise war das jetzt gerade das, was du angesprochen hast. Ja, das teilen wir auf jeden Fall.

Dann stellt sich natürlich die Frage: Ich habe eine Studie mit einem möglichst unverzerrten Schätzer, möglichst unverzerrt, nach Möglichkeiten dieser Studie. Das kann natürlich schon ein Erkenntnisgewinn sein. Ob der dann dahingehend mündet, zu sagen, Zusatznutzen nicht belegt oder, was weiß ich, nicht quantifizierbar oder doch ein bisschen, das ist ja eine andere Frage. Aber Erkenntnisgewinn kann das natürlich schon sein. Insofern kann man sich auch so dieser Frage nähern.

Dieses Paradigma, was es vielleicht in der Vergangenheit gab, eine Studie ist unethisch, wenn es irgendwie keine Fallzahlplanung, Sound und Safe, gibt, das kann man sich möglicherweise

auch im Zusammenhang mit einer AbD ein Stück weit abschminken, zumal wir es ja hier eben nicht mit einer klassischen klinischen Studie zu tun haben in einer experimentellen Situation, sondern mit einer Beobachtungsstudie, für die es übrigens in anderen Zusammenhängen relativ häufig keine Fallzahlplanung gibt. Das ist eher die Ausnahme. Das kann man ja an der Stelle auch noch mal sagen. Das spricht nicht dagegen, sich über die Realisierbarkeit der Studien natürlich Gedanken zu machen. Das ist ja keine Frage.

Aber wenn ich jetzt noch mal den Gedanken von Volker Vervölgyi aufgreifen darf, die Frage der Durchführung ja oder nein, ist die Frage: Wie wertvoll oder wertlos ist das Ergebnis?

**Moderator Thomas Kaiser:** Danke. Ich glaube, es ist tatsächlich wichtig, konzeptionell zwischen Realisierbarkeit und Fallzahlplanung zu unterscheiden. Also, Fallzahlschätzung ist ein bisschen was anderes als Fallzahlplanung. - Dann sind wir bei Frau Sternberg.

**Kati Sternberg:** Jetzt haben Sie in der Zwischenzeit schon zwei Themen aufgemacht. Ich würde gerne zu beiden was sagen.

Die erste Frage war, weil Sie direkt nach Feedback zur Darstellung der Szenarienplanung gefragt haben. Da kann ich vielleicht ein wenig aus dem Nähkästchen plaudern, dass die Darstellung bei uns gerade unter Nichtstatistikern zu großer Verwirrung geführt hat, weil jeder diese klassische Fallzahlplanung gewohnt ist und am Ende eine Fallzahl sieht und daraus entsprechend schlussfolgert, das ist die Zahl, die ich brauche. Und das wird von Nichtstatistikern häufig immer noch so gemacht. Ich glaube, das haben Sie auch schon gemerkt in den Anhörungen, dass es da regelmäßig zur Verwirrung führt, weil jeder annimmt, das wäre jetzt die Fallzahl, die ich brauche, und diese Abschätzung als klassische Fallzahlplanung realisiert. Das ist das eine, was tatsächlich etwas schwierig ist, wo wir sicherlich mehr Erziehung noch bräuchten.

Das andere ist letztendlich, den aufdeckbaren Effekt in Relation zu setzen zu dem, was tatsächlich realisierbar ist. Das ist ja in Ihrem Konzept im Moment nicht enthalten. Da entsteht eine große Lücke, die wir an dieser Stelle irgendwie schließen müssen. Ist denn dieser aufdeckbare Effekt, der ja ein theoretischer ist, tatsächlich realisierbar? Da kommt auch wieder die Szenarienplanung rein, die eben schon erwähnt worden ist. Auch an der Stelle der Szenarienplanung hat der statistisch nicht so versierte Leser das Gefühl, dass es eine realistische Variation an Eingabemöglichkeiten in die Szenarienplanung ist. Wenn wir uns aber mal genau anschauen, was in den letzten Konzepten drinsteht, dann sind das sehr theoretische Variationsmöglichkeiten. Da wird über die - darüber hatten wir schon diskutiert - theoretische Patientenzahl ein wenig hin und her variiert, aber das bleibt immer noch diese theoretische Patientenzahl. Da werden die Effektschätzer der einzelnen Behandlungsgruppen variiert, die aber auch nicht in Relation gesetzt, was wirklich realisierbare Effektschätzer in den einzelnen Behandlungsgruppen überhaupt wären. Da hat man manchmal das Gefühl, das

wird einfach irgendwas zwischen 50 % Responserate und 95 %, und das variiere ich einmal wie wild umher. Also, diese Szenarienplanung ist für unser Gefühl ein sehr theoretisches Konzept mit ganz wenigen Berührungspunkten zur Praxis, und das wird tatsächlich auch schwierig verstanden.

Der zweite Punkt, den Sie inzwischen ja schon aufgemacht haben, ist die Frage, wie viel Evidenz wir denn tatsächlich durch so eine AbD generieren. Da möchte ich noch mal daran erinnern, dass es sich hier um AbDs dreht, die wir für die Nutzenbewertung generieren. Wenn ich mir an der Stelle vorstelle, was ich beispielsweise aus einem Effektschätzer mit einem Konfidenzintervall mache, die Therapien aber nicht untereinander vergleiche bzw. tatsächlich keine Überlegenheit zeigen konnte, dann stellt sich mir die Frage, was ich dann damit mache. Ziel der AbD ist es doch eigentlich, Therapieoptionen untereinander zu vergleichen. Wenn ich hier keinen statistisch signifikanten Überlegenheitstest produzieren kann, wenn das vorab schon absehbar ist, dann ist die Frage, was ich dann für die Nutzenbewertung schlussfolgere und ob ich dann in die Gefahr komme, dass jemand Gleichwertigkeit schlussfolgert aufgrund eines nicht statistischen Tests. Das halte ich an der Stelle für sehr gefährlich.

**Moderator Thomas Kaiser:** Frau Sternberg, das provoziert mich natürlich ein bisschen dazu, dass die AbD ja nur die Korrektur des unzureichenden Wissens zum Zeitpunkt des Eingangs des neuen Arzneimittels in die Versorgung ist. Viele, glaube ich, die jetzt heute hier versammelt sind, ich weiß nicht, ob alle, aber sicherlich ganz viele wären natürlich sehr glücklich, wenn die Quote von ca. 80 % der randomisierten Studien, die es ja durchaus für seltene Erkrankungen gibt in der Zulassung, so wäre, dass sie auch die Frage beantworten würde, nämlich des direkten Vergleichs, sei es zu den aktiven vorhandenen Therapien oder zu BSC. Also, die Gefahr ist ja genau die, dass in der Situation des frühen Marktzugangs ohne die vergleichenden Daten etwas gemacht wird mit einer Überlegenheitsannahme, die zwar Werbung, aber nicht datengestützt ist. Sonst gäbe es die AbD nicht. Wenn klar wäre, dass etwas überlegen ist, wird es keine AbD geben. - Herr Rasch.

**Andrej Rasch:** Lieber Herr Kaiser, nun haben wir aber die AbD, und wir haben nun mal die Situation hin und wieder, dass wir eben diese nicht vergleichenden Daten haben. Es wäre letztlich unser Appell, jetzt auch im Anschluss an die anderen, einfach diese Fragestellung der Fallzahlschätzung tatsächlich gefühlt nicht so sekundär zu behandeln. Warum? Man hat jetzt vor zwei Monaten auch in der Plenumssitzung des G-BA gesehen, zu welchen Diskussionen das führen kann. Es geht letztlich um die Akzeptanz dieses Instruments per se. Es geht auch um nichts weniger als um die Zukunft dieses Instruments, weil man tatsächlich ... Also, man muss deswegen, glaube ich, schon in diesen essenziellen Punkt vielleicht tatsächlich Mühe und Not investieren, so wie man das jetzt auch bei der Confounderadjustierung tut. Ich kann dem nur zustimmen, am Ende muss man schon auch das Gefühl vermitteln, eine faire und realistische Chance zu haben, einen Effekt zu haben für die Quantifizierbarkeit des

Zusatznutzens. Das ist die primäre Aufgabe der AbD. Ich würde tatsächlich doch vehement verneinen, zu sagen, dass das Ergebnis eines nicht belegten Zusatznutzens sozusagen dieses Ziel erfüllt, dass damit sozusagen die Quantifizierbarkeit auch gegeben ist. Das kann natürlich rauskommen, aber es darf nicht das Gefühl vermittelt werden, dass eine AbD dann zustande kommt, die absehbar nicht dieses Ziel einer Quantifizierung des Zusatznutzens erzielen kann.

**Moderator Thomas Kaiser:** Herr Rasch, danke für den Blickwinkel. Ich denke - zumindest habe ich das jetzt so wahrgenommen -, dass wir alle darin übereinstimmen, dass dieses Thema der Fallzahl - ich nenne es ganz bewusst nicht Fallzahlplanung, sondern Fallzahl - auch in Verbindung mit Realisierbarkeit auf jeden Fall etwas ist, was wichtig ist. Man sollte nicht in ein Forschungsvorhaben gehen, das am Ende zu einer Datenerhebung ohne verwertbare Ergebnisse führt, wo es vermutlich nicht - da geht es jetzt gar nicht ums IQWiG, sondern da geht es insgesamt um alle Beteiligten - so eine Übereinstimmung gibt bei ihrer klaren Perspektive, das muss auf jeden Fall die Quantifizierung im Sinne eines Zusatznutzens sein, es muss um Quantifizierung des Zusatznutzens gehen, aber in der AM-NutzenV steckt auch der nicht belegte Zusatznutzen drin. Aber das, wie gesagt, will ich jetzt gar nicht in den Vordergrund rücken. Ich glaube, der entscheidende Punkt ist tatsächlich, hier lohnt es genauso wie bei den Confoundern, gewissen Aufwand zu investieren, um nicht in ein für alle Beteiligten nicht sinnvolles Forschungsvorhaben einzugehen. Da gebe ich Ihnen völlig recht.

Gut. Angesichts der fortgeschrittenen Zeit würde ich jetzt zum nächsten Tagesordnungspunkt gehen, der ja durchaus nicht ganz irrelevant ist, nämlich

#### 2.4.4 Tagesordnungspunkt 3: Schwellenwert für die verschobene Nullhypothese

Ein bisschen mit Blick auf die Uhr: Die letzten beiden Punkte, TOP 4 und 5, sind aus unserer Sicht relativ kurz abhakbar, sodass wir uns dafür eine gewisse Zeit nehmen können, denn dieser TOP 3 ist sicherlich auch ein relevanter.

Wer möchte das von uns einführen?

**Caroline Wöhl:** Im Rapid Report A19-43 haben wir für die verschobene Nullhypothese in einer AbD einen Schwellenwert festgesetzt, der unterhalb des Wertes für den dramatischen Effekt liegen soll - das ist Ihnen ja allen bekannt -, also beispielsweise in einem Bereich von 2 bis 5 für das relative Risiko, um der Gefahr systematischer Verzerrungen aufgrund des nicht randomisierten Vergleiches Rechnung zu tragen.

In vielen Stellungnahmen wurde diese vorgeschlagene Schwelle als zu streng kritisiert. Es wurde mitunter auch vorgeschlagen, vollständig auf die verschobene Nullhypothese zu verzichten, teilweise bei gleichzeitigem Absenken der Irrtumswahrscheinlichkeit.

Jetzt muss man an der Stelle aber sagen, dass international unumstritten ist, dass selbst bei sorgfältigster Analyse, also auch unter Erfüllung sehr hoher Qualitätsanforderungen, gerade aufgrund dieser fehlenden Randomisierung immer relevante Unsicherheiten verbleiben, also nur um jetzt hier auch mal unbekannte Confounder genannt zu haben. Von daher stellt aus unserer Sicht der vollständige Verzicht auf eine verschobene Nullhypothese keine Option dar.

Der MD Bund hat in seiner Stellungnahme angeregt, über eine weniger ambitionierte Schwelle nachzudenken, und hat beispielsweise eine Schwelle von 1,5 bzw. 0,67 in den Raum gestellt. Daran anknüpfend hat sich uns die Frage gestellt, welche Alternativen Sie für konkrete Schwellen begründet nennen können, also idealerweise unter Verweis auf entsprechende Literatur.

**Moderator Thomas Kaiser:** Das ist keine ganz einfach zu lösende Aufgabe, aber wir alle stehen ja vor der Situation eines nicht einfach zu lösenden Forschungsvorhabens einer AbD. Deswegen bitte.

**Stefan Lange:** Da der MD Bund angesprochen ist, meldet sich jetzt der MD Bund. So habe ich das jedenfalls verstanden. Natürlich kann ich dafür keine Literatur nennen. Wenn ich mich recht entsinne, war im seinerzeitigen Rapid Report für die 0,5 bzw. 2 bis 5 auch nicht unbedingt eine Literaturstelle genannt. Aber das ist ja vielleicht erst mal sekundär, da kann man sich ja irgendwie auch dran halten. Das ist auch meines Erachtens gar nicht der entscheidende Punkt, sondern ich glaube ... Frau Wöhl, Sie haben gesagt, es ist internationale Einigkeit, dass relevante Unsicherheit besteht. Ich glaube, den Satz kann man nicht bestreiten. Man kann sich halt höchstens streiten über die Frage, wie hoch denn die Unsicherheit ist. Es gibt ja nun schon eine Plethora von Studien, die nun verglichen haben, was denn rauskommt, wenn ich eine Fragestellung mit einer RCT beantworte oder wenn ich eine mit einer Nicht-RCT beantworte, bis hin jetzt zu dem neuesten und möglicherweise auch interessantesten Ansatz, mit dem Target Trial Emulations, wo es auch schon entsprechende Vergleiche gibt. Da könnte man natürlich Empirie bemühen und überlegen, was denn da so im Durchschnitt an Unterschied rauskommt. Das ist vermutlich nicht so viel. Dann sagen viele, das spielt aber auch gar keine Rolle, ob was rauskommt oder ob nichts rauskommt, weil wir sehen, dass hier die Varianz größer ist. Das verstehe ich. Diese größere Varianz könnte man ja dann auch noch mal mitberücksichtigen aus solchen Nicht-RCTs.

Umgekehrt muss man natürlich trotzdem sagen, wenn man so einen wahnsinnigen Aufwand betreibt - ich übertreibe jetzt etwas -, einerseits zur Identifikation, dann aber auch zur Berücksichtigung von Confoundern und zur Generierung einer sehr, sehr hohen Datenqualität, die dann wahrscheinlich sogar höher ist als in vielen Investigator-Initiated Trials, randomisierten Trials, dann, glaube ich, sollte man überlegen, ob das dann eben nicht auch bei der Definition der Schwelle eine Berücksichtigung finden kann. Weg vom dramatischen

Effekt ist es in der Tat, aber das ist ja auch eine Schwelle, die ich weiß nicht wie oft in der Medizin bisher überschritten worden ist, ich glaube, nicht allzu oft.

Ein weiterer Gedanke, den wir gehabt haben, war, ob man nicht auch in der Planungsphase einer AbD und dann in der konkreten Ausgestaltung einer AbD Methoden nutzen kann, um den Bias innerhalb der AbD abzuschätzen. Mein Lieblingsbeispiel sind immer die Negativkontrollen, jetzt hier konkret die Negativ-Outcome-Kontrollen. Natürlich stößt das auch an Grenzen. Dann kommt die nächste Frage: Wie definiert ihr denn die Negative? Wo kommt die denn her? Wie kommt ihr denn auf die Annahmen? - Alles geschenkt, das ist klar. Je expliziter man wird - das kennen wir alle -, desto vielfältiger werden die da gestellten Fragen, und trotzdem ist es ja einfach nur eine Idee, um zu versuchen, vielleicht eine AbD noch etwas realisierbarer zu machen, weil natürlich diese Schwelle von 2 bzw. 0,5 doch schon eine starke Hürde darstellt, glaube ich, das kann man so sagen.

Vielleicht erst mal so viel.

**Moderator Thomas Kaiser:** Vielen Dank für die Einschätzung. Um deinen Hinweis zu bestätigen: In der Tat, für die 2 bis 5 gab es auch keine dezidierte Literaturstelle. Es ist eben auch schwierig, weil dann wiederum bei Empirie ja auch die Frage wäre: Worauf bezieht sich die Empirie? Sind das seltene, nicht seltene Erkrankungen? Ist es diese Erkrankung, ist es jene Erkrankung? Und Qualität ist ja auch kein 0/1-Begriff, sondern da stecken auch noch andere Dinge der Einordnung hinter. Insofern schwierig mit der Literatur, nichtsdestotrotz, es hätte ja sein können, dass da irgendeine bahnbrechende Arbeit durchgeführt worden wäre in Essen. - Dann wären wir bei Herrn Eisele.

**Lewin Eisele:** Wir hatten in unserer Stellungnahme auch diese hohe Hürde kritisiert und als mögliche Alternative eine E-Value-basierte Herangehensweise vorgeschlagen. Das wäre eine Kombination aus Signifikanztests ohne verschobene Nullhypothese und ein Kriterium für den E-Value, um sozusagen eine Hürde für residuelles Confounding einzubauen. Also, in unserer Stellungnahme haben wir da die Details ein bisschen mehr skizziert. Nur so als Anregung.

Moderator Thomas Kaiser: Vielen Dank. Wir haben uns ja sogar auch damals im Rapid Report schon oder kurz danach mit dem E-Value beschäftigt. - Guido Skipka möchte dazu was sagen.

**Guido Skipka:** Die E-Values sind uns natürlich auch nicht unbekannt. Wir haben uns das angesehen, ganz speziell ich habe mir das auch mal angesehen. Zum Beispiel für das relative Risiko ist der E-Value eine lineare Funktion vom Schätzer, zweimal RR minus ein halb ungefähr. Was ich sagen will: Es ist keine andere Information im E-Value als in dem geschätzten Effekt. Wenn man sich jetzt über Schwellen unterhält für den geschätzten Effekt, kann man das natürlich eins zu eins umrechnen in eine Schwelle für den E-Value. Also, das hilft uns in dem Sinne jetzt nicht, weil es kein anderes Konzept ist, das ist halt eine andere Interpretation, die

kann man gehen, aber der Frage, welchen Schwellenwert wir anlegen, kommt man so auch nicht näher, weil man es halt umrechnen müsste.

**Moderator Thomas Kaiser:** Vielen Dank. - Herr Eisele noch mal gerne dazu.

**Lewin Eisele:** Ohne jetzt zu tief ins Detail zu gehen, aber unser Vorschlag war, da eben quasi so einen vom Kontext der Fragestellung abhängigen Schwellenwert für den E-Value zu nehmen, der letztlich ein Stück weit auf der Confounderanalyse basiert, also sozusagen, wenn Sie wollen so, einen evidenzbasierten Referenzrahmen aus der Confounderanalyse, um dann den E-Value sozusagen damit abzugleichen. Das war so ein bisschen die Idee, wie gesagt, jetzt ohne hier ins Detail zu gehen. Wir haben das in unserer Stellungnahme ein bisschen näher skizziert, wie so etwas aussehen könnte.

**Moderator Thomas Kaiser:** Vielen Dank auch noch mal für den Hinweis und auch tatsächlich dezidiert danke dafür, dass Sie da so eine Ausarbeitung gemacht haben, weil genau darum geht es ja auch, dass wir uns solche Dinge noch mal anschauen, ob man daraus nicht - selbst wenn es nicht für das E-Value wäre - von dem Konzeptionellen her noch mal was mitnehmen kann, um sich der tatsächlich in dem konkreten Forschungsvorhaben anzulegenden Schwelle noch mal besser nähern zu können. Also, es sind zwei Sachverhalte. Das eine ist, kontextabhängig mit einer Schwelle zu arbeiten - das steckte allerdings in 2 bis 5 auch schon drin, aber dem kann man sich ja vielleicht ein bisschen besser nähern -, und das andere ist: Ist nicht schon die Ausgangsschwelle, die formuliert worden ist in dem Rapid Report und jetzt auch in dem hiesigen Rapid Report, also, ich spreche jetzt von den zweien, A19-43 und dem jetzigen, nicht eine zu ambitionierte? So ist ja die Formulierung, die genutzt wurde. Deswegen noch mal vielen Dank, Herr Eisele, dafür.

Dann wäre die Frage, ob es weitere Redebeiträge dazu gibt, denn das sehe ich derzeit nicht. - Herr Rasch.

**Andrey Rasch:** Ich kann mich nur den Vorrednern anschließen. Wir haben das ja auch nicht zum ersten Mal angemerkt. Diese Schwelle, die hier gewählt wurde, von 0,2, 0,5 oder 2 und 5, je nachdem, aus welcher Richtung, entspricht ja letztlich aufgrund der verschobenen Nullhypothese mehr oder weniger ziemlich genau aus meiner Sicht der Voraussetzung eines dramatischen Effekts, mit dem wir sonst im AMNOG zu tun haben. In Kombination aber mit der umfassenden Confounderadjustierung ist es aus unserer Sicht sogar noch sozusagen eine Anforderung in der AbD, die sogar über den Anspruch an einen dramatischen Effekt für historische nichtadjustierte Vergleiche noch hinausgeht. Also, man hat sogar sozusagen noch einen draufgesetzt.

Wie man sich dem nähert, okay, da gibt es mehrere Wege, die wurden hier aufgezeigt. Natürlich sind es dann so die üblichen Fragen - Herr Kaiser, Sie haben es ja selbst gesagt -: Ist

diese Schwelle per se vielleicht nicht zu hoch gegriffen? Muss man mit der Grenze des Konfidenzintervalls arbeiten, oder könnte man nicht zumindest die Konfidenzintervallgrenze dann etwas anheben? Da gibt es natürlich mehrere Wege. Wie empirisch ableitbar die sind, sei natürlich dahingestellt, weil, wie auch schon gesagt wurde, die 2 und 5 sind ja auch mehr oder weniger gefühlt gesetzt worden, was auch okay ist. Aber an dieser Stelle erscheint das einfach im Kontext der AbD zu hoch gegriffen, auch wie gesagt in der Konstellation mit der Confounderadjustierung und der verschobenen Nullhypothese.

**Moderator Thomas Kaiser:** Herr Rasch, vielen Dank. Ich habe nur eine Nachfrage, weil Sie ganz am Anfang Ihres Redebeitrags gesagt haben, 2 bis 5 wäre quasi schon der dramatische Effekt analog dazu, und dann würde ja noch etwas obendrauf gesetzt. Ich verstehe das, wenn man das so sagt, 2 bis 5 ist dramatischer Effekt, weil wir ja hier auch noch sagen, wir brauchen auch eine gute Datenqualität, was beim dramatischen Effekt nicht ganz irrelevant ist, aber in den Hintergrund tritt. Jetzt ganz spezifisch bezogen auf die Schwelle von 2. Damit ist nicht mein Thema jetzt gerade, die Schwelle 2 wäre gesetzt oder so, sondern es ist ja genau eine Diskussion, die wir gerade haben, ob das zu ambitioniert ist. Aber können Sie das noch mal erläutern, warum Sie die Schwelle, relatives Risiko von 2, und mir ist klar Konfidenzintervallgrenze mit dem dramatischen Effekt nach dem Verständnis von Paul Glasziou - das ist ja die Ausgangsarbeit -, in Beziehung setzen?

**Andrej Rasch:** Weil wir davon ausgehen, dass der Anspruch an den dramatischen Effekt von 5 bis 10 nicht mehr mit der Nullhypothese verknüpft wird seitens IQWiG, sondern eben sich an den Punktschätzer richtet. In dieser Konstellation, wenn man dann die 2 bis 5 nimmt, aber dann mit der verschobenen Nullhypothese arbeitet, läuft das de facto darauf hinaus, dass das 5 bis 10 für den normalen dramatischen Effekt entspricht.

**Moderator Thomas Kaiser:** Okay, das habe ich verstanden. Das heißt, die Annahme ist, dass die Präzision quasi keine große Rolle spielt bei dem dramatischen Effekt. Okay, danke. - Stefan.

**Stefan Lange:** Es war ja eben die Frage nach der Literaturstelle. Ich kann mich ziemlich ärgern, weil ich vor Urzeiten mal eine Arbeit gefunden habe, die ich leider nicht mehr identifizieren kann, aber vielleicht fällt sie hier irgendjemand anderen aus dem geschätzten Auditorium noch in die Hände oder ein, wo sehr schön gezeigt wurde, dass man sich tatsächlich überlegen kann, ob man in Abhängigkeit von einer Prävalenz, also, es geht wieder in Richtung sehr seltener Erkrankungen, sich nicht Gedanken über ein angehobenes Irrtumsniveau machen sollte, weil aus der Berücksichtigung von Fehlern erster und zweiter Art im Grunde genommen bessere Ergebnisse insgesamt resultieren, wenn man dann eben den Fehler erster Art etwas relaxiert. Das basierte natürlich im Wesentlichen auf Modellannahmen, aber ich fand das in der Zeit sehr interessant, aber ich finde es hier nicht mehr.

Natürlich, Fehler erster Art und Hypothesenschwelle hängen ja eng miteinander zusammen. Das gilt natürlich bei den Unterlegenheitsstudien umso mehr, hier gilt es natürlich aber auch bei einer Schwelle für eine Überlegenheitsstudie. Das führt mich zu den Gedanken, ohne jetzt tatsächlich diese konkrete Literaturstelle nennen zu können, ob man nicht dennoch auch in die Überlegung der zu wählenden Schwelle die Prävalenz der Erkrankung mit einfließen lässt als ein Kriterium, vielleicht auch dann in Richtung - das ist ein bisschen Wischiwaschi-multikriterielle-Definition - eines solchen Schwellenwertes. Aber das liegt mir tatsächlich persönlich sehr am Herzen, weil ich irgendwann denke, wenn wir uns denn einig sind, dass es eine Restriktion mit der realisierbaren Fallzahl gibt, und das ist bei manchen AbDs ja so - ich hoffe, dass ich da richtig liege -, wo festgestellt wird, wir werden nicht mehr als 120, 150, von mir aus 200 Menschen rekrutieren können für die AbD - das liegt eben im Wesentlichen an der Seltenheit der Erkrankung oder der spezifischen Subgruppe oder was auch immer -, dann, finde ich, könnte das auch in diese Überlegung mit einbezogen werden. Denn wenn man das weiterdenkt, irgendwann sagt man, bei 100 hören wir irgendwie auf, dann können wir gar nichts mehr machen, was eigentlich auch doof ist. Oder wenn man im Bereich der randomisierten Studien sagt, ich habe weltweit nur 50 Fälle, auch dann müsste man ja überlegen, ob ich starr an den 5 % Irrtumsniveau festhalte. Ich glaube, irgendwann stößt das an Grenzen. Also, das ist nur noch so ein Gedanke, den ich hier gerne mit einbringen würde, ob es nicht noch andere Kriterien gibt neben der Datenqualität, die sicher eine sehr, sehr große Rolle spielt, neben der Frage eines möglichen residuellen Confounding, das auch eine große Rolle spielt, eben noch diese Aspekte wie die Seltenheit oder sehr große Seltenheit der Erkrankung und gegebenenfalls noch irgendwas anderes im Zusammenhang mit Nutzen-Schaden-Abwägung. Das kann ja eventuell auch eine Rolle spielen.

**Moderator Thomas Kaiser:** Vielen Dank. Das ist ja ein Thema, das das IQWiG ja auch, glaube ich, 2014 im Rapid Report mal bearbeitet hatte: Was gibt es auch für Möglichkeiten bei Studien seltener Erkrankungen? Da war die andere Irrtumswahrscheinlichkeit ja auch ein Punkt. Ich glaube, das ist auch in diesem ASTERIX-Projekt - das ist nicht der Gallier, sondern das ist der „STERN“, glaube ich, der dem Ganzen einen Namen gegeben hat - auf EU-Ebene untersucht worden. Das gucken wir uns einfach noch mal an, also was da für Themen auch sind, weil das ist so ein Ansatz in der Studie gewesen. Okay. Vielen Dank.

Gibt es weitere Meldungen? Wenn das nicht der Fall ist, dann kommen wir zum

#### **2.4.5 Tagesordnungspunkt 4: Alternative Methoden zu Propensity-Score-Verfahren bei kleinen Fallzahlen und vielen Confoundern**

Bitte sehr.

**Guido Skipka:** Wir haben in einem Rapid Report Evidenz zusammengetragen, inwieweit Propensity-Score-basierte Verfahren bei kleinen Fallzahlen noch funktionieren oder auch

nicht funktionieren. Wir haben dazu ein paar Arbeiten gefunden. Wir haben das zusammengeschrieben. Die Quintessenz war, wenn die Fallzahl sehr klein ist und die Zahl der Confounder sehr groß, dann kann es dazu kommen, dass Schätzungen nicht mehr möglich sind. Das sind Konvergenzprobleme bei diesen iterativen Verfahren. Das ist erst mal so, wie es ist.

Es gab ein paar Stellungnahmen zu diesem Abschnitt. Zum einen wurde angemerkt, dass wir uns doch nicht so auf diese Propensity-Score-basierten Verfahren beschränken sollten. Wir sollten uns öffnen. Es gäbe auch noch andere Verfahren. Die wären es vielleicht auch wert, angeschaut zu werden und beachtet zu werden.

Das ist vielleicht ein Missverständnis, was ich vielleicht hier ausräumen kann. Wir sehen diese Propensity-Score-basierten Verfahren erst mal so als Einstieg, als regelhafte Möglichkeit, um dieser Confounderproblematik zu begegnen. Aber wir sehen das nicht ausschließlich bei diesen Verfahren. Der Eindruck entsteht vielleicht, weil das immer so unser Hauptaugenmerk ist. Aber das kann ich hiermit ausräumen. Also, wir verschließen uns anderen Verfahren nicht. Wenn man im Rahmen der AbDs, sei es wegen des Kontexts oder weil man spezielle Annahmen treffen kann, andere Verfahren plant, durchführt und gut dokumentiert, werden wir die akzeptieren und daraus auch die entsprechenden positiven Schlüsse ziehen. Also, die Sorge kann ich Ihnen vielleicht hier einfach mal nehmen.

Ganz konkret haben wir noch eine etwas ausführlichere Stellungnahme von Sanofi-Aventis bekommen, die auch diese weiterführenden Verfahren auflistet und auch mit einigen Literaturstellen unterlegt hat. Da waren zum Beispiel G-Computation-Verfahren oder auch Targeted Maximum Likelihood Estimation, Random-Forest-Verfahren. Das sind jetzt nur mal so ein paar Stichworte, Alternativen, um der Confounderproblematik zu begegnen. Wir haben uns diese Literatur angeschaut. Das waren, glaube ich, neun Artikel und ein ganzes Buch, sofern wir sie nicht schon kannten. In der Stellungnahme von Sanofi-Aventis wurde halt behauptet, dass diese Verfahren besonders gut geeignet wären, bei kleinen Fallzahlen noch zu stabilen Modellschätzungen zu kommen.

Jetzt komme ich zu unserer Frage. Wir haben uns wie gesagt diese ganzen Literaturstellen angeschaut und haben das daraus nicht ableiten können. Also, diese Zitate, die Sie uns mitgegeben haben, beschreiben halt diese alternativen Verfahren, erst mal in der Theorie, in der Mathematik. Zum Teil werden auch ein paar Simulationsergebnisse oder Beispiele gezeigt. Das sind aber dann auch Simulationen mit N gleich 30.000. Das würde ich jetzt nicht als kleine Fallzahl bezeichnen. Also, wir können das daraus erst mal nicht ableiten. Ich habe als Statistiker da auch etwas Schwierigkeiten, weil ich denke, wenn man sehr wenige Patienten und sehr viele Confounder hat, dann ist das nicht nur ein Problem von Propensity-Score-basierten Verfahren, dass das dann irgendwann nicht mehr berechenbar ist. Das liegt in der Natur der Sache. Das wird für die anderen Verfahren vermutlich ohne weitere Annahmen auch gelten. Von daher wären wir doch sehr interessiert an diesen weiteren Verfahren, ob Sie

tatsächlich Simulationsergebnisse speziell für diese Frage kennen, wo man also wirklich eine kleine Fallzahl hat. Da würde ich mal so sagen, N gleich 200 oder weniger, wo gezeigt wird, dass andere Verfahren doch deutlich stabiler sind als Propensity-Score-basierte Verfahren, das würde uns schon sehr interessieren.

**Moderator Thomas Kaiser:** Guido, vielen Dank. Das ging direkt an die Firma Sanofi-Aventis. Ich sehe allerdings Markos Dintsios, der sich im Übergang der Fragestellung gemeldet hat. Markos, geht das zu der Frage jetzt von Guido Skipka oder war das ein anderer Punkt? Weil dann würde ich nämlich erst mal die Vertreterin von Sanofi-Aventis den Vortritt geben.

**Charalabos-Markos Dintsios:** Das war zum vorigen Punkt. Es war genau im Übergang. Deswegen kannst du mich einfach übergehen. Kein Problem.

**Moderator Thomas Kaiser:** Okay, das heißt, gleich noch mal kurz zurück zu TOP 3. Das machen wir gerne. - Dann aber gerne Frau Vaitsiakhovich.

**Tatsiana Vaitsiakhovich:** Wir begrüßen erst mal ausdrücklich die Möglichkeit, unsere Perspektive einzubringen. Wie Sie schon erläutert haben, in unserer Rückmeldung haben wir insbesondere betont, dass aus unserer Sicht eine allgemeine methodische Orientierung hilfreich wäre. Das ist sehr schön zu hören, dass Sie tatsächlich wahrscheinlich jetzt breiter gehen werden und andere Methoden auch akzeptieren werden potenziell.

Zu Ihrer Frage, ob wir besondere Erfahrungen haben, was kleine Patientenzahl angeht, und auch, wie sich andere Methoden in dem Fall im Vergleich zu PS-basierten Methoden zeigen. Ich kann nur berichten, dass es tatsächlich Methoden gibt da, wo Sie unabhängig von der Anzahl von Kovariaten sehr gut mit sehr vielen Formationen umgehen können in kleineren Patientenkollektiven, da, wo es eigentlich nicht notwendig ist, um zu verstehen, ob eine Kovariate ein Confounder ist oder eine andere Art von Variablen, auf die wir nicht unbedingt adjustieren müssen. Es gibt auch tatsächlich Optimierungsverfahren, die nicht PS-Kalkulation oder modellbasierte Verfahren benötigen. Das heißt, wenn wir im Endeffekt Confounder identifiziert haben oder aufgrund von vorhandenen Daten Confounder gesammelt haben, können wir diese Methoden anwenden und berücksichtigen. Positivität ist da. Das heißt, wir können auch alle Kovariaten behalten und nicht unbedingt PS-Modell anwenden und dann sehen, eigentlich müssen wir Kovariaten oder Variablen ausschließen, so wie es im Rapid Report beschrieben ist. Man kann alle behalten und gucken, dass die Effektschätzer in dem Fall auch zu bekommen sind unter Berücksichtigung von großer Anzahl von Kovariaten und kleinen kollektiven Mengen.

Es ist natürlich immer die Frage: Was machen wir, wenn wir die Patientenzahl noch mal reduzieren müssen, so wie in PS-Methoden? In dem Fall haben sich gewichtete Methoden tatsächlich besser gezeigt, da, wo Gewichte nicht nur aufgrund von PS, Propensity Score,

berechnet werden können, aber es gibt andere optimierungsisierte Methoden. Das haben wir alles der Literaturliste beigelegt.

**Moderator Thomas Kaiser:** Vielen Dank für den Hinweis. Trotzdem noch mal eine Nachfrage von mir dazu. Herr Skipka hatte erwähnt, wir haben uns diese Literatur ausführlich angeschaut, und das, was Sie gerade gesagt haben, geht aus den Papern nicht hervor. Deswegen die Frage: Welche möglicherweise zusätzliche Literatur haben Sie, wo das, was Sie gerade beschrieben haben, mitgestützt wird, weil die Paper selber stützen das nicht?

**Tatsiana Vaitsiakhovich:** Vielen Dank für die Frage. Ich glaube, hier werden wir dazu fragen, dass wir unterschiedliche Methoden in dem Fall ausprobieren können. Es gibt keine fertige Lösung zurzeit, was zumindest mir bekannt ist. Ich möchte nur dazu betonen, dass die Propensity-Score-basierten Methoden in dem Fall sehr selten helfen werden. Wir werden mit sehr großen Problemen konfrontiert. Die anderen Methoden können hier helfen, diese Probleme umzugehen. Was im Endeffekt besser ist, das ist heutzutage mir persönlich nicht bekannt.

**Moderator Thomas Kaiser:** Vielen Dank noch mal für die Klarstellung. - Guido Skipka.

**Guido Skipka:** Noch mal: Wir verschließen uns anderen Verfahren als Propensity-Score-basierten Verfahren nicht, wenn sie denn adäquat sind. Ich habe eben bei Ihnen jetzt ein bisschen rausgehört, man versucht dann mal aus diesem Strauß von Verfahren zu gucken, was funktioniert, was nicht. Das hat natürlich so ein bisschen die Gefahr der Beliebigkeit. Also, Sie müssten sich schon bemühen, im Studienprotokoll dann zumindest so eine Art Hierarchie der Verfahren anzugeben, damit man nicht völlig in die Beliebigkeit abdriftet.

**Moderator Thomas Kaiser:** Genau. In der Publikation von Desai zu den Propensity-Score-Verfahren ist ja auch so eine allerdings auf Propensity-Score-Verfahren beschränkte Entscheidungsstruktur beschrieben worden. Das wäre dann halt die Zielsetzung, dass man das, was Guido Skipka gerade so als Ausprobieren beschrieben hat, dann eben auch als so einen systematischen, gestuften, vielleicht hierarchischen Weg beschreibt und sagt, unter den und den Bedingungen würde man dann jetzt bei dem Verfahren bleiben, weil es gut geeignet ist. Das ist ja nachvollziehbar, dass man ohne Kenntnis der Daten dann tatsächlich gegebenenfalls nicht das optimale Verfahren wählen würde, wenn man sich auf eins festlegt. Aber, ich glaube, so eine tatsächliche Entscheidungsstruktur, eine hierarchische Entscheidungsstruktur, die sollte man dann schon prospektiv abbilden. Denn sonst geht es zur - Guido hat es gesagt - Beliebigkeit, aber das hat ja auch ein bisschen was mit Data Mining oder Ergebnisfischen zu tun.

Möchten Sie darauf noch mal antworten?

**Tatsiana Vaitsiakhovich:** Ich glaube, ja. Vielen Dank noch mal, dass ich mich hier melden kann. Ich glaube, man muss die Methoden wählen und die Verfahren beschreiben da, wo die tatsächlich outcome-blinded sind. Das ist im Rahmen von Beobachtungsstudien immer der Fall. Es gibt auch unterschiedliche Prüfmethoden, wie man das tatsächlich assessen kann, um zu verstehen, ob das geeignet oder nicht geeignet ist. So wie in Propensity-basierten Methoden setzen wir auch bestimmte Schwellenwerte. Was wir sehr oft gesehen haben, wenn dieser Schwellenwert überschritten ist, wird der Confounder entzogen. Da gibt es systematische Assessments da, wo wir nicht Outcome-Daten sehen, um zu entscheiden, welche Methode am besten funktionieren würde.

**Moderator Thomas Kaiser:** Gut. Vielen Dank noch mal für die Klarstellung. - Wenn jetzt keine weiteren Anmerkungen zu diesem TOP 4 sind - das scheint nicht der Fall zu sein -, dann Markos Dintsios gerne noch mal ein Statement zu TOP 3 ergänzend.

**Charalabos-Markos Dintsios:** Sorry, das war genau beim Übergang. - Ausgehend von dem, was Stefan Lange gesagt hat, würde ich mal den Vorschlag unterbreiten wollen, dass man selbstreferenziell ein bisschen vorgeht, wie zum Beispiel aus der Stressanalyse bekannt den Alpha-Fehler mal floaten lässt auf das Doppelte, auf 10 %, und aus dem, was bereits gemacht wurde und in Zukunft noch kommen wird, sich anschauen sollte, wie dann die Beziehung ausfällt und ob das theoretisch in einem Bereich liegt, den man noch tolerieren könnte, da wir ja nicht so viel Empirie von draußen haben, was die Methode anbelangt, sozusagen sich selber auf das beziehen, was wir hier im System sehen. Dann können wir ja sehen, ob so ein Vorschlag, den Alpha-Fehler floaten zu lassen - Stressanalyse heißt ja meistens halbieren und verdoppeln; halbieren macht in diesem Kontext überhaupt keinen Sinn, verdoppeln vielleicht schon von 5 auf 10 % -, ... ein bisschen Erkenntnis gewinnen, wie sich das aus den Verfahren gestaltet, die wir haben, einfach um ein bisschen Orientierung zu gewinnen. Das als Vorschlag.

**Moderator Thomas Kaiser:** Vielen Dank für diesen unterstützenden Hinweis, gegebenenfalls eben bezogen auch auf die Ausgangssituation der Fallzahl, das in der Irrtumswahrscheinlichkeit zu berücksichtigen. Das werden wir auf jeden Fall noch mal mitnehmen.

Gut. Dann wären wir schon beim letzten Tagesordnungspunkt, nämlich

#### **2.4.6 Tagesordnungspunkt 5: Hürden bei Nutzung materieller Anreize als Instrument zur Sicherstellung der Vollständigkeit der Daten**

**Caroline Wöhl:** Wir hatten im Rapid Report unter anderem diese materiellen Anreize als Möglichkeit genannt, um sowohl die teilnehmenden Zentren, aber auch die teilnehmenden Patientinnen und Patienten zur Datenerhebung zu motivieren und darüber dann den Anteil fehlender Werte möglichst gering zu halten. Das war ein Aspekt unter vielen, die wir im Rapid Report aufgelistet haben. Solche Anreize werden ja auch in Registern und auch in nicht interventionellen Studien bereits zielführend eingesetzt. Also, wir hatten auch ein paar

Referenzen zitiert. Um in diesem Rahmen zwei Beispiele zu nennen: Im RABBIT-Register wird gemäß Registergutachten den beteiligten Zentren eine finanzielle Aufwandsentschädigung, jeweils in Abhängigkeit der Qualität der dokumentierten Datensätze, zur Verfügung gestellt. Und dann ist uns noch die nicht interventionelle ActiMON-Studie bekannt. „ActiMON“ steht hier für „Aktivitätsmonitoring bei Jugendlichen und jungen Erwachsenen mit Rheuma“. Dort haben die Teilnehmenden Sachprämien unter anderem in Form von Amazon-Gutscheinen für ihre Teilnahme erhalten. Die Teilnahme umfasst auch das Ausfüllen von PRO-Fragebogen. In einzelnen Stellungnahmen, unter anderem vom vfa, wurde jetzt aber der Einsatz materieller Anreize kritisch hinterfragt. Aus unserer Sicht sollte der Anspruch in einer AbD analog zum Aufsetzen klinischer Studien eigentlich immer darauf ausgerichtet sein, darüber qualitativ hochwertige prospektive Daten zu erhalten. Daher die Frage an Sie, welche Vorschläge Sie haben, jenseits materieller Anreize für eine hohe Datenqualität zu sorgen, also für möglichst vollständige Datensätze?

**Moderator Thomas Kaiser:** Das war die Frage. - Herr Bussilliat.

**Paul Bussilliat:** Erst mal, glaube ich, haben wir hier in der Runde einen Konsens, dass die Vollständigkeit der Daten sehr wichtig ist. Unsere Anmerkung in der Stellungnahme sollte jetzt auch nicht falsch interpretiert werden, dass wir uns dem komplett verwehren. Also, es kann sicherlich in einzelnen Fällen - Sie haben die Beispiele ja auch genannt - sinnvoll sein, materielle Anreize in Form von beispielsweise Gutscheinen zu setzen. Wir wollten einfach nur die Frage an der Stelle aufwerfen, ob man hier nicht auch wie gesagt über Alternativen nachdenken kann, um hier die Compliance beim Ausfüllen der Fragebögen sicherzustellen, oder ob das überhaupt ein Problem ist, das sich jetzt schon abzeichnet. Ich glaube, gerade im Kontext der AbD sprechen wir doch über ganz besondere Therapiesituationen. Da stellt sich die Frage, ob hier wirklich die Probleme auftreten, dass die Patientinnen und Patienten diese Fragebögen eben nicht ausfüllen.

Das wollten wir an der Stelle einfach noch mal einwerfen, sind aber der Meinung insgesamt, dass das von Einzelfall zu Einzelfall geprüft werden kann und da eine generelle Empfehlung vielleicht an der Stelle gar nicht notwendig ist.

**Moderator Thomas Kaiser:** Vielen Dank. Nach Ihrem Beitrag ein Input von meiner Seite - das kennt man auch ganz gut aus der Forschung -: Je mehr der Sinn von denjenigen gesehen wird für die Erhebung, bei denen erhoben wird, umso eher sind sie bereit, diese Erhebung zu machen, also beispielsweise auch, dass tatsächlich mit den Informationen umgegangen wird, also nicht einfach nur irgendeine Datenerhebung, die keine Bedeutung hat, aber wenn für die Menschen natürlich erhoben wird, wie es ihnen geht, wie es ihnen auch im Verlauf geht, und das wird auch betrachtet, dann gibt es natürlich auch eine große Bereitschaft, tatsächlich diese Dinge zu erheben. Manchmal gibt es so ein bisschen den Punkt - ich weiß nicht genau, ob ich das jetzt richtig bei Ihnen rausgehört habe, aber es könnte sein .... Ich habe es jedenfalls

in manchen Anhörungen von manchen insbesondere ärztlichen Vertretern gehört, das sei ja alles überkandidelt und zu schwierig auch in schwierigen Situationen für die Menschen. Da zeigt die Forschung etwas anderes, aber eben nur dann, wenn man auch diese Ergebnisse, die man erhebt, ernst nimmt und mit den Menschen dann bespricht und Konsequenzen daraus zieht.

Jetzt habe ich Frau Regierer, die ja auch mindestens indirekt angesprochen wurde, nämlich mit einer umfangreichen Erfahrung in diesem Zusammenhang.

**Anne Regierer:** Am DRFZ führen wir ja die genannte RABBIT-Studie durch, und auch das ActiMON-Projekt war im DRFZ. Ich wollte nur noch einmal bestätigen, dass aus unserer Sicht die Aufwandsentschädigung für die Einrichtungen, also sprich für die Ärztinnen und Ärzte und Dokumentare, extrem relevant ist. Bei zunehmendem Fachkräftemangel ist es aus unserer Sicht extrem relevant, dass die Leistungen, die die Kollegen und Kolleginnen machen, angemessen sozusagen entschädigt werden. Unsere Erfahrung mit sozusagen Gutscheinen für die Patientinnen und Patienten ist sehr positiv. Von daher ist es aus unserer Sicht sehr wichtig, dass man das irgendwie gewährleisten kann, um die Motivation zu steigern.

**Moderator Thomas Kaiser:** Vielen Dank auch noch mal für den Hinweis aus der Praxis. Ohne jetzt schon das abschließende Wort zu sprechen, weil wir haben ja noch drei Sekunden: Im Grunde genommen ist das auch noch mal ein Hinweis darauf, dass man auf verschiedenen Ebenen, an verschiedenen Stellen etwas dazu beitragen sollte, dass das Forschungsvorhaben, wenn man es denn startet, auch sinnvoll durchgeführt wird und dann auch tatsächlich aussagekräftige Ergebnisse liefert. Es ist wirklich nichts ineffizienter, als ein inadäquates Forschungsvorhaben zu starten mit viel Aufwand, und am Ende kann man wenig daraus sagen. Deswegen, glaube ich, muss man alle diese verschiedenen Aspekte sowohl in der Vorbereitung - das war ja unser erster Tagesordnungspunkt - und dann auch in der Durchführung beachten. Es lohnt sich manchmal mehr Aufwand, sei es personeller, sei es finanzieller, zu betreiben, um ein gutes Ergebnis zu bekommen, denn mit zu wenig Aufwand dann gar nichts zu haben, das ist die allerschlechteste Lösung aus meiner Sicht.

Jetzt habe ich doch schon ein halbes Schlusswort gesprochen, will das aber nicht als Schlusswort stehen lassen, möchte Ihnen noch die Gelegenheit geben, weil es jetzt zum TOP 5 keine weitere Meldung gibt, noch zu

#### **2.4.7 Tagesordnungspunkt 6: Verschiedenes**

etwas anzubringen.

Wie gesagt, Ihre Stellungnahmen haben wir gelesen, aber wenn es noch etwas Dringendes gibt, dann gerne. - Das muss aber auch nicht sein, denn wir sind jetzt schon ein wenig über der Zeit, sind aber fast mit einer Punktlandung fertig geworden.

Ich bedanke mich ganz, ganz herzlich bei Ihnen allen, nicht nur für die heutige sehr konstruktive Erörterung, die uns noch mal viele wichtige Informationen für die Überarbeitung des Rapid Reports gegeben hat, sondern auch für die Stellungnahmen, die Sie geschrieben haben, auch für die verschiedenen Vorschläge, die wir uns anschauen werden. Wenn unser gemeinsames Interesse ist, in einer defizitären Evidenzsituation möglichst viel dazu beizutragen, dass das verändert wird, dann, glaube ich, erreichen wir einiges. Insofern noch mal vielen Dank dafür.

Ich wünsche Ihnen einen schönen Tag und bis bald.

## Anhang A Dokumentation der Stellungnahmen

# Inhaltsverzeichnis

	Seite
<b>A.1 Stellungnahmen von Organisationen, Institutionen und Firmen.....</b>	<b>A 2</b>
A.1.1 AbbVie Deutschland GmbH & Co. KG.....	A 2
A.1.2 AMS Advanced Medical Services GmbH .....	A 12
A.1.3 Bayer Vital GmbH .....	A 18
A.1.4 Bristol-Myers Squibb GmbH & Co. KGaA (BMS).....	A 22
A.1.5 Bundesverband der Pharmazeutischen Industrie e. V. (BPI) .....	A 30
A.1.6 CSL Behring GmbH .....	A 36
A.1.7 Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM) .	A 42
A.1.8 Deutsche Gesellschaft für Innere Medizin e. V. (DGIM).....	A 44
A.1.9 Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ) und Gesellschaft für Neuropädiatrie e. V. (GNP).....	A 47
A.1.10 Deutsche Gesellschaft für Nephrologie e. V. (DGfN).....	A 51
A.1.11 Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU) .....	A 54
A.1.12 Deutsche Krankenhausgesellschaft e. V. (DKG).....	A 59
A.1.13 Ecker + Ecker GmbH .....	A 67
A.1.14 IGES Institut GmbH .....	A 72
A.1.15 iOMEDICO AG .....	A 78
A.1.16 Janssen-Cilag GmbH.....	A 81
A.1.17 Lilly Deutschland GmbH .....	A 92
A.1.18 Medizinischer Dienst Bund (KÖR) .....	A 97
A.1.19 Merck Healthcare Germany GmbH.....	A 107
A.1.20 MSD Sharp & Dohme GmbH.....	A 111
A.1.21 Pfizer Pharma GmbH.....	A 117
A.1.22 Roche Pharma AG .....	A 126
A.1.23 Sanofi-Aventis Deutschland GmbH.....	A 132
A.1.24 Servier Forschung und Pharma-Entwicklung GmbH .....	A 141
A.1.25 SmartStep Consulting GmbH .....	A 146
A.1.26 Verband Forschender Arzneimittelhersteller e. V. (vfa) .....	A 153
A.1.27 Vertex Pharmaceuticals (Germany) GmbH.....	A 164

## **A.1 Stellungnahmen von Organisationen, Institutionen und Firmen**

### **A.1.1 AbbVie Deutschland GmbH & Co. KG**

#### **Autorinnen und Autoren**

- Gothe, Henrike
- Mentrup, Sabine
- Musiolik, Katharina
- Sternberg, Kati
- Thaa, Bastian

## **Stellungnahme zum vorläufigen Rapid Report**

**Berichtnr: A25-13**

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Sternberg, Kati; Dr.
Musiolik, Katharina
Gothe, Henrike; Dr.
Thaa, Bastian; Dr.
Mentrup, Sabine
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b> <input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: AbbVie Deutschland GmbH & Co. KG <input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.*

Anmerkung: In dieser Stellungnahme wird aus Gründen der Lesbarkeit das generische Maskulinum verwendet. Dies bezieht sich jedoch inhaltlich gleichermaßen auf alle Geschlechter.

### **5.3 Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen [S. 12ff]**

#### Anmerkung

*„Die Fallzahlplanung für eine RCT basiert in der Regel auf Informationen aus Pilotstudien oder vergleichenden Phase-II-Studien, die eine relativ sichere Bestimmung der notwendigen Fallzahl erlauben. Ein Aufgreifkriterium für die Forderung einer AbD ist jedoch gerade, dass keine vergleichenden Daten im zu bewertenden Anwendungsgebiet vorliegen. Daher verfolgt die Konzepterstellung vielmehr das Ziel, eine Einschätzung zur grundsätzlichen Machbarkeit einer AbD zu geben.“ [S.13]*

Es ist anzumerken, dass auch im Falle der Fallzahlplanung für eine RCT in der Praxis nicht immer Daten vorliegen, die exakt den Bedingungen der geplanten RCT entsprechen. In solchen Fällen wird die bestmögliche vorliegende Evidenz genutzt, um eine methodisch möglichst robuste, aber ethisch vertretbare Fallzahl zu berechnen. Dabei wird in Kauf genommen und entsprechend als Unsicherheit berücksichtigt, dass es zwischen den zugrunde gelegten Studiendaten (z. B. Phase II Studien) und der zu planenden Studie Abweichungen (z. B. bezüglich der Populations-Charakteristika) geben kann. Dies kann sowohl die im klinischen Entwicklungsprogramm bislang generierten Daten zur Intervention als auch den geplanten Komparator betreffen, zu dem in der Regel nur veröffentlichte Studiendaten bekannt sind.

Auch für die Intervention können und werden regelhaft neben Informationen aus bereits vorliegenden Studien, weitere Informationen aus dem Anwendungsgebiet herangezogen. Daraus ergibt sich eine Fallzahlabschätzung auf möglichst umfassenden Informationen, die im Zuge einer Interimsanalyse angepasst werden kann. Auch die AbD könnte von diesem pragmatischeren Ansatz profitieren. Zwar sind die Fallzahlen mit erhöhter Unsicherheit behaftet, können jedoch einen ersten groben Aufschluss über die grundsätzliche Machbarkeit vor Beginn der AbD liefern. Diese erste Fallzahlabschätzung könnte in den Interimsanalysen nach Beginn der AbD weiter bestätigt bzw. konkretisiert werden.

### Vorgeschlagene Änderung

Die grundsätzliche Machbarkeit einer AbD sollte immer vorab mittels einer Fallzahlplanung geprüft werden. Das Konzept der orientierenden Betrachtung von Fallzahlszenarien ist dazu nicht geeignet, da dieses Konzept keinerlei Aufschluss über die eigentlich benötigte Fallzahl zur Detektion eines für die Nutzenbewertung relevanten Effektes gibt, und sollte daher gestrichen werden. Zur Fallzahlplanung sollte ein pragmatischer Ansatz verfolgt und analog zu klinischen Studien die beste verfügbare Evidenz genutzt werden. Die eventuell bestehenden Unsicherheiten innerhalb dieser Schätzung können und sollten dann im entsprechenden Kontext (z. B. der Indikation) diskutiert werden.

### Anmerkung

*„Ein Aufgreifkriterium für eine AbD ist eine unzureichende Evidenzlage für ein neues Arzneimittel im Vergleich zum therapeutischen Standard. Mit der AbD werden demnach erstmals vergleichende Daten gewonnen. Vor diesem Hintergrund kann es für die Situation einer AbD sinnvoll sein, eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlschätzung auf dieses Ziel auszurichten. Stattdessen ist denkbar, die Schätzung der notwendigen Fallzahl darauf zu beziehen, dass aus einem gewählten Propensity-Score-Verfahren unter Sicherstellung einer adäquaten Confounderadjustierung verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95 %-Konfidenzintervall resultieren (siehe hierzu auch Abschnitt 5.7), und zwar losgelöst davon, ob sich unter Berücksichtigung der verschobenen Nullhypothese ein Vorteil des zu bewertenden Arzneimittels ableiten lässt. Unter dieser Voraussetzung stellt jedes Ergebnis einer AbD einen Erkenntnisgewinn dar.“ [S. 14]*

Es erscheint fraglich, ob bei der Fallzahlplanung für eine AbD der Fokus auf die Gewinnung von Effektschätzungen ohne den Nachweis eines Vorteils des zu bewertenden Arzneimittels tatsächlich einen Erkenntnisgewinn darstellt.

Wenn die Fallzahlplanung einer AbD nicht mehr darauf ausgerichtet wäre, einen Vorteil des Medikaments zu zeigen, wäre innerhalb der AbD ein nicht signifikanter Überlegenheitstest zu erwarten. Es sei daran erinnert, dass ein nicht signifikanter Überlegenheitstest (oder ein erst gar nicht geplanter Test) keine Schlussfolgerungen zulässt. Ein nicht signifikanter Vergleich von Therapieoptionen erlaubt weder die Schlussfolgerung, dass eine Therapieoption überlegen ist, noch dass sie es nicht ist. Insbesondere kann aus einem solchen Ergebnis auch keine Aussage über die Vergleichbarkeit der Therapieoptionen getroffen werden. Dies wirft die Frage auf, welchen Erkenntnisgewinn Effektschätzungen bieten, wenn diese statistisch nicht gerechtfertigt sind und keine Aussagekraft haben. Möglich wäre die Festlegung einer alternativen verschobenen Hypothese, die nicht auf einen dramatischen Effekt abzielt und entsprechend auch für die Fallzahlplanung verwendet wird. Jedoch ist in diesem Szenario unklar, wie diese Hypothese aussehen soll. Zudem ergibt sich im Rahmen der orientierenden Fallzahlszenarien die Gefahr eines ergebnisgetriebenen Vorgehens, da in dieser Situation eben genau mögliche aufdeckbare Effekte berechnet werden.

Auch eine Planung als Nicht-Unterlegenheitsstudie ist in der klinischen Forschung grundsätzlich möglich, bedarf aber ebenso einer Fallzahlabschätzung wie eine Überlegenheits-Studie und benötigt nicht weniger Patienten. Dies würde die Umsetzung innerhalb einer AbD, in deren Anwendungsgebieten regelhaft wenig Patienten zur Verfügung stehen, zusätzlich erschweren. Das Design einer Nicht-Unterlegenheitsstudie kommt für eine AbD aber auch deshalb nicht infrage, da das Ziel einer AbD die Quantifizierung des Zusatznutzens (sprich der potenziellen Überlegenheit einer Therapie) ist.

Zudem existieren oft bereits Effektgrößen aus einarmigen Studien oder Effektschätzer aus vergleichenden Studien, z.B. wenn der Komparator nicht Teil der ZVT ist. Diese Effektschätzer bieten ein höheres Maß an Zuverlässigkeit, da Vergleichsstudien eine klassische Fallzahlplanung durchlaufen und entsprechend eine ausreichende Anzahl an Patienten einschließen müssen. Auch bei der Fallzahlplanung einarmiger Studien besteht das Ziel, ausreichend Teilnehmer einzuschließen, um verlässliche Aussagen über Effektgrößen zu treffen und seltene Nebenwirkungen zu identifizieren. Es ist daher fraglich, inwieweit ein zusätzlicher Effektschätzer, der keinen Vergleich im Sinne der evidenzbasierten Nutzenbewertung mit ZVT-Therapieoptionen zulässt und im Gegensatz zu klinischen Studien keine konkrete Fallzahlplanung erfordert, überhaupt einen Erkenntnisgewinn liefert und im Verhältnis zum Aufwand angemessen ist.

Das Ziel einer AbD ist es, Erkenntnisse zu sammeln, die im Rahmen von Nutzenbewertungen nach § 35a SGB V genutzt werden können. Angesichts der bisher stark evidenzbasierten Methodik des IQWIG ist es überraschend, dass solche Effektschätzer, die keine Aussage bzgl. Signifikanz (also Überlegenheit) liefern, überhaupt Teil einer Nutzenbewertung werden könnten. Die bisherigen Methoden des IQWIG bieten hierfür keinen Ansatz. Zudem liegen in

diesen Anwendungsbereichen, wie bereits erwähnt, oft Daten aus einarmigen Studien oder vergleichenden Studien, deren Komparator nicht Teil der ZVT ist, vor. Diese Studien, mit einem höheren Evidenzgrad als eine AbD, könnten theoretisch und sollten vorrangig für solche Bewertungen verwendet werden. Dazu sollten methodische Grundlagen (außerhalb der kaum akzeptierten indirekten Vergleiche und dramatischer Effekte) geschaffen werden, bevor Daten mit geringerem Evidenzgrad in die Nutzenbewertung eingehen.

Ob eine AbD das Ziel, über die Nutzenbewertung hinaus Therapieempfehlungen abzuleiten, erfüllen kann, ist ebenfalls zu bezweifeln. Eine reine Identifizierung von Effektschätzern ohne Möglichkeit eines statistisch sinnvollen Vergleichs bietet evidenzbasiert wenig Erkenntnis. Daraus resultierende Schlussfolgerungen zum Vorteil oder zur Gleichwertigkeit von Therapieoptionen als Unterstützung bei medizinischen Therapieentscheidungen sind, wie oben dargelegt, unzulässig. Schlimmer noch, es besteht die Gefahr, dass die Ergebnisse zu Vergleichen führen, die die Integrität der evidenzbasierten Ableitung gefährden. Insofern stellt sich sogar die Frage, ob hier indirekt der nicht evidenzbasierten Abwägung von Therapieoptionen Vorschub geleistet wird.

Ebenso fraglich ist die ethische Komponente für die Patienten, die bei der Fallzahlplanung klinischer Studien eine Rolle spielt und auch in der AbD berücksichtigt werden muss. Es ist keineswegs so, dass innerhalb einer AbD nur Routinedaten erhoben werden. Im Gegenteil, die höheren Anforderungen an eine AbD (z.B. im Hinblick auf Patient Reported Outcomes) zur Generierung nutzbarer Daten für die Nutzenbewertung erfordern ein größeres Maß an Mitarbeit der Patienten, die im Allgemeinen schwer krank sind und sich in einer Situation mit einhergehender Unsicherheit und psychischem Stress befinden. Von diesen Patienten in dieser schwierigen Therapiesituation wird ein erhöhter Aufwand verlangt, ohne dass jedoch, wie in klinischen Studien, vorab zumindest ein quantifizierbarer Vorteil für die Allgemeinheit (ausgedrückt durch ein Fallzahlplanung und der Abschätzung zur Realisierbarkeit der Fallzahl) absehbar ist.

### Vorgeschlagene Änderung

Insgesamt muss die Fallzahlplanung einer AbD auf den Nachweis eines Vorteils einer Therapieoption ausgerichtet sein. Der Abschnitt zur Fallzahlplanung auf Basis eines (nicht vorhandenen) Evidenzgewinns unabhängig von einem Therapieoptionenvorteil sollte gestrichen werden, und stattdessen die Bedeutung der Ausrichtung der Fallzahlplanung auf einen statistischen Test betont werden.

## Anmerkung

*„Um in jenen Fällen, in denen eine orientierende Fallzahlschätzung nicht sinnvoll möglich ist, dennoch Aussagen zum Umfang der Datenerhebung treffen zu können, erfolgt eine orientierende Betrachtung von Fallzahlszenarien. Diese folgt im Wesentlichen dem Prinzip der orientierenden Fallzahlschätzung.“ [S.20]*

Ausgehend von der Prämisse, dass die Diskussion zur Machbarkeit einer AbD wichtig, notwendig und in der Verfahrensordnung des G-BA (5. Kapitel, § 54, „Beurteilung der Erforderlichkeit der Forderung einer anwendungsbegleitenden Datenerhebung nach § 35a Absatz 3b Satz 1 SGB V“) vorgesehen ist, und diese Machbarkeit von der benötigten Fallzahl abhängt, ist das Konzept der „orientierenden Betrachtung von Fallzahlszenarien“ fragwürdig. Die durchgeführten Berechnungen sind keineswegs eine orientierende Betrachtung der Fallzahlszenarien, sondern vielmehr generische Berechnungen aufdeckbarer Effekte, die vom Kontext der konkreten AbD und der Indikation losgelöst sind. Bei dieser Darstellung des IQWiG fehlt die tatsächliche Einordnung der Ergebnisse in den Kontext der spezifischen Indikation. Solange die „aufdeckbaren Effekte“ nicht in Beziehung zu den erwartbaren Effekten aus der besten verfügbaren Evidenz gesetzt werden, sind sie nicht interpretierbar. Dadurch bietet die Herangehensweise keine Informationen darüber, wie viele Patienten für eine AbD benötigt werden — ein entscheidender Faktor, um zu bewerten, ob eine AbD in einem angemessenen Zeitraum, der Erkenntnisse für die Bewertung des Zusatznutzens zulässt, realisierbar ist. Auch hier besteht die Befürchtung, dass sich das IQWiG von dem evidenzbasierten Ansatz innerhalb einer AbD loslöst und stattdessen einem nicht vorhandenen Erkenntnisgewinn außerhalb des statistischen Testens Vorschub leistet.

Die orientierende Betrachtung von Fallzahlszenarien liefert schlussendlich keine Aussage über die Durchführbarkeit einer AbD, da die betrachteten Szenarien theoretischer Natur sind und durch die getätigten Annahmen noch mehr durch Unsicherheiten belastet sind als eine Fallzahlplanung basierend auf Studiendaten. Es werden Annahmen getroffen; diese können im Rahmen der AbD aber nur schwerlich bis gar nicht diskutiert werden, dazu gehört das Verhältnis zwischen Interventions- und Kontrollarm und insbesondere, ob ausreichend Patienten in Registern rekrutierbar und somit für eine AbD verfügbar sind. Die Frage, ob die verfügbare Zahl an Patienten geeignet ist, Daten für die Nutzenbewertung zu generieren, wird unter diesen zusätzlichen methodischen Unsicherheiten gerade nicht beantwortet. Zudem ergibt sich die Frage, wie die Eignung der Annahmen und sich ergebenden Szenarien bewertet werden soll. Im Rapid Report werden keinerlei Kriterien zur Beurteilung der orientierenden Fallzahlszenarien genannt, sodass die Entscheidung der Machbarkeit der AbD im Allgemeinen auf dem potenziell bestdurchführbaren Szenario beruht. Ob dieses auch in einem angemessenen Zeitrahmen realisierbar ist und eine AbD in der Realität gelingen kann, wird gerade nicht diskutiert.

### Vorgeschlagene Änderung

Eine orientierende Betrachtung von Fallzahlszenarien, die ohne die Einordnung der Ergebnisse (aufdeckbarer Effekt versus realistische, in einer AbD zeigbare Effekte) und ohne Diskussion der zur Abschätzung genutzten Parameter (z.B. Population im Anwendungsgebiet versus realistisch in die AbD rekrutierbare Patienten) erfolgt, kann keinerlei Aussagen zum notwendigen Umfang der Datenerhebung treffen und ist daher nicht sinnvoll. Eine zweckmäßige Einschätzung kann nur durch eine Fallzahlberechnung erzielt werden, die zwingend vor Beginn einer AbD notwendig ist. Das Konzept der orientierenden Betrachtung von Fallzahlszenarien bietet keinen Erkenntnisgewinn für die Diskussion zur Machbarkeit einer AbD und sollte daher vollständig gestrichen werden.

### Anmerkung

*„Die Informationen zu den Patientenzahlen liegen aufgrund der Unsicherheit der Abschätzung üblicherweise in Form einer Spanne vor. Diese Spanne wird herangezogen, um ein Spektrum unterschiedlicher Patientenzahlen festzulegen, die in die weiteren Berechnungen eingehen. Wie viele Patientinnen und Patienten im zu bewertenden Anwendungsgebiet tatsächlich für eine AbD zur Verfügung stehen, ist dabei in der Regel nicht abschätzbar, da dies von unterschiedlichen Faktoren beeinflusst wird (u. a. von der Teilnahmebereitschaft der Zentren und der Patientinnen und Patienten).“ [S.20]*

Vollkommen korrekt weist das IQWiG darauf hin, dass die Anzahl der Patienten, die tatsächlich in eine AbD aufgenommen werden können, schwer abschätzbar ist und nicht der Anzahl der Patienten im Anwendungsgebiet entspricht. Insofern ist es fraglich, warum das Konzept der orientierenden Betrachtung von Fallzahlszenarien dennoch die im Rahmen der Nutzenbewertung bestimmte Patientenzahl im Anwendungsgebiet nutzt und welche Schlussfolgerungen aus dieser orientierenden Betrachtung der Fallzahlen überhaupt möglich sind. Dies gilt insbesondere, wenn in vielen Fällen von vornherein klar ist, dass die theoretische Patientenzahl im Anwendungsgebiet keinesfalls der Patientenzahl entspricht, die tatsächlich in angemessener Zeit in eine AbD rekrutierbar sind. Es wird somit implizit davon ausgegangen, dass diese Patientenzahl ausreicht, um in einer AbD geeignete Evidenz für die Nutzenbewertung zu generieren, was keinesfalls mit Sicherheit gegeben ist.

### Vorgeschlagene Änderung

Auch wenn sich Ungewissheiten nicht vollständig auflösen lassen, muss zur Abwägung der Machbarkeit einer AbD realistisch abgeschätzt werden, wie viele Patienten aus dem Anwendungsgebiet in angemessener Zeit im Sinne der Nutzenbewertung in die AbD

eingeschlossen werden können. Jegliche orientierende Abschätzung in Form von Fallzahlszenarien auf Basis der theoretisch verfügbaren Patientenzahl im Anwendungsgebiet bietet keinen Erkenntnisgewinn und sollte daher gestrichen werden.

### Anmerkung

*„Unabhängig vom gewählten methodischen Vorgehen für die Konzepterstellung (orientierende Fallzahlschätzung oder orientierende Betrachtung von Fallzahlszenarien) sollte der pU des zu bewertenden Arzneimittels die (finale) Fallzahlplanung der AbD zum Zeitpunkt der Zwischenanalysen auf Basis der dann vorhandenen vorläufigen Ergebnisse vornehmen [56]. Dieses Vorgehen wird als grundsätzlich geeignet beurteilt, da die Schätzung der notwendigen Fallzahl bei versorgungsnahen Datenerhebungen – in Abgrenzung zu klinischen Studien – nicht mit Hinblick auf eine aktive Rekrutierung erfolgt. Stattdessen werden möglichst alle Patientinnen und Patienten einer Registerpopulation, die die Einschlusskriterien für die AbD erfüllen, nach ihrer Einwilligung in die AbD eingeschlossen. Dies erfolgt auch vor dem Hintergrund, dass, je nach gewähltem Propensity-Score-Verfahren, ggf. ein relevanter Anteil der Patientinnen und Patienten nicht in der Auswertung berücksichtigt werden kann. Wie viele Patientinnen und Patienten nicht in die Analysepopulation eingehen, ist dabei im Vorfeld nicht abschätzbar.“ [S. 24f]*

Der Vorschlag, Interimsanalysen zur Fallzahlplanung zu verwenden, ist begrüßenswert und bei der Durchführung einer AbD auch bereits regelhaft vorgesehen. Allerdings ist es auch im Falle geplanter Interimsanalysen sinnvoll, vor Beginn der AbD mit Hilfe der verfügbaren Evidenz zu prüfen, ob eine mögliche Fallzahl zumindest grob innerhalb einer AbD realisierbar ist. Erst wenn dies der Fall ist, sollte die AbD durchgeführt und diese Fallzahl mit Hilfe einer Interimsanalyse besser eingegrenzt bzw. angepasst werden. Dazu ist jedoch zumindest eine Fallzahlplanung vor Beginn der AbD notwendig; eine orientierende Betrachtung von Fallzahlszenarien kann dies nicht leisten. Selbst eine konkrete Fallzahlabschätzung vor Start einer AbD gewährleistet nicht, dass im Rahmen einer Interimsanalyse eine Fallzahlplanung durchgeführt werden kann: Weder in der laufenden AbD zu Onasemnogen-Abeparvovec (Zolgensma®, spinale Muskelatrophie) noch zu Brexucabtagen autoleucel (Tecartus®, Mantelzelllymphom) waren die bisher vorgelegten Zwischenanalysen „reif“ genug dafür, obwohl das IQWiG im Vorfeld der Forderung einer AbD im AbD-Konzept eine Fallzahlabschätzung durchgeführt hatte. Eine Interimsanalyse kann also lediglich eine Möglichkeit sein, die Fallzahlplanung unter den entsprechenden Umständen zu konkretisieren, es ist jedoch nicht sicher zu sagen, dass dies mit einem sinnstiftenden Ergebnis durchgeführt werden kann. Dies gilt umso mehr, wenn vor Beginn der AbD lediglich „orientierende Fallzahlszenarien“ vorliegen. Eine (finale) Fallzahlplanung auf Basis der

Zwischenanalysen wird in Situationen, in den nur orientierende Fallzahlszenarien vor Beginn der AbD betrachtet wurden, nur selten oder gar nicht möglich sein.

#### Vorgeschlagene Änderung

Interimsanalysen zur Aktualisierung der Fallzahlplanung sind wichtig. Sie ersetzen jedoch nicht eine Fallzahlplanung vor Beginn der AbD, die abschätzen muss, ob eine AbD überhaupt durchführbar ist. Das Konzept der orientierenden Betrachtung von Fallzahlszenarien ist dazu nicht geeignet und sollte gestrichen werden.

#### **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

### A.1.2 AMS Advanced Medical Services GmbH

#### **Autorinnen und Autoren**

- Bogner, Kathrin
- Ebentheuer, Lisa-Marie
- Reiter, Bennet
- Richter, Swantje
- Wüstner, Stefanie

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Ebentheuer, Lisa-Marie
Reiter, Bennet; Dr. rer. nat.
Wüstner, Stefanie; Dr. rer. nat.
Bogner, Kathrin; Dr. rer. soc.
Richter, Swantje
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b> <input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: AMS Advanced Medical Services GmbH <input type="checkbox"/> als Privatperson(en)

***Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.***

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

Wie das IQWiG in seinem Rapid Report feststellt, ist die derzeitige Vorgehensweise zur adäquaten Identifikation von Confoundern in Anlehnung an Pufulete et al., mit erheblichem Aufwand verbunden. Das Institut beabsichtigt daher, in einem nachgelagerten Schritt zu prüfen, ob Anpassungen zur Reduzierung des Aufwands möglich sind. In dem Arbeitspapier GA23-02 des IQWiG zur systematischen Confounderidentifikation schätzen die externen Sachverständigen die systematische Literaturrecherche nach RCT und Kohortenstudien als zeitaufwendigsten Arbeitsschritt ein, der mehr als die Hälfte der erforderlichen Personentage in Anspruch nimmt [1].

Auch aus Sicht von AMS besteht hier das größte Potenzial für eine sinnvolle Aufwandsreduzierung. Ein effizienter Ansatz wäre, die Recherche vorrangig auf Sekundärliteratur wie systematische Reviews und Meta-Analysen auszurichten, anstatt auf Primärliteratur. Diese Literaturquellen enthalten bereits aggregierte Informationen bezüglich potenzieller Confounder und decken ein breites Spektrum an Evidenz ab (über RCT und Kohortenstudien hinaus). Dies ist insbesondere für die Confounderidentifikation im Rahmen der AbD relevant, da für die betreffenden Indikationen aufgrund niedriger Patientenzahlen meist kaum RCT und Kohortenstudien verfügbar sind. Zwar sieht das IQWiG in diesem Fall eine Ausweitung der Recherche auf einarmige Studien vor, dies würde den Aufwand jedoch nochmals deutlich erhöhen. Zusätzlich kann bei einer Recherche nach Sekundärliteratur der Aufwand durch eine zeitliche Einschränkung sinnvoll eingegrenzt werden, da neuere Sekundärliteratur auch Evidenz älterer Primärliteratur beinhaltet. Lediglich für den Zeitraum ab der Veröffentlichung der aktuellsten Sekundärliteratur müsste zusätzlich nach Primärliteratur gesucht werden um eine Berücksichtigung neuester Erkenntnisse zu gewährleisten.

Fazit:

Aus Sicht von AMS stellt eine (ggf. zeitlich eingeschränkte) systematische Recherche nach Sekundärliteratur in Kombination mit einer ergänzenden Recherche nach neuesten RCT eine sinnvolle Reduzierung des Aufwands ohne nennenswerten Informationsverlust dar und sollte bei künftigen Überlegungen zur Aufwandsreduzierung der Confounderidentifikation in Betracht gezogen werden.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.6.2.1. (S. 43)	<p><b>Anmerkung</b></p> <p>Im vorläufigen Rapid Report wird der Begriff der Überlappung mehrfach aufgegriffen, z.B. im Kontext fehlender Werte und der Anwendung von MICE (multiple Imputation mittels Chained Equations). Dabei wird ausgeführt, dass sich „gute Überlappung“ auf eine ausreichend große überlappende Fläche bezieht. Auch auf eine „ausreichende Balanciertheit“ zwischen den beiden Behandlungsgruppen wird eingegangen.</p> <p>Allerdings fehlen konkrete Entscheidungskriterien, ab wann eine Überlappung oder eine Balanciertheit als ausreichend anzusehen ist. In bisherigen AbD-Verfahren wurden hierzu unterschiedliche Maßstäbe angelegt.</p> <p>So wurden meist Überlappungen von über 50 % als adäquat angesehen, wie beispielsweise in den Verfahren zu Brexucabtagen-Autoleucel, Onasemnogen-Abeparvovec oder Risdiplam [2-4].</p> <p>Bezüglich der Balanciertheit schwankten die in den Studienunterlagen genannten und für die Datenerhebung berücksichtigten Schwellenwerte für die standardisierte Mittelwertsdifferenz (Standardized Mean Difference, SMD). Im Verfahren zu Etranacogen Dezaparvovec wurde idealerweise eine Balanciertheit der SMD von &lt; 0,1 angestrebt, jedoch SMD von &lt; 0,25 toleriert [5]. Für Onasemnogen-Abeparvovec und Risdiplam wurde eine ausreichende Balanciertheit mit SMD von &lt; 0,2 beschrieben [2, 4], während bei Brexucabtagen-Autoleucel ein Wert der SMD von &lt; 0,25 als ausreichend galt [3]. Diese Bandbreite zeigt, dass die Entscheidung aktuell fallweise erfolgt, ohne dass nachvollziehbar ist, welche Grenzwerte in welchem Kontext, bzw. in Abhängigkeit welcher weiteren methodischen Gegebenheiten, als adäquat angesehen werden. Dies erschwert die methodische Planung einer AbD und birgt die Gefahr, dass vergleichbare Fragestellungen auf Basis unterschiedlicher Maßstäbe beurteilt werden.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>AMS schlägt vor, im Rapid Report klare Entscheidungshilfen für die Bewertung von Überlappung und Balanciertheit zu geben. Dabei könnte dargestellt werden, unter welchen methodischen und inhaltlichen Voraussetzungen bestimmte Schwellenwerte geeignet sind und in welchen Fällen Abweichungen begründet werden sollten. Eine solche Orientierung würde dazu beitragen, die Transparenz und Vergleichbarkeit der Bewertungen zu erhöhen und die Planung zukünftiger AbDs deutlich erleichtern.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b>
	<p><i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i></p>
5.7 (S. 50)	<p><b>Anmerkung</b></p> <p>Im vorläufigen Rapid Report wird darauf hingewiesen, dass „Zur Sicherstellung einer adäquaten Confounderadjustierung [...] eine ausreichend hohe Anzahl von Patientinnen und Patienten erforderlich“ ist und sich die notwendige Patientenzahl mit zunehmender Anzahl potenziell relevanter Confounder erhöht.</p> <p>Die im Abschnitt 5.7 zitierten Simulationsstudien zeigen zwar die Verzerrungen, die mit kleinen Fallzahlen einhergehen, betrachten jedoch lediglich bis zu 10 Confounder. Das sind, wie auch vom IQWiG angemerkt, deutlich weniger Confounder, als in laufenden AbD-Verfahren häufig berücksichtigt werden.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>In diesem Kontext begrüßt AMS den Vorschlag, bei fehlender Konvergenz des Modells die Anzahl an Confoundern gezielt zu reduzieren, in dem sukzessive die als weniger relevant eingestuften Confounder aus dem Modell entfernt werden, immer unter Berücksichtigung des spezifischen Einzelfalls.</p> <p>Kritisch sehen wir jedoch den Vorschlag, in solchen Fällen eine „weitere Verschiebung der Nullhypothese“ vorzunehmen. Dieses Vorgehen ist nicht zielführend, da so nur extrem große Effekte nachweisbar wären, sodass selbst klinisch relevante Ergebnisse gegebenenfalls statistisch nicht mehr erkannt werden. AMS schlägt vor, diesen Zusatz zu entfernen.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

#### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Systematische Confounderidentifikation in der Indikation schubförmig remittierende multiple Sklerose (RRMS), Arbeitspapier Projekt: GA23-02, Version 1.0. 2025. Verfügbar unter: [https://www.iqwig.de/download/ga23-02\\_systematische-confounderidentifikation-indikation-rrms\\_arbeitspapier\\_v1-0.pdf](https://www.iqwig.de/download/ga23-02_systematische-confounderidentifikation-indikation-rrms_arbeitspapier_v1-0.pdf). [Zugriff am: 18.08.2025]
2. Novartis Gene Therapies Inc. Routine data collection and evaluations of onasemnogene abeparvovec in Germany, Study Protocol, Protocol Number: COAV101A1DE01, Version: 4.01. 2024. Verfügbar unter: [https://www.g-ba.de/downloads/40-268-10961/2024-06-06\\_AM-RL-](https://www.g-ba.de/downloads/40-268-10961/2024-06-06_AM-RL-)

XII Onasemnogen-Abeparvovec 2020-AbD-001 Ueberpruefung-SP-SAP Studienunterlagen.pdf. [Zugriff am: 18.08.2025]

3. Gilead Sciences GmbH. STUDY PROJECT PLAN 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, Project Plan Number: RW-X19-2206, Version: 4.0. 2025. Verfügbar unter: [https://www.g-ba.de/downloads/40-268-11585/2025-06-18\\_AM-RL-XII\\_BrexCel\\_2021-AbD-008\\_Feststellung\\_Studienunterlagen.pdf](https://www.g-ba.de/downloads/40-268-11585/2025-06-18_AM-RL-XII_BrexCel_2021-AbD-008_Feststellung_Studienunterlagen.pdf). [Zugriff am: 18.08.2025]
4. F. Hoffmann-La Roche Ltd. EVALUATION OF A REAL WORLD DATA COLLECTION FOR THE REASSESSMENT OF THE ADDITIONAL BENEFIT OF EVRYSDI® (RISDIPLAM), PROTOCOL NUMBER: ML44661, VERSION NUMBER: 3.0. 2019. Verfügbar unter: [https://www.g-ba.de/downloads/40-268-10838/2024-09-19\\_AM-RL-XII\\_Risdiplam\\_AbD-004\\_Feststellung\\_Studienunterlagen.pdf](https://www.g-ba.de/downloads/40-268-10838/2024-09-19_AM-RL-XII_Risdiplam_AbD-004_Feststellung_Studienunterlagen.pdf). [Zugriff am: 18.08.2025]
5. CSL Behring GmbH. STUDY PROTOCOL Routine Practice Data Collection and Evaluation of etranacogene dezaparvovec (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 G-BA, Protocol Version: original v3.0. 2024. Verfügbar unter: [https://www.g-ba.de/downloads/40-268-10667/2024-07-18\\_AM-RL-XII\\_Etranacogen-Dezaparvovec\\_2022-AbD-005\\_Studienunterlagen.pdf](https://www.g-ba.de/downloads/40-268-10667/2024-07-18_AM-RL-XII_Etranacogen-Dezaparvovec_2022-AbD-005_Studienunterlagen.pdf). [Zugriff am: 18.08.2025]

### **A.1.3 Bayer Vital GmbH**

#### **Autorinnen und Autoren**

- Dintsios, Charalabos-Markos

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> Bitte nutzen Sie pro Person 1 Zeile.  Dintsios, Charalabos-Markos; Dr. Dr.
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b> <input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Bayer Vital GmbH  <input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.3.1 (S. 14)	<p><u>Anmerkung</u></p> <p>In der wissenschaftlichen Ausarbeitung wird ausgeführt: „Ein Aufgreifkriterium für eine AbD ist eine unzureichende Evidenzlage für ein neues Arzneimittel im Vergleich zum therapeutischen Standard. Mit der AbD werden demnach erstmals vergleichende Daten gewonnen. Vor diesem Hintergrund kann es für die Situation einer AbD sinnvoll sein, eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlsschätzung auf dieses Ziel auszurichten. Stattdessen ist denkbar, die Schätzung der notwendigen Fallzahl darauf zu beziehen, dass aus einem gewählten Propensity-Score-Verfahren unter Sicherstellung einer adäquaten Confounderadjustierung verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95 %-Konfidenzintervall resultieren (siehe hierzu auch Abschnitt 5.7), und zwar losgelöst davon, ob sich unter Berücksichtigung der verschobenen Nullhypothese ein Vorteil des zu bewertenden Arzneimittels ableiten lässt. Unter dieser Voraussetzung stellt jedes Ergebnis einer AbD einen Erkenntnisgewinn dar“.</p> <p>Dies mag wissenschaftlich zwar stimmen, verkennt aber die gesetzlich intendierte Veranlassung einer AbD mit dem expliziten Ziel einer Quantifizierung des Zusatznutzens von bspw. Orphan Drugs im Rahmen einer anwendungsbegleitenden Datenerhebung nach dem Gesetz für mehr Sicherheit in der Arzneimittelversorgung. Die entsprechenden Argumente sind der Gesetzesbegründung zu entnehmen und auch in der Verfahrensordnung des G-BA so festgehalten.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Ab „vor diesem Hintergrund ...“ Passus löschen.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
5.3.1 (S. 16ff.) Gesamter Abschnitt „verschobene Nullhypothese“	<p><u>Anmerkung</u></p> <p>Da anders als in einer RCT über Confounderkontrolle nur ‘obvious bias’ kontrolliert werden kann, ein Teil potenziellen Bias aber als ‘hidden’ und damit als nicht kontrollierbar auftreten kann, ist damit zu rechnen, dass keine idealtypisierte Strukturgleichheit der Vergleichsarme über eine AbD erzielt werden kann, sondern letztere über entsprechende statistische Verfahren approximiert wird. Dies induziert eine inhärente Heterogenität der Vergleichsarme, die somit zu einer geringeren Präzision der daraus abgeleiteten Effekte führt. Die verschobene Nullhypothese ist bereits für Effektschätzer aus RCTs sehr konservativ angesetzt wie den Stellungnahmen unterschiedlicher Akteure bei ihrer Einführung durch das IQWiG zu entnehmen ist. Im Rahmen der AbD gestaltet sich der ganze Ansatz noch konservativer und impliziert zur Unterschreitung der endpunktspezifischen Schwellenwerte des oberen Konfidenzintervalls quasi dramatische Effekte. Aufgrund der obengenannten inhärent bedingten größeren Heterogenität zwischen den Studienarmen im Rahmen der AbD ist die verschobene Nullhypothese hier nicht zielführend und es sollte in Anerkennung dieser Konstellation primär auf die Effektschätzer als solche, insofern statistisch signifikant, zur Ableitung eines endpunktbezogenen Zusatznutzens abgestellt werden.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Die verschobene Nullhypothese nicht im Rahmen der AbD zur Ableitung eines endpunktbezogenen Zusatznutzens anwenden und direkt aus den Effektschätzern den endpunktspezifischen Zusatznutzen ableiten.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

#### **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

#### **A.1.4 Bristol-Myers Squibb GmbH & Co. KGaA (BMS)**

##### **Autorinnen und Autoren**

- Bluhmki, Tobias
- Böhm, Doris
- Ellis, Aneurin
- Hofmann-Xu, Liping
- Pedretti, Annina
- Streicher, Sarah

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Dr. Tobias Bluhmki
Doris Böhm
Liping Hofmann-Xu
Aneurin Ellis
Sarah Streicher
Annina Pedretti
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Bristol-Myers Squibb GmbH & Co. KGaA
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

Beim vorliegenden vorläufigen Rapid Report des IQWiG handelt es sich um eine wissenschaftliche Ausarbeitung zu ausgewählten methodischen Aspekten zur Generierung versorgungsnaher Daten im Rahmen einer anwendungsbegleitenden Datenerhebung (AbD). Die Bristol-Myers Squibb GmbH & Co. KGaA (im Folgenden BMS) begrüßt ausdrücklich die Möglichkeit der Stellungnahme und der Diskussion der betroffenen Akteure vor Finalisierung des Berichts. Da eine Kommentierung des Rapid Report A19-43 [1] seinerzeit nicht möglich war und der aktuelle vorläufige Rapid Report ausdrücklich als weiterführende wissenschaftliche Ausarbeitung dessen gilt, bezieht sich die Stellungnahme von BMS auch auf grundlegende Aspekte und Vorgaben aus dem Bericht A19-43.

Gemäß Beauftragung durch den G-BA vertieft das IQWiG ausgewählte Themenbereiche zur Methodik von AbD. BMS erkennt diese Auseinandersetzung ausdrücklich an und würdigt den vorliegenden vorläufigen Rapid Report als wichtigen Beitrag zur Weiterentwicklung der Methodik zur Generierung versorgungsnaher Daten.

Gleichzeitig sieht BMS die grundsätzliche Machbarkeit einer AbD in bestimmten Konstellationen kritisch. Vor diesem Hintergrund möchte BMS insbesondere zur Schätzung notwendiger Fallzahlen (orientierende Fallzahlenschätzung bzw. orientierende Betrachtung von Fallzahlszenarien) Stellung nehmen, da diese entscheidend für die Bewertung der grundsätzlichen Machbarkeit der AbD ist. In diesem Zusammenhang stellt die vom IQWiG erhobene Forderung, der Datenanalyse einen statistischen Test mit verschobener Nullhypothese zugrunde zu legen, aus Sicht von BMS weiterhin eine methodische Herausforderung dar. Insbesondere bei seltenen Erkrankungen kann dies zu hohen Fallzahlen führen, die in der Versorgungspraxis in einem angemessenen Zeitrahmen schwer zu erreichen sind. Zudem werden in AbD-Verfahren oft mehrere wirksame und verträgliche Therapieoptionen miteinander verglichen, sodass die strikten Anforderungen an die Behandlungseffekte möglicherweise nicht der Realität entsprechen.

Aus Sicht von BMS sollte – gemäß Gesetzesgrundlage – die valide Quantifizierung des Zusatznutzens stets das übergeordnete Ziel einer AbD darstellen [2]. In diesem Zusammenhang erscheint die Aussage des IQWiG, dass „[...] jedes Ergebnis einer AbD einen

*Erkenntnisgewinn dar[stellt] [...] losgelöst davon, ob sich [...] ein Vorteil des zu bewertenden Arzneimittels ableiten lässt“ (S. 14) diskussionswürdig.*

Im Folgenden wird auf die genannten Punkte näher eingegangen.

Kapitel / Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung
	<i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
Abschnitt 5.2 (S. 9 ff) Abschnitt 5.7 (S. 50 ff)	<u>Anmerkung</u> <p>Die im vorläufigen Rapid Report dargestellten Inhalte zur Confounderidentifikation basieren im Wesentlichen auf dem Arbeitspapier GA23-02 des IQWiG [3]. BMS begrüßt ausdrücklich, dass die Thematik vertieft und ein konkretes methodisches Vorgehen diskutiert wird.</p> <p>Auch wenn der Aufwand dieses Ansatzes fallspezifisch schwer abschätzbar ist, teilt BMS die Einschätzung des IQWiG hinsichtlich der Herausforderungen, die mit einer hohen Anzahl an Confoundern bei gleichzeitig kleinen Fallzahlen in AbD-Verfahren verbunden sind. Das IQWiG selbst führt dazu aus, dass „<i>[...] unklar [bleibt], in welchen Fällen (und unter welchen Voraussetzungen) Propensity-Score-Analysen in einer AbD zu interpretierbaren Ergebnissen führen werden</i>“ (S. 55).</p> <p>Das im Arbeitspapier GA23-02 diskutierte Beispiel zur Indikation schubförmig remittierende multiple Sklerose (RRMS) verdeutlicht diese Problematik: Trotz methodischer „Verdichtung“ der ursprünglich identifizierten 132 potenziellen Confounder verbleiben 28, die gemäß IQWiG-Methodik vollständig erhoben und inklusive möglicher Wechselwirkungen in die Analyse einbezogen werden sollen [4].</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Mit dem Ziel eine valide Quantifizierung des Zusatznutzen zu ermöglichen, stellen Propensity-Score-Analysen in Anwendungsgebieten mit kleinen Patientenkollektiven auch aus Sicht von BMS eine besondere Herausforderung dar – insbesondere im Hinblick auf die Datenerhebung und -verfügbarkeit sowie die Modellstabilität. BMS wünscht sich hier eine weiterführende Diskussion zu alternativen Methoden.</p>
Abschnitt 5.3.1 (S. 12 ff)	<u>Anmerkung</u> <p><i>Verschobene Nullhypothese</i></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Das IQWiG hält im vorliegenden vorläufigen Rapid Report weiterhin an statistischen Tests mit verschobener Nullhypothese fest. Aus Sicht von BMS ist dieser methodische Ansatz kritisch zu bewerten. Der vorgeschlagene Schwellenwert für das relative Risiko im Bereich von 2 bis 5, bezogen auf die untere Grenze des Konfidenzintervalls, ist wissenschaftlich nicht etabliert und berücksichtigt weder die Art der Erkrankung noch den zugrunde liegenden Endpunkt. Vor dem Hintergrund der für die AbD vorgesehenen parallelen, prospektiven Datenerhebung, einer einheitlichen Messung von Endpunkten und der Berücksichtigung von Confoundern erscheint dieser Ansatz methodisch sehr anspruchsvoll und nicht verhältnismäßig.</p> <p>BMS ist der Auffassung, dass verbleibende Unsicherheiten nach erfolgter Adjustierung nicht pauschal durch die Forderung großer Effektstärken kompensiert werden sollten. Insbesondere kann bei einem Vergleich wirksamer und verträglicher Therapieoptionen im Versorgungsalltag der tatsächlich zu erwartende Behandlungseffekt geringer ausfallen. Stattdessen sollten die getroffenen Annahmen zur Strukturgleichheit kritisch überprüft und durch geeignete Sensitivitätsanalysen ergänzt werden.</p> <p><i>Verteilungsverhältnis zwischen den Behandlungsgruppen</i></p> <p>Das IQWiG gibt an, dass im Rahmen der AbD-Konzepterstellung – sofern keine belastbaren Informationen zu den Versorgungsanteilen der jeweiligen Arzneimittel vorliegen – Verteilungsverhältnisse in der Regel zwischen 5:1 und 1:5 angenommen werden, um deren Auswirkungen auf die benötigte Fallzahl bzw. Effektstärke zu untersuchen. BMS begrüßt diese differenzierte Betrachtungsweise als praktisch relevant und sachgerecht.</p> <p>BMS möchte jedoch darauf hinweisen, dass stark asymmetrische Verteilungsverhältnisse wie 1:5 (oder extremer) praktische Herausforderung und Risiken für die Validität kausaler Schlussfolgerungen mit sich bringen. So kann es vorkommen, dass bestimmte Kombinationen von Patientenmerkmalen – insbesondere seltene Konstellationen – im unterrepräsentierten Arm kaum oder gar nicht vertreten sind. Dies führt zu extremen Propensity Scores nahe 0 oder 1, was instabile Gewichtungen oder mangelnde Überlappung der zugrundeliegenden Verteilungen zur Folge haben kann. In der Konsequenz stellt sich die Frage, inwieweit unter solchen Bedingungen eine Verletzung der Positivitätsannahme vorliegt. Dies kann die Aussagekraft der Ergebnisse erheblich beeinträchtigen.</p> <p><u>Vorgeschlagene Änderung</u></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>BMS begrüßt grundsätzlich die Generierung versorgungsnahe Evidenz und sieht sie als sinnvolle Ergänzung zur Nutzenbewertung.</p> <p>Aus Sicht von BMS sollte die Konzepterstellung zur Einschätzung der grundsätzlichen Machbarkeit einer AbD deutlich differenzierter erfolgen. Neben der methodischen Ausgestaltung sind zudem Rahmenbedingungen wie die Unterschiedlichkeit der zu vergleichenden Therapiekonzepte, Präferenzen von Ärzt:innen und Patient:innen und das zu erwartende Verteilungsverhältnis zwischen den Behandlungsgruppen zu berücksichtigen. Das vorgeschlagene Vorgehen des IQWiG, in dem Fall, dass Studienergebnisse keine hinreichende Einschätzung der Effekte zulassen, Fallzahlszenarien zu beschreiben und eine „aufdeckbare Effektgröße“ anzugeben, ohne deren Erreichbarkeit einzuordnen greift aus Sicht von BMS zu kurz.</p> <p>Die valide Quantifizierung des Zusatznutzens sollte stets das übergeordnete Ziel einer AbD sein. Entsprechend möchte BMS anregen, eine Weiterentwicklung der regelhaft angewandten Methodik in Betracht zu ziehen und die konzeptionellen Rahmenbedingungen bereits frühzeitig kritisch zu diskutieren.</p>
Abschnitt 5.1 (S. 6 ff.)  Abschnitt 5.3.2 (S. 18 ff.)	<p><u>Anmerkung</u></p> <p>Das IQWiG beschreibt das Konzept der Target Trial Emulation als „Best-Practice“-Ansatz – eine Einschätzung, die auch von BMS geteilt wird. Dabei ist zu beachten, dass das Design der emulierten hypothetischen randomisierten kontrollierten Studie (Target Trial), die die relevante Fragestellung behandelt und an dem sich die AbD orientiert, notwendigerweise einem pragmatischen Ansatz folgt [5,6]. Im Gegensatz zu klassischen hochkontrollierten randomisierten klinischen Studien sind pragmatische Studien durch größere Heterogenität (etwa in Patientencharakteristika oder -adhärenz) gekennzeichnet, was typischerweise zu einem höheren Maß an „Hintergrundrauschen“ (<i>noise</i>) und somit auch tendenziell zu kleineren Behandlungseffekten führen kann. Dies gilt umso mehr für die nicht-randomisierten Emulationen solcher Studien. Vor diesem Hintergrund erscheint die Annahme großer Effektstärken in der Fallzahlplanung methodisch fragwürdig: Sie kann zu einer systematischen Unterschätzung der erforderlichen Fallzahl führen und damit Fehleinschätzungen begünstigen – etwa hinsichtlich der Dauer oder der generellen Machbarkeit der AbD.</p> <p><u>Vorgeschlagene Änderung</u></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b>
	<p><i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i></p> <p>Eine umfassende Berücksichtigung dieser Aspekte erscheint notwendig, um die Vorteile der Target Trial Emulation adäquat und praktikabel in die Nutzenbewertung zu integrieren. Insbesondere sollte die Validität der für die Fallzahlplanung angenommenen Effektgrößen kritisch hinterfragt werden (s. auch vorherige Anmerkung). Dies gilt auch für eine Bewertung der Größe des „aufdeckbaren Effektes“ im Rahmen der orientierenden Betrachtung von Fallzahlszenarien.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. 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: 11.08.2025]. URL: [https://www.iqwig.de/download/a19-43\\_versorgungsnahe-daten-zum-zwecke-der-nutzenbewertung\\_rapid-report\\_v1-1.pdf](https://www.iqwig.de/download/a19-43_versorgungsnahe-daten-zum-zwecke-der-nutzenbewertung_rapid-report_v1-1.pdf).
2. Deutscher Bundestag (2019): Beschlussempfehlung und Bericht Ausschuss für Gesundheit (14. Ausschuss) a) zu dem Gesetzentwurf der Bundesregierung – Drucksache 19/8753 – Entwurf eines Gesetzes für mehr Sicherheit in der Arzneimittelversorgung [online]. 2019 [Zugriff am 14.08.2025]. URL: <https://dserver.bundestag.de/btd/19/106/1910681.pdf>
3. Gesundheitswesen. Systematische Confounderidentifikation in der Indikation schubförmiger remittierende multiple Sklerose (RRMS); Arbeitspapier [online]. 2025 [Zugriff: 11.08.2025]. URL: <https://doi.org/10.60584/GA23-02>.
4. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 14.08.2025]. URL: [https://www.iqwig.de/methoden/allgemeine-methoden\\_version-7-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf).

5. Matthews AA, Danaei G, Islam N et al. Target trial emulation: applying principles of randomised trials to observational studies. *BMJ* 2022; 378: e071108. <https://doi.org/10.1136/bmj-2022-071108>.
6. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* 2016; 183(8): 758-764. <https://doi.org/10.1093/aje/kwv254>.

### **A.1.5 Bundesverband der Pharmazeutischen Industrie e. V. (BPI)**

#### **Autorinnen und Autoren**

- Wilken, Matthias

## **Stellungnahme zum vorläufigen Rapid Report**

**Berichtnr: A25-13**

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

**Name, Vorname; Titel des/der Stellungnehmenden**

*Bitte nutzen Sie pro Person 1 Zeile.*

Wilken, Matthias; Dr.

**Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)**

- im Namen folgender Institution / Organisation: BPI e. V.
- als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

In den nunmehr über vier Jahren seit Einleitung der ersten Verfahren zur Festlegung einer anwendungsbegleitenden Datenerhebung (AbD) konnten durch die Verfahrensbeteiligten eine Reihe von Erfahrungen gesammelt werden. Dabei zeigte sich, dass die Planung und Umsetzung einer AbD von hoher Komplexität gekennzeichnet ist und die Frage der Umsetzbarkeit unter Realbedingungen häufig ein zentrales Diskussionsthema darstellt. Dies ist insofern erwartbar, als der Anwendungsbereich der AbD genau diejenigen Arzneimittelinnovationen umfasst, für die z. B. aufgrund der Seltenheit der Erkrankung oder eines besonders hohen medizinischen Bedarfs die Generierung von Evidenz aus randomisierten kontrollierten Studien erschwert war.

Es ist hervorheben, dass das übergeordnete Ziel einer AbD die Quantifizierung eines Zusatznutzen ist. Die Aussage des IQWiG im vorliegenden vorläufigen Rapid Report, dass „[...] jedes Ergebnis einer AbD einen Erkenntnisgewinn dar[stellt] [...] losgelöst davon, ob sich [...] ein Vorteil des zu bewertenden Arzneimittels ableiten lässt“ (S. 14) steht nicht im Einklang mit der Gesetzesbegründung zum Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV) zu ihrer Einführung. Dort heißt es: „Ziel der anwendungsbegleitenden Datenerhebung ist es, eine valide Quantifizierung des Zusatznutzens zu erreichen. Diesem Ziel muss der Gemeinsame Bundesausschuss bei der Festlegung der Anforderungen Rechnung tragen. Gleichzeitig müssen die Anforderungen unter Berücksichtigung des Grundsatzes der Verhältnismäßigkeit geeignet und angemessen sein.“ (Drucksache 19/10681, S. 85). Dies bedeutet im Umkehrschluss auch, dass die Forderung einer AbD unterbleiben muss, sofern nicht von Anfang an realistische Chancen bestehen, dass diese zu einer Quantifizierbarkeit des Zusatznutzens führen wird. Es ist eine Aufgabe des IQWiG, den G-BA in dieser Hinsicht verfahrensbezogen zu beraten.

Vor dem Hintergrund der vielfältigen Fragen bzgl. für die Durchführbarkeit einer AbD zur Generierung von Evidenz für die Nutzenbewertung ist es nachvollziehbar, dass der gesetzliche Rahmen der AbD vorsieht, dass der G-BA „insbesondere Vorgaben zur Methodik sowie zu patientenrelevanten Endpunkten und deren Erfassung“ bestimmen muss (§ 35a Abs. 3b, Satz 4 SGB V). Gemäß Gesetzesbegründung zum GSAV hat der G-BA „so konkret wie möglich festzulegen, in welcher Form die anwendungsbegleitende Datenerhebung durchgeführt

werden soll“. Tatsächlich aber enthalten häufig weder die Konzepte des IQWiG noch die Forderungsbeschlüsse des G-BA hinreichend konkrete Vorgaben zu zentralen Aspekten der Studienplanung und bleiben somit hinter den gesetzlichen Anforderungen zurück.

Mit aktuellen Erfahrungen, Analysen und Weiterentwicklungsvorschlägen beschäftigt sich die Publikation „Machbarkeit und Praktikabilität der anwendungsbegleitenden Datenerhebung“.<sup>1</sup> Der BPI bittet um ihre Berücksichtigung bei der Erstellung des Rapid Reports. Diese Publikation äußert sich umfassend zu verschiedenen Herausforderungen, so dass nachfolgend nur auf ausgewählte Aspekte eingegangen wird.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b>
5.3, Fallzahlschätzung	<p><b>Anmerkung</b></p> <p>Die AbD verfolgt das Ziel, Evidenzlücken zu schließen und den Zusatznutzen eines Arzneimittels zu quantifizieren. Im vorläufigen Rapid Report fehlen jedoch Hinweise, wie die beschriebenen Ansätze zur Fallzahlschätzung konkret in eine belastbare Quantifizierung des Zusatznutzens überführt werden können.</p> <p>Der vorläufige Rapid Report weist zudem darauf hin, dass im Rahmen einer AbD auch ein Erkenntnisgewinn als Studienziel ausreichend sein könne. Diese Perspektive steht nicht mit der gesetzlichen Vorgabe nach §35a SGB V im Einklang die explizit auf die Quantifizierung des Zusatznutzens abzielt.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Es sollte ergänzend darauf eingegangen werden, dass die Auswahl und Priorisierung der Methodik stets dem Ziel dienen, eine Quantifizierung des Zusatznutzen zu ermöglichen.</p> <p>Dazu können insbesondere gehören:</p> <ul style="list-style-type: none"> <li>• eine realistische Fallzahlplanung,</li> <li>• die Operationalisierung geeigneter Endpunkte mit klarer Bewertungsstrategie,</li> <li>• ggf. die Berücksichtigung alternativer Schwellenwerte,</li> <li>• Hinweise an den G-BA, sofern eine Quantifizierung methodisch oder organisatorisch nicht realistisch erreichbar ist.</li> </ul>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.5 Patientenberichtete Endpunkte	<p><b>Anmerkung</b></p> <p>Das IQWiG stellt fest, dass PRO-Daten im Kontext von AbD in der Versorgungspraxis erhoben werden. Damit gehe einher, dass sich die Bedingungen für die Datenerhebungen deutlich von jenen in einer klinischen Studie unterscheiden. Das IQWiG zeigt nachfolgend Wege auf, wie unter diesen Rahmenbedingungen eine Erhebung von PRO mit möglichst wenig Aufwand gelingen kann. Es bedarf an dieser Stelle Ergänzungen.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>In vergangenen Verfahren ergaben sich bezüglich der Verwendung von PRO Herausforderungen, die auch bereits im G-BA Workshop 2024 diskutiert wurden. Es wäre hilfreich, wenn das IQWiG sich auch dazu äußert, wie damit umzugehen ist, wenn keine spezifischen Fragebögen vorliegen bzw. diese nicht in relevanten Sprachen vorliegen und/oder nicht validiert sind. Zudem wäre es hilfreich, wenn das IQWiG Stellung bezieht, wie Schwierigkeiten bei der Erhebung über einheitliche Zeitpunkte insbesondere im SoC-Arm aufgelöst werden können, da SoC-Patienten mit hoher Wahrscheinlichkeit nicht in gleichem Umfang ein Interesse an der Teilnahme haben.</p>
5.6 Fehlende Werte	<p><b>Anmerkung</b></p> <p>Das IQWiG schlägt eine Vergütung des Aufwands der teilnehmenden Zentren durch Zahlung einer Aufwandsentschädigung vor, um fehlende Werte zu vermeiden. Pharmazeutischen Unternehmen (pU) sind durch Vergütung auf Fair Market Value-Basis Grenzen gesetzt. Es ist zweifelhaft, inwiefern Zentren hierdurch incentiviert werden. Der Vorschlag des IQWiG für finanzielle Zuwendungen oder Sachprämien für Patienten durch den pU ist ebenfalls rechtlich zweifelhaft. Insgesamt sind pU gesetzlich enge Grenzen gesetzt.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Das IQWiG wird gebeten, den Passus zu überarbeiten und dabei auf die gesetzlichen Vorgaben bezüglich Zuwendungen durch den pU stärker einzugehen. Es sollten keine Optionen in den Raum gestellt werden, die rechtlich nicht möglich sind.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

## **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

---

- i Wilken, M., Berkemeier, F.: „Machbarkeit und Praktikabilität der anwendungsbegleitenden Datenerhebung (AbD)“, in: „Monitor Versorgungsforschung“ (02/24), S. 62–68.  
<http://doi.org/10.24945/MVF.02.24.1866-0533.2595>

### **A.1.6 CSL Behring GmbH**

#### **Autorinnen und Autoren**

- Jülich, Fabian

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Jülich, Fabian
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: CSL Behring GmbH
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.*

### **5.3 - Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen, 5.3.1 Verschobene Nullhypothese [S. 16]**

#### Anmerkung

Ein wesentlicher Aspekt der Fallzahlsschätzung gemäß Rapid Report betrifft das Konzept der verschobenen Nullhypothese, d.h. eine Hypothesenschwelle mit einem Rate Ratio = 0,5 bzw. 2 (RR0 = 0,5; RR0 = 2). Die verschobene Nullhypothese wird laut Rapid Report eingeführt, um mögliche Verzerrungen durch eine fehlende Randomisierung zu kompensieren. Es wird dabei vorgegeben, dass ein Konfidenzintervall vollständig außerhalb der verschobenen Hypothesengrenze liegen muss, um einen Effekt mit ausreicher Aussagesicherheit ableiten zu können. Dieses Vorgehen ist sehr konservativ mit Blick auf die Aussagesicherheit und führt zu der Vervielfachung der benötigten Anzahl an Patienten, die rekrutiert und ausgewertet werden müssen (Wilken & Berkemeier, 2024). Dies wiederum kann in vielen Fällen dazu führen, dass relevante Effekte von innovativen Therapien nicht als solche anerkannt werden bzw. die Durchführung einer ausreichend „gepowerten“ Studie sich schlussendlich als nicht realisierbar erweist. Gerade bei AbDs mit kleinen Patientenzahlen (Orphan Diseases) erscheint es fragwürdig, ob die notwendigen Fallzahlen überhaupt erreicht werden können. So war die Frage der Fallzahlsschätzung in allen bisher bewerteten und begonnenen AbD ein zentraler und kritisch diskutierter Aspekt bezüglich der Realisierbarkeit der Studie.

#### Vorgeschlagene Änderung

Aus Sicht von CSL Behring sollte versucht werden einen „Mittelweg“ zu finden, der eine ausreichend hohe Aussagesicherheit auf der einen Seite bei gleichzeitig regelhaft guter Realisierbarkeit auf der anderen Seite gewährleistet. Zu diesem Kontext wurde im Rahmen einer Simulationsstudie ein alternativer Vorschlag untersucht und abgeleitet, der dem bereits bestehenden Prinzip des "dramatischen Effekts" bei nicht adjustierten indirekten Vergleichen folgt (Wilken & Berkemeier, 2024). Dabei wird ein höheres Signifikanzniveau angesetzt, d. h.  $p < 0,01$  an Stelle von  $<0,05$ , jedoch ohne eine Verschiebung der Nullhypothese. Dieses Vorgehen würde die Anzahl an benötigten Patienten deutlich reduzieren (mitunter um >90%) und die Realisierbarkeit von AbD somit wesentlich verbessern: „Exemplarisch läge die notwendige Fallzahl z. B. bei einem RR von 0,4 (d. h. einer Risikoreduktion um 60%) und einer Eventrate von 50 im Kontrollarm bei den bisher angewendeten AbD-Effektschwellen bei fast 1.975 Patient:innen. Bei Anwendung des vorgeschlagenen „Mittelwegs“ dagegen wären nur 146 Patient:innen erforderlich [...]“ (Wilken & Berkemeier, 2024). Einzelheiten zu diesem Vorschlag, einschließlich entsprechender Fallzahlkalkulationen, sind im eingereichten Studienprotokoll zur AbD von Etranacogen dezaparvovec beschrieben (CSL Behring, 2024).

### **5.3 - Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen, 5.3.3 Entscheidungsfindung für oder gegen eine orientierende Fallzahlschätzung [S. 25]**

#### Anmerkung

Im Rapid Report wird ausgeführt, dass „je nach gewähltem Propensity-Score-Verfahren, ggf. ein relevanter Anteil der Patientinnen und Patienten nicht in der Auswertung berücksichtigt werden kann. Wie viele Patientinnen und Patienten nicht in die Analysepopulation eingehen, sei dabei im Vorfeld nicht abschätzbar.“

Es ist nachvollziehbar, dass der anteilige „Verlust“ an Patienten bei der Ergebnisauswertung durch eine entsprechende Confounderadjustierung im Vorfeld der AbD nicht zuverlässig abschätzbar ist und dieser eine relevante Größenordnung haben kann.

Es ist jedoch kritisch, dass dieser Umstand bislang bei den orientierenden Fallzahlplanungen im Rahmen der Machbarkeitsanalyse vor Beginn der AbD komplett unberücksichtigt bleibt. Dadurch werden die vom IQWIG ermittelten Fallzahlen bzw. Fallzahl-Szenarien systematisch unterschätzt. Es ist deshalb davon auszugehen, dass es hierdurch regelhaft zu fehlerhaften Schlussfolgerungen bei der Bewertung der Realisierbarkeit von AbD kommt.

## Vorgeschlagene Änderung

Der durch Confounderadjustierung anteilige „Verlust“ an auswertbaren Patienten sollte auf Basis von konservativen Annahmen oder mithilfe von Szenarien bereits bei der Machbarkeitsanalyse einer AbD (vor deren Beginn) bei der Fallzahlplanung berücksichtigt werden, um das Risiko einer Unterschätzung der Fallzahl und fehlerhaft Rückschlüsse zur Realisierbarkeit zu minimieren. Hierzu wurde bereits mündlich ein Vorschlag seitens des IQWIG im G-BA Workshop zu AbD am 26. April 2024 gemacht.

## **5.6 – Fehlende Werte, Fehlende Werte und Confounderadjustierung [ab S. 37]**

### Anmerkung

Die Forderung nach vollständigen Daten für alle Confounder ist realitätsfern, insbesondere bei seltenen Erkrankungen oder komplexen Registerstrukturen. MICE, als vorgeschlagenes Standardverfahren, funktioniert gut bei großen Datensätzen aber bei kleinen Fallzahlen, wie in AbDs, kann es zu Konvergenzproblemen kommen. Die Qualität der Imputation hängt stark von der Verfügbarkeit geeigneter Hilfsvariablen ab. Diese sind in Registern oft nicht vorhanden. Es wird nicht erklärt, wie die Schwellenwerte (z. B. 30 % für Endpunkte, 50 % insgesamt) zustande kommen. Es fehlen Verweise auf Studien oder methodische Standards, die diese Grenzwerte stützen und dadurch wirken sie willkürlich und nicht wissenschaftlich fundiert.

Die Anforderungen an die Confounderadjustierung sind theoretisch korrekt, aber in der Praxis und insbesondere bei seltenen Erkrankungen oft nicht erfüllbar. Die notwendige Fallzahl für eine stabile Adjustierung ist nicht spezifiziert, obwohl sie entscheidend ist. Das IQWiG erkennt an, dass Propensity-Score-Verfahren ursprünglich für große Datensätze entwickelt wurden. Es verweist auf Simulationsstudien, die zeigen, dass diese Verfahren unter bestimmten Bedingungen auch bei kleinen Fallzahlen funktionieren können. Wenn die Anzahl der Confounder zu hoch ist, kann es aber zu Konvergenzproblemen kommen. Das IQWiG nennt keine konkreten Schwellenwerte oder Empfehlungen, wie viele Patienten pro Confounder mindestens erforderlich sind, um eine stabile Adjustierung zu gewährleisten. Ohne Mindestfallzahl kann es passieren, dass die Modelle nicht konvergieren, die Ergebnisse nicht interpretierbar sind oder die Unsicherheit zu groß wird. Das kann zu verzögerten Bewertungen, unnötigen Studienverlängerungen oder sogar Nichtverwertbarkeit führen.

Verschiedene Propensity-Score-Verfahren (Matching, IPTW, Stratifizierung) können auch unterschiedliche Anforderungen an die Fallzahl haben. Das wird nicht berücksichtigt.

### Vorgeschlagene Änderung

Die im Bericht dargestellten Anforderungen an den Umgang mit fehlenden Werten sind methodisch nachvollziehbar, aber in der Praxis schwer umsetzbar. Insbesondere bei seltenen Erkrankungen und komplexen Registerstrukturen sind vollständige Daten für alle relevanten Confounder kaum realisierbar. Die Empfehlung zur Anwendung von MICE ist sinnvoll, sollte aber durch eine realistische Einschätzung der Grenzen bei kleinen Fallzahlen ergänzt werden.  
Die vorgeschlagenen Schwellenwerte für die Verwertbarkeit von Daten erscheinen nicht ausreichend empirisch fundiert. Wir empfehlen, die Anforderungen an Datenvollständigkeit und Imputation stärker kontextabhängig zu gestalten und durch praxisnahe Guidance zur Durchführung von Sensitivitätsanalysen zu ergänzen.

Die im Bericht dargestellten Anforderungen an die Confounderadjustierung sind methodisch fundiert, aber in der Praxis – insbesondere bei kleinen Fallzahlen und komplexen Registerdaten – schwer umsetzbar. Die vorgeschlagene Strategie zur Entfernung weniger relevanter Confounder bei Konvergenzproblemen birgt das Risiko einer Unteradjustierung und systematischen Verzerrung. Wir empfehlen, die Anforderungen an die Confounderadjustierung durch praxisnahe Guidance zu ergänzen – etwa zur Priorisierung von Confoundern und zur Interpretation der Ergebnissicherheit bei eingeschränkter Adjustierung.

### **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

CSL Behring. (2024). *Study protocol: Routine Practice Data Collection and Evaluation of etranacogene dezaparvovec (Hemgenix®)* . [https://www.g-ba.de/downloads/40-268-10667/2024-07-18\\_AM-RL-XII\\_Etranacogen-Dezaparvovec\\_2022-AbD-005\\_Studienunterlagen.pdf](https://www.g-ba.de/downloads/40-268-10667/2024-07-18_AM-RL-XII_Etranacogen-Dezaparvovec_2022-AbD-005_Studienunterlagen.pdf).

Wilken, M., & Berkemeier, F. (2024). Machbarkeit und Praktikabilität der anwendungsbegleitenden Datenerhebung (AbD). *Monitor Versorgungsforschung*, 02.

### **A.1.7 Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM)**

#### **Autorinnen und Autoren**

- Baum, Erika

04. August 2025

**IQWiG-Auftrag A25-13 - vorläufiger Rapid Report zur Stellungnahme  
bis 19.08.2025**

Danke für die Möglichkeit der Kommentierung. Die DEGAM sieht hier keinen Korrekturbedarf.

Es ist aus unserer Sicht aber bedauerlich, dass IQWIG keinen generellen Zugang zu anonymisierten Daten der Krankenversicherungen hat. Das würde die Qualität und Vollständigkeit der auszuwertenden Daten wesentlich verbessern und Ressourcen sparen. Als öffentlich getragene Einrichtung mit höchstem wissenschaftlichem Standard dürfte es hier auch keine Sicherheitsprobleme geben.

Kontakt:

DEGAM-Bundesgeschäftsstelle  
Schumannstraße 9  
10117 Berlin

[REDACTED]

[REDACTED]

### A.1.8 Deutsche Gesellschaft für Innere Medizin e. V. (DGIM)

#### **Autorinnen und Autoren**

- Ertl, Georg
- Führer-Sakel, Dagmar
- Sauerbruch, Tilman

Deutsche Gesellschaft für Innere Medizin e.V. · Postfach 2170 · 65011 Wiesbaden

Institut für Qualität und Wirtschaftlichkeit  
im Gesundheitswesen (IQWiG)  
Siegburger Str. 237  
50679 Köln

## Vorstand und Geschäftsführung

Geschäftsstelle Wiesbaden  
Irenenstraße 1 · 65189 Wiesbaden

[REDACTED]  
Internet: [www.dgim.de](http://www.dgim.de)

Dependance Berlin  
Oranienburger Str. 22 · 10178 Berlin

[REDACTED]  
12.08.2025 / mp

## **Kommentar zum Entwurf der wissenschaftlichen Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zweck der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

Sehr geehrte Damen und Herren,

die DGIM dankt für die Möglichkeit des Kommentars zu diesem Ansatz einer Beurteilung neuer Arzneimittel. Der Entwurf beschränkt sich auf die Nutzenbewertung nach § 35a SGB V. Grundsätzlich verbergen sich aber hier Ansätze, die den Rahmen der Bewertung therapeutischer Maßnahmen insgesamt erweitern. Dies begrüßt die DGIM.

Solche Ansätze ermöglichen, einen Wirkstoff auch über das eingeschränkte Spektrum eines RCT hinaus zu beurteilen (z.B. nicht berücksichtige Altersgruppen oder Schweregrade, regionale Ungleichheiten, Einfluss von Komedikationen etc.). Wir gehen davon aus, dass die im Verteiler genannten wissenschaftlichen Fachgesellschaften und auch die pharmazeutischen Unternehmen hier detailliert kommentieren.

Die einzelnen Schwerpunkt-Gesellschaften der DGIM (n= 11) verfügen über viel eigene und häufig international geprägte Erfahrung:

- in der Organisation und Pflege von Registern (gerade auch für seltene Erkrankungen),
- in der Bildung vergleichbarer Gruppen über propensity score matching zur Erfassung von Therapieeffekten,
- in der Anwendung von Patienten-Apps zum Monitoring,
- im Rückgriff auf Sozialdaten, dort wo sie zugänglich sind,
- in der anwendungsbegleitenden Datenerfassung,
- in der häufig schwierigen Erfassung von Lebensqualität,
- im statistischen Umgang mit Confounder-Variablen,
- in der Verlängerung von RCT in eine Extension-Phase

Hier könnte sich die DGIM vorstellen gemeinsam mit dem IQWiG einen Austausch der Kenntnisse und Argumente zu organisieren, um einige der oben genannten Punkte zu besprechen. Dies könnte auch Ihre Anforderungen an einen statistischen Analyseplan und die Art der Datenerfassung für Register (Dateninfrastruktur) umfassen. Auch unterstützt die DGIM Ihr Anliegen, die Datenerfassungen praktikabel und vom Aufwand her vertretbar zu gestalten.

Sollten Sie einen solchen Austausch mit den Schwerpunkt-Gesellschaften wünschen, so wäre die DGM gerne bereit, das zusammen mit dem IQWiG zu organisieren.

Mit den besten Grüßen

Deutsche Gesellschaft für Innere Medizin e.V.

[REDACTED]  
Univ.-Prof. Dr. Dr. med. Dagmar Führer-Sakel  
Vorstandsvorsitzende

[REDACTED]  
Prof. Dr. med. Georg Ertl  
Generalsekretär

[REDACTED]  
Prof.em. Dr. Tilman Sauerbruch  
Berater

**A.1.9 Deutsche Gesellschaft für Kinder- und Jugendmedizin e. V. (DGKJ) und Gesellschaft für Neuropädiatrie e. V. (GNP)**

**Autorinnen und Autoren**

- Meisel, Roland
- Rascher, Wolfgang
- Schara-Schmidt, Ulrike
- Ziegler, Andreas

# **Stellungnahme zum vorläufigen Rapid Report**

**Berichtnr: A25-13**

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGV**

**Von: Deutsche Gesellschaft für Kinder- und Jugendmedizin e.V. (DGKJ) und Gesellschaft für Neuropädiatrie e.V. (GNP)**

## **Vorab allgemeine Aspekte**

Grundsätzlich begrüßen wir den Rapid Report mit seinen Konkretisierungen bei der orientierenden Fallzahlschätzung, der Bedeutung eines klar definierten Beobachtungsstarts und des Behandlungswechsels sowie die jetzt eindeutig geforderten patientenberichteten Endpunkte in den Registern. Damit wurden viele wichtige Punkte sehr gut aufgegriffen und umfassend dargestellt.

Wir bitten um Verständnis, dass wir uns in der Kürze der Zeit in der Haupturlaubszeit im Sommer auf grundsätzliche Kommentare und Vorschläge beschränken, die aus Sicht der Kinder- und Jugendmedizin wichtig sind. Diese basieren auf der inzwischen mehrjährigen Erfahrung in den laufenden AbD in der Kinder- und Jugendmedizin.

## **Zum Verfahrensablauf: Expertenbeteiligung (Abb. 3: 3-schrittiges Verfahren, S. 10)**

Wir plädieren für die Einbeziehung von medizinischen Expert/-innen mit Erfahrung in der jeweiligen Versorgungspraxis der Indikation und des neuen Verfahrens/Medikamentes während der Generierung von AbD-Vorgaben des G-BA. Das Ziel wäre hier ein intrinsischer Praxis-feasibility check potentieller AbD-Szenarien.

Darüber hinaus sollten geeignete Expert/-innen zur Bewertung der Studienprotokolle und der Konzepte der AbDs von der jeweils zuständigen Fachgesellschaft bestimmt werden. Sie sollten weder identisch mit den Principal Investigators der AbD noch gleichzeitig Registerverantwortliche Personen sein, um mögliche Partikularinteressen auszuschließen.

## **Patientenauswahl**

Die Patientenauswahl für Verum- und Vergleichsgruppe muss auf Basis der Erfassung von in der Versorgungspraxis routinemäßig (nicht-interventionell) erhobener Parameter erfolgen können.

Diese in der Versorgungspraxis routinemäßig (nicht-interventionell) erhobener Parameter müssen eine ausreichende Risikofaktor- bzw. Schweregrad-Adjustierung der Verum- und Vergleichsgruppe für wesentliche Confounder ermöglichen. Wenn nicht beide Anforderungen

erfüllt werden können, ist eine AbD unbeachtet Methoden-wissenschaftlicher Betrachtungen nicht durchführbar.

Darüber hinaus ist eine sinnvolle Einbindung der Patienten bzw. Erziehungsberechtigten zur Erfassung der patient-reported outcomes (PROs) über digitale Tools zu gewährleisten. Hier sollten bei der Protokolldiskussion mögliche Übersetzungen mitbedacht werden, sonst ergibt sich ein Bias, weil je nach Indikation zahlreiche Patienten nicht eingeschlossen werden könnten.

### **Datenerhebung / Erhebungszeitpunkte**

Wir halten die Ausarbeitung eines strukturierten Behandlungsplans für zwingend notwendig. Dieser sollte in einem systematisch und methodisch ausreichend validem Ansatz entwickelt werden, z.B. in einer Kombination aus systematischer Literaturrecherche und anschließendem Delphiverfahren. Der strukturierte Behandlungsplan definiert den standard-of-care in der Regelversorgung inkl. der Erhebungszeitpunkte für systematische Registerdaten und patient-reported outcome measures (PROMs). Dies reduziert die Heterogenität der erhobenen Beobachtungsdaten und liefert Ergebnisdaten, die besser vergleichbar sind.

### **Qualität und Quantität der Daten / Sicherheitsdaten**

Bei der Auswahl geeigneter Register sollte ein Kriterienkatalog aufgebaut werden, der eine Echtzeitmessung systematischer Qualitätssicherungskriterien voraussetzt (z.B. sog. Performanceindikatoren zu Datenvollständigkeit und Datenqualität).

Der Verwendung des Begriffs Adverse event (AE) ist in der Regelversorgung nach Zulassung im Rahmen der AbD und anderer post-marketing-Studien nicht zielführend. Die Sammlung dieser Daten vergleichbar zu einer Phase-III-Studie führt zu einer starken und nicht vergleichbaren Datenverzerrung zwischen unterschiedlichen Anwendungszentren. Der Aufwand steht nicht im Verhältnis zur Datenqualität und ist v.a. auch nicht in der Regelversorgung abbildbar. Es sollten daher explizit Adverse events of special interest (AESI) festgelegt werden, die systematisch graduiert nach Schweregrad (CTCAE-Klassifikation) und nur ab CTCAE-Grad  $\geq 2$  erfasst werden. Bereits in der Packungsbeilage erfasste und erwartbare UAW sollten nur erfasst werden, wenn sie im Schweregrad über das erwartbare Maß hinausgehen. Wichtig ist v.a. die Sammlung von UAWs, bei denen eine mindestens mögliche Assoziation im Zusammenhang mit der Gabe des Arzneimittels besteht (Adverse Drug Reactions, ADR). Darüber hinaus sollten nur diejenigen AEs gesammelt werden, die in der Anwendung erfahrene Ärztinnen und Ärzte als neu, im Schweregrad außergewöhnlich oder im Muster als auffällig bewerten. Die Definition von sog. Negativlisten in Abstimmung zwischen Arzneimittelbehörde, Anwendern, pU und den erfassenden Registern hilft, um diejenigen Ereignisse zu definieren, die in jedem Fall nicht gemeldet werden sollen, z.B. banale Atemwegsinfekte bei Kindern im Winter, Unfälle ohne Hinweis auf Produktassoziation etc.

Wir plädieren für eine **Fokussierung der Datenerhebung** auf Parameter, die folgende Anforderungen kombiniert erfüllen:

- Die Daten sollten in der Versorgungspraxis routinemäßig (nicht-interventionell) in Verum- UND Vergleichstherapie-Gruppe erhebbar sein

- Die Daten sollten valide und reliabel sein für einen vordefiniert und in der Fachwelt akzeptiert Nutzen-relevanten Wirksamkeits- oder kritischen Sicherheitsendpunkt
- Es sollte ein Verzicht auf eine umfassende AE-Erfassung erfolgen, da diese außerhalb klinischer Prüfungen unrealistisch umfassend erfass- und dokumentierbar ist, und damit
  1. kritisch bias-gefährdet ist
  2. unvalide und wertarm bleibt

## **Laufzeit**

AbDs müssen in einer Laufzeit von insgesamt max. 5 Jahren (Rekrutierung, Beobachtung und Auswertung) Unterschiede oder Vergleichbarkeit von Verum- und Vergleichsgruppe bezüglich akzeptiertem Nutzen-relevanten Wirksamkeits- oder kritischen Sicherheitsendpunkt aufzeigen können.

## **Abschätzung der verfügbaren Fallzahlen (Kapitel 5.3)**

Bei Einbezug der verfügbaren Fallzahlen plädieren wir dafür, bereits im Vorfeld folgende Faktoren miteinzubeziehen:

- Soziale und kulturelle inclusive sprachliche Barrieren für den Patienteneinschluss (Compliance)
- Realistische Betrachtung des Anteils der an „AbD-fähigen“-Gesundheitseinrichtungen behandelten Patientenanteile an der identifizierten Gesamtkohorte mit Indikation für die Verum-Therapie
- Teilnehmer/-innen-Selektion durch Risikofaktor- bzw. Schweregrad-Adjustierung der Verum- und Vergleichsgruppe für wesentliche Confounder

## **Therapiewechsel (Kap. 5.4.2):**

Das Beispiel spinale Muskelatrophie zeigt, dass eine überbrückende Therapie (bridging Therapie), wenn auch nur wenige Tage oder Wochen mit anschließendem Therapiewchsel, z.B. auf Zolgensma, die Krankheitstrajektorien erheblich verändern kann. Die Gruppen mit oder ohne bridging Therapie können somit nicht 1:1 und ohne Confounderadjustierung miteinander verglichen werden. Die Festlegung der Interventions- und Comparatorgruppen sollten daher im Vorfeld der AbD nicht nur mit dem pU abgestimmt werden, sondern mit von der jeweiligen med. Fachgesellschaft bestimmten Expert/-innen. Hier muss v.a. die „Praxistauglichkeit“ der gewählten Interventions- und Comparatorstichproben geprüft werden.

### **A.1.10 Deutsche Gesellschaft für Nephrologie e. V. (DGfN)**

#### **Autorinnen und Autoren**

- Weinmann-Menke, Julia



**Vorläufiger Rapid Report Berichtnr: A25-13**  
**Titel: Wissenschaftliche Ausarbeitung**  
**ausgewählter Aspekte zur Generierung versorgungsnaher Daten**  
**und deren Auswertung zum Zwecke der Nutzenbewertung**  
**von Arzneimitteln nach § 35a SGB V**

**Stellungnahme der Deutschen Gesellschaft für Nephrologie  
(DGfN)**

Allgemeine Anmerkungen:

Die DGfN begrüßt, dass das IQWiG seine Methodik weiterentwickelt, und sich im Rahmen der Planung von AbD, intensiv mit Möglichkeiten praktischen Vorgehens zur Evidenzgenerierung in besonderen Umständen auseinandersetzt. Wie auf S. xxiii richtig beschrieben wird, ergibt sich je nach zu bewertender Therapie die Schwierigkeit, eine ausreichende Patientenzahl rekrutieren zu können.

**Auch in anderen Kontexten kann es schwierig sein, Studien ausreichend zu powern, beispielsweise dann, wenn klassische Endpunkte so selten sind, dass sehr lange Beobachtungsdauern nötig werden. Dies ist im Bereich chronischer Nierenkrankheiten (CKD) der Fall.**

Im Frühstadium verlaufen CKD über lange Zeit weitgehend asymptomatisch, sie sind bei frühzeitiger Diagnose an dieser Stelle aber noch gut behandelbar. Im Laufe der fortgeschrittenen CKD können sich Einschränkungen metabolischer Funktionen oder mangelnder Regulation des Flüssigkeits-, Säure/Basen- und Elektrolythaushalts laborchemisch manifestieren<sup>1</sup>.

Die Beeinträchtigung der Lebensqualität ist ein Phänomen der späten Stadien einer CKD<sup>2</sup>. Wenn erst klinische Symptome auftreten<sup>3 4</sup>, ist ein großer Teil der Nierenfunktion bereits irreversibel verloren, und es besteht die Notwendigkeit für eine Nierenersatztherapie (Dialyse oder Nieren-Transplantation)<sup>5</sup>. Eine Vielzahl der Patienten stirbt im Krankheitsverlauf bereits vorher an kardiovaskulären Ereignissen<sup>6</sup>. Daher ist das Ziel aller Behandlungen, die CKD frühzeitig zu erkennen und das Fortschreiten der CKD, den damit einhergehenden Verlust der Nierenfunktion und das Auftreten von Folge- und Begleiterkrankungen zu verlangsamen<sup>7</sup>.

**Das therapeutische Ziel möglichst langer Zeiten gesunden Überlebens über Jahre oder Jahrzehnte ist mit dem Design von klinischen Studien unter Verwendung von terminalen Endpunkten schwierig vereinbar.**

<sup>1</sup> Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. 2024.

<sup>2</sup> Webster AC, Nagler EV, Morton RL, Masson P. Chronic Kidney Disease. Lancet (London, England). 2017;389(10075):1238-52. Epub 2016/11/27.

<sup>3</sup> Masuda T, Nagata D. Recent advances in the management of secondary hypertension: chronic kidney disease. Hypertension research : official journal of the Japanese Society of Hypertension. 2020;43(9):869-75. Epub 2020/06/20.

<sup>4</sup> Shih HM, Wu CJ, Lin SL. Physiology and pathophysiology of renal erythropoietin-producing cells. Journal of the Formosan Medical Association = Taiwan yi zhi. 2018;117(11):955-63. Epub 2018/04/16.

<sup>5</sup> Webster AC, Nagler EV, Morton RL, Masson P. Chronic Kidney Disease. Lancet (London, England). 2017;389(10075):1238-52. Epub 2016/11/27.

<sup>6</sup> Lancet Diab & Endokrinol Volume 3, Issue 7p514-525 July 2015, Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data, Kunihiro Matsushita, MDa · Prof Josef Coresh, MDa ckdpc@jhmi.edu · Yingying Sang, MSA · Prof John Chalmers, MDB · Caroline Fox, MDC · Prof Eliseo Guallar, MDA· et al. doi: 10.1016/S2213-8587(15)00040-6.

<sup>7</sup> Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. 2024.



Klassische Endpunkte, klinische Symptome oder die „quality of life“ sind daher in den einer Behandlung zugänglichen früheren CKD-Stadien ungeeignet. **Surrogatparameter** wie der jährliche Abfall der geschätzten GFR (eGFR, estimated GFR) (sog. eGFR-slope) oder eine Albuminurie erlauben es, eine Veränderung der Nierenfunktion zu monitorieren, und können das Eintreten der klinischen Endpunkte zuverlässig vorhersagen. Die GFR ist hierbei der Marker für die Entgiftungsfunktion der Niere. Die Albuminurie ist ein früher Biomarker der gestörten Schrankenfunktion in den Glomeruli. Beide Parameter können in klinischen Studien stellvertretend für das Erreichen von MAKE (major kidney events), MACE (major cardiovascular events) oder Tod betrachtet werden<sup>8 9</sup>; durch ihre Verwendung kann bei gleichbleibender Power die Beobachtungsdauer erheblich reduziert werden<sup>10</sup>. Inzwischen liegen sowohl für die GFR als auch für die Albuminurie umfangreiche Validierungsstudien vor<sup>11 12 13 14</sup>, die den Kriterien des IQWiG zur Surrogatvalidierung entsprechen. Sowohl die FDA als auch die EMA haben auf die vorliegende Evidenz reagiert und akzeptieren beide Surrogat-Endpunkte anstelle von klassischen klinischen Endpunkten<sup>15 16 17 18 19</sup>.

#### Vorgeschlagene Änderungen:

**Die DGfN bittet daher das IQWiG, die Evidenz zu prüfen, und – auch im Hinblick auf eine zukünftige Aktualisierung des Methodenpapiers – die Möglichkeit zu prüfen, die eGFR-slope und die Albuminurie als Surrogatparameter bei entsprechend darauf abgestimmten Studiendesign in Nutzenbewertungsverfahren zu akzeptieren.**

<sup>8</sup> Hallan SI, Matsushita K, Sang Y, Mahmoodi BK, Black C, Ishani A, et al. Age and association of kidney measures with mortality and end-stage renal disease. *Jama*. 2012;308(22):2349-60. Epub 2012/11/01.

<sup>9</sup> Gansevoort RT, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. *Kidney international*. 2011;80(1):93-104. Epub 2011/02/04.

<sup>10</sup> Greene T, Teng CC, Inker LA, Redd A, Ying J, Woodward M, et al. Utility and validity of estimated GFR-based surrogate time-to-event end points in CKD: a simulation study. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2014;64(6):867-79. Epub 2014/12/03.

<sup>11</sup> Inker LA, Lambers Heerspink HJ, Mondal H, Schmid CH, Tighiouart H, Noubary F, et al. GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2014;64(6):848-59. Epub 2014/12/03.

<sup>12</sup> Coresh J, Turin TC, Matsushita K, Sang Y, Ballew SH, Appel LJ, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *Jama*. 2014;311(24):2518-31. Epub 2014/06/04.

<sup>13</sup> Lambers Heerspink HJ, Tighiouart H, Sang Y, Ballew S, Mondal H, Matsushita K, et al. GFR decline and subsequent risk of established kidney outcomes: a meta-analysis of 37 randomized controlled trials. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2014;64(6):860-6. Epub 2014/12/03.

<sup>14</sup> Levey AS, Inker LA, Matsushita K, Greene T, Willis K, Lewis E, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2014;64(6):821-35. Epub 2014/12/03.

<sup>15</sup> European Medicines Agency (EMA). Guideline on the clinical investigation of medicinal products to prevent development/slow progression of chronic renal insufficiency. 2016.

<sup>16</sup> Levey AS, Gansevoort RT, Coresh J, Inker LA, Heerspink HL, Grams ME, et al. Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD: A Scientific Workshop Sponsored by the National Kidney Foundation in Collaboration With the US Food and Drug Administration and European Medicines Agency. *American Journal of Kidney Diseases*. 2020;75(1):84-104.

<sup>17</sup> Levey AS, Inker LA, Matsushita K, Greene T, Willis K, Lewis E, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2014;64(6):821-35. Epub 2014/12/03.

<sup>18</sup> Thompson A, Smith K, Lawrence J. Change in Estimated GFR and Albuminuria as End Points in Clinical Trials: A Viewpoint From the FDA. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2020;75(1):4-5. Epub 2019/11/02.

<sup>19</sup> European Medicines Agency (EMA). Qualification Opinion for GFR slope as a Validated Surrogate Endpoint for RCT in CKD. 2023.

### **A.1.11 Deutsche Gesellschaft für Orthopädie und Unfallchirurgie e. V. (DGOU)**

#### **Autorinnen und Autoren**

- Gravius, Sascha

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Univ.-Prof. Dr. med. Sascha Gravius, MHBA
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU)
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

Für das Fachgebiet der Orthopädie und Unfallchirurgie, das sich intensiv mit interventionellen Therapien, insbesondere chirurgischen Eingriffen, medizinischen Geräten und Rehabilitationsmaßnahmen befasst, sind die im Bericht dargestellten methodischen Grundlagen zur Generierung und Auswertung versorgungsnaher Daten von herausragender Relevanz. Obwohl der Bericht primär auf die Nutzenbewertung von Arzneimitteln fokussiert ist, sind die diskutierten Konzepte, wie die *Target-Trial-Emulation*, der Umgang mit fehlenden Werten und die Notwendigkeit einer adäquaten Confounderkontrolle, grundsätzlich auf alle vergleichenden Studien ohne Randomisierung übertragbar, die in der Orthopädie und Unfallchirurgie zur Bewertung von Interventionen unerlässlich sind.

### **Spezifische Aspekte im Einzelnen:**

#### **5.1 Einleitende Anmerkungen zu den Anforderungen an vergleichende Studien ohne Randomisierung zum Zwecke der Nutzenbewertung (S. 5)**

##### **Anmerkung:**

Das Konzept der Target-Trial-Emulation ist ein fundamentaler Best-Practice-Ansatz zur Minimierung von Verzerrungen. Für Arzneimittel in der Orthopädie und Unfallchirurgie ist dieser Ansatz entscheidend, da hier traditionelle randomisierte kontrollierte Studien (RCTs) aus praktischen oder ethischen Gründen oft nicht durchführbar sind (z.B. bei seltenen Arzneimittelwirkungen oder bei heterogenen Patientenkolorten nach komplexen Verletzungen).

##### **Vorgeschlagene Änderung:**

Betonung, dass die Target-Trial-Emulation ein geeigneter Ansatz für die evidenzbasierte Bewertung von Arzneimitteln in der Orthopädie und Unfallchirurgie ist und eine detaillierte Studienplanung sowie umfassende, qualitativ hochwertige Daten erfordert, um die fehlende Randomisierung bestmöglich auszugleichen.

#### **5.2 Confounderidentifikation und -auswahl (S. 9)**

##### **Anmerkung:**

Die systematische Identifikation von Confoundern ist für Arzneimittel in der Orthopädie und Unfallchirurgie besonders komplex und ressourcenintensiv. Dies liegt an der Vielzahl patienten- und erkrankungsspezifischer Störgrößen wie Verletzungsschwere, Voroperationen, Begleiterkrankungen, Implantateinsatz oder Rehabilitationsmaßnahmen. Der Bericht selbst weist auf diese Ressourcenintensität hin.

Vorgeschlagene Änderung:

Hervorhebung, dass die umfassende und präspezifizierte Confounderidentifikation bei der Bewertung von Arzneimitteln in der Orthopädie und Unfallchirurgie aufgrund der großen Anzahl komplexer und heterogener potenzieller Störgrößen eine besonders hohe Ressourcenbindung erfordert, die jedoch für valide Ergebnisse unerlässlich ist.

## 5.5 Patientenberichtete Endpunkte (S. 33)

Anmerkung:

Patientenberichtete Endpunkte (PROs) sind für die Nutzenbewertung von Arzneimitteln in der Orthopädie und Unfallchirurgie von großer Bedeutung, da sie direkte patientenrelevante Auswirkungen auf Schmerzintensität, Funktionsfähigkeit und Lebensqualität abbilden. Die Empfehlung zur digitalen Erhebung von PROs (z.B. per App oder Patientenportal) ist hierbei besonders relevant, da sie eine niedrigschwellige, standardisierte und regelmäßige Datenerfassung ermöglicht und die Datenqualität verbessert.

Vorgeschlagene Änderung:

Explizite Nennung der hohen Relevanz der digitalen Erfassung von PROs für die Nutzenbewertung von Arzneimitteln in der Orthopädie und Unfallchirurgie, um patientenrelevante Effekte effizient zu erfassen und in die Versorgungsrealität zu integrieren. Das Feedback an Patienten und Behandler sollte als entscheidender Hebel zur Verbesserung der Versorgungsqualität erwähnt werden.

### 5.6.1 Maßnahmen zur Vermeidung fehlender Werte (S. 37)

Anmerkung:

Die Minimierung fehlender Werte ist entscheidend, insbesondere für die langfristige Bewertung von Arzneimitteln in der Orthopädie und Unfallchirurgie (z.B. bei Osteoporose oder nach Gelenkersatz), wo lange Beobachtungszeiträume notwendig sind. Eine dauerhaft verfügbare und funktionsfähige Dateninfrastruktur in Registern ist hierfür zentral, deren Aufbau und Unterhalt idealerweise durch die pharmazeutischen Unternehmer unterstützt werden sollten.

Vorgeschlagene Änderung:

Stärkere Betonung, dass der Aufbau und Unterhalt einer dauerhaft verfügbaren und funktionsfähigen Dateninfrastruktur in Registern eine zentrale Voraussetzung für die Qualität und Nutzbarkeit versorgungsnaher Daten sind. Dies ist insbesondere für langfristig zu bewertendem Arzneimittel in der Orthopädie und Unfallchirurgie unerlässlich und muss nachhaltig finanziert und personell ausgestattet werden, idealerweise durch eine breite Unterstützung verschiedener Stakeholder.

## 5.7 Propensity-Score-Analysen in Anwendungsgebieten mit kleinen Patientenkollektiven (S. 50)

Anmerkung Propensity-Score-Verfahren sind für bestimmte Arzneimittel in der Orthopädie und Unfallchirurgie (z.B. innovative Therapien für seltene Knochenerkrankungen) relevant, wo kleine Patientenkollektive vorliegen. Obwohl Simulationsstudien die Anwendbarkeit in kleinen Populationen zeigen, bestehen Unsicherheiten hinsichtlich der Interpretierbarkeit, insbesondere bei der typischerweise hohen Anzahl komplexer Confounder in diesem Fachgebiet.

Vorgeschlagene Änderung:

Aufnahme eines Hinweises darauf, dass die Anwendbarkeit und Interpretierbarkeit von Propensity-Score-Analysen für Arzneimittel in der Orthopädie und Unfallchirurgie besonders herausfordernd sein kann, da hier oft kleine Patientenkollektive auf eine Vielzahl komplexer patienten- und verfahrensbezogener Confounder treffen. Die Notwendigkeit weiterer kontextspezifischer Forschung sollte betont werden.

## 6 Fazit (S. 56)

Anmerkung:

Die im Rapid Report zusammengefassten methodischen Erkenntnisse und Empfehlungen (z.B. Target-Trial-Emulation, Confounderidentifikation, digitale PROs) sind uneingeschränkt übertragbar und von essenzieller Bedeutung für die robuste Nutzenbewertung von Arzneimitteln in allen medizinischen Fachgebieten, einschließlich der Orthopädie und Unfallchirurgie. Sie tragen maßgeblich dazu bei, bestehende Evidenzlücken in der Routineversorgung zu schließen.

Vorgeschlagene Änderung: Eine klarere Formulierung im Fazit, die die generelle Übertragbarkeit und die hohe Relevanz der dargestellten methodischen Konzepte auf die Nutzenbewertung aller Arzneimittel im deutschen Gesundheitswesen – inklusive derer in der Orthopädie und Unfallchirurgie – explizit betont, um die breitere Anwendbarkeit und den Wert des Reports hervorzuheben.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.*

## Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

### **A.1.12 Deutsche Krankenhausgesellschaft e. V. (DKG)**

#### **Autorinnen und Autoren**

- Aenderl, Isabel
- Schier, Maike
- Wenzel-Seifert, Katharina

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Wenzel-Seifert, Katharina, Dr.
Schier, Maike, Dr.
Aenderl, Isabel, Dr.
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Deutsche Krankenhausgesellschaft
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.*

### **Kapitel 5.3 Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen S. 12 bis 25**

#### Anmerkung

Der Gemeinsame Bundesausschuss (G-BA) kann bei bestimmten Arzneimitteln vom pharmazeutischen Unternehmer innerhalb einer angemessenen Frist die Vorlage anwendungsbegleitender Datenerhebungen und Auswertungen zum Zweck der Nutzenbewertung fordern. Dabei sieht die Verfahrensordnung des G-BA vor, dass bei der Beurteilung der Erforderlichkeit einer AbD, die Realisierbarkeit und Angemessenheit einer Datenerhebung berücksichtigt werden. Zur Vorbereitung eines Beschlusses nach § 35a Absatz 3b Satz 1 SGB V wird ein Konzept für eine anwendungsbegleitende Datenerhebung und von Auswertungen erstellt, das insbesondere Anforderungen an die Art, die Dauer und den Umfang der Datenerhebung enthalten und Grundlage des vom pharmazeutischen Unternehmer zu erstellenden statistischen Analyseplans und Studienprotokolls darstellen soll.

Mit Blick auf den Umfang der AbD sieht das IQWiG zwei Vorgehensweisen vor: die orientierende Fallzahlschätzung sowie bei unzureichender Evidenzlage (insbesondere im Vergleich zum therapeutischen Standard), die orientierende Betrachtung von Fallzahlszenarien. Grundlage beider Vorgehensweisen ist dabei eine verschobene Nullhypothese. Demnach kann aus Sicht des IQWiG bei einer nicht-randomisierten Studie (aufgrund potentiell unbekannter Confounder) erst ab einer bestimmten Effektstärke eine positive oder negative Aussage zum Nutzen oder Schaden einer Intervention abgeleitet werden. Dies ist laut IQWiG erst dann möglich, wenn das Konfidenzintervall für den Effekt oberhalb bzw. unterhalb einer zu definierenden Schwelle liegt. Und diese Schwelle wird vom IQWiG auf 0,5 bzw. 2 festgelegt.

Bei der orientierenden Betrachtung von Fallzahlszenarien wird vom IQWiG mittels der geschätzten potentiell zur Verfügung stehenden Patientenzahlen, des Signifikanzniveaus sowie der Power die „aufdeckbare“ Effektgröße berechnet. Als aufdeckbar gilt eine Effektgröße dann, wenn das im Rahmen

einer AbD geschätzte 95 %-Konfidenzintervall mit 80 % Power bei gegebener Patientenzahl vollständig oberhalb ( $RR_0 = 2$ ) bzw. unterhalb ( $RR_0 = 0,5$ ) der verschobenen Hypothesengrenze liegt (Seite 21 und 22). Die Patientenzahlen werden dabei anhand der geschätzten potentiell zur Verfügung stehenden Patientenzahlen im zu bewertenden Anwendungsbereich festgelegt.

Das Ergebnis der Fallzahlszenarien ist die aufdeckbare Effektgröße, diese wird vom IQWiG nicht weiter in den gegebenen klinischen Kontext eingeordnet. Da sie durch die verschobene Nullhypothese im Bereich von unter 0,5 bzw. über 2 liegt, wäre es auch wichtig zu wissen, wie wahrscheinlich im gegebenen Kontext das Auftreten der durch dieses Vorgehen aufdeckbaren Effektgröße ist. Auch dies bleibt im methodischen Vorgehen des IQWiG aus.

Als Beispiel dafür, wie sich diese Herangehensweise auf ein AbD-Konzept auswirken kann, zeigt sich gut anhand des AbD-Konzepts A25-06 (zu Loncastuximab tesirin im Anwendungsbereich DLBCL) in dem für das Gesamtüberleben aufdeckbare Effektgrößen mit einer Spanne beim Hazard Ratio zwischen 0,32 und 0,40 berechnet wurden. Im Konzept konstatiert das IQWiG im Fazit lediglich, dass *diese [orientierende Betrachtung von Fallzahlszenarien] die Effektgrößen für eine AbD zum Wirkstoff Loncastuximab tesirin aufzeigen, die auf Basis der verfügbaren Patientenzahlen und unter Berücksichtigung der verschobenen Nullhypothese aufdeckbar sind (AbD Konzept A25-06 Loncastuximab Seite 40).*“ Eine Einordnung in den klinischen Kontext und eine Bewertung, wie realistisch solche Effektschätzer sind, werden im Konzept nicht vorgenommen.

Im Anhang H „Datengrundlage für die Extrapolation zum Endpunkt Gesamtüberleben“ des hier beispielhaft angeführten Konzepts werden vom IQWiG Daten aus Studien zu Loncastuximab und der Vergleichstherapie herangezogen, um die Dauer der AbD und die Ereignisanteile der Vergleichstherapien abschätzen zu können. Zieht man diese Studienergebnisse heran, um abzuschätzen, wie realistisch es ist, dass Effektschätzer zwischen 0,32 und 0,40 erreicht werden können, sieht man, dass die tatsächlich erwartbaren Effekte diese Größenordnung bei weitem verfehlten. In der Tabelle werden für das mediane Gesamtüberleben für Loncastuximab 9,5 Monate angegeben, für die Komparatoren 15,5 bzw. 11,2 Monate. Diese Studienergebnisse wurden zwar vom IQWiG nicht für eine orientierende Fallzahlschätzung herangezogen, da es Unklarheiten gab, wie hoch der jeweilige Anteil der Patientinnen und Patienten war, die für eine CAR-T-Zelltherapie oder eine Stammzelltransplantation nicht infrage kommen und ob sich dieser Anteil zwischen den Studien zu Loncastuximab tesirin und den Vergleichstherapien unterscheidet. Auch wenn es in den Studien also mögliche Unterschiede in den Patientencharakteristika gibt, scheint es insgesamt jedoch sehr unwahrscheinlich zu sein, dass in der AbD Effektschätzer erreichbar sind, die dazu führen, dass unter Berücksichtigung der verschobenen Nullhypothese ein Zusatznutzen abgeleitet werden kann.

Zudem verdeutlicht die orientierende Betrachtung von Fallzahlszenarien unter Berücksichtigung der verschobenen Nullhypothese im Konzept zu Loncastuximab tesirin, dass mit Hilfe einer AbD nur in einem sehr engen Fenster von < 0,5 bis > 0,2 (hier: Hazard Ratio zum Gesamtüberleben) ein Zusatznutzen abgeleitet werden könnte. Ab einem beobachteten Effekt von 0,2 kann laut IQWiG Methodenpapier ein dramatischer Effekt abgeleitet werden (IQWiG Methodenpapier 7.0, Kapitel 3.2.2 Seite 60). Für die Ableitung eines dramatischen Effektes ist es aus Sicht der DKG nicht notwendig, eine aufwendige AbD durchzuführen. Es stellt sich zudem auch die Frage, welche Patientenzahlen erforderlich wären, um eine Hazard Ratio < 0,5 und > 0,4 mit ausreichender Power nachzuweisen. Dieses Intervall der HR wäre ja noch mit der verschobenen Nullhypothese vereinbar.

Vor dem Hintergrund der fehlenden Beantwortung der Frage, ob eine AbD realisierbar ist und wie hoch der tatsächliche Umfang in einer adäquaten Dauer wäre, sieht es die DKG als notwendig an, regelhaft im Rahmen der Erstellung von Konzepten eine orientierende Fallzahlschätzung sowie entsprechende Angaben zur Dauer der AbD vorzulegen.

Dass eine Fallzahlschätzung bei AbDs herausfordernd ist (i.d.R. keine vergleichenden Daten, Heranziehung von Daten aus verschiedenen Studien, geringe Abweichung in der Patientenpopulation) bleibt unbestritten. Dass eine Fallzahlplanung für eine randomisierte kontrollierte Studie dagegen immer auf validere oder vergleichende Daten, zurückgreifen kann (siehe Ausführungen des IQWiG auf Seite 13), halten wir nicht in allen Fällen für zutreffend. Auch in der Planung von Phase-III-Studien kann es Untersicherheiten in der Fallzahlplanung geben, die entsprechend in die Berechnung einfließen müssen. Dennoch müssen im Rahmen der Studienplanung schon aus ethischen Gründen Fallzahlplanungen vorliegen, um nicht unnötig viele oder auch zu wenige Probanden einzuschließen. Andernfalls ist der Einschluss von Patientinnen und Patienten in eine Studie bzw. auch in eine AbD nicht zu rechtfertigen. Ziel jeder Studie, egal ob AbD oder nicht, muss die Generierung statistisch belastbarer Ergebnisse sein, was ohne Fallzahlplanung ad absurdum geführt wird.

Zum Zeitpunkt der Forderung der AbD kann eine Fallzahlschätzung aufgrund der bereits genannten Unsicherheiten natürlich nur einen orientierenden Charakter haben, was vom IQWiG auf Seite 13 auch aus unserer Sicht sachgerecht ausgeführt wird. Im Rahmen dieser orientierenden Fallzahlschätzung sollten, wie auf Seite 15 dargelegt, unterschiedliche Verteilungsverhältnisse dargestellt werden, da der therapeutische Stellenwert des neuen Arzneimittels auch im Kontext der bisherigen Vergleichstherapie nur schwer einzuschätzen ist. Alle mit der orientierenden Fallzahlschätzung verbundenen Unsicherheiten sollten entsprechend adressiert und eine finale Fallzahlplanung im Verlauf der AbD gefordert werden.

Im Zuge der verpflichtenden Erstellung des statistischen Analyseplans und Studienprotokolls ist die orientierende Fallzahlschätzung zudem für den pharmazeutischen Unternehmer als Grundlage für die weitere Planung notwendig. Unbenommen davon ist, dass aufgrund des orientierenden Charakters in der Folge (z. B. zum Zeitpunkt von Zwischenanalysen) eine finale Fallzahlplanung erfolgt, nämlich dann, wenn es erste Daten zu den beobachteten Effekten vorliegen.

Die Beurteilung der Realisierbarkeit und der Angemessenheit und insbesondere auch der Zweck einer AbD sind für die DKG zentrale Aspekte für die Entscheidung, ob eine AbD gefordert werden soll. Dazu bedarf es einer Fallzahlschätzung, die sich an einem realistisch annehmbaren Effekt orientiert. Erst dann kann eingeschätzt werden, ob es angemessen und sinnvoll ist, eine AbD durchzuführen. Dabei gilt es den Zweck gemäß § 35a Abs. 3 SGB V, nämlich eine erneute Nutzenbewertung, nicht aus den Augen zu verlieren. Obwohl nicht Gegenstand der aktuellen Beauftragung, bleibt die Frage zu klären, wie die AbD geeignete Daten generieren soll, die für eine erneute Nutzenbewertung herangezogen können. Gemäß § 5 Abs 7 AM-Nutzen-V ist das Ausmaß des Zusatznutzens und die therapeutische Bedeutung des Zusatznutzens unter Berücksichtigung des Schweregrades der Erkrankung gegenüber dem Nutzen der zweckmäßigen Vergleichstherapie in sechs Stufen (erheblich, beträchtlich, gering, nicht quantifizierbar, kein und geringerer Zusatznutzen) zu quantifizieren. Bei einer AbD ist aufgrund der verschobenen Nullhypothese in der Regel lediglich ein Bereich detektierbar, der in den bisherigen AMNOG-Verfahren im Ausmaß mindestens einem beträchtlichen Zusatznutzen entspräche. Eine

Quantifizierung des Ausmaßes des Zusatznutzens in sechs Stufen wie in der AM Nutzen-V vorgesehen, ist somit schwer vorstellbar.

### Vorgeschlagene Änderung

Streichung der Ausführungen zu orientierender Betrachtung der Fallzahlszenarien

## **Kapitel 5.2 Confounderidentifikation S. 9-12)**

### Anmerkung

Die adäquate Identifikation und Auswahl von Confoundern spielt eine zentrale Rolle bei der Planung nicht randomisierter vergleichender Studien, in denen eine Adjustierung für potentielle Confounder erfolgen soll. Das IQWiG hat dazu ein Arbeitspapier (GA23-02, Version 1.0) erstellt, in dem einer Publikation von Pufulete 2022 (Pufulete M et al., Confounders and co-interventions identified in non-randomized studies of interventions. J Clin Epidemiol. 2022; 148:115–23) folgend eine systematische Confounderidentifikation in der Indikation schubförmig remittierende multiple Sklerose (RRMS) an einem Vergleich zweier medikamentöser Therapien durchgeführt wurde. Ziel dieses Projektes war es, die grundsätzliche Machbarkeit des von Pufulete 2022 beschriebenen Vorgehens einzuschätzen. Im vorliegenden Rapid Report beschreibt das IQWiG kurz ein 3-schrittiges Verfahren aus Confounderidentifikation mittels Extraktion aller potentiellen Confounder durch systematische Literaturrecherche und Experteninterviews (1. Schritt), inhaltlicher Zusammenführung der Confounder-Variablen (2. Schritt) und Bewertung der Relevanz der Confounder durch klinische Expertinnen und Experten (3. Schritt). Da es sich bei der RRMS um eine Indikation handelt, für die es einen sehr umfangreichen Evidenzkörper gibt, war es möglich, sich in dem Projekt bei den Datenquellen auf RCTs und Kohortenstudien zu beschränken. Im ersten sehr aufwendigen Schritt wurden 160 potentielle Confounder identifiziert. Basierend auf der Fragestellung konnten 28 Variablen ausgeschlossen werden. Durch Zusammenführung konnte die Zahl der potentiellen Confounder letztendlich auf 28 potentielle Confounder reduziert werden.

Im Rapid Report stellt das IQWiG fest, dass das im Arbeitspapier GA23-02 beschriebene Vorgehen der Confounderidentifikation und -auswahl zwar durchführbar, aber auch sehr aufwendig sei. Nach Angaben der mit der Durchführung des Projektes beauftragten externen Experten wurden für die Identifizierung der 160 Confounder 62 Personentage (à 8 Arbeitsstunden) benötigt. In einem nächsten Schritt soll daher ergänzend zum Arbeitspapier und außerhalb des vorliegenden Rapid Reports geprüft werden, ob weitere Anpassungen möglich sind, um die Machbarkeit durch Aufwandreduktion zu erhöhen.

### Vorgeschlagene Änderung

Wir stimmen dem IQWiG zu, dass es notwendig ist, den mit der Durchführung einer Confounderidentifikation und -auswahl verbundenen Arbeitsaufwand zu reduzieren. Darüber hinaus ist es aus unserer Sicht auch notwendig, eine weitere Reduktion der Anzahl der letztendlich ausgewählten potentiellen Confounder anzustreben. Für eine Adjustierung im Rahmen dafür geeigneter statistischer Verfahren (z.B. Propensity Score Matching) erscheint die Zahl von 28 Variablen sehr hoch zu sein. Es fällt zudem auf, dass die Qualität der für die Confounderidentifikation herangezogenen Studien in dem vom IQWiG herangezogenen Verfahren nicht bewertet und berücksichtigt wird. Im Arbeitspapier wurde auf eine Indikation zurückgegriffen, für die es viele RCTs und Kohortenstudien gibt. In den Anwendungsgebieten, für die vom G-BA AbDs gefordert werden, ist das meist nicht der Fall und es müssen Studien niedriger Evidenz, herangezogen werden. Aus unserer Sicht sollte die Qualität der Quelle, aus der Confounder extrahiert werden, bei der Bewertung der Relevanz mit einbezogen werden.

## Kapitel 5.5 Patientenberichtete Endpunkte S. 33-37

### Anmerkung

Das IQWiG führt aus, dass auch für die Erhebung von Daten zu patientenberichteten Endpunkten (PROs) in einer AbD, Qualitätsanforderungen erfüllt werden müssen, damit interpretierbare Ergebnisse geliefert werden, die für eine Nutzenbewertung verwertbar sind. Die Datenerhebung ist dadurch erschwert, dass sie im Kontext der Versorgungspraxis erfolgen muss, der sich deutlich von den kontrollierten Bedingungen in einer klinischen Studie unterscheiden. Das betrifft insbesondere unregelmäßige Visiten auf unterschiedlichen Versorgungsebenen (Hausarzt, Versorgung in Zentren). Darüber hinaus ist in vielen Registern, die ansonsten für AbDs infrage kämen, keine Datenerhebung integriert und muss ggf. erst aufgebaut werden.

Das IQWiG schlägt, vor, dass wie bei der AbD zu Brexucabtagene Autoleucel in der Indikation rezidiviertes oder refraktäres Mantelzell-Lymphom ein externes Institut mit der Erhebung der PRO-Daten beauftragt wird, wenn die Erhebung der PRO-Daten nicht direkt über Register erfolgen kann. Dadurch entfällt die Kopplung der Erhebung dieser Daten mit regulären Arztvisiten und der damit verbundene Aufwand der Register können geringer gehalten werden. Die Befragungen könnten durch die postalische Versendung der Fragebögen oder digital durchgeführt werden. Wir sehen es trotzdem insbesondere bei Patienten und Patientinnen des Vergleichsarmes, die weniger motiviert sein könnten, oder bei Patientinnen und Patienten, die durch Symptomatik oder unerwünschte Nebenwirkungen in ihrem Allgemeinbefinden stark beeinträchtigt sind, als schwieriger an, solche Daten auf diesem Wege möglichst lückenlos zu erfassen, als wenn die Befragung im zeitlichen Zusammenhang mit dem gewohnten Arztkontakt steht. Die Verwertbarkeit von PRO-Daten ist daran gebunden, dass die Rücklaufquoten über 70% liegen. Dieses Qualitätskriterium ist unter den erschwerten Bedingungen einer AbD nur schwer zu erfüllen.

Das IQWiG beschreibt darüber hinaus, dass für die Erhebungszeitpunkte angemessene Toleranzfenster, die nicht aneinandergrenzen, präspezifiziert werden sollen (Seite 35). Es könnte zudem sinnvoll sein, asymmetrische Toleranzfenster zu wählen, damit die Erhebungszeitpunkte mit den

jeweiligen Therapiephasen übereinstimmen. An dieser Stelle adressiert das IQWiG ein Problem, das auch bei der Bewertung von PRO-Daten, die in RCTs erhoben wurden, und im Rahmen des AMNOG-Verfahrens für die Nutzenbewertung eingereicht wurden, wiederholt aufgetreten ist. Problemen traten auf, wenn zumeist im Interventionsarm komplexe Therapieregime wie bei der Behandlung mit CAR-T-Zellen oder mit Gentherapien durchgeführt wurden und mit Standardtherapien, die wie zum Beispiel Chemo-Immuntherapien völlig anderen zeitlich strukturierten Therapieschemata folgen, verglichen werden mussten. In solchen in Rahmen von AbDs nicht seltenen Fallkonstellationen ist es aus unserer Sicht äußerst schwierig sinnvolle Erhebungszeitpunkte festzulegen, die nicht nur die unterschiedlichen Längen der Therapiephasen innerhalb eines Armes, sondern auch zwischen den Armen und die mit diesen Therapiephasen verbundenen Belastungen der Patientinnen und Patienten berücksichtigt. In den bislang durchgeführten Nutzenbewertungen von Therapieregimen mit CAR-T-Zellen auf Basis von RCTs, für die PRO-Daten erhoben wurden, waren diese Daten bedauerlicherweise aus den genannten Gründen nicht verwertbar. Zudem gab es im Rahmen von onkologischen AMNOG-Verfahren, in denen deutlich weniger komplexe Therapieregime miteinander verglichen wurden, dann Probleme, wenn in den Studienarmen Chemo(immun)Therapien mit unterschiedlicher Zykluslänge eingesetzt wurden. Auch in solchen Fällen war aus Sicht des IQWiG eine genaue Abstimmung der Erhebungszeitpunkte mit den Zeitpunkten der Therapiephasen erforderlich, damit die Belastung durch die Behandlung im Zyklusverlauf in den Studienarmen nicht ungleich abgebildet wird. Das IQWiG sah die Vergleichbarkeit als nicht gegeben an, wenn die Erhebung nicht in beiden Studienarmen zu Zeitpunkten vergleichbarer Belastung (z. B. vor Beginn des nächsten Zyklus) erfolgte. In solchen Fällen wurden die Daten nicht herangezogen (z. B. im AMNOG-Verfahren zu Pembrolizumab in Kombination mit Axitinib im Anwendungsgebiet Nierenzellkarzinom, Erstlinie mit Beschluss vom 14. Mai 2020). Unter den Bedingungen einer AbD, in der mit Toleranzfenstern gearbeitet werden muss, halten wir solche Qualitätsanforderungen für nicht umsetzbar.

### **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

### **A.1.13 Ecker + Ecker GmbH**

#### **Autorinnen und Autoren**

- Ahrens, Lucas
- Beuermann, Janik
- Leismann, Janine
- Oldeland, Jens

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Leismann, Janine
Ahrens, Lucas; Dr.
Oldeland, Jens; Dr.
Beuermann, Janik
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Ecker + Ecker GmbH
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

Die Ecker + Ecker GmbH ist als Unternehmensberatung auf das Gesundheitswesen spezialisiert. Als renommierte Fachberatung bietet sie seit mehr als 13 Jahren maßgeschneiderte Lösungen für den Market Access von Arzneimitteln und Medizinprodukten. Schwerpunkte des Leistungsspektrums in Deutschland sind die frühe Nutzenbewertung und Preisverhandlungen im Rahmen des AMNOG, sowie Preis- und Festbetragsanalysen im Produktlebenszyklus. Zusammen mit der 2023 gegründeten Schwesterfirma Accessus Health GmbH unterstützt Ecker + Ecker länderübergreifend zur europäischen Nutzenbewertung (EU-HTA).

Vor über 5 Jahren forderte der G-BA die erste anwendungsbegleitende Datenerhebung (AbD) für den Wirkstoff Onasemnogen-Abeparvovec. Ziel war und bleibt bei allen anschließenden Forderungen das Schließen von Evidenzlücken nach dem Markteintritt eines Wirkstoffs. Im Laufe dieser Zeit wurden die methodischen Probleme der AbD deutlich sichtbar, zum Beispiel die oft fehlende Fallzahlabschätzung oder der unklare Umgang mit – im Rahmen der AbD naturgemäß vorkommenden – sehr kleinen Patientenkollektive.

Insofern begrüßt die Ecker + Ecker GmbH das Bemühen von G-BA und IQWiG sehr, die Limitationen der AbD zu adressieren und methodische Ansätze zu identifizieren, die diese Schwierigkeiten potenziell lindern können. Gerade die Beschreibung zum methodischen Umgang mit fehlenden Werten im Kontext von Propensity-Score-Analysen stufen wir als hilfreich ein, um die AbD so auszugestalten, dass sie die eigentliche Zielsetzung erfüllen kann: die Generierung belastbarer Evidenz.

Darüber hinaus werden im Folgenden einige spezifische Anpassungen vorgeschlagen, um die Evidenzgenerierung im Rahmen der AbD weiter zu festigen.

Kapitel / Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.3.2 (S. 20)	<u>Anmerkung</u> Das IQWiG schlägt als Alternative zu einer Fallzahlplanung eine „orientieren Betrachtung von Fallzahlszenarien“ vor. Dabei wird „anders als bei der orientierenden Fallzahlabschätzung [...] die im Anwendungsgebiet zur Verfügung stehende Anzahl von Patientinnen und Patienten als Grundlage genommen, um die aufdeckbaren Effektgrößen zu ermitteln.“

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Dieser Ansatz ist aus Sicht des IQWiG verständlich, birgt aber erhebliche Gefahren: Nur weil eine aufdeckbare Effektgröße berechnet wurde, bedeutet dies nicht, dass dieser Effekt in der Realität auch beobachtet werden kann. In dem vom IQWiG genannten Beispiel auf S. 21 fehlt jegliche Einordnung der ermittelten Kenngröße – dies ist aber absolut relevant, um die Erfolgsaussichten und damit die Sinnhaftigkeit einer AbD beurteilen zu können.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Bei denen durch das IQWiG im Rahmen der AbD-Konzepterstellung ermittelten Fallzahlszenarien sollte immer auch eine Einordnung zur Größenordnung der angenommenen Patientenzahlen und der ermittelten aufdeckbaren Effekte vorgenommen werden. Dabei sollte der Fokus darauf liegen, ob die beschriebenen Effekte im Rahmen einer AbD realistisch zu beobachten sind.</p>
5.3.3 (S. 24)	<p><b>Anmerkung</b></p> <p>Das IQWiG schlägt vor, dass „der pU des zu bewertenden Arzneimittels die (finale) Fallzahlplanung der AbD zum Zeitpunkt der Zwischenanalysen auf Basis der dann vorhandenen vorläufigen Ergebnisse vornehmen“ sollte. Dieses Vorgehen sei grundsätzlich geeignet, „da die Schätzung der notwendigen Fallzahl bei versorgungsnahen Datenerhebungen – in Abgrenzung zu klinischen Studien – nicht mit Hinblick auf eine aktive Rekrutierung erfolgt.“</p> <p>Dieser Aussage stimmen wir nicht zu. Die Machbarkeit einer ausreichenden Rekrutierung von Patienten für die AbD und damit eine solide Fallzahlplanung muss ein zentraler Bestandteil des IQWiG-Konzepts im Rahmen der Forderung einer AbD durch den G-BA sein. Die AbD wird naturgemäß oft in Indikationen gefordert, die durch sehr kleine Patientenpopulationen definiert werden. In solchen Fällen eine AbD zu starten ohne belastbar abschätzen zu können, ob die Rekrutierung einer ausreichenden Patientenzahl überhaupt gelingen kann, ist fahrlässig.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Eine Fallzahlplanung darf nicht erst im Rahmen einer Zwischenanalyse einer laufenden AbD erfolgen. Vielmehr muss eine belastbare Fallzahlplanung Teil der Entscheidungsfindung über die Forderung einer AbD sein, also zu einem viel früheren Zeitpunkt vorliegen als vom IQWiG vorgeschlagen. Sollte eine Fallzahlplanung vor Start der AbD nicht durchführbar sein oder aufzeigen, dass eine unrealistisch hohe Fallzahl benötigt würde, um einen notwendigen Effekt nachzuweisen, sollte geprüft werden, inwiefern dies nicht ein negativer Indikator für die Machbarkeit einer AbD ist, anstatt die Entscheidung über eine Machbarkeit in die schon laufende AbD zu verlagern.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

## **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

#### **A.1.14 IGES Institut GmbH**

##### **Autorinnen und Autoren**

- Berkemeier, Fabian
- Sussmann, Stephanie

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> Bitte nutzen Sie pro Person 1 Zeile.
Berkemeier, Fabian; Geschäftsführer
Sussmann, Stephanie; Bereichsleiterin Value Strategy & RWE
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: IGES Institut GmbH
<input type="checkbox"/> als Privatperson(en)

***Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.***

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

IGES begrüßt die Auseinandersetzung des IQWiG mit den Erfahrungen aus der AbD im Rahmen des neuen vorläufigen Rapid Reports und nimmt als erfahrener Dienstleister mit langjähriger Expertise auf dem Gebiet nachfolgend Stellung.

Es wird positiv angemerkt, dass dem von IGES eingebrachten Prevalent New User Design in der Hämophilie seitens des IQWiG gefolgt wird und das genannte Design in Zusammenhang mit der Wahl des Indexdatums unterstützt wird.

Das übergeordnete Ziel der anwendungsbegleitenden Datenerhebung (AbD) ist die Quantifizierung eines Zusatznutzens. Im Rapid Report wird jedoch nicht ausreichend betont, dass die methodischen Anforderungen diesem Hauptzweck dienen sollten. Die Aussage des IQWiG, dass „[...] jedes Ergebnis einer AbD einen Erkenntnisgewinn dar[stellt] [...] losgelöst davon, ob sich [...] ein Vorteil des zu bewertenden Arzneimittels ableiten lässt“ (S. 14) sollte dabei im Kontext der bisher angewendeten Effektschwellen zur Anerkennung eines quantifizierbaren Zusatznutzens diskutiert werden (1). Eine Auseinandersetzung mit den Anforderungen an den Nachweis eines quantifizierbaren Zusatznutzens fehlt im Rapid Report weitgehend, obwohl die bisher zur Anwendung kommenden Schwellen ein zentraler Grund für lange Laufzeiten zur Patientenrekrutierung oder einer Unmöglichkeit von AbD darstellen, was letztlich einer Evidenzgenerierung entgegensteht. Die Aussage, wonach jedes Ergebnis einer AbD einen Erkenntnisgewinn dar[stelle], ist daher zumindest in Fällen, in denen die aktuell zur Anwendung kommenden Effektschwellen der Umsetzbarkeit einer AbD entgegenstehen, im Widerspruch zum grundlegenden Ziel der AbD im Rahmen der Nutzenbewertung. Der eigentliche Zweck einer AbD ist es, die für die Nutzenbewertung erforderliche Evidenz zu generieren, die bei der Zulassung noch nicht vorlag.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.*

**Abschnitt 5.3 Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen. S. 12 ff.**

## Anmerkung

Ein zentraler Punkt für die Realisierbarkeit einer AbD und deren mittelfristige Akzeptanz stellen die Anforderungen an den Nachweis eines quantifizierbaren Zusatznutzens dar. Die derzeit geforderten Effektschwellen, d. h. die Anforderungen an das Ausmaß und die Sicherheit des relativen Behandlungseffekts, stellen dabei eine höchst anspruchsvolle Hürde in der Konzeption und Durchführung dar.

Die AbD verkörpert einen Mittelweg zwischen nicht adjustierten indirekten Vergleichen – also der „naiven“ Verwendung einzelner Studienarme unterschiedlicher Studien für einen Vergleich – und dem Goldstandard der RCT. Die aktuell geforderten Effektschwellen für AbD werden diesem Charakter einer AbD als Mittelweg zwischen nicht adjustierten indirekten Vergleichen und einer RCT bisher nicht gerecht.

Bei nicht adjustierten indirekten Vergleichen im Rahmen der Nutzenbewertung findet bisher das Konzept eines dramatischen Effekts Anwendung. Das IQWiG sieht dabei vor, dass ein Effekt in nicht adjustierten indirekten Vergleichen als dramatisch angesehen werden kann, wenn er auf dem Niveau von 1% signifikant ist und als relatives Risiko (RR) ausgedrückt den Wert 10 übersteigt (2). Die Herleitung dieser Grenze basiert auf der Operationalisierung eines dramatischen Effekts von Glasziou et al. (2007), wonach ab einem beobachteten RR von 5 bis 10 der Therapieeffekt nicht mehr plausibel allein durch Störgrößeneinflüsse erklärt werden kann (3). Die Schwellen beziehen sich dabei sowohl im Methodenpapier des IQWiG als auch in der zugrundeliegenden Literatur auf den Effektschätzer und dessen Signifikanzniveau. Vereinfacht ausgedrückt wird daher bei nicht adjustierten indirekten Vergleichen in der Nutzenbewertung bisher für den Nachweis eines Zusatznutzens gefordert, dass das bewertete Arzneimittel mit 99-prozentiger Sicherheit einen Vorteil hat und dieser im wahrscheinlichsten Fall mindestens einen Faktor zwischen 5 und 10 hat.

Bei RCTs gelten ebenso Schwellenwerte zur Einordnung des Zusatznutzens. Für Endpunkte der Dimensionen Mortalität, Lebensqualität sowie schwere Symptome und Nebenwirkungen sieht das IQWiG dabei vor, dass die Grenzen des 95%-Konfidenzintervalls den Nulleffekt nicht einschließen dürfen. Vereinfacht ausgedrückt wird daher bei RCTs in der Nutzenbewertung für den Nachweis eines Zusatznutzens gefordert, dass das bewertete Arzneimittel mit 95-prozentiger Sicherheit einen Vorteil hat, wie groß dieser Vorteil im wahrscheinlichsten Fall ist, spielt dagegen keine Rolle.

Für AbD wird derzeit zum Nachweis eines quantifizierbaren Zusatznutzens das Testen einer verschobenen Nullhypothese gefordert (2), wonach eine Schwelle des RR von 2 bis 5 vorgesehen ist. Diese wird explizit aus den Schwellenwerten des dramatischen Effekts hergeleitet, wobei aufgrund der hohen Qualitätsanforderungen einer AbD an Datenerhebung und Ergebnisadjustierung eine Herabsetzung des Schwellenwerts von 5–10 auf 2–5 für das RR

erfolgt. Anders als beim „dramatischen Effekt“, wo die IQWiG-Methoden eine Anwendung der Schwellenwerte auf den Effektschätzer vorsehen (d. h. die Mitte des Konfidenzintervalls), fordert das IQWiG für die AbD eine Anwendung der Schwellenwerte auf die Grenzen des 95%-Konfidenzintervalls, ohne im Rahmen der Herleitung näher auf diesen hochrelevanten Unterschied einzugehen und diesen zu begründen.

Eine Simulationsstudie von Wilken und Berkemeier (2024) zeigt dabei, dass die bisher zur Anwendung kommenden Effektschwellen für AbD im Bereich seltener Erkrankungen fast ausschließlich erreicht werden, wenn gleichzeitig auch das Kriterium eines „dramatischen Effekts“ erfüllt ist (4). Entsprechend findet aktuell ein Kriterium Anwendung, das in seiner Erreichbarkeit ähnlich strikt zu bewerten ist als jenes, das seit Jahren für komplett nicht adjustierte Vergleiche zur Anwendung kommt. Dies erscheint angesichts der umfangreichen Anforderungen zur Sicherstellung der Datenqualität und Adjustierung von Confoundern nicht zielführend und beeinträchtigt letztlich die Generierung von Evidenz für die Nutzenbewertung. Aufgrund der bisher verwendeten Effektschwellen bei AbD und der oftmals sehr hohen resultierenden Schätzungen für die benötigten Fallzahlen ist die Durchführung von AbD teilweise unmöglich, in anderen Fällen zumindest höchst fraglich und erfordert selbst bei angenommener Machbarkeit häufig viele Jahre.

Zur Adressierung des Ziels einer möglichst guten Evidenzgenerierung für die Nutzenbewertung in vertretbaren Zeiträumen erscheint es IGES daher zielführend, mögliche Alternativen zu den bisher Anwendung findenden Effektschwellen wissenschaftlich zu untersuchen. Die Anwendung von Effektschwellen, welche tatsächlich dem Charakter einer AbD als „Mittelweg“ zwischen dem Goldstandard RCT einerseits und naiven Vergleichen andererseits gerecht werden, würde die Durchführbarkeit von AbD stark verbessern und damit die Generierung von Evidenz oftmals überhaupt erst ermöglichen und/oder erheblich beschleunigen.

## Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Versorgungsnahe Daten zum Zwecke der Nutzenbewertung - Rapid Report: Version 1.1 (A19-43). Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG); 2020 [Stand: 19.08.2025]. Verfügbar unter: [https://www.iqwig.de/download/a19-43\\_versorgungsnahe-daten-zum-zwecke-der-nutzenbewertung\\_rapid-report\\_v1-1.pdf](https://www.iqwig.de/download/a19-43_versorgungsnahe-daten-zum-zwecke-der-nutzenbewertung_rapid-report_v1-1.pdf).
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 7.0. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG); 2023 [Stand: 19.08.2025]. Verfügbar unter: [https://www.iqwig.de/methoden/allgemeine-methoden\\_version-7-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf).
3. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ 2007; 334(7589):349–51 [Stand: 19.08.2025]. Verfügbar unter: <https://www.bmjjournals.org/content/334/7589/349.short>.
4. Wilken M, Berkemeier F. Machbarkeit und Praktikabilität der anwendungsbegleitenden Datenerhebung (AbD). MVF 2024; 2024(02):62–8 [Stand: 19.08.2025]. Verfügbar unter: <https://www.monitor-versorgungsforschung.de/abstract/machbarkeit-und-praktikabilitaet-der-anwendungsbegleitenden-datenerhebung-abd-erfahrungen-analysen-und-weiterentwicklungsvorschlaege/>.

### **A.1.15 iOMEDICO AG**

#### **Autorinnen und Autoren**

- Haug, Nina

## **Stellungnahme**

### **1. Abschnitt 5.6.2.1 (S. 46)**

#### Anmerkung:

Im Absatz unter „Beschreibung geeigneter Sensitivitätsanalysen“ wird erwähnt, dass auxiliare Variablen, welche mit Confoundern assoziiert sind, die Qualität von Imputationen erhöhen können. Jedoch besteht hierbei das Risiko von Collider Bias [1].

#### Vorgeschlagene Änderung:

Ich schlage vor, in diesem Zusammenhang auf das Risiko von Collider Bias bei multipler Imputation hinzuweisen und die Studie [1] zu zitieren.

### **2. Abschnitt 5.7 (S. 52)**

#### Anmerkung:

Hier wird im zweiten Absatz auf die Varianzschätzung mittels Bootstrapping bei der Verwendung von Inverse Probability of Treatment Weighting (IPTW) zur Confounderadjustierung verwiesen. Falls zusätzlich fehlende Daten imputiert werden müssen, stellt sich die Frage der Reihenfolge von a) Ziehen der Bootstrap-Samples und b) Erzeugung multipler Imputationen. In [2] werden diesbezüglich unterschiedliche Varianten im Rahmen einer Simulationsstudie miteinander verglichen. Insbesondere wird die Variante, bei der zunächst Bootstrap-Samples gezogen werden und anschließend jedes Sample mehrfach imputiert wird (Boot MI), mit der Variante verglichen, bei der zunächst multiple Imputationen erzeugt werden und anschließend Bootstrap-Samples jeder Imputation gezogen werden (MI Boot). In dieser Studie erwies sich Boot MI als die zuverlässigste, wenn auch rechenintensivste Variante. Allerdings lieferte MI Boot ebenfalls valide Ergebnisse, sofern der Anteil fehlender Daten nicht zu hoch war.

#### Vorgeschlagene Änderung:

Ich rege an, an dieser Stelle auf die Fragestellung der Reihenfolge von Bootstrapping und Generierung multipler Imputationen einzugehen und die Arbeit [2] zu zitieren.

### **3. Abschnitt 5.7 (S. 51)**

#### Anmerkung:

Im Absatz unter „Simulationsstudien zu Propensity-Score-Analysen bei kleinen Fallzahlen“ wird auf unterschiedliche Methoden zur Confounderadjustierung mittels Propensity Scores eingegangen. In [3] werden theoretische Argumente präsentiert, welche gegen die Verwendung von Propensity Scores zum Matching sprechen.

#### Vorgeschlagene Änderung:

Ich schlage vor, an dieser Stelle auf die in [3] vorgebrachten Argumente hinzuweisen und diese gegebenenfalls einzuordnen.

## **Referenzen**

- [1] E. Curnow, K. Tilling, J. E. Heron, R. P. Cornish, and J. R. Carpenter, “Multiple imputation of missing data under missing at random: including a collider as an auxiliary variable in the imputation model can induce bias,” *Front. Epidemiol.*, vol. 3, Sep. 2023, doi: 10.3389/fepid.2023.1237447.
- [2] M. Schomaker and C. Heumann, “Bootstrap inference when using multiple imputation,” *Statistics in Medicine*, vol. 37, no. 14, pp. 2252–2266, Jun. 2018, doi: 10.1002/sim.7654.

- [3] G. King and R. Nielsen, “Why Propensity Scores Should Not Be Used for Matching,” *Political Analysis*, vol. 27, no. 4, pp. 435–454, 2019.

### **A.1.16 Janssen-Cilag GmbH**

#### **Autorinnen und Autoren**

- Eisele, Lewin
- Huschens, Susanne
- Mehnert, Angelika
- Perings, Dorothea
- Sindern, Jörn

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

Name, Vorname; Titel des/der Stellungnehmenden

Bitte nutzen Sie pro Person 1 Zeile.

Dr. Lewin Eisele

Dr. Susanne Huschens

Dr. Angelika Mehnert

Dr. Dorothea Perings

Dr. Jörn Sindern

Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)

- im Namen folgender Institution / Organisation: Johnson & Johnson, Johnson & Johnson Platz 1, 41470 Neuss
- als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Johnson & Johnson begrüßt die durch den G-BA eingeräumte Gelegenheit zur Stellungnahme zu dem Entwurf des Rapid Reports des IQWiG zu dem Thema „Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“ und geht auf die folgenden Spezifischen Aspekte ein:

- Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen
- Grundlagen der Fallzahlsschätzung – Verschobene Nullhypothese
- Interpretierbarkeit von Analysen mit fehlenden und / oder imputierten Werten

### **Kapitel 5.3, Abschnitt 5.3 [S. 12-25]**

#### **Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen**

##### Anmerkung:

Der Entwurf des „Rapid Reports zur wissenschaftlichen Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung nach § 35a SGB V“ soll das Ziel verfolgen, das methodische Vorgehen bei der anwendungsbegleitenden Datenerhebung (AbD) zu erörtern. Gemäß der in der Beschlussempfehlung und des Berichts des Ausschusses für Gesundheit getätigten Formulierung „[ist es] das Ziel der anwendungsbegleitenden Datenerhebung, eine valide Quantifizierung des Zusatznutzens zu erreichen. Diesem Ziel muss der Gemeinsame Bundesausschuss bei der Festlegung der Anforderungen Rechnung tragen.“ (1). Der Zweck der anwendungsbegleitenden Datenerhebung ist nach § 35a SGB V ausschließlich die Nutzenbewertung.

In Kapitel 5.3 setzt sich das IQWiG mit den Herausforderungen aufgrund der i. d. R. nur in geringem Umfang vorliegenden Informationen auseinander. Es diskutiert die gemeinsame methodische Grundlage der beiden Herangehensweisen „orientierende Fallzahlsschätzung“ und „orientierende Betrachtung von Fallzahlszenarien“ sowie deren konzeptionelle Unterschiede. Beiden Herangehensweisen ist gemeinsam, dass sie die Größen Fallzahl, Power, Effektgröße und Signifikanzniveau in Beziehung setzen. Jedoch beantworten beide Herangehensweisen unterschiedliche Fragestellungen: Die orientierende Fallzahlsschätzung zielt darauf ab, unter Vorgabe der Parameter Power, Effektgröße und Signifikanzniveau eine Aussage zur benötigten Fallzahl zu machen, während die orientierende Betrachtung von Fallzahlszenarien aus den vorgegebenen Parametern Power, Fallzahl und Signifikanzniveau eine Aussage zur nachweisbaren Effektgröße macht (S. 23, Abbildung 8).

Das IQWiG beschreibt auf S. 12 des Rapid Reports die regelhafte Beauftragung durch den G-BA zur Erstellung von Konzepten für AbDs und deren Auswertung wie folgt:

*„Die Konzepte sollen auch die Anforderungen an die Dauer und den Umfang der jeweiligen Datenerhebung beschreiben. Dabei geht es um die Beantwortung der Frage, ob die geplante AbD prinzipiell realisierbar und sinnvoll durchführbar ist. Auf Basis der vorliegenden Erkenntnisse zum zu bewertenden Arzneimittel und zum Komparator sowie der Anzahl der im jeweiligen Anwendungsgebiet zur Verfügung stehenden Patientinnen und Patienten wird abgeschätzt, ob innerhalb eines akzeptablen Zeitraums ausreichend Patientinnen und Patienten bzw. Ereignisse (d. h. die notwendige Fallzahl bzw. Ereigniszahl) eingeschlossen bzw. beobachtet werden können, um mit der AbD grundsätzlich aussagekräftige Ergebnisse für eine Nutzenbewertung generieren zu können.“*

Daraus geht hervor, dass die Fallzahl eine wesentliche Größe ist, um die Durchführbarkeit einer AbD beurteilen zu können. In einem Konzept für eine AbD ist daher die Fallzahl als eine für die Frage der Durchführbarkeit einer AbD wesentliche abhängige Größe zu diskutieren. Folglich ist eine orientierende Fallzahlschätzung prinzipiell geeignet, um den Auftrag des G-BA zur Erstellung eines AbD-Konzepts für eine der Nutzenbewertung dienende Datenerhebung zu erfüllen.

Es ist jedoch fraglich, ob auch die Herangehensweise der orientierenden Fallzahlszenarien ebenso geeignet ist, den Auftrag zur Beurteilung der Durchführbarkeit einer AbD zu erfüllen. In den orientierenden Fallzahlszenarien ist die Fallzahl einer der vorgegebenen Parameter und die nachweisbare Effektstärke die abhängige Größe.

Die Irrelevanz der theoretisch nachweisbaren Effektstärke im Rahmen eines AbD-Konzepts ergibt sich vor allem daraus, dass eine Annahme über eine zu erwartende Effektstärke eine zwingend zu machende Annahme ist, um über die Herangehensweise einer (orientierenden) Fallzahlschätzung eine sinnvolle Diskussion der benötigten Fallzahl, sowie der sich in einem konkreten Versorgungskontext daraus ergebenden erwarteten Dauer einer AbD, führen zu können.

Zudem ist eine theoretisch nachweisbare Effektstärke eine generische Größe ohne Bezug zu einer konkreten Fragestellung in einem Indikationsgebiet mit Intervention und zugehörenden Komparatoren.

Tatsächlich nennt das IQWiG als Rationale für die Sinnhaftigkeit der orientierenden Fallzahlszenarien auch ein anderes Ziel als das der Nutzenbewertung. So wird auf S. 14 des Rapid Reports geschrieben:

*„Vor diesem Hintergrund kann es für die Situation einer AbD sinnvoll sein, eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlschätzung auf dieses Ziel auszurichten. Stattdessen ist denkbar, die Schätzung der notwendigen Fallzahl darauf zu beziehen, dass aus einem gewählten Propensity-Score-Verfahren unter Sicherstellung einer adäquaten Confounderadjustierung verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95 %-Konfidenzintervall resultieren (siehe hierzu auch Abschnitt 5.7), und zwar losgelöst davon, ob sich unter Berücksichtigung der verschobenen Nullhypothese ein Vorteil des zu bewertenden Arzneimittels ableiten lässt. Unter dieser Voraussetzung stellt jedes Ergebnis einer AbD einen Erkenntnisgewinn dar.“*

Wenn bereits vor Durchführung einer AbD feststeht, dass ein Vorteil des zu bewertenden Arzneimittels und damit ein Zusatznutzen nicht nachweisbar ist, kann keine ergebnisoffene Nutzenbewertung durchgeführt werden. Eine solche AbD ist damit nicht geeignet, den Auftrag

des Gesetzgebers zu erfüllen, dass eine AbD der Nutzenbewertung dienen soll. Ebenso wenig ist die Herangehensweise der orientierenden Fallzahlszenarien, mit denen tatsächlich Effektstärkenszenarien diskutiert werden, als methodische Grundlage für die Erstellung eines AbD-Konzepts geeignet.

Aus Sicht des IQWiG ist in Abwesenheit vergleichender Daten jede aus vergleichender Evidenz gewonnene Erkenntnis ein Gewinn. Es wird nicht in Abrede gestellt, dass jeder Erkenntnisgewinn zu begrüßen ist. Jedoch kann ein AbD-Konzept, welches nicht geeignet ist, die Evidenz für eine Nutzenbewertung zu liefern, nicht als Grundlage für die Forderung einer AbD dienen.

#### Vorgeschlagene Änderung

Die Grundlage eines AbD-Konzepts muss mindestens eine orientierende Fallzahlschätzung sein. Wenn das IQWiG zu dem Schluss kommt, dass eine orientierende Fallzahlschätzung nicht möglich ist, kann in einem Konzept nicht gleichzeitig auch auf die Durchführbarkeit einer AbD geschlussfolgert werden. Die Methode der orientierenden Fallzahlszenarien ist nicht geeignet, Aufschluss über die Durchführbarkeit einer AbD zu geben, und kann folglich nicht Grundlage eines AbD-Konzepts sein. Ein AbD-Konzept, das nicht geeignet ist, die Evidenz für eine Nutzenbewertung zu liefern, kann nicht als Grundlage für die Forderung einer AbD dienen.

### **Kapitel 5, Abschnitt 5.3.1 [S. 16]**

#### **Grundlagen der Fallzahlschätzung – Verschobene Nullhypothese (Alternativmethoden zur Schätzung der Unsicherheit)**

*„Ein weiterer wesentlicher Aspekt der Fallzahlschätzung in einer AbD betrifft die Ergebnisunsicherheit aufgrund der fehlenden Randomisierung. Aus einer nicht randomisierten Studie kann z. B. 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 zum Schaden ergibt sich dann, wenn das Konfidenzintervall für den beobachteten Effekt oberhalb bzw. unterhalb einer zu definierenden Schwelle liegt (Test auf verschobene Nullhypothese H0). Da bei der Generierung von versorgungsnahen Daten (Studien ohne Randomisierung) die Erfüllung umfangreicher Qualitätsanforderungen Voraussetzung für die Betrachtung der Effekte ist, sollte dieser Schwellenwert deutlich unterhalb des Wertes für den „dramatischen Effekt“ (RR von 5 bis 10 [47]) liegen, z. B. in einem Bereich von 2 bis 5 für das RR bezogen auf die untere bzw. die obere Konfidenzintervallgrenze (bzw. 0,2 bis 0,5 z. B. bei mortalitätssenkenden Interventionen) [34].“*

#### Anmerkung

Zum Umgang mit der Unsicherheit bei nicht-randomisierten Vergleichen, z. B. durch residuelles Confounding, greift das IQWiG auf die Testung der geschätzten Effekte gegen eine verschobene Nullhypothese zurück. Die Verwendung dieses Kriteriums mit einem praktisch fixen, willkürlich gewählten Schwellenwert, stellt aus Sicht von Johnson & Johnson ein nicht geeignetes Kriterium dar, um die Validität von Analyseergebnissen aus nicht-randomisierten Vergleichen zu beurteilen.

Die Verwendung des Kriteriums der verschobenen Nullhypothese ist insbesondere für die Erstellung eines AbD-Konzepts als problematisch anzusehen, weil a priori nur für solche Szenarien eine Aussage zum Zusatznutzen theoretisch möglich ist, in denen von einem relativen Risiko von 2 bis 5 (bzw. 0,2 bis 0,5 z. B. bei mortalitätssenkenden Interventionen) ausgegangen werden kann. Eine AbD kann somit prinzipiell nur dann die Datengrundlage für eine Quantifizierung des Zusatznutzens unter Annahme der verschobenen Nullhypothese in einer Nutzenbewertung darstellen, wenn dieser sehr große Unterschied in der Effektstärke Grundlage der Fallzahlplanung ist. Im Allgemeinen können jedoch auch kleinere relative Unterschiede mit einem patientenrelevanten Zusatznutzen verbunden sein. Damit ist das Kriterium der verschobenen Nullhypothese prinzipiell nicht als ein generelles Kriterium für die AbD geeignet.

Wenn die Methode der verschobenen Nullhypothese nicht als allgemeines Kriterium geeignet ist, jedoch als alleiniges Kriterium zum Umgang mit Unsicherheit vorgegeben ist, wird dadurch

das Verfahren der anwendungsbegleitenden Datenerhebung aus Sicht von Johnson & Johnson prinzipiell infrage gestellt. Es sollten daher geeignete Alternativen zur Verwendung der verschobenen Nullhypothese diskutiert werden.

Aus Sicht von Johnson & Johnson ist der E-Value (2) ein geeignetes Instrument, um Aussagen über die Validität von Analyseergebnissen aus nicht-randomisierten Studien zu beurteilen und eine Entscheidung für das Heranziehen solcher Ergebnisse in einer Nutzenbewertung zu treffen.

Der E-Value entspricht der minimalen Assoziationsstärke, die ein nicht gemessener bzw. nicht berücksichtigter Confounder sowohl mit der Behandlung als auch mit dem analysierten Endpunkt haben müsste, um das beobachtete Analyseergebnis zu erklären. Ein hoher Wert des E-Value deutet darauf hin, dass ein beträchtliches, nicht-gemessenes Confounding erforderlich wäre, um einen beobachteten Behandlungseffekt infrage zu stellen (2).

Als Grundlage für die Berechnung des E-Value dient der aus den Daten geschätzte Effekt auf den Endpunkt. Ein Kritikpunkt am E-Value ist, dass durch die Berechnung ein annähernd linearer Zusammenhang zwischen dem beobachteten Effektschätzer und dem berechneten E-Value besteht und dass ein gegebener Effektschätzer immer zu dem gleichen E-Value führt (3). Kern der Kritik ist damit die fehlende Abhängigkeit vom spezifischen Kontext der Fragestellung bzw. Indikation, in dem der E-Value berechnet wird.

Um den Kontext für die Interpretierbarkeit des E-Value herzustellen, schlägt Johnson & Johnson vor, Informationen aus der im Rahmen einer AbD ohnehin geforderten Confounder-Analyse zu verwenden. Diese stellt eine Synthese von umfassender und systematischer Identifikation von Confoundern und Expertenwissen durch klinische Experten im Anwendungsgebiet der AbD dar. Die von Johnson & Johnson durchgeführten Confounder-Analysen beinhalten dabei u. a. zu jedem relevanten Confounder Informationen zur Stärke des Zusammenhangs der Confounder mit dem Endpunkt, die aus den identifizierten Publikationen extrahiert wurden (siehe z.B. in den Verfahren 2022-01-15-D-788: Amivantamab und 2023-02-15-D-919: Ciltacabtagene autoleucel). Diese Informationen liefern damit einen evidenzbasierten Referenzrahmen, z. B. in Form einer empirischen Verteilung der Assoziationsstärken. Zusammen mit der tatsächlichen Verteilung der Confounder innerhalb der Analysepopulation kann damit eine vom konkreten Kontext der Fragestellung der Nutzenbewertung abhängige Interpretation erfolgen. Das bedeutet, dass ein für die Fragestellung spezifischer Schwellenwert für den E-Value hergeleitet werden kann, ab dem es sehr unwahrscheinlich ist, dass residuelles Confounding die Analyseergebnisse infrage stellen würde. Durch die Möglichkeit, umfangreiche Evidenz aus für die Fragestellung spezifischen Confounder-Analysen in das Kriterium zum Umgang mit Unsicherheit einfließen zu lassen, hebt sich das Konzept einer E-Value-basierten Herangehensweise von dem pauschal operierenden Konzept der verschobenen Nullhypothese ab.

### Vorgeschlagene Änderung

Die Methode der verschobenen Nullhypothese ist aufgrund der nicht gegebenen allgemeinen Anwendbarkeit als alleiniges Kriterium für den Umgang mit Unsicherheit in einer AbD kritisch zu hinterfragen. Mit der oben skizzierten E-Value-basierten Herangehensweise liegt ein alternatives Konzept für ein Kriterium vor, das allgemein anwendbar ist und fragestellungs- und Indikationsspezifische Information aus der Confounder-Analyse nutzt. Das Konzept des E-Values sollte im Rapid Report diskutiert werden.

## Kapitel 5, Abschnitt 5.6.2.2 [S. 48]

### Interpretierbarkeit von Analysen mit fehlenden und / oder imputierten Werten

*„Nicht adäquat ist ein Ausschluss von relevanten Confoundern aufgrund von zu hohen Anteilen fehlender Werte. In diesen Situationen lässt sich ein Zusatznutzen lediglich bei sehr großen Effekten im Sinne eines dramatischen Effekts ableiten.“*

Es gibt Situationen, in denen für einen Confounder ein hoher Anteil an fehlenden Werten nicht zu vermeiden ist. Beispielsweise kann hier der Confounder Zytogenetisches Risiko genannt werden, der im Rahmen einer Confounder-Analyse von Johnson & Johnson als relevanter Confounder in einer früheren Nutzenbewertung im Anwendungsgebiet des Multiplen Myeloms identifiziert wurde (siehe Verfahren 2023-02-15-D-919: Ciltacabtagene autoleucel). Für die Bestimmung des Zytogenetischen Risikos ist eine Knochenmarkpunktion erforderlich, die einen invasiven Eingriff darstellt, der in den hinteren Therapielinien nicht in der Regelversorgung durchgeführt wird.

Aus Sicht von Johnson & Johnson kann in einem solchen Fall eine E-Value-basierte Herangehensweise verwendet werden, um zu begründen, dass Analyse-Ergebnisse unter Ausschluss eines relevanten Confounders, z. B. aufgrund eines hohen Anteils fehlender Werte, immer noch als interpretierbar anzusehen sind.

Hierzu können Informationen zum ausgeschlossenen Confounder aus der Confounder-Analyse herangezogen werden (siehe Punkt oben). Diese liefern u. a. einen Überblick über die Größenordnung des Zusammenhangs des Confounders mit Endpunkten. Für die Verteilung des Confounders in der Analysepopulation kann bzgl. der fehlenden Werte z. B. ein worst-case Szenario angenommen werden. Damit wird ein evidenzbasierter Referenzrahmen geschaffen, mit dem die Festlegung eines für die betrachtete Fragestellung spezifischen Schwellenwertes für den E-Value möglich ist, ab dem es als sehr unwahrscheinlich angesehen werden kann, dass die Hinzunahme des ausgeschlossenen Confounders die beobachteten Analyseergebnisse infrage stellen würde. Dies erlaubt die Entscheidung, ob die Analyseergebnisse unter Ausschluss des relevanten Confounders für die Ableitung eines Zusatznutzens herangezogen werden können, ohne dass das Vorliegen eines dramatischen Effektes vorausgesetzt werden muss.

### Vorgeschlagene Änderung

Eine E-Value-basierte Herangehensweise bietet aus Sicht von Johnson & Johnson die Möglichkeit für eine Begründung, Analyseergebnisse auch unter Ausschluss eines relevanten Confounders für die Nutzenbewertung heranziehen zu können, ohne dass das Vorliegen eines

dramatischen Effektes vorausgesetzt werden muss. Das Konzept des E-Values sollte im Rapid Report diskutiert werden.

### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. Deutscher Bundestag. BT-Drucksache 19/10681 (Beschlussempfehlung und Bericht – Ausschuss für Gesundheit). 05.06.2019. [abgerufen am 11.08.2025]. Verfügbar unter: <https://dserver.bundestag.de/btd/19/106/1910681.pdf>.
2. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268-274.
3. Ioannidis, John P A; Tan, Yuan Jin; Blum, Manuel R. Limitations and Misinterpretations of E-Values for Sensitivity Analyses of Observational Studies. Ann Intern Med. 2019;170(2):108-111.

### **A.1.17 Lilly Deutschland GmbH**

#### **Autorinnen und Autoren**

- Bocuk, Derya
- Zschocke, Juergen

**Stellungnahmen zum vorläufigen IQWiG Rapid Report „[A25-13] Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“**

Die anwendungsbegleitende Datenerhebung (AbD) nach §35a SGB V ist ein zentrales Instrument, um Evidenzlücken bei Arzneimitteln – insbesondere Orphan Drugs oder bedingt zugelassenen Therapien – zu schließen.

Der Gemeinsame Bundesausschuss (G-BA) hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit einer wissenschaftlichen Ausarbeitung zur Methodik bei der Generierung und Auswertung versorgungsnaher Daten im Kontext der sogenannten anwendungsbegleitenden Datenerhebung (AbD) beauftragt. Im Rahmen eines vorläufigen Rapid Reports hat sich das IQWiG mit sieben Themen, die für die Durchführung solcher Studien von Bedeutung sind, wie etwa der Identifikation und Auswahl sogenannter Confounder, der Festlegung des Beobachtungsstarts, der Erhebung patientenberichteter Endpunkte oder dem Umgang mit fehlenden Werten befasst. Der IQWiG Rapid Report bietet eine wissenschaftliche Ausarbeitung zur Methodik der Generierung und Auswertung versorgungsnaher Daten im Kontext nicht-randomisierter vergleichender Studien. Ziel ist es, die Umsetzbarkeit und Aussagekraft von AbD-Konzepten zu verbessern und die Anforderungen an Studiendesign, Datenerhebung und Auswertung zu präzisieren.

Mit Veröffentlichung zum 30. Juli 2025 bittet das IQWiG interessierte Personen, Institutionen und (Fach-)Gesellschaften aufgerufen, bis zum 19. August 2025 Stellungnahmen zum vorläufigen Bericht abzugeben.

Lilly Deutschland GmbH (nachfolgend Lilly) gibt im Folgenden als pharmazeutisches Unternehmen eine Stellungnahme zum vorläufigen Rapid Report ab.

**Allgemeine Anmerkungen**

Zunächst begrüßt Lilly, dass im Rahmen der Beteiligung von interessierten Personen, Institutionen und (Fach-)Gesellschaften Wissenschaft, Versorgungspraxis sowie Industrie zusammengeführt werden, um auf diese Weise einen Austausch von Fachexpertise zu ermöglichen. Des Weiteren ist es aus Sicht von Lilly erfreulich, dass der Gesetzgeber grundsätzlich die Nutzung von Versorgungsdaten ermöglicht, die im Rahmen einer AbD generiert werden. Dies kann zu einer Verbesserung der Evidenzgrundlage für Arzneimittel zur Behandlung seltener Leiden oder Arzneimittel mit bedingter oder unter außergewöhnlichen Umständen erteilter Zulassung beitragen, indem die Daten aus der Versorgung von Patientinnen und Patienten im Vergleich zur Versorgung mit anderen Arzneimitteln oder Behandlungsmöglichkeiten (Standardtherapien) betrachtet werden. Dadurch können Patientinnen und Patienten weiterhin zügig Zugang zu lebensrettenden Medikamenten erhalten, während noch weitere, ergänzende Daten erhoben werden. Der G-BA ist dabei gesetzlich verpflichtet, die Vorgaben der AbD klar zu definieren, Zulassungsauflagen zu berücksichtigen sowie resultierende Daten auch für die Nutzenbewertung heranzuziehen.

Aus Sicht von Lilly ist eine praxistaugliche Ausgestaltung und Umsetzung der AbD essenziell für die Realisierung einer adäquaten Evidenzgenerierung. Zum aktuellen Zeitpunkt werfen jedoch die methodischen Vorgaben des G-BA und IQWiG auf Basis der bisher begrenzt verfügbaren Erfahrung mit AbD Fragen zur Umsetzbarkeit auf und wirken teilweise überhöht und praxisfern, so dass daraus eine Überforderung aller Beteiligten resultieren kann.

In diesem Zusammenhang sollte der G-BA frühzeitig und verbindlich zum Thema AbD beraten, um die Planungssicherheit für forschende Pharmaunternehmen gewährleisten zu können. Mögliche

Forderungen zur Einleitung einer AbD sollten frühzeitig mit dem betroffenen pU kommuniziert werden. Mit dem Ziel einer besseren Planbarkeit sollten in einem gemeinsamen Austausch zwischen G-BA, Zulassungsbehörde und pU die individuellen, indikationsspezifischen Anforderungen der Zulassungsbehörden und des Herstellers berücksichtigt werden.

### **Anmerkungen zu spezifischen Aspekten aus dem vorläufigen Rapid Report**

#### ***Schätzung von Fallzahlen – Methodik, Herausforderungen und Kritik***

##### Methodische Grundlagen:

Der Rapid Report beschreibt zwei methodische Ansätze zur Schätzung notwendiger Fallzahlen:

- **Orientierende Fallzahlsschätzung:** Wird durchgeführt, wenn ausreichende Vorabinformationen zur Intervention und zum Komparator vorliegen. Sie basiert auf Annahmen zu Effektgröße, Signifikanzniveau und Power, analog zur klassischen Fallzahlplanung in RCTs.
- **Orientierende Betrachtung von Fallzahlszenarien:** Kommt zum Einsatz, wenn keine ausreichenden Vorabinformationen vorliegen. Hier wird die im Anwendungsgebiet verfügbare Patientenzahl als Ausgangspunkt genommen, um aufdeckbare Effektgrößen zu bestimmen.

Beide Vorgehensweisen sind zwar methodisch nachvollziehbar, verdeutlichen jedoch die mit der Evidenzsituation bei neu eingeführten Arzneimitteln verbundenen Unsicherheiten. Die entsprechenden Konzepte liefern in der Regel lediglich grobe Näherungswerte und dienen vorrangig der Machbarkeitsanalyse einer AbD.

Die Umsetzung der Fallzahlsschätzung ist in der Praxis mit erheblichen Unsicherheiten behaftet:

- Häufig fehlen belastbare Daten zur Effektgröße und zum Basisrisiko.
- Die Verteilung der Patienten auf die Behandlungsgruppen ist schwer vorherzusagen.
- Die Rekrutierung ausreichender Patientenzahlen ist insbesondere bei seltenen Erkrankungen oft nicht realistisch.

Der Report empfiehlt daher, die finale Fallzahlplanung erst im Verlauf der AbD anhand von Zwischenanalysen vorzunehmen und – wo möglich – internationale Register einzubinden, um die Rekrutierung zu erleichtern.

##### Kommentar:

Das Kapitel zur Fallzahlsschätzung stellt einen kritischen Punkt der AbD dar, da die Unsicherheiten in der Fallzahlplanung die Umsetzbarkeit der AbD potenziell infrage stellen. Die Darstellung der beiden Varianten der Fallzahlsschätzung im Bericht bleibt unpräzise und klare Kriterien für die Auswahl der jeweiligen Methode werden nicht genannt.

Weiterhin wird das folgende Statement im Bericht kritisch bewertet: „[...] Vor diesem Hintergrund kann es für die Situation einer AbD sinnvoll sein, eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlsschätzung auf dieses Ziel auszurichten. Stattdessen ist denkbar, die Schätzung der notwendigen Fallzahl darauf zu beziehen, dass aus einem gewählten Propensity-Score-Verfahren unter Sicherstellung einer adäquaten Confounderadjustierung verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95 %-Konfidenzintervall resultieren [...], und zwar losgelöst davon, ob sich unter

Berücksichtigung der verschobenen Nullhypothese ein Vorteil des zu bewertenden Arzneimittels ableiten lässt. Unter dieser Voraussetzung stellt jedes Ergebnis einer AbD einen Erkenntnisgewinn dar.“

Aus dem Bericht geht allerdings nicht hervor, inwiefern der Erkenntnisgewinn bei Nicht-Berücksichtigung der verschobenen Nullhypothese auch in die Ableitung des Zusatznutzens einfließen könnte. Denn in einem der folgenden Abschnitte wird ausgeführt, dass der Schwellenwert für positive oder negative Effekte deutlich unterhalb des Wertes für den „dramatischen Effekt“ liegen muss.

Damit ist auch die Ableitung eines Zusatznutzens bei Nicht-Überlegenheitsstudien im Rahmen der AbD nahezu auszuschließen. Dies trifft umso mehr bei kleinen Fallzahlen und einer damit verbundenen hohen Varianz zu.

Zusammenfassend wird die Flexibilität des IQWiG begrüßt, in bestimmten Situationen auf die Testung auf die verschobene Nullhypothese zu verzichten. Allerdings hat dies gemäß Rapid Report keine Konsequenz auf die Ableitung des Zusatznutzens, der gemäß IQWiG erst bei sehr hohen Effektgrößen beansprucht werden kann.

### ***Patientenberichtete Endpunkte (PRO) – Anforderungen und Chancen***

#### Methodische Grundlagen:

Der G-BA fordert für AbDs regelmäßig die Erhebung von PROs zur Symptomatik und Lebensqualität. Der Rapid Report betont:

- Die Notwendigkeit, validierte und möglichst indikationsspezifische Instrumente einzusetzen.
- Die Erhebung sollte mehrfach jährlich und in standardisierten Intervallen erfolgen.
- Digitale Tools (Apps, Patientenportale) werden als praktikable und effiziente Lösung empfohlen, um die Erhebung unabhängig von Arztbesuchen zu ermöglichen.

Die PRO-Abschnitte befassen sich mit der Digitalisierung der Datenerhebung, der Motivation von Patientinnen und Patienten sowie der Integration in die Versorgungspraxis. Hervorzuheben ist:

- Befürwortung der digitalen PRO-Erhebung.
- Die Option, PRO-Daten für das Symptomtracking und die Beratung von Patientinnen und Patienten in Echtzeit einzusetzen.

#### Kommentar

Die Befürwortung der digitalen PRO-erhebung wird begrüßt, da sie einen effizienteren und flexibleren Umgang mit Fragebögen, beispielsweise deren Anpassung an Online-Formate oder Validitätsaspekte, ermöglicht.

#### ***Weitere methodische Aspekte:***

Das Fehlen eines Komparators im Studiendesign ist eine spezielle Herausforderung (beispielsweise bei seltenen Erkrankungen), da eine RCT oftmals aus ethischen Gründen nicht durchführbar ist. Ergänzende Evidenzgenerierung mittels Beobachtungsstudien oder AbDs können allerdings weitreichende Limitationen mit sich bringen (Vergleichbarkeit der Kohorten, Confounder-Problematik, Stichprobengröße), und es ist im jeweiligen Einzelfall zu klären, wie verlässlich Aussagen zur Wirksamkeit und Sicherheit abgeleitet werden können. Aus Sicht von Lilly sollten diese Aspekte stärker in der Methodik berücksichtigt werden, um einen realistisch umsetzbaren und verhältnismäßigen Rahmen für die AbD zu definieren.

Grundsätzlich ist aus Sicht von Lilly die Teilnahme aller qualifizierten Zentren und Indikationsregister an der AbD zu fördern, um eine Grundlage zur Evidenzgenerierung zu schaffen. Zur Gewährleistung einer hohen Datenqualität verlangt das IQWiG in diesem Zusammenhang, wesentliche Erweiterungen der bestehenden Indikationsregister, beispielsweise bzgl. der Confounder-Erhebung und -Adjustierung, der Erhebungsstruktur und Maßnahmen zur Vollständigkeit der Daten. Die hohen Qualitätsanforderungen werden jedoch in deutschen Patientenregistern derzeit selten erfüllt, so dass derartige Forderungen in der Realität mit einem überdurchschnittlich hohen Aufwand verbunden sein können und u.a. deshalb oftmals schwer umsetzbar sind.

Die Erweiterung des Datensatzes um relevante Confounder, die derzeit in den verfügbaren Registern nicht ausreichend erfasst werden, sowie die Sicherstellung einer ausreichenden Fallzahl zur Beobachtung eines signifikanten Effekts stellen häufig erhebliche Herausforderungen für die Umsetzung von AbD dar. Insbesondere in Abhängigkeit vom Therapiegebiet, der Indikation und der Prävalenz in Deutschland erweist sich dieses Ziel oftmals als sehr ambitioniert und die Realisierbarkeit muss kritisch hinterfragt werden. Der Nachweis des Zusatznutzens, vor allem bei seltenen Erkrankungen, ist somit unter derartigen methodischen Bedingungen kaum realistisch möglich.

## Fazit

Zusammengefasst ist aus Sicht von Lilly die Berücksichtigung der Praxistauglichkeit, eine Flexibilität aller Beteiligten, das Ausmaß der Anforderungen an Auswertungen sowie eine Bereitschaft zur Digitalisierung essenziell für eine erfolgreiche Umsetzung von AbDs. Die teils überhöhten und praxisfernen methodischen Anforderungen sollten nicht die Umsetzbarkeit von AbDs gefährden. Die Anforderungen müssen realisierbar, transparent und nachvollziehbar sein, so dass eine valide und für die Nutzenbewertung verwertbare Evidenzgrundlage zur Beantwortung der konkreten Fragestellung (unter Berücksichtigung der Besonderheiten des Anwendungsgebiets des Arzneimittels) geschaffen werden kann. Dabei ist auch zu berücksichtigen, dass selbst bei gegebener Praxistauglichkeit und Senkung der Anforderungen, eine kleine Stichprobe verbunden mit der Testung auf die verschobene Nullhypothese eine Ableitung eines Zusatznutzens erheblich erschwert.

### **A.1.18 Medizinischer Dienst Bund (KöR)**

#### **Autorinnen und Autoren**

- Gehrman, Ulrich
- Janatzek, Sandra
- Lange, Stefan
- Morche, Johannes

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Lange, Stefan; PD Dr.
Gehrmann, Ulrich
Janatzek, Sandra; Dr.
Morche, Johannes; Dr.
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> <b>im Namen folgender Institution / Organisation: Medizinischer Dienst Bund</b>
<input type="checkbox"/> <b>als Privatperson(en)</b>

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.2 (S. 9-12)	<p><u>Anmerkung</u></p> <p>Wir begrüßen die Ankündigung des IQWiG, in einem nächsten Schritt zu prüfen, welche weiteren Optionen zur Aufwandsreduktion hinsichtlich der Identifikation und Erhebung von Confoundern bestehen [1].</p> <p>Wir regen an, in diese Überlegungen auch einzubeziehen, ob 1) eine (weitere) Reduktion von zu erhebenden und in Auswertungen zu berücksichtigenden Variablen aufgrund der Korrelation von identifizierten potenziellen Confoundern und 2) Anpassungen an Schritt 1 des vom IQWiG beschriebenen Vorgehens zur Identifikation von Confoundern möglich sind. Diese Anregungen erfolgen insbesondere unter Berücksichtigung der vom IQWiG in Kapitel 5.6.1 des Rapid Reports beschriebenen Maßnahmen zur Vermeidung fehlender Werte.</p> <p>Zu 1)</p> <p>Der Einfluss von ungemessenem Confounding ist am stärksten, wenn keine Korrelation zwischen gemessenen und nicht gemessenen Confoundern vorliegt [2]. Daher möchten wir anregen, die Korrelation von potenziellen Confoundern bei der Auswahl der für eine AbD zu erhebenden Variablen stärker zu berücksichtigen. Wir begrüßen die vom IQWiG empfohlene Zusammenfassung von potenziellen Confounder-Variablen, welche ein inhaltlich gleiches Konstrukt messen oder sich inhaltlich überlappen. Wir regen an, für das im Arbeitspapier betrachtete Beispiel möglichst weitere inhaltliche Zusammenführungen durchzuführen bzw. über dieses Beispiel hinaus zu empfehlen, dass eine weitreichende Zusammenführung erfolgen soll. Auch potenzielle Confounder-Variablen, die zunächst einmal nicht das gleiche Konstrukt messen sollen, können miteinander korreliert sein. Daher könnte in einem zusätzlichen Schritt durch Berücksichtigung dieser Korrelationen (auch anhand von Auswertungen in Registern) eine weitere Reduktion von Variablen erfolgen.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Zu 2)</p> <p>Gemäß Ausführungen im Rapid Report und im Arbeitspapier [1] sollten unter Schritt 1 bei randomisierten oder nicht vergleichenden Studien alle Patientencharakteristika und bei vergleichenden Beobachtungsstudien alle Adjustierungsvariablen als potenzielle Confounder angesehen und extrahiert werden. Zudem wird ausgeführt, dass der (anschließende) Ausschluss eines potenziellen Confounders aus inhaltlichen Überlegungen literaturgestützt begründet werden muss. Uns ist unklar, wie solch ein literaturgestützter Ausschluss erfolgen kann bzw. welche Untersuchungen hierzu erforderlich sind. Nach unserem Verständnis wird nach derzeitiger Empfehlung im Arbeitspapier eine Variable, die lediglich in einer einzigen (oder in sehr wenigen) gesichteten randomisierten oder nicht vergleichenden Studien als Patientencharakteristikum aufgeführt wird und dementsprechend in nahezu allen gesichteten Quellen nicht als ein solches genannt wird, dennoch als potenzieller Confounder angesehen, der erhoben und in der Auswertung berücksichtigt werden sollte (wenn keine inhaltliche Zusammenführung mit einem anderen potenziellen Confounder erfolgt). Wir regen an, diese Empfehlung aufgrund folgender Überlegungen zu überdenken.</p> <p>Durch die in Schritt 1 des IQWiG vorgeschlagene Vorgehensweise werden zunächst sehr viele potenzielle Confounder identifiziert, deren Anzahl in Schritt 2 durch eine inhaltliche Zusammenführung reduziert wird. Dieses Vorgehen ist, wie vom IQWiG beschrieben, aufwendig. Zudem resultiert dieses Vorgehen im aufgeführten Beispiel in einer hohen Anzahl (<math>n = 28</math>) an potenziellen Confoundern, die für die Durchführung einer Target-Trial-Emulation erhoben werden müssen. Je mehr Variablen erhoben werden, desto höher ist das Risiko, dass für mindestens eine dieser Variablen Werte fehlen werden. Zudem kann eine (zu) hohe Anzahl an Adjustierungsvariablen zu Problemen in den statistischen Analysen führen. Es ist somit sinnvoll, die Datenerhebung auf diejenigen Variablen zu beschränken, die für eine adäquate Auswertung erforderlich sind.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Um die Praktikabilität zu erhöhen, könnte aus unserer Sicht ein aufwandsärmeres Vorgehen, das in weniger zu erhebenden (und in Analysen zu berücksichtigenden) Variablen resultiert, hilfreich sein. Hierzu wäre denkbar, in Schritt 1 auf Faktoren zu fokussieren, die die Therapieentscheidung beeinflussen. Neben Interviews mit klinischen Expertinnen und Experten wäre es aus unserer Sicht sinnvoll, regelhaft Leitlinien (falls vorhanden) nach Faktoren, die die Therapieentscheidung beeinflussen, zu durchsuchen [3, 4]. Um die wechselseitigen Beziehungen der identifizierten Variablen und den Einfluss dieser Variablen auf die Exposition bzw. Therapieentscheidung und das Outcome zu visualisieren und für die Auswahl potenzieller Confounder zu berücksichtigen, kommt die Anwendung eines gerichteten azyklischen Graphen (DAG) in Frage [5]. Bekannte instrumentelle Variablen sollten nicht als potenzielle Confounder angesehen und nicht als Adjustierungsfaktoren berücksichtigt werden [6, 7]. Die externen Sachverständigen stellten in ihrem Bericht fest, dass „der datengesteuerte Ansatz sehr zeitaufwändig [sei]. In Anbetracht der begrenzten Effizienz bei der Identifizierung der relevanten Confounder hat dieser Ansatz ein schlechtes Kosten-Nutzen-Verhältnis.“ [1]. Vor diesem Hintergrund ist ein Anspruch auf Vollständigkeit der Literaturrecherche und -analyse aus unserer Sicht zu überdenken. So könnte beispielsweise die Recherche in lediglich einer bibliographischen Datenbank (Medline) durchgeführt oder eine zeitliche Beschränkung der Recherche vorgenommen werden. Zudem schlagen wir vor, das Sättigungskriterium für RCTs und einarmige Studien zu reduzieren (beispielsweise auf einen Wert von 5). Alternativ (bei weiterer Anwendung eines Sättigungskriteriums von 10) möchten wir anregen, Variablen, die lediglich in einer Quelle als Patientencharakteristikum aufgeführt werden, nur dann als potenzielle Confounder anzusehen, wenn dies inhaltlich plausibel erscheint. Bei Nennung in lediglich einer gesichteten Quelle müsste somit keine literaturgestützte Begründung dafür erfolgen, dass diese Variable nicht als potenzieller Confounder angesehen wird.</p> <p>Das im Rapid Report und im Arbeitspapier des IQWiG beschriebene Vorgehen zur Confounderidentifikation soll dazu dienen, Bias, der durch messbare Confounder entsteht, zu reduzieren. Um Auswirkungen durch residuelles / ungemessenes Confounding besser abschätzen zu können, kann es aus unserer Sicht zudem sinnvoll sein, beispielsweise Methoden der Quantitativen Bias-Analyse [8] oder Negativ-Kontrollen [9, 10] anzuwenden.</p> <p><b>Vorgeschlagene Änderung</b></p> <ul style="list-style-type: none"> <li>• Stärkere Berücksichtigung von Korrelationen zwischen Variablen zur Reduktion der Anzahl potenzieller Confounder</li> </ul>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<ul style="list-style-type: none"> <li>• Überdenken von oben genannten Vorschlägen zur Anpassung von Schritt 1 (der Identifikation von Confoundern)</li> </ul>
5.3.1 (S. 14)	<p><u>Anmerkung</u></p> <p>Das IQWiG beschreibt, dass „... es für die Situation einer AbD sinnvoll sein [kann], eine vergleichende Studie <b>nicht zwingend</b> mit dem Ziel durchzuführen, <b>einen Vorteil</b> des neuen Arzneimittels <b>zu zeigen</b> und die Fallzahlsschätzung auf dieses Ziel auszurichten“ (eigene Hervorhebung). Und weiter: „Stattdessen ist denkbar, die Schätzung der notwendigen Fallzahl darauf zu beziehen, dass aus einem gewählten Propensity-Score-Verfahren unter Sicherstellung einer adäquaten Confounderadjustierung <b>verwertbare</b> Ergebnisse im Sinne <b>einer Effektschätzung</b> mit zugehörigem 95 %-Konfidenzintervall resultieren (siehe hierzu auch Abschnitt 5.7), und zwar losgelöst davon, ob sich unter Berücksichtigung der verschobenen Nullhypothese ein Vorteil des zu bewertenden Arzneimittels ableiten lässt. Unter dieser Voraussetzung <b>stellt jedes Ergebnis einer AbD einen Erkenntnisgewinn dar</b>“ (eigene Hervorhebung).</p> <p>Wir teilen die Ansicht, dass Daten aus qualitativ hochwertigen nicht randomisierten Studien vor allem beim Fehlen von Daten aus aussagekräftigen randomisierten Studien einen Erkenntnisgewinn darstellen können. Es stellt sich allerdings die Frage, ob es dann nicht konsequenter und informativer wäre, bei der orientierenden Betrachtung von Fallzahlszenarien (Seite 20 ff.) nicht „aufdeckbare Effekte“, sondern bei gegebenen Fallzahlen für bestimmte Konstellationen (Basisrisiko und Effektschätzungen) die erzielbare Präzision (z. B. Breite von 95%-Konfidenzintervallen) darzustellen.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Bei der orientierenden Betrachtung von Fallzahlszenarien bei gegebenen Fallzahlen für bestimmte Konstellationen (Basisrisiko und Effektschätzungen) die erzielbare Präzision (z. B. Breite von 95%-Konfidenzintervallen) darstellen.</p>
5.3.1 S. 16	<p><u>Anmerkung</u></p> <p>Das IQWiG stellt fest: „Eine (positive oder negative) Aussage zum Nutzen oder zum Schaden ergibt sich dann, wenn das Konfidenzintervall für den beobachteten Effekt oberhalb bzw. unterhalb einer zu definierenden Schwelle liegt (Test auf verschobene Nullhypothese H0). Da bei der Generierung von versorgungsnahen Daten (Studien ohne Randomisierung) die Erfüllung umfangreicher Qualitätsanforderungen Voraussetzung für die Betrachtung der Effekte ist, sollte dieser Schwellenwert deutlich unterhalb des Wertes für den</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>„dramatischen Effekt“ (RR von 5 bis 10 [47]) liegen, z. B. in einem Bereich von 2 bis 5 für das RR bezogen auf die untere bzw. die obere Konfidenzintervallgrenze (bzw. 0,2 bis 0,5 z. B. bei mortalitätssenkenden Interventionen) [34]“ (eigene Hervorhebung).</p> <p>Grundsätzlich unterstützen wir die Idee, statistische Hypothesentests mit einer verschobenen Nullhypothese durchzuführen, um Mindesteffektstärken statistisch abzusichern [11, 12]. Ebenso sinnvoll ist es, bei nicht randomisierten Studien ein „Handicap“ in die Formulierung der zu prüfenden Hypothese(n) einzubauen, um der den nicht randomisierten Studien inhärenten Unsicherheit zu begegnen.</p> <p>Gleichzeitig hat das IQWiG unseres Wissens nach bei seinen bisherigen Konzepten für eine AbD noch nie von dem „z. B.“ dergestalt Gebrauch gemacht, dass eine weniger ambitionierte Schwelle als 2 (bzw. 0,5) vorgeschlagen wurde (beispielsweise 1,5 bzw. 0,67). Wir regen an, über dermaßen abgesenkte Schwellen vor dem Hintergrund einer „Erfüllung umfangreicher Qualitätsanforderungen als Voraussetzung“ nachzudenken. Alternativ wäre denkbar, auf Basis geeigneter Methoden (beispielsweise Nutzung von Negativ-Kontrollen [hier: Outcomes, für die keine kausale Beeinflussung durch die zu prüfende Intervention anzunehmen ist]) eine quantitative Bias-Abschätzung vorzunehmen, um daraus eine Schwelle abzuleiten [9, 10]. Dies wäre in einem Studienprotokoll adäquat a priori zu planen.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Erweiterung der Methodik um Kriterien für die „Erfüllung umfangreicher Qualitätsanforderungen“, die gegebenenfalls geeignet sein könnten, in statistische Hypothesenprüfungen abgesenkte Schwellen jenseits von 2 (bzw. 0,5) aufzunehmen. Alternativ könnte die Möglichkeit eröffnet werden, auf Basis von geeigneten Bias-Abschätzungen Schwellen in Studien festzulegen.</p>
Glossar S. xiv und 5.6.2.1 (S. 42-43)	<p><b>Anmerkung</b></p> <p><i>Definition „Missing at Random“ (MAR):</i> Im Glossar steht auf S. xiv: „Unter dieser Annahme hängt die Wahrscheinlichkeit für das Fehlen eines Wertes von anderen beobachteten Werten einer Person ab, nicht aber vom fehlenden Wert selbst.“</p> <p>Ein gleichbedeutender Satz steht in Kap. 5.6.2.1 auf S. 42. Diese Definition ist nicht adäquat, stattdessen sollte es im Glossar heißen:</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>„Unter dieser Annahme hängt die Wahrscheinlichkeit für das Fehlen eines Wertes <b>bedingt auf die beobachteten Werte</b> der betreffenden Person nicht vom fehlenden Wert selbst ab.“</p> <p>Entsprechend sollte auch der Satz in Kap. 5.6.2.1 auf S. 42 angepasst werden.</p> <p>Bei der weiteren Beschreibung von MAR wird zwar jeweils gesagt, dass die Berücksichtigung der beobachteten Daten zu einer unverzerrten Schätzung führen kann, aber dies gehört bereits in die Definition.</p> <p>S. zum Beispiel in Carpenter et al 2023 [13] , S. 13.</p> <p><i>Definition „Missing not at Random“ (MNAR):</i></p> <p>Im Glossar steht auf S. xiv:</p> <p>„Die Wahrscheinlichkeit für das Fehlen eines Wertes hängt vom tatsächlichen, fehlenden Wert der betrachten Zielgröße selbst ab.“</p> <p>Derselbe Satz steht in Kap. 5.6.2.1 auf S. 43.</p> <p>Auch diese Definition ist nicht ganz ausreichend, stattdessen sollte es jeweils heißen:</p> <p>„Die Wahrscheinlichkeit für das Fehlen eines Wertes hängt vom tatsächlichen, fehlenden Wert der betrachten Zielgröße selbst ab, <b>und diese Abhängigkeit verschwindet auch nicht durch Bedingen auf die beobachteten Werte der betreffenden Person.</b>“</p> <p>Bei der weiteren Beschreibung von MNAR wird zwar gesagt, dass die Verzerrung nicht korrigiert werden kann, aber dies gehört bereits in die Definition.</p> <p>S. zum Beispiel in Carpenter et al 2023 [13] , S. 17.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Anpassung der Definitionen im Glossar und im Fließtext entsprechend obiger Anmerkungen</p>
5.6.2.2 (S. 47)	<p><u>Anmerkung</u></p> <p>Im Abschnitt „Schwellenwerte für fehlende Werte in Endpunkten“ wird folgendes ausgeführt:</p> <p>„Wenn für &gt; 30 % der Studienpopulation keine Werte vorliegen oder beträgt der Unterschied zwischen den Behandlungsgruppen &gt; 15 Prozentpunkte, so werden die Ergebnisse für den entsprechenden Endpunkt im Regelfall nicht berücksichtigt [34]. Ist der Anteil kleiner als 30 %, so wird endpunktspezifisch über die Konsequenzen hinsichtlich der Ergebnissicherheit diskutiert. Im Falle von Endpunkten mit Messwiederholungen, wie z. B. PROs, sollten für mindestens 70 % der Patientinnen und Patienten sowohl der Wert zu</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Studienbeginn als auch mindestens ein beobachteter Wert zu einem weiteren Zeitpunkt vorliegen. <u>Bei Analysen zu einem bestimmten Zeitpunkt sollten zu diesem Zeitpunkt Werte für mindestens 50 % der Patientinnen und Patienten vorliegen.</u> Diese Schwellenwerte dienen als Orientierung, wenn nicht mit ausreichender Sicherheit davon ausgegangen werden kann, dass die Werte zufällig fehlen.“</p> <p>Wir vermuten, dass der unterstrichene Satz als zusätzliche notwendige Bedingung für Auswertungen mit Messwiederholungen gemeint ist. Insbesondere vermuten wir, dass auch bei Erfüllen dieser Bedingung weiterhin die 70 %-Regel der ausgewerteten Patientinnen und Patienten gelten soll. Dies ist jedoch nicht ganz eindeutig formuliert. Um Missverständnissen vorzubeugen, regen wir daher an, den unterstrichenen Satz als zusätzliche (und somit nicht alternative) Bedingung entsprechend zu präzisieren.</p> <p><u>Vorgeschlagene Änderung</u> Sprachliche Präzisierung, um möglichen Missverständnissen vorzubeugen</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

## Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

- [1] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Systematische Confounderidentifikation in der Indikation schubförmig remittierende multiple Sklerose (RRMS); Arbeitspapier [online]. 2025. Letzter Aufruf: 15.08.2025. <https://doi.org/10.60584/GA23-02>
- [2] Schulz, J., Moodie, E.E.M., Shortreed, S.M. No unmeasured confounding: Known unknowns or... not? Am J Epidemiol, 2023; 192 (9): 1604-1605
- [3] Starke, P. Harm Reduction of Switching from Metformin Plus Sulfonylureas to Metformin Plus DPP4s in Older Adults: A Target Trial Emulation using German Routine Claims Data. 2023. Universitätsmedizin Göttingen, Institut für medizinische Statistik; Masterarbeit.
- [4] Starke, P., Thürmann, P., Grobe, T., Friede, T., Mathes, T. Real-World Harm Reduction of Metformin Plus DPP4 Inhibitors versus Metformin Plus Sulfonylureas in Older Adults: A Target Trial Emulation Using German Claims Data. Drugs Aging, 2025; 42 (7): 655-663
- [5] Feeney, T., Hartwig, F.P., Davies, N.M. How to use directed acyclic graphs: guide for clinical researchers. BMJ, 2025; 388: e078226

- [6] Myers, J.A., Rassen, J.A., Gagne, J.J., Huybrechts, K.F., Schneeweiss, S., et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*, 2011; 174 (11): 1213-1222
- [7] VanderWeele, T.J. Principles of confounder selection. *Eur J Epidemiol*, 2019; 34 (3): 211-219
- [8] Brown, J.P., Hunnicutt, J.N., Ali, M.S., Bhaskaran, K., Cole, A., et al. Quantifying possible bias in clinical and epidemiological studies with quantitative bias analysis: common approaches and limitations. *BMJ*, 2024; 385: e076365
- [9] Lipsitch, M., Tchetgen Tchetgen, E., Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 2010; 21 (3): 383-388
- [10] Hwang, H., Quiroz, J.C., Gallego, B. Assessing the effectiveness of empirical calibration under different bias scenarios. *BMC Med Res Methodol*, 2022; 22: 208
- [11] Victor, N. On clinically relevant differences and shifted null hypotheses. *Methods Inf Med*, 1987; 26 (3): 109-116
- [12] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. 2023. Köln: IQWIG; Version 7.0; Letzter Aufruf: 15.08.2025.  
[https://www.iqwig.de/methoden/allgemeine-methoden\\_version-7-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf)
- [13] Carpenter, J.R., Kenward, M.G., Bartlett, J.W., Morris, T.P., Quartagno, M., et al. Multiple Imputation and its Application. 2nd edition. Chichester: John Wiley & Sons, Ltd. 2023

### **A.1.19 Merck Healthcare Germany GmbH**

#### **Autorinnen und Autoren**

- Giesl, Nicole
- Osowski, Ulrike

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> Bitte nutzen Sie pro Person 1 Zeile.
Dr. med Osowski, Ulrike; Head of Health Technology Assessment & HEOR
Giesl, Nicole; Manager Health Technology Assessment & HEOR
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Merck Healthcare Germany GmbH
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

#### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
z. B. 5.3.1 (S. 13f)	<p><u>Anmerkung</u></p> <p>Das IQWiG führt aus, dass im Gegensatz zu klinischen Studien bei einer AbD keine Fallzahlplanung durchgeführt werden kann, da die notwendigen Vorabinformationen oftmals nicht vorliegen. Folglich soll auf eine orientierende Fallzahlabschätzung zurückgegriffen werden. Weiterhin wird ausgeführt, dass es sinnvoll sein kann, eine vergleichende Studie im Rahmen einer AbD nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlabschätzung darauf auszurichten. Stattdessen soll sich die Schätzung der Fallzahl lediglich darauf beziehen, verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95%-Konfidenzintervall zu generieren.</p> <p>Dies ist in Frage zu stellen, da jede Datenerhebung darauf abzielen sollte klare und verwertbare Ergebnisse zu liefern. Daten sollten generell nicht nur zur Erfüllung formaler Anforderungen gesammelt werden, sondern um einen tatsächlichen Erkenntnisgewinn zu erzielen. Die Generierung von lediglich statistisch signifikanten Ergebnissen, ohne ein Erreichen der Schwellenwerte für eine Aussage anzustreben, ist zudem unter ethischen Gesichtspunkten problematisch. Immerhin wird den Patienten ein zusätzlicher Aufwand durch eine Datenerhebung zugemutet. Dieser sollte in jedem Fall mit einem Erkenntnisgewinn für die medizinische Gesellschaft einhergehen und daher auf die Demonstration eines Vorteils ausgerichtet werden.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p><u>Vorgeschlagene Änderung</u></p> <p>„Ein Aufgreifkriterium für eine AbD ist eine unzureichende Evidenzlage für ein neues Arzneimittel im Vergleich zum therapeutischen Standard. <del>Mit der AbD werden demnach erstmals vergleichende Daten gewonnen. Vor diesem Hintergrund kann es für die Situation einer AbD sinnvoll sein, eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlschätzung auf dieses Ziel auszurichten. Stattdessen ist denkbar, die Schätzung der notwendigen Fallzahl darauf zu beziehen, dass aus einem gewählten Propensity Score Verfahren unter Sicherstellung einer adäquaten Confounderadjustierung verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95 %-Konfidenzintervall resultieren (siehe hierzu auch Abschnitt 5.7), und zwar losgelöst davon, ob sich unter Berücksichtigung der verschobenen Nullhypothese ein Vorteil des zu bewertenden Arzneimittels ableiten lässt. Unter dieser Voraussetzung stellt jedes Ergebnis einer AbD einen Erkenntnisgewinn dar. Eine AbD ist zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlschätzung ist auf dieses Ziel auszurichten.““</del></p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

#### Literaturverzeichnis

Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.

### **A.1.20 MSD Sharp & Dohme GmbH**

#### **Autorinnen und Autoren**

- Colonna, Vanessa
- Eberle, Sonja
- Pfitzer, Lisa
- Ullraum, Anna

## **Stellungnahme zum vorläufigen Rapid Report**

**Berichtnr: A25-13**

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Eberle, Sonja
Pfitzer, Lisa; Dr.
Colonna, Vanessa; Dr.
Ullraum, Anna
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: MSD Sharp & Dohme GmbH
<input type="checkbox"/> als Privatperson(en)

***Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.***

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

MSD begrüßt grundsätzlich die zunehmende Anerkennung der Bedeutung von versorgungsnahen Daten (Real World Evidence) für die Nutzenbewertung von Arzneimitteln. Die Einbeziehung von RWE kann wertvolle ergänzende Erkenntnisse liefern, die über die Ergebnisse kontrollierter klinischer Studien hinausgehen und somit eine realitätsnähere Bewertung des Nutzens und der Sicherheit von Therapien ermöglichen.

Gleichzeitig möchten wir jedoch auf wesentliche Herausforderungen und Limitationen des derzeitigen Konzepts der Anwendungsbegleitenden Datenerhebung (AbD) hinweisen, die in der Praxis vielfach zu erheblichen Problemen führen und die Umsetzung in den meisten Fällen verhindern. Die überwiegende Zahl der bisherigen AbD-Verfahren wurde aus methodischen oder logistischen Gründen eingestellt, weil eine praxisorientierte Durchführung nicht möglich ist.

MSD plädiert daher für eine pragmatische und flexible Weiterentwicklung des gesamten AbD-Konzepts, die den besonderen Herausforderungen der Real World Evidence Rechnung trägt. Nur so kann die Einbindung versorgungsnaher Daten in die Nutzenbewertung tatsächlich einen Mehrwert für Patienten, Leistungserbringer und das Gesundheitssystem leisten.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b>
5.2 (S. 9-12)	<p><b><u>Anmerkung zur Confounderidentifikation und -auswahl:</u></b></p> <p>MSD begrüßt die Bemühungen des IQWiG, im Rahmen des Arbeitspapiers GA23-02 einen methodisch fundierten Ansatz zur Confounderidentifikation und -auswahl zu entwickeln. Die systematische Berücksichtigung von Confoundern ist essenziell, um Verzerrungen in der Auswertung versorgungsnaher Daten zu minimieren und valide Aussagen zur Wirksamkeit und Sicherheit von Interventionen zu ermöglichen.</p> <p>Gleichwohl erscheint das im Arbeitspapier vorgeschlagene Vorgehen, das theoretisch nachvollziehbar und methodisch stringent ist, in der praktischen Umsetzung im Rahmen einer AbD kaum realisierbar. Die einzelnen Schritte zur Identifikation relevanter Confounder sind mit einem erheblichen Zeit- und</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Ressourcenaufwand verbunden. Vor dem Hintergrund des gesetzlich vorgegebenen engen Zeitrahmens für die Durchführung einer AbD – der nur wenige Monate bis zur Fertigstellung des Studienprotokolls und des statistischen Analyseplans umfasst – ist der im Arbeitspapier geforderte Aufwand zur systematischen Confounderidentifikation in der Praxis kaum zu bewältigen. Diese Diskrepanz zwischen theoretischem Anspruch und praktischer Umsetzbarkeit sollte bei der weiteren Ausgestaltung der Anforderungen berücksichtigt werden, um eine realistische und effiziente Durchführung von AbD-Verfahren zu gewährleisten.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>MSD weist darauf hin, dass das vorgeschlagene Vorgehen in der Praxis aufgrund des hohen Ressourcenaufwands und des engen gesetzlichen Zeitrahmens für die Erstellung von Studienprotokoll und Analyseplan häufig nicht umsetzbar ist. Es bedarf alternativer, pragmatischer Ansätze zum Umgang mit Confoundern, um deren Umsetzbarkeit innerhalb von AbD-Verfahren in der Praxis zu verbessern.</p>
5.3 (S. 12-25)	<p><b>Anmerkung zur Fallzahlschätzung:</b></p> <p>Die Forderung, dass der pharmazeutische Unternehmer die finale Fallzahlplanung zum Zeitpunkt der Zwischenanalysen auf Basis vorläufiger Ergebnisse anpassen soll, ist kritisch zu bewerten. Ein derartiges adaptives Vorgehen verschiebt lediglich den Zeitpunkt der Entscheidung über die Realisierbarkeit der AbD nach hinten, ohne das grundsätzliche Problem zu lösen.</p> <p>Im Rahmen der orientierenden Fallzahlschätzung ist es unerlässlich, unter Einbeziehung klinischer Expert:innen vor Beginn der AbD folgende Aspekte kritisch zu bewerten:</p> <ul style="list-style-type: none"> <li>• Ist die erwartete Effektgröße realistisch und klinisch relevant?</li> <li>• Sind die geschätzten Patientenzahlen realistisch und praktisch umsetzbar, insbesondere wenn die Fallzahlschätzung mehrstufig erfolgt oder wenn mehrere Therapieoptionen außerhalb der AbD bestehen, die den Anteil geeigneter Patient:innen reduzieren?</li> </ul> <p>Nur durch eine solche realitätsnahe und interdisziplinär abgestimmte Fallzahlplanung kann die AbD als Instrument der Nutzenbewertung praktikabel und zielführend gestaltet werden.</p> <p><b>Vorgeschlagene Änderung:</b></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	MSD schlägt daher vor, dass eine realistische mit klinischen Expert:innen und allen beteiligten Institutionen abgestimmte Fallzahlschätzung bereits vor Beginn der AbD vorliegen muss, um eine fundierte und praxisnahe Planung zu gewährleisten. Im Einzelfall sollte auch erwogen werden, ob von starren methodischen Vorgaben abgewichen werden kann und flexiblere statistische Annahmen zur Fallzahlschätzung angewendet werden können.
5.5 (S. 33-37)	<p><u>Anmerkung zu den patientenberichteten Endpunkten:</u></p> <p>Die bestehenden bürokratischen Hürden stellen derzeit eine wesentliche Herausforderung bei der Erfassung von patientenberichteten Endpunkten (Patient-Reported Outcomes, PRO) dar und sind im gesetzlich vorgegebenen engen Zeitrahmen für Planung und Durchführung einer AbD nur schwer zu bewältigen. Die zeitlichen Restriktionen erschweren es, eine umfassende und qualitativ hochwertige PRO-Erfassung sicherzustellen.</p> <p>Die Publikation von Singer et al. (Singer et al. 2024) zur Erfassung von PRO-Daten im Rahmen von AbD-Verfahren bei CAR-T-Therapien zeigt, dass die PRO-Erhebung grundsätzlich im Kontext von Registern möglich ist. Gleichzeitig wird jedoch deutlich, dass die Umsetzung mit einem erheblichen personellen, logistischen und vor allem bürokratischen Aufwand verbunden ist.</p> <p><u>Vorgeschlagene Änderung:</u></p> <p>Aus den oben genannten Gründen ist es aus Sicht von MSD notwendig, die Rahmenbedingungen für AbD insbesondere im Hinblick auf die Schwierigkeiten bei der PRO-Erhebung zu überdenken und zu vereinfachen. Nur so kann eine praktikable Integration patientenberichteter Endpunkte gewährleistet werden, die den Anforderungen an eine aussagekräftige Nutzenbewertung gerecht wird.</p>
5.7 (S. 50-55)	<p><u>Anmerkung zu den Propensity-Score-Analysen:</u></p> <p>MSD sieht noch offene Fragen bzgl. Propensity-Score-Analysen zur Confounderadjustierung im Rahmen von AbD insbesondere in Indikationsgebieten mit kleinen Patientenkollektiven. Während im Rapid Report an mehreren Stellen das Propensity-Score-Verfahren als geeignete Methode hervorgehoben wird, wird gleichzeitig die praktische Umsetzbarkeit in solchen Settings infrage gestellt. Diese Diskrepanz führt zu Unsicherheiten hinsichtlich der methodischen Anforderungen.</p> <p>Darüber hinaus wird im Rapid Report empfohlen, die Anzahl der Confounder schrittweise zu reduzieren, um die Analyse handhabbar zu machen. Dieses Vorgehen steht jedoch im Widerspruch zum aufwändigen und umfassenden</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Prozess der Confounderidentifikation, wie er im Arbeitspapier GA23-02 beschrieben wird, bei dem in der Regel eine Vielzahl relevanter Confounder berücksichtigt werden soll. Die Reduktion der Confounderzahl kann zu einer unzureichenden Kontrolle von Verzerrungen führen und somit die Validität der Ergebnisse beeinträchtigen.</p> <p><u>Vorgeschlagene Änderung:</u></p> <p>Vor diesem Hintergrund wäre aus Sicht von MSD eine klarere methodische Orientierung wünschenswert, die sowohl die praktischen Herausforderungen kleiner Patientenkollektive als auch die Anforderungen an eine valide Confounderadjustierung berücksichtigt.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

#### **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. Singer, S., Bayer, O., Schranz, M. et al. „Patient-reported outcomes“ in medizinischen Registern. *Onkologie* **30**, 304–311 (2024). <https://doi.org/10.1007/s00761-024-01494-2>

### **A.1.21 Pfizer Pharma GmbH**

#### **Autorinnen und Autoren**

- Annacker, Vincent
- Böhme, Sarah
- Kauffmann, Stephanie
- Kühne, Felicitas
- Kürschner, Niclas
- Schmitter, Sarah

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> Bitte nutzen Sie pro Person 1 Zeile.
Annacker, Vincent
Böhme, Sarah
Kauffmann, Stephanie
Kühne, Felicitas; Dr.
Kürschner, Niclas
Schmitter, Sarah
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Pfizer Pharma GmbH
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

Die Pfizer Pharma GmbH (nachfolgend: Pfizer) begrüßt die Möglichkeit zur wissenschaftlichen Diskussion des Entwurfes für den Rapid Report „[A25-13] Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“. Pfizer beteiligt sich an der wissenschaftlichen Diskussion und nimmt zu folgenden Punkten Stellung:

- Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen
- Grundlagen der Fallzahlsschätzung mit verschobener Nullhypothese
- Confounderidentifikation- und Analyse
- Propensity-Score-Analysen in Erkrankungen mit kleinen Patientenkollektiven

Kapitel / Abschnitt (Seite) im Dokument	Stellungnahme mit Begründung sowie vorgeschlagene Änderung <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
5.3 Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformationen (S. 12 ff)	<p><u>Anmerkung</u></p> <p>Das IQWiG schlägt vor, zur Einschätzung der Machbarkeit der anwendungsbegleitenden Datenerhebung (AbD) eine orientierende Betrachtung von Fallzahlszenarien durchzuführen, wenn keine hinreichenden Informationen für eine orientierende Fallzahlsschätzung vorliegen.</p> <p>Pfizer weist darauf hin, dass bei einer orientierenden Betrachtung von Fallzahlszenarien wie vom IQWiG angemerkt, die zugrundeliegenden Parameter der Methodik identisch sind, sich aber dennoch durch Austausch von Fallzahl und Effektgröße eine grundsätzlich andere Lesart ergibt.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<p>Bei der orientierenden Fallzahlschätzung wird auf Basis der vorliegenden Evidenz zur Effektgröße und der weiteren Parameter wie Signifikanzniveau und Power die nötige Stichprobengröße geschätzt. Diese ermittelte Stichprobengröße wird dann bei der Beurteilung der Machbarkeit einer AbD in den Kontext der jährlich zu erwartenden Patientenzahl in der entsprechenden Erkrankung gesetzt [1]. Dagegen ist bei einer orientierenden Betrachtung von Fallzahlszenarien die theoretische Anzahl von Patient:innen, die für eine Teilnahme an einer AbD infrage kommen, ausschlaggebend. Dabei wird, anders als bei einer orientierenden Fallzahlschätzung, vorab keine plausible bzw. zu erwartende Effektgröße festgelegt, sondern basierend auf der in der Erkrankung zur Verfügung stehenden Patientenzahl berechnet, welche Effektgrößen aufdeckbar sind. Ob diese Effektgrößen auch nur ansatzweise realistisch sind, bleibt unberücksichtigt.</p> <p>Anhand der im Rapid Report beispielhaft dargestellten Fallzahlszenarien aus dem AbD-Konzept A25-07 zu Epcoritamab [2] zeigt sich diese Problematik. Es ist fraglich, ob diese Szenarien auf realistischen Annahmen hinsichtlich der rekrutierbaren Stichprobengrößen beruhen und ob die aufdeckbaren Effektgrößen tatsächlich zu erwarten sind.</p> <p>Denn auch bei der Annahme, dass die gezeigten Effekte in nicht-randomisierten Beobachtungsstudien grundsätzlich einer systematischen Verzerrung unterliegen, kann nicht von Effektgrößen dieses Ausmaßes (aufdeckbares Hazard Ratio von 0,36 bei N=600) ausgegangen werden. Zumal bereits für einen Großteil der verzerrenden Faktoren adjustiert wird.</p> <p>Pfizer ist der Ansicht, dass die Fallzahlszenarien auf realistischen Annahmen bezüglich der Größe an der AbD vermutlich teilnehmenden Patientenzahl beruhen sollten. Gleches gilt für die orientierende Fallzahlschätzung, die auf realistischen Annahmen bezüglich der Effektstärke beruhen sollte. Dementsprechend ist es nicht zielführend, von einer Vollerhebung der gesamten Population auszugehen.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<p>Offen bleibt auch, inwieweit die Information über die aufdeckbaren Effekte für die Beurteilung der Machbarkeit der AbD überhaupt herangezogen werden kann. Im genannten Beispiel sind bei Einschluss aller Patienten in die AbD, die das IQWiG den Fallzahlszenarien zugrunde legt, ausschließlich sehr deutliche Effektstärken aufdeckbar (das gesamte 95%-Konfidenzintervall des Hazard Ratios müsste unterhalb von 0,3 bis max. 0,4 liegen). Würden weniger Patient:innen als tatsächlich in der Erkrankung vorhanden rekrutiert werden, so müsste der Effekt noch größer sein, um überhaupt aufdeckbar zu sein. Auch wenn davon auszugehen ist, dass eine Überlegenheit gegenüber der zVT vorliegt, ist es fraglich, ob ein dermaßen großer Effekt zu erwarten ist. Dennoch wurde die AbD zu Epcoritamab als machbar eingestuft [2].</p> <p>Weiterhin bleibt offen, inwiefern es adäquat ist, die gemäß IQWiG für eine Fallzahlschätzung unzureichende Evidenz der Abschätzung von Ereignisanteilen als Bestandteil einer orientierenden Betrachtung von Fallzahlszenarien für Intervention und Vergleich zugrunde zu legen. Die Evidenz zur Abschätzung der Ereignisanteile könnte dann auch im Rahmen einer orientierenden Fallzahlschätzung mit verschiedenen Szenarien hinsichtlich der Ereignisanteile genutzt werden, um die Machbarkeit der AbD einschätzen zu können.</p> <p>Insgesamt ist Pfizer der Ansicht, dass die Beurteilung der Realisierbarkeit und Angemessenheit einer AbD gemäß § 54 Abs.1 Nr. 3 VerfO. über die Betrachtung von Fallzahlszenarien nahezu unmöglich ist.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Für die Beurteilung der Realisierbarkeit und Angemessenheit einer AbD gemäß § 54 Abs. 1 Nr. 3 VerfO ist im Rahmen der Konzepterstellung regelhaft eine orientierende Fallzahlschätzung durchzuführen. Bei Unsicherheiten hinsichtlich der zu erwartenden Effekte können für die Fallzahlschätzung verschiedene Szenarien für den Effektschätzer verwendet werden. Diese sollten einen realistischen Bereich abdecken.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
5.3.1 Grundlagen der Fallzahlschätzung mit verschobener Nullhypothese (S.16ff)	<p><u>Anmerkung:</u></p> <p>Das IQWiG geht bei der Fallzahlschätzung im Rahmen einer AbD von einer verschobenen Nullhypothese von 2 bis 5 für das Relative Risiko (bzw. 0,2 bis 0,5 bei mortalitätsenkenden Interventionen) aus. Aus Sicht von Pfizer ist dieses strikte Kriterium überflüssig, da die Anpassung für mögliche Verzerrungen zwischen dem Interventionsarm und dem Kontrollarm bereits durch die Berücksichtigung aller relevanten Confounder erfolgt. Es ist zu bezweifeln, dass der Einfluss möglicher weiterer unbekannter Confounder so groß ist, dass die Effektschätzung maßgeblich verzerrt ist. Daher sollte die verbleibende Rest-Unsicherheit nicht dazu führen, dass das gesamte 95%-Konfidenzintervall über dem vorgegebenen Schwellenwert von 2 bis 5 (bzw. unter 0,2 bis 0,5) liegen muss. Dieser Rest-Unsicherheit wird bereits dadurch Rechnung getragen, dass die Aussagesicherheit per se herabgestuft wird.</p> <p>Es bleibt zudem unklar, wie bei Anwendung der verschobenen Nullhypothese eine Quantifizierung des Zusatznutzens bei erneuter Nutzenbewertung erfolgt bzw. welche Schwellenwerte für die Quantifizierung des Zusatznutzens herangezogen werden.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Pfizer schlägt daher vor, im Fall einer geplanten AbD mit einer hohen Kontrolle von Verzerrungen von einer verschobenen Nullhypothese abzusehen.</p>
5.2 Confounderidentifikation und -auswahl (S. 9 ff)	<p><u>Anmerkung</u></p> <p>Pfizer begrüßt es erstmalig zum Thema „Confounderidentifikation und- Auswahl“ Stellung nehmen zu können.</p> <p>Das IQWiG verweist in seinen Ausführungen auf das Arbeitspapier GA23-02 [3], in dem es auf den Ansatz von Puflete et al. 2022 [4] zurückgreift.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
	<p>Pufulete et al. haben in ihrer Arbeit versucht für Confounder in einer Population, die verschiedene duale Thrombozytenaggregationshemmer erhält, zu kontrollieren - somit für eine Erkrankung mit einer hohen Patientenzahl. Auch das IQWiG hat in seinem Arbeitspapier mit der schubförmig remittierenden multiplen Sklerose eine Erkrankung mit eher höheren Fallzahlen betrachtet. Daher stellt sich die Frage, inwieweit sich der Ansatz auf Erkrankungen mit wenigen Patienten, die meist Gegenstand einer AbD sind, übertragen lässt. Ergänzend zu dieser übergeordneten Problematik möchte Pfizer im Folgenden auf weitere Aspekte eingehen.</p> <p>Als Ergebnis der im Arbeitspapier GA23-02 durchgeföhrten systematischen Confounderidentifikation in der Erkrankung schubförmig remittierende multiple Sklerose (RRMS) wurden 160 bzw. (nach Zusammenführung) 28 Confounder identifiziert. Über den Ansatz nach Pufulete et al. ergibt sich demnach selbst nach Zusammenführung eine hohe Anzahl an Kovariaten.</p> <p>Pfizer bezweifelt, dass die Erhebung zahlreicher potenziell relevanter Confounder, welche bei einer AbD im Rahmen einer Adjustierung zu berücksichtigen wären, in Registern bzw. Studienzentren mit der entsprechenden Datenqualität praktisch umsetzbar ist (z. B. fehlende Werte). Damit einher geht ein erhöhter Dokumentationsaufwand, welcher bereits in vergangenen AbD-Verfahren von Registern grundsätzlich als problematisch erachtet wurde [5, 6].</p> <p>Weiterhin sollen gemäß den Ausführungen im Arbeitspapier GA23-02 Confoundervariablen, die ein inhaltlich gleiches Konstrukt messen oder sich inhaltlich überlappen, zusammengeführt werden. Pfizer hinterfragt, ob dies inhaltlich in allen Fällen zielführend möglich ist.</p> <p>Das Bestreben zur Aufwandsreduktion mit Blick auf die ressourcenintensive systematische Confounderidentifikation wird von Pfizer dennoch begrüßt.</p> <p><u>Anmerkung:</u></p> <p>Des Weiteren möchte Pfizer darauf hinweisen, dass im vorliegenden Rapid Report eine Auseinandersetzung mit dem Konzept der Directed Acyclic Graphs (DAGs) fehlt.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p><b>Vorgeschlagene Änderung:</b></p> <p>DAGs sind ein hilfreicher, effektiver und validierter Ansatz, um Confounding zu vermeiden. Zu diesem Zwecke umfasst er unter anderem die Möglichkeit, ein Minimalset an Variablen zu definieren, welche die sogenannten „backdoopaths“ schließen [7, 8]. Pfizer ist daher der Ansicht, dass dieser Ansatz im Rapid Report berücksichtigt und diskutiert werden sollte.</p>
5.7 Propensity-Score-Analysen in Erkrankungen mit kleinen Patientenkollektiven (S. 50 ff)	<p><b>Anmerkung:</b></p> <p>Pfizer begrüßt die Ausführungen zu Propensity-Score-Analysen, deren Fokus auch der Auftragsspezifisierung des G-BA entsprechen.</p> <p><b>Vorgeschlagene Änderung:</b></p> <p>Im Rahmen der sich ständig weiterentwickelnden Methodendiskussion sollte auch der mögliche Einsatz anderer adäquater Analysemethoden eruiert bzw. die Möglichkeit eines Einsatzes offen gelassen werden.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). *Konzept für eine anwendungsbegleitende Datenerhebung –Brexucabtagen autoleucel.* 2023 Besucht am: 15 August 2025]; Available from: [https://www.g-ba.de/downloads/40-268-10466/2022-07-21\\_AM-RL-XII\\_Autologe-Anti-CD19-transduzierte-CD3-positive-Zellen\\_AbD-008\\_Forderung\\_IQWiG.pdf](https://www.g-ba.de/downloads/40-268-10466/2022-07-21_AM-RL-XII_Autologe-Anti-CD19-transduzierte-CD3-positive-Zellen_AbD-008_Forderung_IQWiG.pdf).
2. Institut für Wirtschaftlichkeit und Qualität im Gesundheitswesen (IQWiG). *Epcoritamab (DLBCL), Bewertung gemäß § 35a SGB V, AbD-Konzept.* 2025 Besucht am: 15 August 2025]; Available from: [https://www.g-ba.de/downloads/40-268-11711/2025-07-17\\_AM-RL-XII\\_Epcoritamab\\_2022-AbD-018\\_Forderung\\_Konzept-IQWiG.pdf](https://www.g-ba.de/downloads/40-268-11711/2025-07-17_AM-RL-XII_Epcoritamab_2022-AbD-018_Forderung_Konzept-IQWiG.pdf).
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). *Arbeitspapier - Systematische Confounderidentifikation in der Indikation schubförmig*

*remittierende multiple Sklerose (RRMS)*). 2025 Besucht am: 15 August 2025]; Available from: [https://www.iqwig.de/download/ga23-02\\_systematische-confounderidentifikation-indikation-rrms\\_arbeitspapier\\_v1-0.pdf](https://www.iqwig.de/download/ga23-02_systematische-confounderidentifikation-indikation-rrms_arbeitspapier_v1-0.pdf).

4. Pufulete, M., et al., *Confounders and co-interventions identified in non-randomized studies of interventions*. Journal of Clinical Epidemiology, 2022. **148**: p. 115-123.
5. Gemeinsamer Bundesausschuss (G-BA). *Zusammenfassende Dokumentation über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buch Sozialgesetzbuch (SGB V) Talquetamab (rezidiviertes und refraktäres multiples Myelom, mind. 3 Vortherapien); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen*. 2024 Besucht am: 15 August 2025]; Available from: [https://www.g-ba.de/downloads/40-268-10656/2024-07-18\\_AM-RL-XII\\_Talquetamab\\_2023-AbD-005\\_Forderung\\_ZD.pdf](https://www.g-ba.de/downloads/40-268-10656/2024-07-18_AM-RL-XII_Talquetamab_2023-AbD-005_Forderung_ZD.pdf).
6. Gemeinsamer Bundesausschuss (G-BA). *Zusammenfassende Dokumentation über die Einstellung eines Beratungsverfahrens nach § 35a Abs. 3b des Fünften Buches Sozialgesetzbuch (SGB V) Brexucabtagen-Autoleucel (rezidivierte oder refraktäre B-ZellVorläufer akute lymphatische Leukämie) Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen*. 2023 Besucht am: 15 August 2025]; Available from: [https://www.g-ba.de/downloads/40-268-9654/2023-07-20\\_AM-RL-XII\\_Einstellung-Beratungsverfahren\\_Brexucabtagen-Autoleucel-2022-AbD-008\\_ZD.pdf](https://www.g-ba.de/downloads/40-268-9654/2023-07-20_AM-RL-XII_Einstellung-Beratungsverfahren_Brexucabtagen-Autoleucel-2022-AbD-008_ZD.pdf).
7. Greenland, S., J. Pearl, and J.M. Robins, *Causal Diagrams for Epidemiologic Research*. Epidemiology, 1999. **10**(1).
8. Tennant, P.W.G., et al., *Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations*. Int J Epidemiol, 2021. **50**(2): p. 620-632.

### **A.1.22 Roche Pharma AG**

#### **Autorinnen und Autoren**

- Knoerzer, Dietrich

## **Stellungnahme zum vorläufigen Rapid Report**

**Berichtnr: A25-13**

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

**Name, Vorname; Titel des/der Stellungnehmenden**

*Bitte nutzen Sie pro Person 1 Zeile.*

Dietrich Knoerzer

**Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)**

- im Namen folgender Institution / Organisation: Roche Pharma AG
- als Privatperson(en)

***Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.***

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

Der Entwurf des Rapid Report ‚Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V‘ wird im folgenden verkürzt als Rapid Report (RR) referenziert.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.*

#### **Kapitel 5.1. S. 8, Positivität)**

##### Anmerkung

Im Entwurf des Rapid Reports finden sich auf S.8 (Sicherstellung der Positivität) Ausführungen zur Positivität, zu denen im Folgenden im Zusammenhang mit S. 24 (Gewährleistung einer ausreichenden Rekrutierung der Studienpopulation) Stellung genommen wird.

Wenn das neue Medikament in der Behandlungsrealität zu einer Einnischung der Patienten führt, ist es nahezu unmöglich Positivität zu erreichen. So können z.B. Patienten mit Ausgangssituation x regelhaft das eine Medikament bekommen, jene mit y regelhaft ein anderes Medikament zugewiesen bekommen.

Dies ist nicht beeinflussbar und kann auch nicht durch bloßes zusätzliches Rekrutieren gelöst werden.

Ein pauschaler Verweis darauf, dass es „im Verantwortungsbereich des pU“ liege, eine ausreichend große Patientenzahl zu rekrutieren, genügt daher nicht, um das inhaltlich/methodische Problem zu adressieren.

##### Vorgeschlagene Änderung

Der Rapid Report sollte auch solche Situationen beschreiben und einen Weg aufzeigen, wie bei nicht herstellbarer Positivität vorgegangen werden soll, und welche Kriterien zu welchen Entscheidungen führen (s.a. RR 2.2.).

Für die Fragestellung wäre eine Art gesamthafter Entscheidungsbaum im Rapid report (ähnlich jenen im RR zur Surrogatvalidierung) hilfreich.

### **Kapitel 5.3.2., S. 20, orientierende Fallzahlserien. Anmerkung 1**

#### Anmerkung

Die ‚Orientierende Betrachtung von Fallzahlserien‘ (RR S. 20) setzt voraus, dass es eine Abschätzung der Patientenpopulation gibt, die durch verschiedene Herangehensweisen angenähert werden. Wird durch das Medikament oder eine sonstige Entscheidung im Gesundheitssystem (z. B. für/gegen ein Screening-Verfahren) die Population so verändert, dass einzelne Teilpopulationen gar nicht mehr entstehen können, hilft eine solche (retrospektive) Patientenpopulationsbestimmung und ihre Fortschreibung nicht.

Eine Internationalisierung der Patientenrekrutierung (RR S. 24) ist hier keine Hilfe: Sollte der Behandlungsalgorithmus dort wesentlich anders sein (z.B. Screening-Verfahren nicht verfügbar), können die Daten aus inhaltlichen Gründen nicht übertragen werden. Ist der Behandlungsalgorithmus gleich (z.B. Screening-Verfahren ebenfalls eingeführt), gibt es keine auch dort keine Daten, die in die AbD integriert werden könnten.

#### Vorgeschlagene Änderung

Es sollte im Rapid Report klar beschrieben werden, wie in solchen Situationen vorgegangen werden soll und welche Kriterien zu welchen Entscheidungen führen.

Für die Fragestellung wäre eine Art gesamthafter Entscheidungsbaum im Rapid report (ähnlich jenen im RR zur Surrogatvalidierung) hilfreich.

## **Kapitel 5.3.2., S. 20, orientierende Fallzahlserien. Anmerkung 2**

### Anmerkung

Die ‚Orientierende Betrachtung von Fallzahlserien‘ (RR S. 20) kann zu unerreichbaren Effektannahmen führen, so dass eine Fallzahlerhöhung für die Fragestellung keine Lösung ist.

Auch hier wäre eine Beschreibung des dann zu erfolgenden Vorgehens hilfreich.

### Vorgeschlagene Änderung

Für die Fragestellung wäre eine Art gesamthafter Entscheidungsbaum im Rapid report (ähnlich jenen im RR zur Surrogatvalidierung) hilfreich.

## **Kapitel 5.5, S. 33, Patientenberichtete Endpunkte**

### Anmerkung

Gemäß Rapid report sollen für die Patientenberichteten Endpunkte Fragebögen verwandt werden, die gemäß etablierter Standards validiert wurden (RR S. 33, Kap. 5.5.). Wenn in der AbD auf Instrumente zurückgegriffen muss, die in dieser Indikation nicht gängige Praxis sind, wird die AbD interventionell.

Es stellt sich die Frage, wie mit Situationen umgegangen werden soll, in denen die klinische Routine keine Fragebögen verwendet, die gemäß der zitierten Standards validiert wurden. Es geht bei dieser Frage insbesondere um die Implikationen für die AbD, die sich durch Änderung der regulatorischen Einordnung ergeben.

### Vorgeschlagene Änderung

Hier ist eine Klärung der Szenarien und der verwandten Kriterien durch das IQWiG notwendig und der Ablauf einer interventionellen AbD deutlich zu beschreiben.

## **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

### **A.1.23 Sanofi-Aventis Deutschland GmbH**

#### **Autorinnen und Autoren**

- Höller, Alexandra
- Vaitsiakhovich, Tatsiana
- Zietze, Henny Anna

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> Bitte nutzen Sie pro Person 1 Zeile.
Höller, Alexandra
Vaitsiakhovich, Tatsiana
Zietze, Henny Anna
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Sanofi-Aventis Deutschland GmbH
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

### **Analytische und statistische Verfahren**

#### Anmerkung

Der vorliegende Bericht A25-13 behandelt zentrale methodische Aspekte anwendungsbegleitender Datenerhebungen (AbD). Dabei liegt ein besonderer Fokus auf spezifischen analytischen und statistischen Verfahren wie Propensity-Score-Analysen (PS) und Multiple Imputation mittels Chained Equations (MICE), die für vergleichende Studien ohne Randomisierung zur Schätzung kausaler Effekte und den Umgang mit fehlenden Daten eingesetzt werden.

Diese methodische Tiefe ist nachvollziehbar und wertvoll. Aus wissenschaftlicher Perspektive könnte es jedoch hilfreich sein, einen breiteren methodischen Rahmen zu erwägen. Anstelle der Hervorhebung einzelner Verfahren wäre es zielführender, einen offenen, kontextabhängigen Ansatz für die Auswahl geeigneter statistischer Methoden zu empfehlen. Dies würde der Vielfalt etablierter Ansätze zur Schätzung kausaler Effekte und zur Imputation fehlender Werte besser Rechnung tragen – insbesondere in Anwendungsfeldern mit kleinen Patientenkollektiven, wie AbD.

#### Vorgeschlagene Änderung

Es könnte erwogen werden, einen breiteren methodischen Rahmen zu empfehlen, der die Auswahl geeigneter statistischer Verfahren offen und kontextabhängig gestaltet, anstatt einzelne Methoden explizit hervorzuheben.

Die Wahl geeigneter statistischer Verfahren sollte sich stets an den spezifischen Forschungsfragen und dem jeweiligen Kontext orientieren [Cousineau, 2023], [Shi, 2024], [Zubizarreta, 2023].

Neben etablierten Ansätzen existiert eine Vielzahl neuartiger Methoden zur kausalen Inferenz, die – abhängig von Datenstruktur und Zielsetzung – wertvolle Alternativen zu PS-Verfahren (Matching, Gewichtung, Stratifizierung oder Regressions-adjustierung) darstellen

können. Dazu zählen optimierungsbasierte Verfahren [Cousineau, 2023], [Zubizarreta, 2012], [Cho, 2018], [Privitera, 2024], G-Computation [Shi, 2024], Targeted Maximum Likelihood Estimation (TMLE) [van der Laan, 2006], Random-Forest-basierte Ansätze [Breiman, 2001], usw [Zubizarreta, 2023].

Für kleinere Patientenkollektive, die für AbD besonders relevant sind, ist eine sorgfältige Auswahl der statistischen Methode von zentraler Bedeutung. Es erscheint empfehlenswert, diese Auswahl auf Basis der erzielten Balance zwischen den Vergleichsgruppen vorzunehmen. Die Methoden von [Hainmueller, 2012], [Zubizarreta, 2012], [Cho, 2018] und [Privitera, 2024] ermöglichen eine exakte Kovariatenbalance zwischen den Behandlungsgruppen, ohne dass ein PS-Modell erforderlich ist. Sie eignen sich besonders für kleine Stichproben und reduzieren die Modellabhängigkeit bei der Schätzung von Behandlungseffekten.

Abschließend lässt sich festhalten, dass eine methodisch offene, evidenzbasierte und kontextabhängige Herangehensweise entscheidend ist, um die Aussagekraft und Validität der Analysen zu stärken.

## Datenquellen

### Anmerkung

Das IQWiG bevorzugt für anwendungsbegleitende Datenerhebungen primär bestehende Patientenregister als Datenquelle. Der Bericht legt den Fokus auf diese Register, da sie als geeignetes Instrument zur Erhebung versorgungsnaher Daten gelten. Andere Datenquellen, insbesondere GKV-Routinedaten, werden derzeit als nicht ausreichend geeignet eingeschätzt, um die Anforderungen an Datenqualität im Rahmen der Nutzenbewertung zu erfüllen. Diese Einschätzung basiert auf methodischen und praktischen Erwägungen hinsichtlich der Validität, Vollständigkeit und Relevanz der Daten für die Bewertung von Arzneimitteln gemäß § 35a SGB V.

### Vorgeschlagene Änderung

Die Fokussierung auf Patientenregister als primäre Datenquelle erscheint zu einseitig. Statt ein bestimmtes Quelletyp (Register) zu bevorzugen, wäre es zielführender, allgemeine Kriterien für „fit-for-purpose“-Datenquellen zu benennen und deren Eignung systematisch zu bewerten [FDA, 2023]. Routinedaten, andere Quellen als auch Daten-Linkage könnten unter bestimmten Voraussetzungen geeignet sein. Eine methodenoffene Bewertung würde der Vielfalt verfügbarer Datenquellen besser gerecht. Die Einschätzung zur Datenqualität sollte evidenzbasiert und kontextabhängig erfolgen, nicht pauschal [NICE, 2022].

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.*

## **Glossar**

### Anmerkung

Die Glossarsektion des Berichts A25-13 enthält hilfreiche Definitionen zentraler Begriffe, jedoch erscheint die Darstellung in Teilen zu stark auf spezifische statistische Methoden fokussiert (z.B. „Estimand“, „Konvergenz“, „Überlappung“).

### Vorgeschlagene Änderung

Das Glossar sollte Begriffe in ihrer allgemeinen, methodenunabhängigen Definition einführen. Eine Verknüpfung mit spezifischen Verfahren kann zu Missverständnissen führen und die Übertragbarkeit auf andere methodische Ansätze einschränken. Eine neutrale und kontextoffene Darstellung der Begriffe würde die Verständlichkeit und Anwendbarkeit verbessern – insbesondere für Leserinnen und Leser mit unterschiedlichem methodischem Hintergrund.

## **Confounderidentifikation und -auswahl (Abschnitt 5.2 [S. 9ff.])**

### Anmerkung

Für den Themenbereich der Confounderidentifikation und -auswahl wurde im vorliegenden Bericht auf das IQWiG-Arbeitspapier A25-13 zurückgegriffen. Die Darstellung im vorliegenden Bericht konzentriert sich auf die daraus abgeleiteten Empfehlungen zur systematischen Identifikation relevanter Confounder für nicht-randomisierte vergleichende Studien. Positiv hervorzuheben ist, dass die Autoren zutreffend auf den hohen methodischen Aufwand bei der Auswahl geeigneter Confounder hinweisen.

Gleichzeitig sollte berücksichtigt werden, dass moderne Verfahren existieren, die mit einer großen Anzahl von Kovariaten umgehen können, ohne eine strikte Unterscheidung zwischen Confoundern und anderen Variablenarten (z. B. Prädiktoren für die Behandlung oder für das Outcome) zu erfordern. Beispiele hierfür finden sich bei [Cousineau, 2023], [van der Laan, 2006], [Zubizarreta, 2023]. Der Einsatz solcher Verfahren kann den Aufwand für die Confounderidentifikation reduzieren, ohne die Validität der kausalen Schätzung zu gefährden.

### Vorgeschlagene Änderung

Ergänzung des Berichts um eine methodenoffene Perspektive, die moderne Verfahren zur Berücksichtigung von Confoundern beinhaltet. Insbesondere sollten Verfahren genannt werden, die mögliche Alternativen zur klassischen Confounderdefinition und -auswahl darstellen. Dies würde dem aktuellen Stand der Methodik Rechnung tragen und den Bericht inhaltlich erweitern.

### **Schätzung notwendiger Fallzahlen vor dem Hintergrund unzureichender Vorabinformation (Abschnitt 5.3 [S. 12ff.])**

#### Anmerkung

Das IQWiG schlägt im Bericht A25-13 vor, „eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlschätzung auf dieses Ziel auszurichten“, sondern die Schätzung der Fallzahl darauf zu beziehen, dass „verwertbare Ergebnisse im Sinne einer Effektschätzung mit zugehörigem 95 %-Konfidenzintervall resultieren“ (siehe Kapitel 5.3, S. 54). Diese Formulierung markiert eine konzeptionelle Verschiebung in der Zielsetzung von AbD. Während der Gemeinsame Bundesausschuss (G-BA) die AbD als Instrument versteht, um „mehr über den Nutzen und Schaden des neuen Arzneimittels herauszufinden“ und damit eine entscheidungsrelevante Evidenzbasis für die Nutzenbewertung zu schaffen, verlagert der hier vorgeschlagene Ansatz den Fokus auf die generelle statistische Verwertbarkeit – unabhängig von einem konkreten Bewertungsziel.

Aus wissenschaftlicher Sicht ist eine Fallzahlschätzung, die sich primär an der Möglichkeit einer Effektschätzung orientiert, nicht ausreichend zielgerichtet. Sie birgt das Risiko, dass Studien durchgeführt werden, deren Ergebnisse zwar methodisch verwertbar sind, aber keine klare Aussagekraft im Kontext der Nutzenbewertung entfalten. Dies kann zu einer Entkopplung von Studiendesign und Bewertungszweck führen und die Effizienz der Evidenzgenerierung im Rahmen des § 35a SGB V beeinträchtigen.

#### Vorgeschlagene Änderung

Die Fallzahlschätzung sollte sich am Erkenntnisinteresse der AbD orientieren und insbesondere darauf abzielen, eine belastbare Aussage zur Nutzenbewertung des neuen Arzneimittels zu ermöglichen. Eine Effektschätzung ohne Bezug zu einem konkreten Bewertungsziel sollte nicht als alleinige Rechtfertigung für die Durchführung einer AbD dienen.

## **Verschobene Nullhypothese (Abschnitt 5.3 [S. 16])**

### Anmerkung

Das IQWiG schlägt im Bericht A25-13 die Anwendung einer verschobenen Nullhypothese als Strategie zur Ergebnisinterpretation in AbD-Studien vor. Da in nicht-randomisierten Studien trotz sorgfältiger Identifikation von Confoundern ein Restrisiko für residuale oder ungemessenes Confounding besteht, soll ein Effekt erst dann als belastbar gelten, wenn er eine definierte Schwelle überschreitet – etwa ein Hazard Ratio (HR) von 2. Diese Vorgehensweise soll die Ergebnissicherheit erhöhen, indem nur starke Effekte als robust gegenüber potenziellen Verzerrungen interpretiert werden.

### Vorgeschlagene Änderung

Die im Bericht formulierte Schwellenanforderung eines starken Behandlungseffekts (z. B. HR = 2) erscheint jedoch im Kontext der modernen Arzneimittelentwicklung nicht realistisch [Rothwell, 2018]. Sie ignoriert die typischen Effektgrößen, wie sie in aktuellen klinischen Studien beobachtet werden, und kann dazu führen, dass klinisch relevante, aber moderate Effekte nicht angemessen berücksichtigt werden [Moon, 2018].

## **Modellkonvergenz bei der Schätzung von Propensity Scores (Abschnitt 5.7 [S. 50ff.])**

### Anmerkung

Im Bericht wird zutreffend darauf hingewiesen, dass bei der Anwendung von PS-basierten Matchingverfahren in Studien mit kleinen Fallzahlen die Anzahl der in das Modell aufgenommenen Confounder begrenzt sein kann. Andernfalls kann es zu Problemen bei der Modellkonvergenz und in der Folge zu einer Nicht-Berechenbarkeit der Effektschätzungen kommen.

Die im Bericht vorgeschlagene Strategie (z. B. Kapitel 5.7, S. 53), weniger wichtige Confounder schrittweise aus dem Modell zu entfernen, stellt jedoch einen potenziell riskanten Ansatz dar, der vor allem aus der Praxis der PS-Modellierung stammt. Diese Vorgehensweise wird häufig gewählt, um Konvergenzprobleme bei einer großen Anzahl von Kovariaten zu vermeiden. Wie es in [VanderWeele, 2019] betont ist, kann eine rein statistisch motivierte Variablenelektion jedoch zu Verzerrungen führen, wenn relevante Confounder versehentlich ausgeschlossen werden.

Alternativ stehen andere Verfahren zur Verfügung, die auch mit einer großen Anzahl von Kovariaten umgehen können, ohne die Modellstabilität zu gefährden. Dazu zählen baumbasierte Verfahren wie Random Forests [Breiman, 2001] sowie Regularisierungsansätze wie Lasso-Regression [Franklin, 2015], sehe auch [Zubizarreta, 2023].

Eine methodenoffene Perspektive würde hier zu robusteren und flexibleren Lösungen führen und die Validität der kausalen Schätzungen stärken.

### Vorgeschlagene Änderung

Überarbeitung der Empfehlung zur schrittweisen Entfernung von Confoundern bei Konvergenzproblemen zugunsten einer Darstellung alternativer Verfahren, die auch bei hoher Kovariatenanzahl stabile Modellschätzungen ermöglichen. Eine solche Ergänzung würde die methodische Vielfalt abbilden und die Validität der Effektschätzungen stärken.

### **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

1. [[Breiman, 2001] Breiman L. Random Forests. Machine Learning, 2001, 45:5–32. <https://doi.org/10.1023/A:1010933404324>
2. [Cho, 2018] Cho WKT. An evolutionary algorithm for subset selection in causal inference models. J Operational Research Society, 2018, 69(4):630–644.
3. [Cousineau, 2023] Cousineau M et al. Estimating causal effects with optimization-based methods: A review and empirical comparison. Eur. J. Oper. Res. 2023, 304(2):367–380. <https://doi.org/10.1016/j.ejor.2022.01.046>
4. [FDA, 2023] FDA. Guidance for Industry, Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products, 2023. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug>
5. [Franklin, 2015] Franklin JM et al. Regularized Regression Versus the High-Dimensional Propensity Score for Confounding Adjustment in Secondary Database Analyses. Am J Epidemiol. 2015, 182(7):651-9. doi:10.1093/aje/kwv108
6. [Hainmueller, 2012] Hainmueller J et al. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis, 2012, 20:25–46. doi:10.1093/pan/mpr025
7. [Moon, 2018] Moon DH et al. Defining a Clinically Meaningful Benefit in Cancer Clinical Trials: From the Perspectives of the Clinical Trialist, Patient, and Society. JNCI Cancer Spectrum, 2018, 2(4):pky039. <https://doi.org/10.1093/jncics/pky039>

8. [NICE, 2022] NICE. Real-world evidence framework, 2022. <https://www.nice.org.uk/corporate/ecd9/chapter/assessing-data-suitability>
9. [Privitera, 2024] Privitera S et al. An evolutionary algorithm for the direct optimization of covariate balance between nonrandomized populations. *Pharm Stat.* 2024, 23(3):288-307. doi:10.1002/pst.2352
10. [Rothwell, 2018] Rothwell JC et al. A study of target effect sizes in randomised controlled trials published in the Health Technology Assessment journal. *Trials*, 2018, 19:544. <https://doi.org/10.1186/s13063-018-2886-y>
11. [Shi, 2024] Shi AX et al. Comprehensive Review and Tutorial on Confounding Adjustment Methods for Estimating Treatment Effects Using Observational Data. *Appl. Sci.* 2024, 14:3662. <https://doi.org/10.3390/app14093662>
12. [van der Laan, 2006] van der Laan MJ et al. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2006, 2(1). <https://doi.org/10.2202/1557-4679.1043>
13. [VanderWeele, 2019] VanderWeele TJ. Principles of confounder selection. *European Journal of Epidemiology*, 2019, 34(3):211–219. <https://doi.org/10.1007/s10654-019-00494-6>
14. [Zubizarreta, 2012] Zubizarreta JR. Using mixed integer programming for matching in an observational study of kidney failure after surgery. *J Amer Stat Assoc*, 2012, 107(500):1360–1371.
15. [Zubizarreta, 2023] Zubizarreta JR, Stuart EA, Small DS, Rosenbaum PR (Eds.). *Handbook of Matching and Weighting Adjustments for Causal Inference* (1st ed.). Chapman and Hall/CRC, 2023.

#### **A.1.24 Servier Forschung und Pharma-Entwicklung GmbH**

##### **Autorinnen und Autoren**

- Kucka, Annett

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> Bitte nutzen Sie pro Person 1 Zeile.
Kucka, Annett
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Servier Forschung und Pharma-Entwicklung GmbH
<input type="checkbox"/> als Privatperson(en)

***Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.***

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte hier ein. Spezifische Aspekte adressieren Sie bitte unter „Anmerkung“ und „Vorgeschlagene Änderung“.

*Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.*

### **Allgemeiner Abschnitt**

Servier begrüßt grundsätzlich die Initiative des IQWiG, mit dem Rapid Report A25-13 eine strukturierte methodische Grundlage für die Generierung und Auswertung versorgungsnaher Daten im Kontext der Nutzenbewertung bereitzustellen. Die systematische Darstellung relevanter Aspekte stellt einen wichtigen Beitrag zur Weiterentwicklung der evidenzbasierten Bewertung von Arzneimitteln dar. Die folgenden Anmerkungen zielen darauf ab, einzelne methodische Punkte im Sinne der praktischen Umsetzbarkeit weiter zu schärfen.

### **Kapitel 5.3, S. 12-24, Fallzahlsschätzung**

#### **Anmerkung**

Die Nutzenbewertung mit begleitender Datenerhebung nach §35a SGB V verfolgt das Ziel, Evidenzlücken zu schließen und den Zusatznutzen eines Arzneimittels quantifizierbar darzustellen. Der Rapid Report A25-13 benennt diese Zielsetzung, geht jedoch nicht im Detail darauf ein, wie die beschriebenen Ansätze zur Fallzahlsschätzung konkret in eine belastbare Quantifizierung des Zusatznutzens überführt werden können.

Der Rapid Report weist zudem darauf hin, dass im Rahmen einer AbD auch ein Erkenntnisgewinn als Studienziel ausreichend sein kann. Diese Perspektive unterscheidet sich jedoch von der gesetzlichen Zielvorgabe nach §35a SGB V, die explizit auf die Ermöglichung einer Quantifizierung des Zusatznutzens abzielt.

Im aktuellen Nutzenbewertungsverfahren wird die Effektstärke in der Regel anhand von Konfidenzintervallen beurteilt. Dabei gilt die Überschreitung eines vordefinierten Schwellenwerts durch die Untergrenze des 95%-Konfidenzintervalls als maßgeblich für die Ableitung eines Zusatznutzens. Für versorgungsnahe Studien mit potenziell erhöhter Unsicherheit erscheint eine transparente Darstellung der Bewertungsgrundlage erforderlich.

### Vorgeschlagene Änderung

Ergänzung eines expliziten Hinweises, dass die Auswahl und Priorisierung der Methodik stets dem Ziel dienen sollten, eine Quantifizierung des Zusatznutzen zu ermöglichen.

Dazu können insbesondere gehören:

- Eine realistische Fallzahlplanung
- Die Operationalisierung geeigneter Endpunkte mit klarer Bewertungsstrategie,
- ggf. die Berücksichtigung alternativer Schwellwerte,
- sowie die Prüfung von Alternativen zur AbD, falls eine belastbare Quantifizierung methodisch oder organisatorisch nicht erreichbar ist.

Zudem könnte auf die Relevanz der Zielsetzung gemäß §35a SGB V hingewiesen werden, um eine einheitliche und nachvollziehbare Herangehensweise an die Fallzahlplanung in AbD-Studien zu fördern.

### **Kapitel 5.2, S. 9-11, Confounderidentifikation und -auswahl**

#### Anmerkung

Die im Rapid Report dargestellten methodischen Empfehlungen zur Confounderidentifikation beziehen sich primär auf strukturierte Literaturrecherchen und Expertenbefragungen. Diese Verfahren sind etabliert, ihre Aussagekraft kann jedoch nach Studiensituation begrenzt sein.

Die ausschließliche Bezugnahme auf GA23-02 und den Ansatz nach Pufulete 2022 wird aus Sicht aktueller internationaler Methodenguidelines, wie Good Practices for Real-World Data, FDA Real-World Evidence Program Guidance und EMA Guideline, als nicht ausreichend angesehen. Es fehlt eine Einordnung möglicher Limitationen dieser Vorgehensweise sowie

eine Diskussion alternativer oder zusätzlicher Verfahren, die zur Akzeptanz der Confounderadjustierung beitragen können.

Darüber hinaus wird im Rapid Report zutreffend beschrieben, dass in bestimmten Versorgungssituationen eine vollständige Identifikation und Adjustierung potenzieller Confounder aufgrund begrenzter Datenverfügbarkeit nicht immer realisierbar ist. In solchen Fällen kann eine transparente Darstellung methodischer Restriktionen zur Einordnung der Ergebnisse beitragen, ohne den Erkenntnisgewinn grundsätzlich infrage zu stellen. Ist jedoch absehbar, dass ein relevanter Erkenntnisgewinn nicht erzielt werden kann, sollte in diesen Fällen eine Prüfung von Alternativen zur AbD ermöglicht werden.

### Vorgeschlagene Änderung

Ergänzung um eine Einordnung, dass die Identifikation und Adjustierung von Confoundern in versorgungsnahen Studien methodische Herausforderungen mit sich bringen kann, die Diskussion dieser Herausforderungen den Erkenntnisgewinn jedoch nicht infrage stellen. Hinweise zum Umgang mit residualer Verzerrung sowie die Möglichkeit der Nutzung datengetriebener Verfahren könnten zur methodischen Absicherung beitragen.

### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

Rapid Report A25-13, IQWiG, Version 1.0, 28.07.2025

Pufulete M., et al. Confounders and co-interventions identified in non-randomized studies of interventions. J Clin Epidemiol 2022; 148:115-123

Berger ML, Sox H, Willke RJ, et al. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness. Value in Health 2017; 1003-1008

Framework for FDA's Real-World Evidence Program. 2018

EMA. Guideline on registry-based studies. EMA/426390/2021

Van der Laan MJ, Rose S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer 2011

### **A.1.25 SmartStep Consulting GmbH**

#### **Autorinnen und Autoren**

- Hild, Clara
- Struck, Karolin
- Volmer, Timm

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Volmer, Timm; Dr.
Struck, Karolin;
Hild, Clara;
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> <b>im Namen folgender Institution / Organisation:</b> SmartStep Consulting GmbH
<input type="checkbox"/> <b>als Privatperson(en)</b>

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

#### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigelegt werden.</i>
5.1 (S. 7)	<u>Anmerkung</u> Die Einschlusskriterien für die Studie mit versorgungsnahen Daten (in dem Fall die Anwendungsbegleitende Datenerhebung [AbD]) sollten laut Tabelle 1 identisch zu einer hypothetischen randomisierten Studie (RCT) sein. Die Selektion der Patient*innen, deren Daten in die AbD einfließen können, würden dementsprechend zwei wesentlichen Einschränkungen unterliegen: Erstens müssen die Leistungserbringer*innen (Ärzt*innen oder Versorgungszentren) an der geforderten AbD, gemäß der Beschränkung der Versorgungsbefugnis teilnehmen (§ 35a Absatz 3b Satz 2 SGB V). Zweitens müssen die Patient*innen den Ein- und Ausschlusskriterien unterliegen, die analog zu den Vorgaben einer hypothetischen RCT definiert wären. Im Besonderen im Fall von Orphan-Drug(OD)-Verfahren, stellt es sich als herausfordernd dar, überhaupt ausreichend Patient*innen einzuschließen und die geforderten Fallzahlen zu erfüllen. Dementsprechend sollte die Wahl der Kriterien möglichst wenige Patient*innen ausschließen.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Zudem soll eine AbD Daten zum Nutzen des Arzneimittels in der Versorgung im deutschen Gesundheitskontext liefern. Eine Restriktion der Patient*innen über das Anwendungsgebiet gemäß Zulassung hinaus, wie es der Fall wäre, wenn Einschlusskriterien einer RCT übernommen bzw. simuliert werden würden, scheint dem zu widersprechen. Vielmehr scheint gerade für den Fall des Einsatzes bei Kindern und Jugendlichen eine Restriktion auf das zugelassene Anwendungsgebiet die Möglichkeit der umfassenden Datensammlung auch zum Zwecke eines eventuell später angemessenen Evidenztransfers (besonders bei anstehender Zulassungserweiterung) sehr einschränkend und nur bedingt sachdienlich. Diese genau abzuwägende Entscheidung gehört in die Hand des behandelnden Arztes. Wenn es eine positive Behandlungsentscheidung gibt, sollte diese auch vollständig dokumentiert und nicht ausgeschlossen werden. Schlussendlich sollte die Fallzahlplanung weiterhin die Patient*innen des Anwendungsgebiets umfassen und für die Nutzenbewertung der Evidenz aus der AbD sollten die entsprechenden Patient*innen berücksichtigt werden.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Die Einschlusskriterien sollten dem Anwendungsgebiet gemäß Zulassung entsprechen und maximal weitere verschreibungsrelevante Aspekte der Fachinformation umfassen. Dadurch kann gewährleistet werden, dass alle Patient*innen, die vom Anwendungsgebiet umfasst werden und von dem Arzneimittel profitieren könnten, tatsächlich behandelt werden können. Wenn Zulassungserweiterungen abzusehen sind, sollte überprüft werden, ob diese Patient*innen ebenfalls in die AbD eingeschlossen werden können.</p>
5.2 (S. 12)	<p><b>Anmerkung</b></p> <p>Die vollständige Erhebung aller relevanter Confounder kann zu sehr umfangreichen Fragebögen zu Studienbeginn führen. Dies erhöht die Wahrscheinlichkeit für fehlende Werte, da entweder nicht alle nötigen Informationen vorliegen oder der Aufwand so hoch wird, dass die Befragung vorzeitig abgebrochen wird.</p> <p><b>Vorgeschlagene Änderung</b></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Für sehr umfangreiche Sets von Confoundern kann eine Reduktion auf die wichtigsten Elemente die Datenqualität erhöhen und das Risiko fehlender Werte gesenkt werden. Diese Reduktion des Confounderset kann zum Beispiel durch die Verwendung gerichteter azyklischer Graphen, also von Modellen, die Variablen über gerichtete Kanten ohne zirkuläre Abhängigkeiten miteinander verknüpfen, vorgenommen und diskutiert werden.</p>
5.3.1 (S. 16)	<p><u>Anmerkung</u></p> <p>Die Fallzahlplanung mit verschobenen Nullhypotesen erwartet einen positiven Effekt im Bereich 2 – 5 für das RR. Dieser Ansatz ist zur Kontrolle des Bias aufgrund unbekannter Confounder etabliert und sinnvoll. Es wird nicht diskutiert, wie mit zu erwartenden Effekten von 1 – 2 oder einer zu kleinen Fallzahl in einer der Behandlungsgruppen umgegangen werden soll. Da AbDs häufig bei OD-Verfahren angefordert werden, ist die Überschreitung der vorgeschlagenen Grenzen von 2 – 5 in der Effektstärke in Zusammenspiel mit nur wenigen verfügbaren Patient*innen und freier Behandlungswahl (im Gegensatz zum RCT-Setting) nur in seltenen Fällen zu erwarten.</p> <p>In Szenarien, in denen sich schon in der pivotalen Studie große Effekte zeigen, und dementsprechend auch große Effekte in einer AbD erwartet würden, ist es wahrscheinlich, dass viele Patient*innen die Behandlung zur neuen Option wechseln und kaum noch Fälle unter der bisherigen Standardtherapie auftreten. Zusätzlich ist es möglich, dass das Confounderprofil zwischen den Behandlungsgruppen deutlich unbalanciert ist. Dadurch ist selbst bei großen Effekten in den häufigsten Anwendungsfeldern der AbD ein großes Konfidenzintervall zu erwarten und ein Nachweis eines Unterschieds zu einer verschobenen Hypothesengrenze nur schwer erreichbar.</p> <p>Die Formulierung einseitiger Hypothesen ist zudem in vielen, aber nicht allen Fällen angebracht. Vor allem für Lebensqualität und Sicherheitssignale ist oft nicht im Vorhinein abschätzbar, ob ein Effekt vorliegt und welche Richtung dieser ausschlägt, da bisherige Studien mit der zu betrachtenden Intervention diese zum Teil nicht erheben oder die Studiendauern zu kurz waren.</p> <p><u>Vorgeschlagene Änderung</u></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Es sollte auf alle zu erwartende Szenarien eingegangen werden, und das Vorgehen für entweder zu geringe zu erwartende Effekte unterhalb eines RR von 2, oder zu große Konfidenzintervalle erläutert werden. Gerade für diese Fälle kann ein einheitliches und präspezifiziertes Vorgehen dabei helfen, aussagekräftige Evidenz im Rahmen einer AbD zu generieren.</p> <p>Es sollte auch auf Szenarien eingegangen werden, in denen die Verwendung einer einseitigen Hypothese nicht angemessen ist.</p>
5.6.1 (S. 41)	<p><u>Anmerkung</u></p> <p>Die Empfehlung des Einsatzes materieller Anreize in Form finanzieller Zuwendungen oder Sachprämien zur Vermeidung fehlender Werte im Rahmen der AbD ist kritisch zu betrachten. Direkte finanzielle Zuwendungen oder Sachprämien können zwar möglicherweise einen wichtigen Anreiz für Patient*innen darstellen, Fragebögen vollständig auszufüllen und schließlich zu einer höheren Qualität der AbD führen. Trotz dessen könnten diese Anreize eine unzulässige Beeinflussung der Patient*innen darstellen und die Neutralität der Datenerhebung gefährden, insofern sie nicht eingeschränkt werden.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Für direkte finanzielle Zuwendungen oder Sachprämien an die Patient*innen sollten folgende Rahmenbedingungen gesetzt werden:</p> <ul style="list-style-type: none"> <li>➔ Alle Patient*innen sollten unabhängig von der erhaltenen Behandlung gleichermaßen entschädigt werden.</li> <li>➔ Die Höhe der finanziellen Zuwendungen oder Sachprämien sollte auf Basis des geschätzten Zeitaufwands für die Erhebung der Endpunkte, beispielsweise durch Fragebögen, begrenzt werden. Denkbar wäre hier eine Obergrenze in Euro pro Stunde, die zudem die Vollständigkeit des Fragebogens oder der erhobenen Daten mit einbezieht.</li> <li>➔ Es muss sichergestellt werden, dass die Antworten und Aussagen der Patient*innen nicht durch die Zuwendungen beeinflusst werden. Die Zuwendungen sollten vollständig unabhängig von den gegebenen Antworten erfolgen und ausschließlich auf der Vollständigkeit der erhobenen Daten basieren.</li> </ul>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

## **Literaturverzeichnis**

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

### **A.1.26 Verband Forschender Arzneimittelhersteller e. V. (vfa)**

#### **Autorinnen und Autoren**

- Bussilliat, Paul
- Rasch, Andrej
- Werner, Sebastian

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

<b>Name, Vorname; Titel des/der Stellungnehmenden</b> <i>Bitte nutzen Sie pro Person 1 Zeile.</i>
Rasch, Andrej; Dr.
Bussilliat, Paul
Werner, Sebastian; Dr.
<b>Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)</b>
<input checked="" type="checkbox"/> im Namen folgender Institution / Organisation: Verband Forschender Arzneimittelhersteller e.V.
<input type="checkbox"/> als Privatperson(en)

**Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.**

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

Im vorliegenden vorläufigen Rapid Report „Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“ (Berichtnr: A25-13) finden sich wichtige Aspekte, die für die forschenden Pharmaunternehmen von Bedeutung sind. Vor diesem Hintergrund ist die Möglichkeit der Kommentierung zu begrüßen.

Die dargelegten Anforderungen sind sehr weitreichend und stellen die Beteiligten bei der AbD vor Herausforderungen, die sich bereits deutlich abzeichnen. Aus Sicht des vfa sollten die Anforderungen verhältnismäßig sein und die Intention der AbD – die Quantifizierung des Zusatznutzens – im Zentrum stehen. Beim Beschluss, ob eine AbD durchgeführt werden soll, muss mit ausreichender Sicherheit angenommen werden können, dass diese ihr Ziel erreicht und praktisch umsetzbar ist. Das ist die notwendige Voraussetzung, die nur erfüllt werden kann, wenn im Zuge der Konzepterstellung und des Fachaustauschs mögliche Zweifel behoben werden können – insbesondere bezüglich der avisierten Fallzahl, was eine realistische Schätzung erforderlich macht. Eine AbD muss der gesetzlichen Bestimmung gerecht werden, sie hat keinen Selbstzweck.

Die folgenden Ausführungen beziehen sich vor allem auf methodische Vorgaben, die praktische Umsetzbarkeit und deren Vereinbarkeit mit den ursprünglichen Zielen der AbD. Dabei werden zentrale Aspekte wie die Confounderidentifikation, die Fallzahlschätzung, die Wahl des Beobachtungsbeginns, der Umgang mit fehlenden Daten und Adjustierungsverfahren adressiert.

Zu den o.g. Aspekten wird im Detail Stellung genommen.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.2 (S. 9ff.)	<u>Anmerkung</u>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Die vorgeschlagene systematische Confounderidentifikation ist in der praktischen Umsetzung sehr aufwändig. Der Aufwand sollte jedoch in einem angemessenen Verhältnis zum Erkenntnisgewinn stehen. Der im referenzierten Arbeitspapier GA23-02 exemplarisch ermittelte Aufwand von 62 Personentagen, zuzüglich des Aufwands für das Projektmanagement (eigene Schätzung: etwa 8 bis 12 Personentage) und für die Erstellung des Berichts (eigene Schätzung: etwa 5 bis 10 Personentage) ist kaum zu rechtfertigen. Die im Arbeitspapier schlussendlich ermittelten 28 potenziellen Confounder hinterlassen darüber hinaus erhebliche Zweifel an der Umsetzbarkeit bzgl. der Erhebung im Register. Bereits die Anzahl relevanter Confounder in den laufenden AbD-Verfahren stellt sich als herausfordernd dar. Die bisherige Höchstanzahl von 22 zu erhebenden Confoundern erscheint bereits über das hinauszugehen, was in einem vertretbaren Verhältnis zum Aufwand steht. Eine hohe Anzahl zunächst zu erhebender Confounder ist auch vor dem Hintergrund der, auf S. 53 des vorliegenden Rapid Reports ausgeführten, schrittweisen Entfernung von Confoundern aus dem Modell im Rahmen der Propensity-Score-Analyse, nicht vertretbar.</p> <p>Insgesamt bleibt unklar, ob das methodische Vorgehen der Fallstudie im Bereich der multiplen Sklerose auf die meist sehr seltenen Erkrankungen im Kontext der AbD übertragbar ist.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Eine zu hohe Anzahl zu erhebender Confounder muss im Sinne der Umsetzbarkeit der AbD vermieden werden. Vor diesem Hintergrund erscheint das Vorgehen zur systematischen Confounderidentifikation noch nicht ausgereift und praktisch nicht durchführbar. Möchte man auch anhand nicht-randomisierter Studien qualitativ hochwertige Evidenz in einem relevanten Zeitraum generieren, ist ein praktikables Verfahren zur Confounderidentifikation notwendig. Nach Möglichkeit sollten Scheinkorrelationen identifiziert und nicht berücksichtigt werden, gesicherte Kausalzusammenhänge wären regelhaft zu berücksichtigen.</p> <p>Möglichkeiten zur weiteren Reduktion sollten, wie auf Seite 12 angeschnitten, weiterführend diskutiert werden. Ob das dargestellte Verfahren im Bereich seltener Erkrankungen anwendbar ist, wäre ebenfalls noch zu prüfen.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Die Ausführungen haben auch Relevanz abseits der AbD. So würden beispielsweise bei der Darstellung eines historischen Vergleichs die zeitintensiven Schritte der Confounderidentifikation ebenfalls erfolgen, müssen aber frühzeitig initiiert werden, damit die Analysen auch ins Dossier integriert werden können. Im Gegensatz zu einer ABD steht der pU vor der Dossiereinreichung aber vor der Herausforderung, dass die systematische Recherche nicht älter als 3 Monate sein darf und somit wiederholt werden muss. Für den Fall, dass dann neue Confounder identifiziert werden, müssten diese auch wieder von Experten evaluiert und in den Analysen im Dossier entsprechend berücksichtigt werden, was in diesem kurzen Zeitraum sehr herausfordernd ist. Es sollte daher auch entsprechend geprüft werden, ob diese Schritte mit anderen Dossieranforderungen vereinbar sind.</p>
5.1.2 (S. 11)	<p><b>Anmerkung</b></p> <p>Die Auswahl der Expertinnen und Experten erscheint nicht strukturiert vorgenommen worden zu sein. Es sollte klarer werden, wie Expertenmeinungen strukturiert herangezogen werden können.</p> <p>Für pharmazeutische Unternehmer ist die Einbindung von externen klinischen Expertinnen und Experten nicht trivial, da jeweils Vereinbarungen und Verträge abgeschlossen werden müssen. Selbst die vom IQWiG beauftragten externen Sachverständigen, die sicherlich in der wissenschaftlichen Community gut vernetzt sind, konnten nicht die für notwendig erachtete Anzahl an klinischen Expert:innen rekrutieren.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Weiterführende Diskussion zur Strukturierung der Expertenidentifikation, z.B. könnte durch eine Versorgungsstrukturanalyse sichergestellt werden, dass auch Ärztinnen und Ärzte Einfluss nehmen können, die nicht an Leitlinien arbeiten, sondern auch außerhalb der großen Zentren die Versorgungsrealität prägen. Mittels Patientenpfades könnte ermittelt werden, welche Facharztgruppen an der Diagnose und Behandlung teilhaben.</p>
5.3.1 (S. 14)	<p><b>Anmerkung</b></p> <p>Es wird ausgeführt, es könne bei „[...] einer AbD sinnvoll sein, eine vergleichende Studie nicht zwingend mit dem Ziel durchzuführen, einen Vorteil des neuen Arzneimittels zu zeigen und die Fallzahlschätzung auf dieses Ziel auszurichten“ und weiter: es stelle „[...] jedes Ergebnis einer AbD einen Erkenntnisgewinn dar“.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Die Durchführung einer AbD, die nicht das Ziel hat, einen Vorteil des neuen Arzneimittels zu zeigen, widerspricht jedoch ihrer Intention (Vgl. RefE des GSAV, S. 50: um eine „bessere Datenbasis zur Bewertung des Zusatznutzens zu erhalten, kann der G-BA [eine AbD] verlangen“ sowie § 130b SGB V, in dem die Folgen skizziert sind, wenn sich „anhand der gewonnenen Daten keine Quantifizierung des Zusatznutzens belegen lässt“). Das Ziel der AbD ist folglich die Quantifizierung des Zusatznutzens, d.h. die Beantwortung der Frage nach der Überlegenheit der neuen Therapie gegenüber der Vergleichstherapie. Nicht jedes Ergebnis einer AbD stellt somit einen gesetzlich intendierten Erkenntnisgewinn dar, sondern nur jenes, das auf eine Quantifizierung des Zusatznutzens abzielt. Zudem tangiert eine solche nicht zielführende Datenerhebung im Versorgungskontext auch ethische Fragen, nämlich inwiefern sie für Leistungserbringer:innen und Patient:innen eine zumutbare Belastung darstellt. Insgesamt ist die Durchführung solcher AbD strikt abzulehnen.</p> <p><b>Vorgeschlagene Änderung</b> Streichung des Absatzes</p>
5.3.1 (S. 16f.)	<p><b>Anmerkung</b></p> <p>Die Ausführungen zur verschobenen Nullhypothese sind kritisch einzuordnen. Nicht sachgerecht ist die Anforderung eines dramatischen Effekts, da sich die Vorgaben auf Konfidenzintervallgrenzen und nicht auf die Punktschätzer beziehen, z. B. in einem Bereich von 2 bis 5 für das relative Risiko (bzw. 0,2 bis 0,5 bei mortalitätssenkenden Interventionen) bezogen auf die untere bzw. obere Konfidenzintervallgrenze.</p> <p>Diese Anforderungen unterscheiden sich somit de facto nicht von den üblichen Anforderungen der Nutzenbewertung gemäß dem Abschnitt 3.2.2 der Allgemeinen Methoden (Dramatischer Effekt), wonach ein dramatischer Effekt vorliegt, wenn das beobachtete relative Risiko von 5 bis 10 vorliegt. Die AbD-Vorgaben für die Effekte berücksichtigen somit weder die zusätzlich zu erfolgende Confounder-Adjustierung noch die Besonderheiten der Therapiesituation (insb. die kleinen Patientenzahlen), die gemäß Auftrag des G-BA zur AbD jeweils vorliegt. Unter der zusätzlichen Berücksichtigung der Vorgaben an die Confounder-Adjustierung sind die Anforderungen an die Effekte in der AbD damit sogar konservativer als in der Nutzenbewertung nach §35 SGB V.</p> <p>Aus Sicht des vfa müssen die Vorgaben für die Generierung und Auswertung von Versorgungsdaten zweckmäßig, aber auch angemessen und praktisch durchführbar sein. Das spiegelt sich hier nicht wider.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<p><b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b></p> <p><i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i></p>
	<p><b>Vorgeschlagene Änderung</b></p> <p>Die methodischen Richtlinien sollten bezüglich ihrer Praxistauglichkeit erneut einer kritischen Prüfung unterzogen werden. Insb. muss von den der Anforderung eines dramatischen Effekts (im Bereich von 2 bis 5 bezogen auf die untere bzw. obere Konfidenzintervallgrenze) Abstand genommen werden, da es sich hier um eine praxisferne, überhöhte Anforderung handelt</p> <p>Alternativ könnte beispielsweise als Kriterium der Überschreitung des Schwellenwertes eines Relativen Risikos von 2 der Punktschätzer anstelle der Grenze des 95%-Konfidenzintervalls verwendet wird. Die notwendige Effektschwelle kann so eher erreicht werden, zudem wäre oft eine geringere Patientenanzahl ausreichend. Das steigert die Umsetzbarkeit der AbD und erhöht die Chancen einer Quantifizierbarkeit des Zusatznutzens.</p>
5.3.2 (S. 18ff.)	<p><b>Anmerkung</b></p> <p>Eine orientierende Fallzahlsschätzung sollte stets Startpunkt der Überlegungen zur Beauftragung einer AbD sein. Der zunehmende Verzicht auf orientierende Fallzahlsschätzungen in Ermangelung gesicherter Vorabinformationen ist daher kritisch zu sehen. Eine verlässliche Fallzahlplanung stellt das Fundament für eine erfolgreiche, ergo umsetzbare und zielführende, AbD dar.</p> <p>Es bleibt unklar, wie aus der "orientierenden Betrachtung von Fallzahlszenarien" eine konkrete Fallzahl abgeleitet werden kann; die Szenarien sind hypothetisch und können nicht direkt zur Studienplanung verwendet werden.</p> <p>Der Stellenwert von Non-Inferiority-/Equivalence-Designs sowie adaptiven und sequenziellen Ansätzen bleibt unklar.</p> <p>Insgesamt wird im Abschnitt zur Fallzahlsschätzung zwar auf die Durchführbarkeit eingegangen, der direkte Zusammenhang mit der Anzahl an Confoundern bzw. deren Skalenniveau wird dort jedoch nicht erläutert.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Ergänzung eines Hinweises, dass auch unter Unsicherheiten sinnvolle orientierende Fallzahlsschätzungen möglich sind und zur Diskussion beitragen. Die Verbindung zw. Fallzahlsschätzung und Anzahl Confounder, bzw. deren Skalenniveau, muss genauer untersucht und erläutert werden, da sich sonst Widersprüche zwischen der Fallzahlplanung und der Confounderidentifikation bzw. Modellwahl ergeben können, die derzeit nicht aufgelöst werden.</p>
5.4.1 (S. 25 ff.)	<p><b>Anmerkung</b></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Den Tag der Therapieentscheidung als Beobachtungsbeginn (<math>t_0</math>) festzulegen, ist in der Versorgungspraxis häufig nicht realisierbar. Die uneingeschränkte Darstellung und hohe Gewichtung der ITT-Analyse erscheint im Kontext anwendungsbegleitender Datenerhebungen daher kritisch, insbesondere bei begrenzter Interpretierbarkeit durch Therapie-Switches. In der Versorgungsrealität kommen Behandlungswechsel in vielen Indikationen wesentlich häufiger vor und sind früher zu erwarten als im Kontext von RCT. Für den Wechsel vom Komparator auf die Intervention bedeutet dies, dass die erwarteten Effekte deutlich kleiner werden.</p> <p>Analysen, die den Behandlungswechseln Rechnung tragen, werden vom IQWiG lediglich als Sensitivitätsanalysen zugelassen. Hierzu ist festzustellen, dass gemäß „ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials“ solche Analysen, die nicht mehr den Treatment policy estimand bedienen, gar keine Sensitivitätsanalysen darstellen, sondern als „supplementary“ Analysis zu interpretieren sind. Sensitivitätsanalysen haben das Ziel, die Robustheit der Hauptanalysen zu unterstützen, indem bestimmte Modellparameter und andere unsichere Faktoren variiert werden. Im Fall von Behandlungswechseln in wesentlichem Umfang kann es jedoch dazu kommen, dass die Hauptanalyse keinen signifikanten Effekt nachweisen kann, hingegen Analysen, die einen anderen Estimand bedienen, dies sehr wohl erreichen. Es wird aus dem Rapid Report jedoch nicht ersichtlich, wie mit solchen Fällen umgegangen werden soll.</p> <p>Im Abschnitt wird sich zudem den verschiedenen Möglichkeiten für die Wahl des Startzeitpunkts der Beobachtung methodisch gewidmet. Bei der Modifikation durch Yang 2022, als Ansatz zur Reduktion zeitbezogener Verzerrungen, werden deutliche Limitationen durch die starke Verringerung der Fallzahl aufgezeigt. In der abschließenden Einordnung wird das PNU-Design empfohlen, sofern das ACNU-Design an seine Grenzen stößt. Die notwendige Voraussetzung, die Sicherstellung der Vollständigkeit und Richtigkeit der Daten zu den Vortherapien im jeweiligen Register, erscheint jedoch häufig nicht erfüllbar. Insgesamt ist von einem hohen Erhebungsaufwand auszugehen (siehe S. 29).</p> <p><u>Vorgeschlagene Änderung</u></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Es bedarf auch hier einer kritischen Reflexion, ob der notwendige Aufwand in einem angemessenen Verhältnis zum Erkenntnisgewinn steht und die Annahmen in der Praxis überhaupt erfüllbar sind. Insgesamt stellen die vorgeschlagenen Verfahren, um eine Übereinstimmung des Beobachtungsbegins, der richtigen Einschlusskriterien und der Behandlungszuteilung zu erzielen, eine große Herausforderung dar, die bei kleinen Patientenkollektiven wenig umsetzbar erscheinen. Diese Problematik wird teilweise vom IQWiG benannt, jedoch wird nicht diskutiert, welche Vorgaben bei kleinen Patientenkollektiven sinnvoll sind.</p> <p>Eine differenziertere Betrachtung wäre geboten, da ITT-Analysen durch die Behandlungswechsel im Versorgungskontext nur begrenzt interpretierbar sind. Ausführungen sind zu ergänzen, wie mit Analysen umzugehen ist, die – anders als die Hauptanalyse – einen signifikanten Effekt nachweisen können, weil sie einen anderen Estimand bedienen.</p>
5.5 (S. 33ff.)	<p><u>Anmerkung</u></p> <p>Die positiven Aussagen zu digitalen PRO-Erhebungen werden vom vfa ausdrücklich unterstützt und könnten zunehmend auch im Setting klinischer Studien relevant werden.</p> <p>Zu begrüßen ist zudem, dass das IQWiG für die PRO-Erhebung bei AbDs Toleranzfenster für die Erhebungszeitpunkte vorschlägt. Die PRO-Erhebung kann aufgrund des beachtlichen Aufwands in der Versorgungsrealität zu hohen fehlenden Werten führen. Toleranzfenster können dazu führen, dass mehr Erhebungen auswertbar sind. Allerdings bleibt die Erhebung von PROs außerhalb von klinischen Studien bzw. in Registern eine Herausforderung.</p>
5.6.1 (S. 37ff.)	<p><u>Anmerkung</u></p> <p>Der Einsatz materieller Anreize in Form finanzieller Zuwendungen (Geschenkgutscheine) oder Sachprämien an die Patientinnen und Patienten (S. 41) muss kritisch hinterfragt werden. Insbesondere bei mehreren AbD innerhalb einer Indikation darf die Therapieentscheidung aus Patientensicht nicht durch finanzielle Zuwendungen beeinflusst werden.</p> <p><u>Vorgeschlagene Änderung</u></p> <p>Die Motivation der Patientinnen und Patienten zum Ausfüllen der PRO-Fragebögen kann durchaus zielführend sein. Hier sind auch andere Lösungen zu diskutieren, der Rahmen des Rapid Reports erscheint dafür jedoch ungeeignet.</p>
5.6.2 (S. 41ff.)	<p><u>Anmerkung</u></p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Der vorgeschlagene methodische Umgang mit fehlenden Werten erfordert Mindestanforderungen an Datenqualität, um valide Modelle zu ermöglichen. Alternativen zur multiplen Imputation (z. B. MMRM, IPW, Zensierung) bleiben unberücksichtigt. Die Definition fehlender Werte im Zusammenhang von Estimands wird nicht diskutiert. Zudem wird die Auswahl der Variablen für das Imputationsmodell nicht ausreichend behandelt.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Mindeststandards für die Datenqualität müssen gewährleistet sein, bevor komplexe Imputationsverfahren eingesetzt werden. Alternativen sollten diskutiert werden.</p>
5.6 (S.48)	<p><b>Anmerkung</b></p> <p>Es werden mehrere Schwellenwerte als Orientierung aus den Simulationsstudien abgeleitet, obwohl die Aussagekraft der Ergebnisse der dargestellten Simulationsstudien hinsichtlich der in AbDs zu erwartenden Datensituationen als limitiert beschrieben wird.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Die skizzierten Schwellenwerte sollten einer praktischen Prüfung zur Übertragbarkeit auf die Datensituation der AbD unterzogen werden. Vorher sind sie lediglich unter Vorbehalt anzuwenden.</p>
5.7 (S. 50–56)	<p><b>Anmerkung</b></p> <p>Propensity-Score-Verfahren sind bei kleinen Fallzahlen methodisch anfällig. Es wird zudem darauf verwiesen, dass bisher keine Evidenz vorliegt, um abschätzen zu können, wann Propensity-Score-Analysen in einer AbD zu interpretierbaren Ergebnissen führen.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Auch hier ist weitergehend zu prüfen, ob eine Anwendung im Kontext der AbD, bei erwartbar kleinen Fallzahlen, zu interpretierbaren Ergebnissen führt.</p>
6 (S. 56)	<b>Anmerkung</b>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>Aufbau und unterhalt einer dauerhaft verfügbaren Dateninfrastruktur wird vom IQWIG als förderlich eingeschätzt, wobei dies auch von pUs unterstützt werden könnte. Der vfa erachtet eine nachhaltige Finanzierung für die langfristige Entwicklung und Pflege der Registerlandschaft als erforderlich. Nutzungsentgelte können dabei als eine zusätzliche Finanzierungssäule dienen. Grundlage ist jedoch eine Basisfinanzierung, die beispielsweise durch die öffentliche Hand geleistet werden muss. Dieser Aspekt fehlt hier.</p> <p><b>Vorgeschlagene Änderung</b></p> <p>Ergänzung um die Notwendigkeit einer Basisfinanzierung, die industrieunabhängig auszustalten ist.</p>
	<p><b>Weitere Aspekte, die keiner spezifischen Stelle im Report zuzuordnen sind</b></p> <p>Die PRO Endpunkte bzw. die Vollständigkeit der Daten wurden adressiert, allerdings ist die Problematik in Registern auch teils bei Sicherheitsendpunkten zu beobachten. Adverse Events werden regelhaft nicht so erhoben, wie in klinischen Studien. Register haben zudem oft unterschiedliche Erhebungsarten von Safety-Events.</p>

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

#### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

### **A.1.27 Vertex Pharmaceuticals (Germany) GmbH**

#### **Autorinnen und Autoren**

- Patchev, Alexandre
- Petry, Dennis
- Schmitt, Susanne
- Wöhling, Heike

## Stellungnahme zum vorläufigen Rapid Report

Berichtnr: A25-13

**Titel: Wissenschaftliche Ausarbeitung ausgewählter Aspekte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V**

*Diese und die folgenden Seiten dieses Dokuments oder Ihre selbst erstellten und beigefügten Anlagen werden in dem Dokument „Dokumentation der Anhörung zum (Berichtsplan / Vorbericht)“ auf der Internetseite des IQWiG veröffentlicht.*

Name, Vorname; Titel des/der Stellungnehmenden

Bitte nutzen Sie pro Person 1 Zeile.

Petry, Dennis

Patchev, Alexandre; Dr. Dr.

Wöhling, Heike; Dr.

Schmitt, Susanne

Die Abgabe der Stellungnahme erfolgt (bitte ankreuzen)

- im Namen folgender Institution / Organisation: Vertex Pharmaceuticals (Germany) GmbH
- als Privatperson(en)

***Die folgende Vorlage für eine Stellungnahme kann beliebig durch eigene Anlagen ergänzt oder ersetzt werden. Diese unterliegen denselben Vorgaben und der Veröffentlichung, wie oben dargestellt.***

### **Stellungnahme zu allgemeinen Aspekten (optional)**

Sofern Sie allgemeine Anmerkungen zum Dokument haben, tragen Sie diese bitte vor der Tabelle ein. Spezifische Aspekte adressieren Sie bitte in der Tabelle.

#### **Anmerkung Vertex:**

Die im vorläufigen Rapid Report des IQWiG dargestellten methodischen Anforderungen können nicht pauschal auf alle Therapiearten angewandt werden. So sind die Vorgaben insbesondere in Studien zu Zell- und Gentherapien (cell and gene therapies, CGT) nicht immer sinnvoll umsetzbar oder mit besonderen Herausforderungen verbunden. Die Besonderheiten, die sich in der Studienplanung für solche Therapien ergeben, sind im vorläufigen Rapid Report nur unzureichend abgebildet.

Bei CGT bestehen häufig Wartezeiten von mehreren Wochen oder Monaten zwischen der Therapieentscheidung und dem tatsächlichen Erhalt der Therapie. Dies kann in der Herstellung der CGT selbst begründet sein, wenn beispielsweise zunächst autologe (Stamm-)Zellen gewonnen und aufbereitet werden müssen und/oder es sind spezifische Vorbehandlungen notwendig, die vor Erhalt der Therapie abgeschlossen sein müssen. Ein Behandlungseffekt durch die jeweilige CGT kann daher – abhängig vom genauen Ablauf der Herstellung und Vorbehandlung – mitunter erst mit großer Verzögerung nach der Therapieentscheidung erwartet werden. Aufgrund dieser Besonderheiten stellt sich die Frage, welcher Zeitpunkt in solchen Fällen als Beobachtungsstart festgelegt werden soll. Eine Primäranalyse über den gesamten Zeitraum ab der Therapieentscheidung, wie im vorläufigen Rapid Report gefordert, führt zu einer verzerrten Schätzung des Therapieeffekts über alle Endpunktkatagorien hinweg und ist somit nicht zielführend.

Des Weiteren werden CGT häufig nur für eine eingeschränkte geeignete Patientenpopulation zugelassen. Die Eignung von Patientinnen und Patienten muss durch umfangreiche Tests sichergestellt werden, die in der Routineversorgung nicht standardmäßig durchgeführt werden. Somit kann in diesen Fällen eine Diskrepanz bezüglich bestimmter klinisch relevanter Parameter zwischen den Patientinnen und Patienten in der Interventionsgruppe und der Kontrollgruppe bestehen, bzw. eine große Unsicherheit beim Einschluss von Patientinnen und Patienten in die Kontrollgruppe, für die die Routineversorgung die notwendigen Daten zur Einschätzung der theoretischen Eignung für die CGT nachvollziehbarweise nicht hergibt.

Diese Besonderheit erschwert die Gewährleistung der geforderten Positivität, d. h. der Eignung aller Patientinnen und Patienten für beide untersuchten Therapieoptionen (Intervention und Komparator). Im vorläufigen Rapid Report wird der Umgang mit dieser Besonderheit nicht thematisiert.

Aus Sicht von Vertex bedarf es einer vertieften Auseinandersetzung mit den spezifischen Herausforderungen von CGT, da die methodischen Anforderungen an die AbD nicht ohne Weiteres auf diese Therapieformen angewandt werden können.

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
5.4.1 (S. 25-27)	<p><b>Anmerkung Vertex</b></p> <p>Das IQWiG merkt zurecht an, dass „die richtige Wahl des Startzeitpunkts der Beobachtung (Index-Datum [t0]) eine der zentralen Herausforderungen [...]“ darstellt. Eine falsche Wahl für t0 kann zu verzerrten Ergebnissen führen. Dem stimmt Vertex zu. Im vorläufigen Rapid Report wird vorgeschlagen, dass t0 als Zeitpunkt der Therapieentscheidung gewählt werden sollte.</p> <p>Vertex merkt an, dass dieser Ansatz nicht für alle Therapien gleichermaßen geeignet ist. Im Beispiel von CGT, die auf autologen Zellen als Ausgangsmaterial basieren, liegt regelmäßig eine teils monatelange Wartezeit zwischen Therapieentscheidung und tatsächlicher Verabreichung der Intervention. Diese Wartezeit führt zu methodischen Herausforderungen, da die Patientinnen und Patienten in der Interventionsgruppe während dieser Zeit weiterhin eine der als Komparatoren festgelegten Therapien erhalten, genau wie die Patientinnen und Patienten in der Kontrollgruppe. Zudem kann der erwartete Zeitpunkt des Wirkeintritts von CGT deutlich nach dem Zeitpunkt der Verabreichung liegen, z. B. weil zunächst ein Engraftment der übertragenen Zellen stattfinden muss. Wird dennoch der Zeitpunkt der Therapieentscheidung als t0 definiert, entsteht eine systematische Verzerrung, da die Zeit zwischen Therapieentscheidung und Gabe der Therapie bzw. erwartetem Wirkeintritt als Teil der „exponierten“ Beobachtungszeit in der Interventionsgruppe gewertet wird, ohne dass eine tatsächliche Exposition gegenüber der Therapie vorliegt.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<p><b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b></p> <p><i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i></p>
	<p><b>Vorgeschlagene Änderung</b></p> <p>Vertex schlägt daher vor, bei CGT (und vergleichbaren Interventionen mit Wartezeit zwischen Therapieentscheidung und -verabreichung, bzw. erwartetem Wirkeintritt) von einer einheitlichen Definition des t0 als Zeitpunkt der Therapieentscheidung abzusehen. Stattdessen wird empfohlen, die Analyse in zwei Phasen zu unterteilen und die Vorbereitungsperiode bei CGT separat zu analysieren und darzustellen:</p> <p>Die Vorbereitungsphase, die in der Interventionsgruppe zum Zeitpunkt der Therapieentscheidung (t01) beginnt und entweder mit der tatsächlichen Verabreichung der Therapie oder mit dem Zeitpunkt des erwarteten Wirkeintritts endet (t02), wird sowohl für die Interventionsgruppe als auch für die Kontrollgruppe separat analysiert. In dieser Phase erhalten in der Regel alle Patientinnen und Patienten eine der als Komparatoren festgelegten Therapien. Der Zeitpunkt t02 wird als Startzeitpunkt für die Hauptanalyse verwendet, da die Behandlungsgruppen erst ab diesem Zeitpunkt hinsichtlich der tatsächlichen Exposition unterscheidbar sind und Effekte in der Interventionsgruppe erst dann auf die Intervention selbst zurückgeführt werden können.</p> <p>Durch eine separate Auswertung der Vorbereitungsphase wird vermieden, dass Effekte und Nebenwirkungen aus dieser Phase fälschlicherweise der Intervention zugeschrieben werden. Dieses Vorgehen ist aus Sicht von Vertex notwendig, um eine sinnvolle Interpretation der Ergebnisse zu gewährleisten. Die transparente Präsentation auch der Daten aus der Vorbereitungsphase wäre hiermit ebenfalls gewährleistet.</p>
5.4.2 (S. 30-32)	<p><b>Anmerkung Vertex</b></p> <p>Das IQWiG merkt an, dass ein Wechsel der Behandlung im Verlauf der AbD potenziell zu Verzerrungen führen kann, insbesondere bei einem Wechsel vom Komparator zur Intervention, im vorläufigen Rapid Report bezeichnet als Szenario B. Dennoch wird gefordert, die Auswertung grundsätzlich gemäß des Intention-to-Treat (ITT)-Prinzips durchzuführen und durch Sensitivitätsanalysen zu ergänzen. Aus Sicht von Vertex ist die ITT-Analyse als Hauptanalyse bei bestimmten Therapien, insbesondere bei hochwirksamen, potenziell kurativen CGT, nicht sinnvoll interpretierbar.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<p>In der Regel handelt es sich bei CGT um einmalig zu verabreichende Therapien, so dass Behandlungswechsel von der Intervention zum Komparator (Szenario A) nicht möglich sind. Es ist jedoch zu erwarten, dass ein relevanter Anteil der Patientinnen und Patienten aus der Kontrollgruppe im Verlauf der Beobachtung zur Intervention wechselt (Szenario B). Die Wahrscheinlichkeit hierfür wird zum einen durch die bessere Wirksamkeit und Aussicht auf eine mögliche Heilung unter CGT erhöht. Zum anderen führt die Positivitäts-Anforderung bei der AbD dazu, dass gemäß der Ein- und Ausschlusskriterien auch für die Kontrollgruppe Patientinnen und Patienten ausgewählt werden, die für die Intervention geeignet sind.</p> <p>Die Zuordnung zu den Behandlungsgruppen abhängig von der Beobachtungszeit unter Therapie mit dem Komparator ab Studienstart (Zeit von Beobachtungsstart [<math>t_0</math>] bis Behandlungswechsel), wie vom IQWiG für CGT vorgeschlagen, ist grundsätzlich sinnvoll. Bei Patientinnen und Patienten der Kontrollgruppe, für die nach dieser festgelegten Identifizierungsperiode eine Therapieentscheidung für die neue Intervention getroffen wird, sollte jedoch eine Zensierung zum Zeitpunkt des Wechsels erfolgen.</p> <p>Wird die Analyse nach dem ITT-Prinzip (Treatment Policy Strategy) durchgeführt, d. h. die Daten von Patientinnen und Patienten aus der Kontrollgruppe fließen auch nach einem Wechsel zur Intervention weiterhin in die Analyse der Kontrollgruppe ein, führt dies zu einer systematischen Verzerrung bei der Schätzung von Behandlungseffekten:</p> <p>Bei Wirksamkeits-Endpunkten fließen Effekte der Intervention fälschlicherweise in die Kontrollgruppe ein und auch mögliche Nebenwirkungen der Intervention werden der Kontrollgruppe zugeordnet. Das ist insbesondere bei einem hohen Anteil an Behandlungswechseln relevant, da so die Kontrollgruppe nach dem ITT-Prinzip zunehmend der Interventionsgruppe entspricht und man die Intervention letztlich nicht mehr mit dem Komparator vergleichen würde, sondern mit einer verzögerten Therapie mit der gleichen Intervention. So beantwortet die ITT-Analyse nicht die relevante Fragestellung, sondern bildet eine „Mischgruppe“ ab, die keine echte Vergleichsbasis mehr darstellt. Die ITT-Analyse ist daher nicht sinnvoll zu interpretieren.</p>

<b>Kapitel / Abschnitt (Seite) im Dokument</b>	<b>Stellungnahme mit Begründung sowie vorgeschlagene Änderung</b> <i>Die Begründung sollte durch entsprechende Literatur belegt werden. Die Literaturstellen, die zitiert werden, müssen eindeutig benannt und im Anhang im Volltext beigefügt werden.</i>
	<b>Vorgeschlagene Änderung</b> Aus Sicht von Vertex ist es notwendig, in Szenarien mit hoher Wahrscheinlichkeit für den Behandlungswechsel zur Intervention eine Analyse nach der Hypothetical Estimand Strategy als Hauptanalyse durchzuführen. Dabei werden Patientinnen und Patienten in der Kontrollgruppe zum Zeitpunkt des Behandlungswechsels zur Intervention zensiert, sodass der Effekt unter der Annahme geschätzt wird, dass sie die ursprüngliche Behandlung beibehalten hätten. Dadurch wird der tatsächliche Behandlungseffekt ohne Vermischung der Interventions- und Kontrollgruppe korrekt abgebildet.

(Bitte fügen Sie weitere Zeilen an, falls dies notwendig sein sollte.)

#### Literaturverzeichnis

*Bitte verwenden Sie zur Auflistung der zitierten Literatur möglichst eine nummerierte Referenzliste und behalten Sie diese Nummerierung bei der Benennung der Dateien (Volltexte) bei.*

Nicht zutreffend, da keine Literatur zitiert wurde.