



**Stellungnahme zum Entwurf der Version 7.0 des IQWiG-Papiers
„Allgemeine Methoden“
durch die gemeinsame Präsidiumskommission
„Methodenaspekte in der Arbeit des IQWiG und IQTiG“ der GMDS und IBS-DR**

Im „Allgemeinen Methodenpapier“ werden die gesetzlichen und wissenschaftlichen Grundlagen der Arbeit am Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) erläutert. Die derzeit gültige Fassung ist die Version 6.1 vom 24. Januar 2022. Am 6. Dezember 2022 hat nun das IQWiG einen Entwurf der neuen Version 7.0 vorgelegt. In der Version 7.0 wurden im Vergleich zur Version 6.1 neben redaktionellen Modifikationen auch inhaltliche Änderungen durchgeführt. Grundlegende Überarbeitungen erfolgten in den Kapiteln zu *IQWiG im Gesundheitswesen* (Kapitel 1), *Produkten des Instituts* (Kapitel 2), *Nutzenbewertung medizinischer Interventionen* (Kapitel 3), *Kosten-Nutzen-Bewertung medizinischer Interventionen* (Kapitel 4), *Informationsbeschaffung* (Kapitel 8) und *Informationsbewertung* (Kapitel 9). Zudem soll in Zukunft die Qualitätsbewertung systematischer Übersichten mit AMSTAR 2 statt mit dem Oxman-Guyatt-Index durchgeführt werden; hierzu wurden entsprechende Änderungen in Kapitel 7 „Evidenzbasierte Gesundheitsinformation für Bürgerinnen und Bürger“ eingefügt. Aus biometrischer Sicht sind insbesondere einige Änderungen in Kapiteln 3 und 9 relevant; aus Sicht der Gesundheitsökonomie und der entscheidungsanalytischen Modellierung sind insbesondere einige Textpassagen in den Kapiteln 1.4 und 4 relevant. Auf diese gehen wir unten näher ein.

Die Kommission begrüßt die regelmäßige Überarbeitung der „Allgemeinen Methoden“. Die stetige Fortentwicklung relevanter Methoden macht dies unerlässlich. Wie aus den spezifischen Kommentaren unten deutlich wird, gibt es einige Abschnitte, die einer Aktualisierung bedürfen. Aus Sicht der Kommission ist es auch verständlich, dass es schwierig ist, alle Bereiche immer gleichmäßig aktuell zu halten. Hier wäre insbesondere der Hinweis, sich in den Bereichen Epidemiologie, Versorgungsforschung (z.B. spezielle Studiendesigns wie Stepped Wedge Designs zur Evaluation von Versorgungskonzepten) und Patient Preference Studies zu verstärken oder externe Beratung einzuholen. Die Kommission hat weiterhin wahrgenommen, dass einige Punkte auf Rapid Reports zurückgehen. Diese können als solche nicht kommentiert werden. Wenn Sie ins Methodenpapier eingehen, um z.B. bis dato bestehende Widersprüche zwischen Rapid Report und Methodenpapier aufzulösen, können sie sehr wohl kommentiert werden. Allerdings setzt dies eine ausreichend ausführliche Darstellung im Methodenpapier voraus. Hier gehen wir im weiteren Verlauf beispielhaft auf den Rapid Report A19-43 „Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V“ ein. Die Kommission ist offen für eine fachliche Diskussion mit dem IQWiG in Bezug auf geeignete Methoden und unterstützt das Institut in dieser Hinsicht gerne.

Die Kommission hat zu den folgenden speziellen Punkten des Papiers Kommentierungen durchgeführt:

Abschnitt 1.4 Gesundheitsökonomie und Abschnitt 4 Kosten-Nutzen-Bewertung medizinischer Interventionen

Die Einordnung der Methode 'gesundheitsökonomische Evaluation' in den Gesamtkontext der umfassenderen Disziplin Gesundheitsökonomie in Abschnitt 1.4 wird begrüßt. Ebenso wird die ausführliche Darstellung der Bedeutung der gesundheitsökonomischen Evaluation einschließlich der Begriffsbestimmungen in diesem Abschnitt begrüßt, die der internationalen Terminologie entsprechen.

Abschnitt 3.1.1 Definition des patientenrelevanten medizinischen Nutzens bzw. Schadens

Im Abschnitt 3.1.1 wird als einzige Möglichkeit zur Beteiligung an der Wahl der patientenrelevanten Endpunkte die Einbeziehung einzelner Betroffener erwähnt. Während dies dem allgemeinen Vorgehen zur Einbindung von Betroffenen in die Erstellung von IQWiG-Produkten entspricht, sollten aus Sicht der Kommission neuere Entwicklungen und systematische Möglichkeiten für die Zielsetzungen in diesem Abschnitt diskutiert werden. Die Kommission verweist insbesondere auf die Empfehlungen von IMI PREFER zur Planung, Durchführung, Analyse und Interpretation von Präferenzstudien, die entscheidungsrelevante Informationen z.B. zur Wahl der patientenrelevanten Endpunkte und zur Abwägung zwischen Endpunkten liefern sollen. Im Rahmen der Empfehlungen hat IMI PREFER (siehe The PREFER consortium (2022)) ein umfassendes Rahmenwerk vorgelegt, das eine positive Bewertung vom Committee for Medicinal Products for Human Use (CHMP) basierend auf einer EMA Methodenqualifizierung erhalten hat (siehe EMA (2022)). Diese Qualifizierung basiert auf gemeinsamen EMA/EUnetHTA Qualification Advice. Zudem empfiehlt die Kommission, dass sogenannte „Core Sets of Patient-reported Outcomes“ berücksichtigt werden, sofern diese für die entsprechende Erkrankung/Intervention publiziert wurden (siehe z.B. Stojkov et al (2021)).

Abschnitt 3.2.1 Umgang mit möglichem Reporting Bias

Im Abschnitt 3.2.1 wird der Umgang mit möglichem Reporting Bias beschrieben. Insbesondere wird hier auf zusätzliche Suchen in Registern klinischer Studien eingegangen. In diesem Zusammenhang möchte die Kommission auf neuere Entwicklungen in der statistischen Methodik hinweisen, die auch die Berechnung von Effektschätzern und Konfidenzintervallen unter Berücksichtigung von zusätzlichen Informationen aus solchen Registern berücksichtigen (siehe z.B. Huang et al (2021, 2022)).

Abschnitt 3.3.4 Konzepte für anwendungsbegleitende Datenerhebungen gemäß § 35a Abs. 3b SGB V (AbD-Konzepte)

Dieser Abschnitt ist neu hinzugekommen, um den Anforderungen der anwendungsbegleitenden Datenerhebungen gemäß § 35a Abs. 3b SGB V Rechnung zu tragen. Die „Beschreibung grundsätzlicher Anforderungen an die Auswertung der Daten“ verweist auf die im Rapid Report A19-43 beschriebenen Konzepte zur Auswertung versorgungsnaher Daten zum Zwecke der Nutzenbewertung von Arzneimitteln. Hier wird hauptsächlich auf „die adäquate Adjustierung für Störgrößen (Confounder)“ abgehoben. Aus Sicht der Kommission sollte an dieser Stelle darauf hingewiesen werden, dass beim Vorliegen von zeitabhängigem Confounding ggf. spezifische Kausalinferenz-Verfahren zur Kontrolle dieses Confounding-Typs erforderlich sind wie u.a. die parametrische g-Formula, marginal structural models mit inverse probability weighting, oder structural nested models mit g-estimation (Robins et al, 2004; Abschnitt „Modelling for explanation“ in Jahn et al, 2022). Aus Sicht der Kommission ist es ferner wichtig zu betonen, dass Confounding jedoch nicht die einzige Quelle von Bias beim Vergleich von Therapiegruppen in nicht-randomisierten Studien darstellen. Neben vielen weiteren möglichen

Aspekten ist eine wesentliche Verzerrungsquelle die Festlegung des Zeitpunkts „Null“, zu dem die Einschlusskriterien erfüllt sind, die Patienten den Therapiegruppen zugewiesen werden und die Nachbeobachtung zur Erfassung der Endpunkte beginnt. Dieser Zeitpunkt Null ist in randomisierten Studien in natürlicher Weise durch den Zeitpunkt der Randomisierung gegeben. In nicht-randomisierten Studien ist dies nicht der Fall und nicht adäquate Festlegungen können zu erheblichen Verzerrungen führen; als Stichworte seien hier *time-dependent bias* und *immortal time bias* genannt (siehe z.B. Hernán et al (2016)). Die Kommission empfiehlt daher, diese Aspekte, die u.a. zu einer geeigneten expliziten Nachbildung einer randomisierten Studie gehören wie die *target trial emulation* (Hernán und Robins, 2016; Hernán et al, 2022), die inzwischen auch von anderen HTA Organisationen in deren Methodenpapieren aufgenommen wurden (National Institute for Health and Care Excellence, 2022), stärker zu betonen.

Wie einleitend bemerkt, wünscht sich die Kommission auch Rapid Reports kommentieren zu können. Da im Abschnitt 3.3.4 des Entwurfs der Version 7.0 der Rapid Report A19-43 zitiert wird, nimmt die Kommission dies zum Anlass diesen zu kommentieren. Aus Sicht der Kommission werden die Vor- und Nachteile verschiedener Auswertungsverfahren in diesem in unausgewogener Weise diskutiert. So wird dort z.B. berichtet, dass der Nachteil von multiplen Regressionsanalysen darin besteht, dass „diese Analysen je nach Fallzahl bzw. Ereigniszahl nur eine begrenzte Anzahl an Confoundern aufnehmen können, da sie sonst unzuverlässige Resultate liefern oder mathematisch gar nicht funktionieren“, während „der große Vorteil der Verfahren, die auf Propensity Scores beruhen, ist, dass im Vergleich zur direkten Adjustierung deutlich mehr mögliche Confounder einbezogen werden können“. Das mag in einzelnen Fällen so richtig sein. Allerdings basiert auch der Propensity Score auf einem Regressionsmodell (häufig einer logistischen Regression) und somit ergeben sich dort ebenso Bedingungen für eine Konvergenz. Insbesondere bei kleinen Datensätzen sind die vermeintlichen Vorteile der Propensity Scores mitunter nicht gegeben (siehe Friedrich und Friede, 2020). Aus Sicht der Kommission ist eine ausgewogenere Diskussion, die sich an den Zielen der jeweiligen Auswertung orientiert, in der Abwägung der Auswertungsmethoden sinnvoller.

Aus Sicht der Kommission gelten die grundsätzlichen Erwägungen zur kausalen Inferenz im Zusammenhang mit anwendungsbegleitenden Datenerhebungen (AbD) auch für Vergleiche mit historischen Kontrollen.

Im Abschnitt 3.2.2 wird ein „dramatischer Effekt“ definiert. Dieser liegt vor, wenn der Effekt statistisch signifikant zum Niveau 1% ist und auf der Skala des relativen Risikos eine Größenordnung von 5 bis 10 erreicht. Bei AbDs liegt jedoch ein sogenannter dramatischer Effekt vor, wenn die untere Grenze des 95% Konfidenzintervalls oberhalb von 2 („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“, Abschnitt 3.3.4) liegt. Die Kommission empfiehlt, die Definitionen zu vereinheitlichen.

Abschnitt 3.5 Diagnostische Verfahren

In Abschnitt 3.5 wird darauf eingegangen, dass diagnostische Tests mit einem kontinuierlichen Ergebnis gemeinsam mit entsprechenden Grenzwerten angegeben werden sollen, die eine Einteilung in ein positives oder negatives Ergebnis erlauben. Aus Sicht der Kommission sind an dieser Stelle nicht nur diagnostische Grenzwerte relevant, sondern auch die vollständige Darstellung der Ergebnisse über eine Receiver-Operating-Characteristic-Kurve (ROC-Kurve) und die Berechnung der zugehörigen Area under the curve (AUC).

Die Kommission begrüßt es, dass im Methodenpapier im Abschnitt 3.5.1.1 nun explizit auf Test-treatment-Studien eingegangen wird (Pepe, 2003; Hot et al, 2021). Jedoch sind diese erst der abschließende Schritt bei der Entwicklung und Bewertung neuer diagnostischer Verfahren. Es wäre wünschenswert herauszustellen, dass diagnostische Verfahren bereits vorab in verschiedenen Studienphasen (siehe z.B. Table 8.1 in Pepe (2003)) hinsichtlich ihrer Genauigkeit in Form von beispielsweise Sensitivität und Spezifität geprüft werden müssen, bevor es zu Test-treatment-Studien kommt.

Im Abschnitt 3.5.1.2 zu „Linked Evidence“ findet sich die Abbildung 14, die für sich genommen nur sehr schwer oder gar nicht nachvollziehbar ist. Insbesondere ist nicht klar, wie sich einige Pfade verzweigen. Die Kommission empfiehlt, die Abbildung zu überarbeiten und evtl. auch in einer ausführlicheren Bildunterschrift oder Legende weitere Informationen zu geben, die ein einfacheres Verständnis erlauben. Zudem wird in Abschnitt 3.5.1.2 nicht explizit auf Screeningverfahren eingegangen. Hier sollte aus Sicht der Kommission zumindest kurz angesprochen werden, dass ähnliche Methoden auch zur Nutzen-Schaden-Bewertung von Screeningverfahren eingesetzt werden können, sofern keine aktuelle direkte Evidenz vorliegt. Dies ist insbesondere der Fall, wenn für ein Screeningverfahren der Nutzen bereits als gesichert gilt und ein entsprechendes Screeningprogramm bereits etabliert ist, aber verschiedene Varianten (Startalter, Stopalter, Screeningintervall, Screeningtestkombinationen) des Screenings verglichen werden (siehe z.B. IQWiG (2022)).

Abschnitt 4 Kosten-Nutzen-Bewertung medizinischer Interventionen

Die Kommission begrüßt die Erweiterung im Abschnitt 4 um die tabellarische explizite Beschreibung des Referenzfalls (Tabelle 4), deren Format sich an internationalen Standards bzw. Lehrbücher anlehnt.

Im Abschnitt 4 wird an verschiedenen Stellen auf „Modellierungen“ eingegangen. Es wird vorgeschlagen, in Abschnitt 4 zu Beginn darzulegen, dass es sich beim Begriff „Modellierungen“ um entscheidungsanalytische Modelle handelt, um diese von statistischen Modellen abzugrenzen (z.B.: S. 84 erste Zeile wie folgt ergänzen: „...werden in der Regel Modellierungen *mit entscheidungsanalytischen Modellen* durchgeführt (Siebert, 2003). In Abschnitt 4.9 Gesundheitsökonomisches Studiendesign wäre es hilfreich, im Unterabschnitt ‚Auswahl der Modellierungstechnik‘ neben der Referenz [596] auch auf die internationalen Empfehlungen zu den genannten einzelnen Verfahren hinzuweisen: „Die Konzeption des Modells ist in der Regel ausschlaggebend dafür, welche Modellierungs-technik (Entscheidungsbaum, Markov-Modell, diskrete Ereignissimulation, etc.) zum Einsatz kommt ([596]; Siebert et al, 2012; Karnon et al, 2012; Pitman et al, 2012). Am Ende des einführenden Unterkapitels in Abschnitt 4.9 (S. 99) sollte darauf hingewiesen werden, dass Inputparameter für entscheidungsanalytische Modellierungen, die sich auf Interventionseffekte beziehen, mittels der entsprechenden Verfahren der Kausalinferenz abzuleiten sind (Kühne et al, 2022; Jahn et al, 2022).

Abschnitt 9.1 Qualitätsbewertung von Einzelstudien

Auf den Seiten 158-159 findet sich eine Liste von Stellungnahmen zur Beurteilung der Verwendbarkeit einer Quelle. Die Kommission empfiehlt, diese um die „REporting of studies Conducted using Observational Routinely-collected Data“ (RECORD) Statements zu erweitern (siehe record-statement.org; Benchimol et al (2015); Langan et al (2018)).

Bei der Überarbeitung des Abschnitts 9.1.1 zu Kriterien für den Einschluss von Studien wurde nun die „80% Regel“ aufgenommen, die besagt, dass eine Studie eingeschlossen wird, „falls auf mindestens

80% der eingeschlossenen Patientinnen und Patienten das Kriterium (Population) zutrifft bzw. falls das Einschlusskriterium bezüglich der Prüfintervention (Interventionsgruppe der Studie) und der Vergleichsintervention (Vergleichsgruppe der Studie) jeweils zu mindestens 80% erfüllt ist“. Die Kommission begrüßt dies, da dies konsistent mit dem bisherigen Vorgehen des Instituts ist und in diesem Punkt nun Klarheit schafft.

Abschnitt 9.3 Spezielle biometrische Aspekte

Im Abschnitt 9.3.3 werden Kriterien und Verfahren zur Beurteilung der klinischen Relevanz beschrieben. Analog zum Methodenpapier 6.0 hat der G-BA einen Schwellenwert von 15% der Spannweite der Skala für PROs in die Verfahrensordnung aufgenommen. Dieser Schwellenwert gilt für alle PROs, für die keine MCID validiert ist oder für die die MCID unterhalb dieser Schwelle liegt. Auf der G-BA FAQ Seite findet sich seit kurzem der Hinweis, dass für Auswertungen des Fragebogens EORTC QLQ-C30 sowie den entsprechenden validierten krankheitsspezifischen Ergänzungsmodulen Analysen zum Responsekriterium 10 Punkte im Dossier darzustellen sind (www.g-ba.de/themen/arzneimittel/arzneimittel-richtlinie-anlagen/nutzenbewertung-35a/faqs, Zugriff: 16.1.2023). Dieses wird so bereits in aktuellen Bewertungen vom IQWiG aufgenommen (siehe z.B. die Bewertung zu Selinexor, Verfahren Nr. 886, IQWiG-Berichte Nr. 1480). Hier wird vom IQWiG kommentiert: "Für den EORTC-QLQ-C30 und seine Zusatzmodule sind im Verfahren der frühen Nutzenbewertung jedoch Auswertungen zu einer Responderschwelle von 10 Punkten erforderlich". Der Hersteller hatte nur Analysen mit einer Responderschwelle von 15 Punkten gemäß IQWiG Methodenpapier 6.1 und der Verfahrensordnung vorgelegt. Nach Ansicht der Kommission sollte diese Ausnahme im Methodenpapier 7.0 Berücksichtigung finden.

Im Abschnitt 9.3.6 werden Verfahren zur Adjustierung für Confounder im Rahmen von anwendungsbegleitenden Datenerhebungen beschrieben. Hierzu hat die Kommission bereits oben beziehungsweise auf den Abschnitt 3.3.4 Anmerkungen gemacht.

Im Abschnitt 9.3.7 werden Verfahren für Metaanalysen dargestellt. Unter B) Heterogenität heißt es: „Zu diesem Zweck gibt es spezielle statistische Methoden wie z. B. das I^2 -Maß [355]. Für dieses Maß existieren Untersuchungen, die eine grobe Einschätzung der Heterogenität zulassen (z. B. über die Kategorien wahrscheinlich unbedeutend [0 bis 40 %], mittelmäßig [30 bis 60 %], substanzial [50 bis 90 %] und erheblich [75 bis 100 %] [174]).“ Die Kommission empfiehlt die Hinweise zum I^2 -Maß zu entfernen, da dessen Verwendung über die Jahre zunehmend in Kritik geraten ist. Die Einschätzung des Ausmaßes der Heterogenität kann alternativ über Schätzung der Standardabweichung tau erfolgen; auch für dieses Maß sind Kategorien des Ausmaßes in der Literatur bekannt (siehe z.B. Tabelle 1 in Friede et al (2017) oder Tabelle 2 in Röver et al (2021)).

Unter A) Allgemeines heißt es: „Kontextabhängig kommen auch alternative Verfahren zur metaanalytischen Zusammenfassung in Betracht, wie z. B. bayessche Verfahren“. Hier würde die Kommission begrüßen, wenn das Institut eine Empfehlung zur Wahl der a priori Verteilungen bei Bayes-Verfahren geben könnte. Insbesondere ist hier die a priori Verteilung für den Heterogenitätsparameter tau von Interesse (Röver et al, 2021).

In der Ergänzung zu Abschnitt 9.3.12 wird die Auswirkung systematisch verkürzter Beobachtungszeiten bei der Erhebung von Patient Reported Outcomes (PROs) diskutiert. Dabei wird auf die Operationalisierung der sogenannten "dauerhaften" Veränderung eingegangen. Bei systematisch verkürzten Beobachtungszeiten in einer der beiden Behandlungsgruppen kann es für die Zeit bis zur Verschlechterung zu einer Verzerrung zu Gunsten der Behandlungsgruppe mit der längeren Beobachtungszeit führen, da die Patienten länger die Chance haben, wieder eine Verbesserung zu

erreichen. Im Falle der Zeit bis zur Verbesserung ist von einer Verzerrung zu Gunsten der Behandlungsgruppe mit der kürzeren Beobachtungszeit zu erwarten, da die Patienten kein Risiko haben, dass nach der Verbesserung eine Verschlechterung festgestellt wird.

Die Kommission rät von der Durchführung derartiger vereinfachter Ereigniszeitanalysen bei komplexen Zeitverläufen ab, da sie auf die Zukunft bedingen, was grundsätzlich unterbleiben sollte (siehe etwa Andersen und Keiding, 2012). Wenn man Therapiegruppen bzgl. ihres zeitlichen Verlaufs in bestimmten verschlechterten oder verbesserten Zuständen vergleichen möchte, sollte man nach Ansicht der Kommission geeignete Mehrstadienmodelle verwenden. Diese würden die Zeiten unter Risiko adäquat berücksichtigen, und die in Abschnitt 9.3.12 erwähnten Probleme bei unterschiedlich langen Beobachtungszeiten würden somit nicht bestehen. Ein Beispiel für solche Analysen (in einem anderen inhaltlichen Kontext, aber genauso auf PRO-Zustände übertragbar) findet sich in Szenario 4 in Schmoor et al (2013) (Referenz [638] im Methodenpapier Version 7.0). Inferenzstatistische Vergleiche von Therapiegruppen, wenn notwendig, sollten nach Ansicht der Kommission entweder bezüglich der *transition hazards* oder der *probabilities* über die Zeit durchgeführt werden; siehe hierzu z.B. Bluhmki et al (2018, 2020).

Es sollte im gleichen Zuge auch die Diskussion zur sogenannten "bestätigten" Veränderung aufgenommen werden, wie es auch auf der FAQ-Seite des G-BA beschrieben ist (www.g-ba.de/themen/arzneimittel/arzneimittel-richtlinie-anlagen/nutzenbewertung-35a/faqs). Die "bestätigte" Veränderung ist dabei eine Veränderung, bei der der Schwellenwert der Response für zwei oder mehrere aufeinanderfolgende Erhebungen überschritten wird. Hierbei ist zu berücksichtigen, dass eine erstmalige Veränderung zum letzten Beobachtungszeitpunkt nicht als "bestätigte" Veränderung gewertet werden sollte, um eine Verzerrung zugunsten der Behandlungsgruppe mit der längeren Beobachtungszeit zu vermeiden.

Referenzen

- Andersen PK, Keiding N (2012) Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine* 31: 1074 – 1088.
- Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, RECORD Working Committee (2015) The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine*. DOI:10.1371/journal.pmed.1001885
- Bluhmki T, Schmoor C, Dobler D, Pauly M, Finke J, Schumacher M, Beyersmann J (2018) A wild bootstrap approach for the Aalen–Johansen estimator. *Biometrics* 74: 977–985.
- Bluhmki T, Schmoor C, Finke J, Schumacher M, Beyersmann J (2020) Relapse- and immunosuppression-free survival after hematopoietic stem-cell transplantation: How can we assess treatment success for complex time-to-event endpoints? *Biology of Blood and Marrow Transplantation* 26: 992-997.
- EMA (2022) Qualification Opinion of IMI PREFER. https://www.ema.europa.eu/documents/regulatory-procedural-guideline/qualification-opinion-imi-prefer_en.pdf (Zugriff: 18. Januar 2023).
- Friede T, Röver C, Wandel S, Neuenschwander B (2017) Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods* 8: 79–91.
- Friedrich S, Friede T (2020) Causal inference methods for small non-randomized studies: Methods and an application in COVID-19. *Contemporary Clinical Trials* 99: 106213.
- Hernán MA, Robins JM (2016) Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology* 183(8): 758-764.
- Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I (2016) Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology* 79: 70–75.
- Hernán MA, Wang W, Leaf DE (2022) Target Trial Emulation: A framework for causal inference from observational data. *JAMA* 328: 2446–2447.
- Hot A, Bossuyt PM, Gerke O, Wahl S, Vach W, Zapf A (2021) Randomized test-Treatment studies with an outlook on adaptive designs. *BMC Medical Research Methodology* 21(1): 110.
- Huang A, Komukai S, Friede T, Hattori S (2021) Using clinical trial registries to inform Copas selection model for publication bias in meta-analysis. *Research Synthesis Methods* 12: 658–673.
- Huang A, Morikawa K, Friede T, Hattori S (2022) Adjusting for publication bias in meta-analysis via inverse probability weighting using clinical trial registries. *Biometrics* (in press). <https://doi.org/10.1111/biom.13822>
- IQWiG. Überprüfung der Altersgrenzen im Mammografie-Screening-Programmen. IQWiG Abschlussbericht Auftrag S21-01, Version 1.1, Stand: 16.08.2022. IQWiG-Berichte – Nr. 1383. Internet: https://www.iqwig.de/download/s21-01_altersgrenzen-im-mammografie-screening-programm_abschlussbericht_v1-1.pdf
- Jahn B, Friedrich S, Behnke J, Engel J, Garczarek U, Münnich R, Pauly M, Wilhelm A, Wolkenhauer O, Zwick M, Siebert U, Friede T (2022) On the role of data, statistics and decisions in a pandemic. *Advances in Statistical Analysis* 106(3): 349-382.

Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J (2012) Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Medical Decision Making* 32(5):701-11.

Kühne F, Schomaker M, Stojkov I, Jahn B, Conrads-Frank A, Siebert S, Sroczynski G, Puntischer S, Schmid D, Schnell-Inderst P, Siebert U (2022) Causal Evidence in Health Decision Making: Methodological Approaches of Causal Inference and Health Decision Science. HTA Report Vol. 509-1. *GMS German Medical Science — an Interdisciplinary Journal* 20:Doc12. DOI: 10.3205/000314.

Langan SM,¹ Schmidt SAJ, Wing K, Ehrenstein V, Nicholls SG, Filion KB, Klungel O, Petersen I, Sorensen HT, Dixon WG, Guttmann A, Harron K, Hemkens LG, Moher D, Schneeweiss S, Smeeth L, Sturkenboom M, von Elm E, Wang SV, Benchimol EI (2018) The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ* 363:k3532. DOI:10.1136/bmj.k3532

National Institute for Health and Care Excellence (2022) NICE real-world evidence framework. Corporate document [ECD9]. Published: 23 June 2022. Internet: <https://www.nice.org.uk/corporate/ecd9/resources/nice-realworld-evidence-framework-pdf-1124020816837>

Pepe MS (2003) The statistical evaluation of medical tests for classification and prediction. Oxford University Press, USA, Kapitel 8.

Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, Brisson M; ISPOR-SMDM Modeling Good Research Practices Task Force (2012) Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-5. *Medical Decision Making* 32(5): 712-21.

Robins JM, Hernán MA, Siebert U (2004) Effects of Multiple Interventions. In: Ezzati M, Lopez AD, Rodgers A, Murray CJL (eds.). *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Vol. 1. World Health Organization: Geneva, chapter 28, pages 2191-2230. Volltext: https://apps.who.int/iris/bitstream/handle/10665/42770/9241580313_eng.pdf?sequence=1&isAllowed=y

Röver C, Bender R, Dias S, Schmid C, Schmidli H, Sturtz S, Weber S, Friede T (2021) On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Research Synthesis Methods* 12: 448–474.

Siebert U (2003) When should decision-analytic modeling be used in the economic evaluation of health care? *The European Journal of Health Economics* 4(3): 143-50.

Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, Kuntz KM (2012) State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Medical Decision Making* 32(5): 690-700.

Stojkov I, Conrads-Frank A, Rochau U, Koinig KA, Arvandi M, Puntischer S, van Marrewijk C, Fenaux P, Symeonidis A, Chermat F, Garelius H, Bowen D, Mittelman M, Mora E, de Witte T, Efficace F, Siebert U, Stauder R (2021) Core set of patient-reported outcomes for myelodysplastic syndromes: an EUMDS Delphi study involving patients and hematologists. *Blood Advances* 6(1): 1-12. DOI 10.1182/bloodadvances.2021004568

The PREFER consortium (2022) PREFER Recommendations: Why, when and how to assess and use patient preferences in medical product decision-making. Zenodo. DOI: 10.5281/zenodo.6470922

Über die GMDS e.V.

Die Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. ist mit derzeit ca. 2.000 Mitgliedern die einzige wissenschaftliche Fachgesellschaft in Deutschland, Österreich und der Schweiz, die die fünf Disziplinen Medizinische Informatik, Medizinische Biometrie, Epidemiologie, Medizinische Dokumentation und Medizinische Bioinformatik und Systembiologie gemeinsam vertritt. Sie kooperiert mit einer Reihe benachbarter Fachgesellschaften und Verbänden. Zudem entwickelt sie die Fachgebiete weiter durch sachverständige Repräsentation u. a. bei der Planung von Förderungsmaßnahmen der Öffentlichen Hand, bei Fragen der Standardisierung und Normung, bei der Errichtung von Lehrinstitutionen, bei Ausbildungs-, Weiter- und Fortbildungsfragen und bei gesetzgebenden Maßnahmen.

Deutsche Gesellschaft für Medizinische Informatik,
Biometrie und Epidemiologie (GMDS) e.V.
Industriestraße 154
D-50996 Köln
www.gmds.de

Über die IBS-DR

Die Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR) umfasst die in der Bundesrepublik Deutschland ansässigen Mitglieder der Internationalen Biometrischen Gesellschaft. Die Internationale Biometrischen Gesellschaft fördert die Entwicklung von statistischer und mathematischer Theorie und die Anwendung ihrer Methoden in den Biowissenschaften. Anwendungsgebiete sind unter anderem biomedizinische Wissenschaften und Öffentliches Gesundheitswesen, Ökologie, Umweltwissenschaft, Forstwissenschaft und Landwirtschaft.

Deutsche Region der Internationalen Biometrischen Gesellschaft
c/o Institut für Biometrie, Epidemiologie und Informationsverarbeitung
Tierärztliche Hochschule Hannover
Bünteweg 2
D-30559 Hannover
<http://biometrische-gesellschaft.de/>