Aktuelle Nutzenbewertung von „CARVYKTI“ als Beispiel für einen AMNOG-Reformbedarf

# Die „Target Trial Emulation“ zu CAR-T im AMNOG

Das Konzept *„Target Trial Emulation“* sowie *propensity score*-basierte Adjustierungsverfahren sind empfohlene Methoden, um in indirekten Vergleichen ohne Brückenkomparator systematische Verzerrung zu minimieren. Anhand des Vergleichs der einarmigen Evidenz für das CAR-T Ciltacel mit Daten der Studie LocoMMotion, die in der medizinischen Fachwelt als beste prospektive Studie zum Versorgungsstandard bei vielfach vorbehandeltem multiplem Myelom anerkannt wurde, wird die Anwendung des Konzepts im AMNOG vorgestellt. Der Vergleich und der damit gezeigte dramatische Überlebensvorteil unter Ciltacel, dessen Validität und klinische Bedeutung durch medizinische Experten bestätigt wird, stehen der vollständigen Ablehnung durch den G-BA gegenüber. Diese Diskrepanz lohnt einen Rückblick.

>> Die Chimäre Antigenrezeptor (CAR)-T-Zelltherapie Ciltacabtagen Autoleucel (kurz: Ciltacel, Handelsname: „CARVYKTI“) des Herstellers Janssen-Cilag GmbH, ein Arzneimittel zur Behandlung seltener Leiden (Orphan Drug), ist zur Therapie des rezidivierten und refraktären multiplen Myeloms nach mindestens drei Vortherapien zugelassen. Die Zulassung basiert auf der einarmigen Phase 1b/2 Studie CARTITUDE-1 (CAR1). Für einen validen und aussagekräftigen Vergleich gegenüber dem derzeitigen Versorgungsstandard unter Verwendung des Konzepts target trial emulation wurde die prospektive nicht interventionelle klinische Studie LocoMMotion (LOCO) mit gleichen Einschlusskriterien initiiert. Zusätzlich liegen mit der CARTITUDE-4 (CAR4) erste Daten einer randomisierten kontrollierten Studie (RCT) zu Ciltacel vor. Ein Teil der Studienpopulation fällt in das derzeit zugelassene Anwendungsgebiet und wurde unterstützend mit der *„Target Trial Emulation“* zusammengeführt. Diese Evidenz war Grundlage für die Nutzenbewertung des Gemeinsamen Bundesausschusses (G-BA) mit Beschluss im August 2023. Im Folgenden wird anhand des Endpunktes Gesamtüberleben die Analysestrategie dargestellt und die Bewertung durch den G-BA kritisch diskutiert.

## Die *„Target Trial Emulation“* der Studien CAR1 und LOCO

Die offene, multizentrische, einarmige Studie CAR1 umfasst N=126 Patienten[1], von denen N=124 Patienten die Kriterien des zugelassenen Anwendungsgebietes erfüllen. In der Nutzenbewertung lagen Daten mit einer medianen Nachbeobachtungsdauer von 28,55 Monaten vor.

Die nicht interventionelle, klinische Studie LOCO umfasst N=248 Patienten, von denen N = 110 sowohl die Kriterien des zugelassenen Anwendungsgebietes als auch die in CAR1 berücksichtigten Schwellenwerte für Laborparameter erfüllen und eine Therapie erhalten, die vom G-BA als Vergleichstherapie definiert wurde. Es lagen Daten mit einer medianen Nachbeobachtungsdauer von 26,9 Monaten vor.

Die RCT CAR4 vergleicht Ciltacel mit den Kombinationstherapien Daratumumab-Pomalidomid-Dexamethason bzw. Pomalidomid-Bortezomib-Dexamethason nach ärztlicher Wahl und umfasst N = 208 vs. N = 211 Lenalidomid-refraktäre Patienten mit einer bis drei Vortherapien, von denen N = 20 vs. N = 29 das Kriterium (mind. drei Vortherapien) des zugelassenen Anwendungsgebietes erfüllen. Es lagen Daten der ersten Interimsanalyse mit einer medianen Nachbeobachtungsdauer von 15,9 Monaten vor. Grundsätzlich wird das Zusammenführen von Evidenz aus randomisierten und nicht randomisierten Studien zur Erhöhung der Aussagekraft befürwortet (2). Die gepoolten Analysen geben somit eine Möglichkeit, die Konsistenz der resultierenden Effekte zu beurteilen.

Für die *„Target Trial Emulation“* aus CAR1 und LOCO ergibt sich eine Analysepopulation von N = 124 vs. N = 110 Patienten. Unter zusätzlicher Berücksichtigung der CAR4 sind N = 144 vs. N = 139 Patienten in der Analysepopulation.

## Statistische Methoden zur Adjustierung der Vergleichsgruppen

Es werden die nicht adjustierte sowie verschiedene adjustierte Analysen dargestellt. Die Adjustierung für Confounder erfolgte mittels *propensity score* (PS)-basierten Methoden, welche in der wissenschaftlichen Literatur als Standardverfahren etabliert sind (3). Dabei wurde für jeden Studienteilnehmer ein PS-Wert basierend auf der individuellen Ausprägung der berücksichtigten Confounder berechnet. Die resultierende Verteilung der PS für die Behandlungs- bzw. Kontrollgruppe und der Überlappungsbereich dieser beiden PS-Verteilungen sind ein Anhaltspunkt für die Vergleichbarkeit der Gruppen. Die verschiedenen durchgeführten Methoden adressieren unterschiedliche Aspekte einer potenziellen Verzerrung, sodass die Robustheit und Aussagekraft der Ergebnisse in der Gesamtschau aller Analysen beurteilt werden können.

## Evaluation der Qualität und Aussagekraft einer *„Target Trial Emulation“*

Um die Aussagekraft des Vergleichs von Ciltacel gegenüber dem Versorgungsstandard im Rahmen der *„Target Trial Emulation“* zu beurteilen, müssen neben der klinischen Relevanz die Qualität der Evidenz und insbesondere die aus der Vergleichbarkeit der Studienpopulationen resultierende Positivität geprüft und diskutiert werden.

---

***1:*** Umfasst alle apheresierten Patienten beider Kohorten (USA und Japan)

## „Target Trial Emulation“

Die *„Target Trial Emulation“* sieht die Nachbildung einer RCT auf Basis von nicht randomisierten Daten vor. Das Konzept wurde 2016 von Hernán und Robins entwickelt, um die Limitationen von nicht randomisierten, indirekten Vergleichen zu minimieren (1). Seitdem wird das Konzept z. B. auch unter Verwendung versorgungsnaher Daten vom IQWiG für die Nutzenbewertung empfohlen (2).

Durch ein adäquates Target-Trial-Design unter Berücksichtigung verschiedener Eigenschaften randomisierter Studien (Tab. 1) kann die Strukturgleichheit zwischen den zu vergleichenden Populationen maximiert werden, was maßgeblich für die Aussagekraft eines Vergleiches ist. Während in randomisierten Vergleichen Strukturgleichheit abgesehen von Zufallsschwankungen garantiert ist, muss in indirekten Vergleichen das Ausmaß der Strukturgleichheit geprüft und ggf. durch methodische bzw. statistische Maßnahmen erhöht werden. Im Rahmen der *„Target Trial Emulation“* ist eine umfangreiche Evaluation und Anpassung der Strukturgleichheit möglich, z. B. mithilfe von PS-basierten Methoden.

| Zusammenfassung der Komponenten für die „Target Trial Emulation“ | |
|---|---|
| **Protokollkomponente** | **Beschreibung** |
| Einschlusskriterien | Die Einschlusskriterien der Studien sind weitestgehend identisch. Aus der Studie LOCO wurden Patienten ausgeschlossen, die nicht die in der Studie CAR1 festgelegten Schwellenwerte für Laborparameter erfüllten. |
| Zu untersuchende Behandlung | Die Patienten erhielten Ciltacel zulassungskonform. Aus der Studie LOCO wurden Patienten ausgeschlossen, die keine der vom G-BA definierten adäquaten Vergleichstherapien erhalten haben. |
| Behandlungszuweisung | Um die für einen Vergleich erforderliche Strukturgleichheit annähern zu können, muss eine Adjustierung relevanter Confounder erfolgen (3). Die Identifikation von Confoundern erfolgte in mehreren Schritten mithilfe systematischer Literaturrecherche und Experteninterviews und wird ausführlich im Dossier zur Nutzenbewertung beschrieben. Relevante Confounder (s. Tab. 4-37 im Nutzendossier) wurden in das Adjustierungsset für die vergleichende Analyse eingeschlossen (Hauptadjustierungsset). Im Rahmen einer im Dossier dargestellten Sensitivitätsanalyse wurden darüber hinaus alle identifizierten Confounder in das Adjustierungsset eingeschlossen (4). |
| Beobachtungsdauer | In der Studie CAR1 ist Baseline definiert als Zeitpunkt der Apherese. In der Studie LOCO ist Baseline definiert als Zeitpunkt der ersten Verabreichung der Studienmedikation. Die mediane Beobachtungsdauer ist vergleichbar. |
| Endpunkte | Die in den Studien CAR1 und LOCO erhobenen Endpunkte sind identisch operationalisiert. |
| Interessierender Vergleich | Das Analyse-Set entspricht dem ITT-Ansatz, da alle Studienteilnehmer gemäß Behandlungsgruppe analysiert wurden, unabhängig davon, ob sie diese Behandlung tatsächlich erhalten haben. |
| Analyseplan | Die vergleichende Analyse war in einem statistischen Analyseplan präspezifiziert und darauf ausgerichtet, den Anforderungen der Nutzenbewertung zu entsprechen. |

***Tab. 1:*** Zusammenfassung der Komponenten für die „*Target Trial Emulation*“ aus CAR1 und LOCO.

Positivität besteht, wenn jeder Patient in der Analyse-Population gemäß der Ausprägung seiner Charakteristika eine positive Wahrscheinlichkeit hat, sowohl der Interventionsgruppe als auch der Kontrollgruppe zugeordnet zu werden. Weitere diagnostische Kriterien zur Evaluation der methodischen Qualität des Vergleichs sind Überlappung der PS-Verteilung und Balanciertheit der Patientencharakteristika. Die Überlappung beschreibt den Anteil der Patienten aus beiden Behandlungsarmen, für die unter Berücksichtigung der Confounder PS vergleichbarer Größe berechnet wurden. Eine Abweichung von der Gleichverteilung der PS ist nicht gleichzusetzen mit fehlender Positivität. Balanciertheit drückt die Strukturgleichheit der Vergleichsgruppen im Hinblick auf die Ausprägung der Confounder nach Anwendung PS-basierter Methoden aus.

Vor der Anwendung statistischer Verfahren muss überprüft werden, dass keine strukturelle Verletzung der Positivität vorliegt. Eine strukturelle Verletzung besteht, wenn z. B. aufgrund von Kontraindikationen eine theoretische Unmöglichkeit für eine Therapiezuordnung vorliegt. Eine Verletzung kann auch aus Zufallsschwankungen in der Stichprobe resultieren (5).

Auf Grundlage von Expertenwissen ist kritisch zu prüfen, inwiefern die theoretische Unmöglichkeit einer Therapiezuordnung (strukturelle Verletzung der Positivität) ausgeschlossen werden kann, sodass eine Adjustierung vor dem Hintergrund struktureller Gegebenheiten zulässig ist (6). Die grafische Darstellung der PS-Verteilung sowie die Betrachtung der Patientencharakteristika geben eine Hilfestellung zur Beurteilung von Positivität. Verteilungen mit geringem Überlappungsbereich sowie unterschiedliche Ausprägungen von Charakteristika können Anzeichen einer Verletzung der Positivitätsannahme sein, die durch klinische Experten bewertet werden müssen. Darüber hinaus soll die Verringerung der Verzerrung durch Confounding durch die statistischen Modelle gegenüber dem Einfluss einer möglichen Verletzung der Positivitätsannahme abgewogen werden (6).

## Diskussion von Positivität und Strukturgleichheit im Ciltacel-Vergleich

Das Studienprotokoll der Studie LOCO basiert weitgehend auf dem Protokoll der CAR1. Die Einschlusskriterien wurden größtenteils übernommen. Eine Ausnahme bilden Schwellenwerte für Laborparameter, die *post hoc* auf die LOCO-Population angewendet wurden. Daher ist für den Vergleich zwischen CAR1 und LOCO strukturelle Positivität gegeben. In den PS-Verteilungen des naiven Vergleichs zeigt sich, dass in allen PS-Intervallen jeweils Patienten aus CAR1 und LOCO zu finden sind (Abb. 1 A). Folglich besteht für jeden Patienten in der Analyse eine positive Wahrscheinlichkeit, jede der beiden zu vergleichenden Therapien zu erhalten. Die Adjustierung ist somit zulässig. Die Beschreibung der (Pseudo-)Populationen anhand von Patientencharakteristika zeigt, dass durch Adjustierung eine hohe Balanciertheit zwischen CAR1 und LOCO erreicht werden konnte (4). Die PS-Verteilung (Abb. 1) verdeutlicht die Erhöhung der Strukturgleichheit durch Adjustierung.

## PS-Matching-Ergebnisse unterstreichen Validität des Überlebensvorteils

Mit der Methode des 1:1-PS-Matching werden basierend auf den PS Paare von Patienten aus Kontroll- und Interventionsgruppe gebildet, was der Nachahmung einer 1:1-Randomisierung entspricht. Unter der im PS-Matching gegebenen Strukturgleichheit innerhalb der zu vergleichenden Analyse-Populationen wird der eindeutige Überlebensvorteil von Ciltacel bestätigt (HR: 0,33, 95% KI: [0,18-0,62]). Vor dem Hintergrund der reduzierten Fallzahl beim Matching (124 vs. 110 zu 48 vs. 48) muss evaluiert werden, inwiefern die Ergebnisse der PS-Matching-Population auf die Gesamtpopulation übertragbar sind. Zwischen den Patienten in der gematchten Stichprobe und den nicht gematchten Patienten zeigen sich weder für Ciltacel noch für den Versorgungsstandard signifikante Unterschiede in der Überlebensprognose. Die Übertragbarkeit wird zusätzlich durch die hohe Konsistenz der PS-Matching Ergebnisse, sowohl im Vergleich zu der naiven Analyse als auch im Vergleich zu den PS-gewichteten Analysen, unterstrichen und erlaubt somit belastbare Aussagen über das Vorliegen eines Überlebensvorteils zugunsten von Ciltacel in der Gesamtpopulation.

## Konsistente Analyseergebnisse belegen dramatischen Überlebensvorteil unter Ciltacel

Die Ergebnisse zum Gesamtüberleben zeichnen ein über alle Adjustierungsmethoden und über beide Vergleichskonstel-

| Propensity Score-basierte Adjustierungsmethoden | |
|---|---|
| Average Treatment Effect (ATE) | Patienten des Kontrollarms (LOCO +/- CAR4) und des Interventionsarms (CAR1 +/- CAR4) werden basierend auf ihren PS mittels Inverse Probability Weighting (IPW) gewichtet, um Strukturgleichheit herzustellen. Hierbei wird eine Pseudo-Population generiert, welche die durchschnittlichen Charakteristika beider Vergleichsgruppen widerspiegelt. |
| Average Treatment Effect Among the Treated (ATT) | Nur Patienten des Kontrollarms (LOCO +/- CAR4) werden basierend auf ihren PS mittels IPW gewichtet, um Strukturgleichheit mit dem Interventionsarm herzustellen. Nach der Gewichtung wird auf die ursprüngliche Patientenzahl zurückskaliert. Hierbei wird eine Pseudo-Population generiert, welche die Charakteristika der Interventionsgruppe widerspiegelt. |
| Trunkierung und Trimmung bei ATE- und ATT-Analysen | Sehr große bzw. sehr kleine PS führen zu extremen Gewichten, die die Robustheit der statistischen Modelle beeinträchtigen können. Um extreme Gewichte zu adressieren, werden die Verfahren Trunkierung und Trimmung angewandt. Als Schwellenwerte werden hierbei die 2,5%- bzw. 97,5%-Perzentile der PS-Verteilung herangezogen. Bei der Trunkierung werden die Gewichte der Patienten, die außerhalb dieses Bereichs liegen, durch ein Minimal- bzw. Maximalgewicht entsprechend der 2,5%- bzw. 97,5%-Perzentile ersetzt. Bei der Trimmung werden Patienten, die jenseits der zuvor genannten Perzentile liegen, aus der Analysepopulation ausgeschlossen. Durch das Trimmen wird eine Pseudo-Population gebildet, in der Strukturgleichheit erhöht wird und stellt eine Annäherung an den Überlappungsbereich dar. |
| 1:1-Matching | Beim PS-Matching werden Paare mit jeweils einem Patienten aus dem Ciltacel-Arm und einem Patienten aus dem Vergleichsarm gebildet, wobei jeweils ein Partner mit der optimalen Entfernung des PS gewählt wird. Die Paare bilden eine Pseudo-Population mit Strukturgleichheit, die einen Überlappungsbereich der PS-Verteilung darstellt. |

***Tab. 2:*** Übersicht der *Propensity Score*-basierten Adjustierungsmethoden.

lationen homogenes Bild eines deutlichen Überlebensvorteils zugunsten von Ciltacel (Abb. 2). Die über unterschiedliche Analysepopulationen sehr präzisen und konsistenten Effektschätzer zeigen, dass eine Therapie mit Ciltacel gegenüber dem Versorgungsstandard die Überlebenswahrscheinlichkeit etwa verdreifacht, und liegen damit in der Größenordnung eines dramatischen Effektes. Im Kontext nicht randomisierter Evidenz unterstreicht dies, dass die beobachteten Unterschiede tatsächlich auf die Wirkung der untersuchten Therapie und nicht auf systematische Verzerrung zurückzuführen sind (2). Zudem führen Fallzahlverluste bei Matching und Trimmen nicht zu einer größeren Unsicherheit, da das Ergebnis über alle Adjustierungsmethoden und Stichproben konsistent ist. Der deutliche Überlebensvorteil unter Ciltacel spiegelt sich trotz der kürzeren Nachbeobachtungszeit auch in der relevanten Subpopulation der RCT CAR4 wider (Hazard Ratio 0,42). Somit sind indirekte und direkte Evidenz konsistent. Eine gemeinsame Auswertung mit CAR4 wurde im Rahmen der Stellungnahme zur Nutzenbewertung vorgelegt und zeigt ebenfalls konsistente Ergebnisse.

## Klinische Experten bestätigen Vergleichbarkeit von CAR1 und LOCO

In der Nutzenbewertung durch den Fachbereich Medizin bewertet der G-BA die Studie LOCO aufgrund von Selektionsbias als ungeeignet für einen indirekten Vergleich mit der Studie CAR1. Ausschlaggebend hierfür sei unter anderem, dass vor der Entscheidung für eine CAR-T-Therapie zusätzlich zu den Kriterien des Anwendungsgebiets auch die Eignung für eine CAR-T-Therapie individuell zu beurteilen ist. Der G-BA sieht darin einen Hinweis auf einen Selektionsbias und infolgedessen fehlende Positivität. Als weitere Begründung für die Annahme eines Selektionsbias führt der G-BA auf, dass die Studien CAR1 und LOCO in unterschiedlichen Regionen durchgeführt wurden, und verweist zum anderen auf eine Empfehlung der S3-Leitlinie, Patienten mit multiplem Myelom eine Behandlung im Rahmen klinischer Studien anzubieten. Diese führe möglicherweise zu einer Trennung von Patienten, die entweder für interventionelle klinische Studien oder anwendungsnahe Studien geeignet seien (4).

Die Frage, ob von einem starken Selektionsbias und infolgedessen von fehlender Positivität auszugehen ist, war Gegenstand des Stellungnahmeverfahrens zur Nutzenbewertung. Insbesondere in der mündlichen Anhörung zeichnete sich, abweichend von der Bewertung des G-BA, eine klare Bestätigung der Vergleichbarkeit der Patientenpopulationen innerhalb der *„Target Trial Emulation"* sowie der sich hieraus ergebende Validität des Vergleiches durch die klinischen Experten ab. Des Weiteren ging aus den Stellungnahmen der Experten hervor, dass in der Durchführung der Studien in unterschiedlichen Regionen oder in der Empfehlung zur Teilnahme an klinischen Studien keine Ursachen eines Selektionsbias gesehen werden (4). Von einer strukturellen Verletzung der Positivität ist folglich nicht auszugehen. Diese klinische Sicht ist entscheidend für die Validität des Vergleichs (6, 7). Der G-BA nimmt die gegenteilige Einschätzung der klinischen Experten zur Kenntnis, hält aber dennoch in seiner abschließenden Bewertung an seiner ursprünglichen Annahme einer fehlenden Positivität fest (4).

## Nur ein Teil der vorliegenden Evidenz wird zur Bewertung des Zusatznutzens herangezogen

Infolge der Ablehnung werden adjustierte Analysen vom G-BA allenfalls deskriptiv genannt bzw. punktuell herangezogen. Im Hinblick auf die Kriterien Überlappung und Balanciertheit werden die eingereichten Analysen durch den G-BA weder in ihrer Gesamtheit betrachtet noch abwägend diskutiert. Die Gesamtheit der Evidenz umfasst naive und adjustierte Analysen in verschiedenen Zuschnitten sowie zusätzliche gepoolte Analysen mit der Studie CAR4. Insbesondere durch das PS-Matching wird nahezu perfekte Überlappung und Balanciertheit erreicht.

Die in der wissenschaftlichen Literatur beschriebene Unterscheidung in strukturelle und rein statistische Verletzung der Positivitätsannahme wird nicht vorgenommen, obwohl der G-BA ebendiese Literatur zitiert. Die Evidenz wurde in weiten Teilen auf Nachfrage des G-BA im Stellungnahmeverfahren eingereicht, wird jedoch nicht ausgewogen dargestellt und diskutiert. Im Hinblick auf die Analysen zu PS-Matching und Trimmung, welche in der Literatur explizit empfohlen werden, um möglichen Verletzungen der Positivitätsannahme und resultierenden extremen Gewichten zu begegnen, thematisiert der G-BA lediglich den Fallzahlverlust als weiteren vermeintlichen Beleg für fehlende Positivität, jedoch nicht die Ausführungen zur Übertragbarkeit der Ergebnisse der Pseudo-Populationen auf die Gesamtpopulation. Insbesondere geht der G-BA nicht auf die Konsistenz aller Analysen ein. Auch die dazu konsistenten Ergebnisse aus der direkten Evidenz der CAR4 werden nicht in die Diskussion einbezogen. Limitationen der Analyse, fehlende Werte und unterschiedliche Erhebungszeitpunkte der Adjustierungsvariablen werden herausgestellt, jedoch nicht ins Verhältnis der ausgeprägten und konsistenten

## Zitationshinweis

Jakobs et al.: „Die ‚Target Trial Emulation' zu CAR-T im AMNOG", in „Monitor Versorgungsforschung" (06/23), S. 46–49. http://doi.org/10.24945/MVF.06.23.1866-0533.2563


Effekte gesetzt. Der G-BA bezieht sich nur auf Teile der Evidenz und lässt wichtige Aspekte der Gesamtschau außer Acht. Mit der Ablehnung der Evidenz zur Quantifizierung des Zusatznutzens negiert der G-BA gleichsam das Vorliegen eines Behandlungseffektes auf das Gesamtüberleben überhaupt. Das ist angesichts der vorgelegten Ergebnisse und des Verlaufs des Stellungnahmeverfahrens nicht nachvollziehbar (4).

## Effektstärke muss zur Beurteilung der Aussagesicherheit betrachtet werden

In früheren Nutzenbewertungen hat der G-BA die Größenordnung der Effektstärke betrachtet, um die Aussagesicherheit des indirekten Vergleichs im Hinblick auf potenzielle systematische Verzerrungen zu beurteilen. Ein hinreichend großer Effekt im Sinne eines dramatischen Effektes, wie Ciltacel ihn im Gesamtüberleben zeigt, führte dabei in der bisherigen Bewertungspraxis zu einem Einbezug der indirekten Evidenz in die Beschlussfassung, auch wenn keine Quantifizierung des Zusatznutzens ausgesprochen wurde. Dies gilt auch in Evidenzsituationen, in denen weder eine adäquate Confounder-Identifizierung noch eine Adjustierung des indirekten Vergleiches vorgenommen wurde, und erfolgte ohne Diskussion der Balanciertheit und Positivität (8, 9). Im Verfahren zu Ciltacel geht der G-BA dagegen von derart starker systematischer Verzerrung aus, dass das Vorliegen eines Überlebensvorteils durch eine Behandlung mit Ciltacel grundsätzlich angezweifelt wird.

Die Effektstärke im Gesamtüberleben wird für die Beurteilung der Aussagesicherheit in diesem Verfahren nicht betrachtet (4). <<


von: Janine Jakobs BS MS,
Ronja Kerßenboom BS MA,
Christina Schulat BA MS,
Dr. rer. nat. Jörn Sindern*
* Janssen Deutschland


**PS-Verteilung vor Adjustierung (A) sowie nach Adjustierung (B, C, D)**

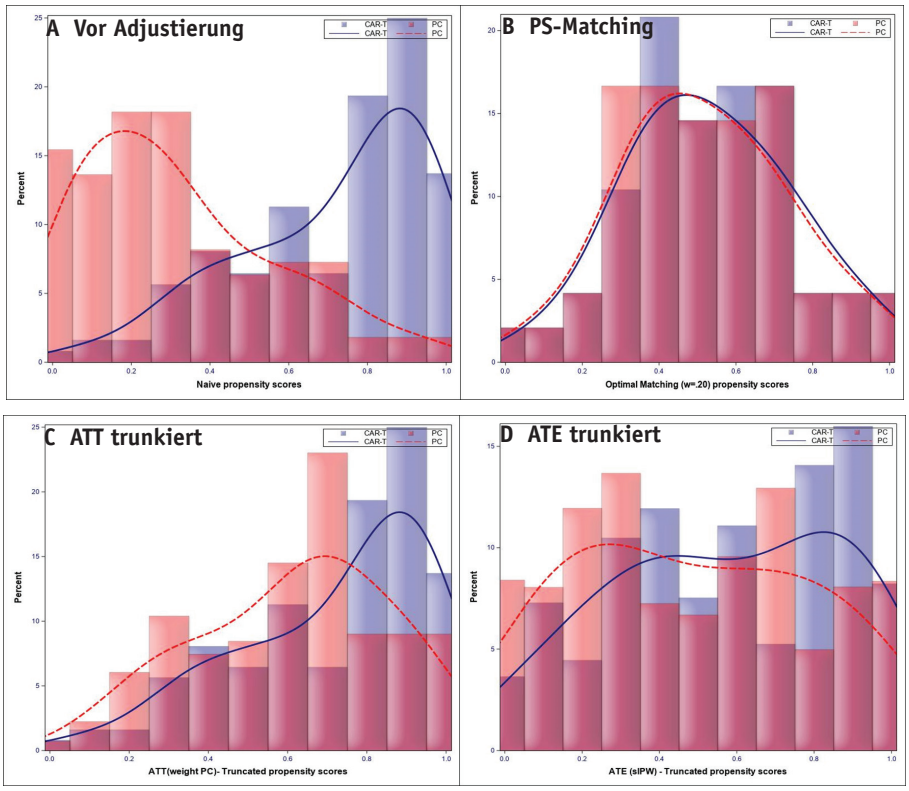



***Abb. 1:*** Die Abbildung zeigt die PS-Verteilung vor Adjustierung (A) sowie nach PS-Matching (B) und trunkierter ATT- (C) bzw. trunkierter ATE-Gewichtung (D). In der naiven Analyse (A) bestehen Unterschiede zwischen CAR1 (Ciltacel) und LOCO (PC; physician's choice), welche durch Adjustierung ausgeglichen werden. Insbesondere aus dem PS-Matching resultiert eine Analyse-Population mit nahezu perfekter Überlappung. Extreme Gewichte in der ATT- und ATE-Analyse wurden durch Trimmung und Trunkierung adressiert.

**Gesamtüberleben von Ciltacel versus Versorgungsstandard**

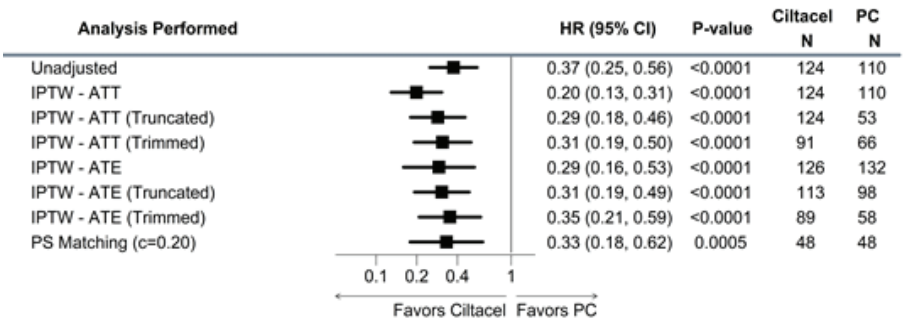



***Abb. 2:*** Gesamtüberleben von Ciltacel versus dem Versorgungsstandard (PC; physician's choice) aus CAR 1 und LOCO. Die Auswertung des Gesamtüberlebens zeigt über alle Adjustierungsmethoden eine hohe Konsistenz. Gewichtete Analysen wurden mit Varianzkorrektur durchgeführt.

## Literatur


1. Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. American journal of epidemiology, 183(8), 758-764.
2. IQWiG. Rapid Report. IQWiG-Berichte – Nr. 863. Version 1.1. Stand: 13.05.2020. 2020. https://www.iqwig.de/download/a19-43_versorgungsnahe-daten-zum-zweckeder-nutzenbewertung_rapid-report_v1-1.pdf
3. Hernán, M. A., & Robins, J. M. (2020). Causal Inference: What If; Chapman & Hall/CRC: Boca Raton, FL, USA, 85-100.Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., & Van Der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. Statistical methods in medical research, 21(1), 31-54.
4. G-BA. Nutzenbewertungsverfahren zum Wirkstoff Ciltacabtagene autoleucel (Multiples Myelom, mind. 3 Vortherapien). https://www.g-ba.de/bewertungsverfahren/nutzenbewertung/924/#beschluesse
5. Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., & Van Der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. Statistical methods in medical research, 21(1), 31-54.
6. Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. American journal of epidemiology, 168(6), 656-664.
7. Zhu, Y., Hubbard, R. A., Chubak, J., Roy, J., & Mitra, N. (2021). Core concepts in pharmacoepidemiology: Violations of the positivity assumption in the causal analysis of observational data: Consequences and statistical approaches. Pharmacoepidemiology and drug safety, 30(11), 1471-1485.
8. G-BA. Tragende Gründe. Axicabtagen-Ciloleucel. 2. Mai 2019. https://www.g-ba.de/downloads/40-268-5741/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_TrG.pdf
9. G-BA. Tragende Gründe Idecabtagen vicleucel. 16. Juni 2022. https://www.g-ba.de/downloads/40-268-8579/2022-06-16_AM-RLXII_Idecabtagen%20vicleucel_D-779_TrG.pdf