

Präv Gesundheitsf  
<https://doi.org/10.1007/s11553-023-01016-9>  
 Eingegangen: 16. November 2022  
 Angenommen: 15. Januar 2023

© Der/die Autor(en) 2023



Tim Friede · Christian Röver · Tim Mathes

Institut für Medizinische Statistik, Universitätsmedizin Göttingen, Göttingen, Deutschland

# Verknüpfung von randomisierten kontrollierten Studien und Real World Data

**Einerseits sind randomisierte kontrollierte Studien („randomized controlled trials“ [RCT]) der Goldstandard für die Evaluation von Interventionen. Andererseits haben sie den Ruf, dass sie oft nur schwer durchzuführen seien und daher unter kleinen Fallzahlen leiden. Zudem wird häufig angeführt, dass RCT aus diesem Grund sowie durch (zu) enge Einschlusskriterien und zu starke Standardisierung vielfach nicht den klinischen Routinebedingungen entsprechen. Beides kann zu Einschränkungen in der Aussagekraft von RCT führen. Zudem sind sog. Real World Data (RWD) zunehmend verfügbar. Lohnt es sich vor diesem Hintergrund überhaupt noch, regelhaft RCT durchzuführen, und, falls ja, wie können diese beiden Evidenzquellen sich sinnvoll ergänzen?**

Bei seltenen Erkrankungen und kleinen Populationen sind RCT von ausreichender Größe manchmal nur schwer oder gar nicht durchzuführen [4]. Dann stellt sich die Frage, ob ein RCT überhaupt sinnvoll ist, wenn die nötige Fallzahl realistischerweise nicht erreicht werden kann. In der Praxis werden vor diesem Hintergrund häufig einarmige Studien durchgeführt; ohne Kontrollgruppen sind diese Studien jedoch nur schwer zu interpretieren. Eine mögliche Lösung besteht in manchen Anwendungen darin, einen kleinen RCT durch RWD zu unterstützen und somit aussagekräftiger zu machen. Im folgenden Abschnitt werden wir für diesen Zweck geeignete statistische Verfahren kurz vorstellen.

Der Kritik der mangelnden Übertragbarkeit von Ergebnissen aus RCT auf Routinepopulationen werden wir uns im zweiten Teil dieses Artikels widmen. Auch hier werden wir sehen, wie RWD genutzt werden können, um in Kombination mit RCT zu robusten Einschätzungen von Therapieeffekten zu gelangen. Die Diskussion wird auch zeigen, dass RWD als Bezeichnung für nichtrandomisierte oder nichtinterventionelle Studien zu kurz greift, da auch randomisierte Interventionsstudien RWD generieren können, sofern sie einen pragmatischen Ansatz verfolgen [2].

Der letzte Abschnitt beschäftigt sich mit der „comprehensive cohort study“. Diese zeichnet sich dadurch aus, dass die Verknüpfung von RCT und Kohortenstudie prospektiv, d. h. schon vor Rekrutierungsbeginn, eingeplant ist.

## Höhere Effizienz in RCT-Auswertung durch den Einbezug von RWD

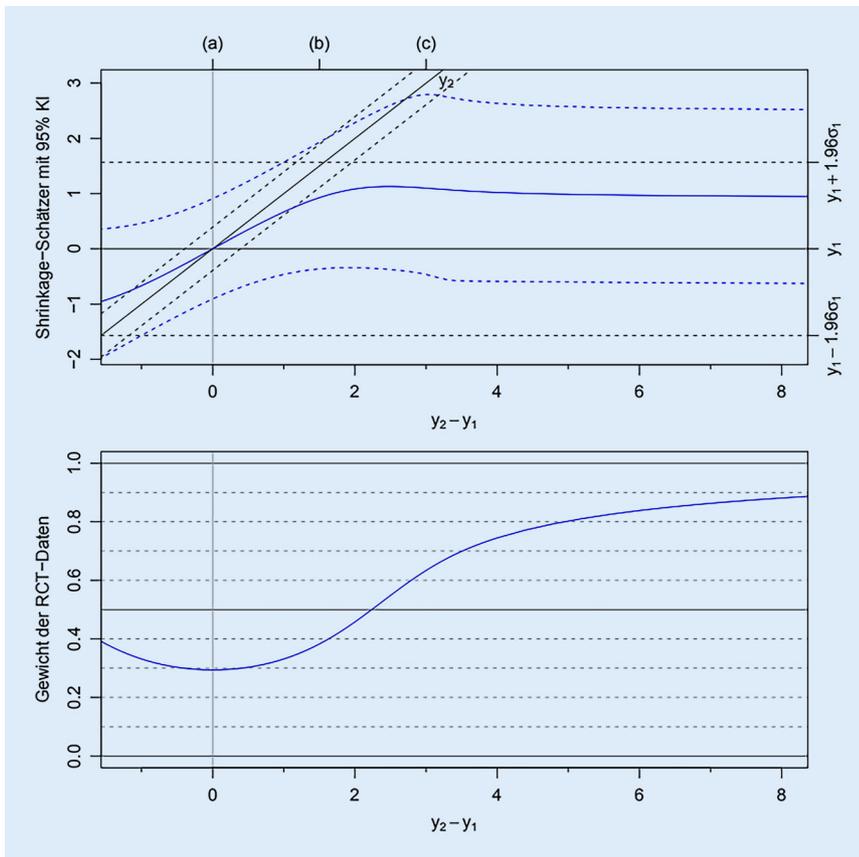
### Anwendungen von hierarchischen Modellen zur Evidenzsynthese

Die Anwendung von hierarchischen Modellen zur Evidenzsynthese ist weit verbreitet. „Hierarchisch“ bedeutet hier, dass Variabilität in den Daten auf verschiedenen Ebenen im Modell berücksichtigt wird, beispielsweise sowohl zwischen Patienten innerhalb einer Studie, als auch zwischen verschiedenen Studien. Gemeinsam ist den hierarchischen Modellen v. a., dass sie potenzielle, leichte Variationen in den Effekten zwischen den Datenquellen vorsehen. Einen einfachen Fall eines hierarchischen Modells stellt

dabei das in der Metaanalyse häufig verwendete Modell mit zufälligen Effekten („random effects [RE] model“) dar [1]. In diesem Zusammenhang haben sich Bayes-Verfahren bewährt [8]. Im Unterschied zur frequentistischen Statistik können bei Bayes-Verfahren vorbestehende Informationen, z. B. Daten aus vorangegangenen Studien oder aber auch Expertenmeinungen, als sog. „A-priori-Informationen“ bei der Auswertung der Daten berücksichtigt werden. Im Zusammenhang mit hierarchischen Modellen sind Bayes-Verfahren besonders gut geeignet, da sie gut mit einer größeren Anzahl von Hierarchieebenen und kleinen Stichprobenumfängen umgehen können [5]. Unterschiede zwischen verschiedenen Analyseansätzen und Vorteile der Bayes-Verfahren zeigen sich insbesondere in Fällen kleiner Stichprobenumfänge oder seltener Ereignisse [3].

Die detaillierte Modellierung separater Datenquellen in einer hierarchischen Struktur erlaubt dann den Rückschluss sowohl auf übergeordnete Parameter, wie einen „Gesamtmittelwert“, als auch die Quantifizierung einzelner untergeordneter Modellkomponenten, wie z. B. einzelne Studieneffekte. In einer Metaanalyse mehrerer Studien kann dabei letztlich die Präzision der einzelnen zugrundeliegenden Studieneffekte erhöht werden. Da hier in der Regel gleichzeitig eine Verschiebung hin zu einem Gesamtmittelwert („shrinkage“) stattfindet, spricht man hier auch von „Shrinkage-Schätzung“ [17].

Bei der Verknüpfung von RWD und RCT-Daten findet man sich häufig im Spezialfall einer Metaanalyse von nur



**Abb. 1** ▲ Illustration des dynamischen Verhaltens eines Shrinkage-Schätzers. Bei Übereinstimmung zwischen RCT („randomized controlled trials“) und Register wird viel Information aus dem Register bezogen; bei wachsender Diskrepanz steigt das Gewicht der RCT-Daten. (Adaptiert nach Abb. 1 in Röver und Friede [18])

2 Studien wieder. In diesem Fall besteht eine alternative Motivation für die Modellierung über das „Referenzmodell“, welches die Schätzung z. B. auf Basis von RWD als eine möglicherweise „verzerrte“ Variante des RCT-Schätzers ansieht, wobei die Abweichung der beiden über ein Maß der Variabilität (auch als Heterogenität bezeichnet) quantifiziert wird [17].

Gelegentlich besteht die Sorge, dass externe RWD (z. B. aus einem großen Register) in einer gemeinsamen Analyse die Daten eines (kleinen) RCT völlig dominieren könnten. Derlei Effekte sollten allerdings nur in Analysen auftreten, in denen die Variabilität der Effekte zwischen den Studien entweder auf 0 gesetzt wird (wie z. B. in Common-effect-Metaanalysemodellen) oder auf 0 geschätzt wird, wie es z. B. in Standardverfahren der Random-effects-Metaanalysemodellen vorkommen kann. In einem hierar-

chischen Modell, und bei Betrachtung des relevanten Shrinkage-Schätzers, verhält sich die Schätzung „dynamisch“ in dem Sinne, dass die Heterogenität der Daten den Einfluss der externen Evidenz bestimmt. Hierdurch bleibt die Gewichtung der einzelnen Datenquellen in jedem Falle begrenzt [18]. Letztlich reproduziert die mathematische Analyse hier, was man auch intuitiv schlussfolgern würde; bei übereinstimmenden Daten werden diese als gegenseitige Bestätigung gesehen, und bei etwaigen Diskrepanzen wird vorsichtiger interpretiert.

Die **Abb. 1** illustriert das dynamische Verhalten eines Shrinkage-Schätzers, wenn ein RCT durch Daten eines Registers unterstützt wird. Bei Übereinstimmung der Daten aus beiden Quellen (Fall a) tragen die Register-Daten erheblich bei, und die Präzision steigt substantiell. Bei wachsender Diskrepanz der Daten (Fälle b–c) wird den RCT-Da-

ten zunehmendes Gewicht beigemessen, und der resultierende Shrinkage-Schätzer wird kaum noch durch das Register beeinflusst [18].

## Beispiele zur gemeinsamen Auswertung von RCT und RWD

Die *Creutzfeldt-Jakob-Krankheit (CJK)* ist eine äußerst seltene Erkrankung. Als mögliche Behandlungsoption bietet sich *Doxycyclin* an, jedoch ist die Datenlage zur Wirksamkeit ausgesprochen dünn. Ein RCT wurde initiiert, dieser wurde allerdings aufgrund der schleppenden Rekrutierung nach Einschluss von nur 12 Patienten vorzeitig beendet. Zur Analyse wurden letztlich weitere, *externe* Daten aus einer Beobachtungsstudie mit 88 Patienten hinzugezogen [26]. Im Rahmen einer Metaanalyse können dabei die Schätzungen basierend auf randomisierten und Beobachtungsdaten separat betrachtet werden, und die Präzision eines Shrinkage-Schätzers für den RCT-Effekt kann, bei gleichzeitiger Würdigung etwaiger Unterschiede, von der Berücksichtigung der Beobachtungsdaten substantiell profitieren; im vorliegenden Fall entspricht der Zugewinn an Information in etwa einer Verdoppelung der Fallzahl des RCT [17, 18].

Das *Alport-Syndrom* ist eine fortschreitende erbliche Erkrankung, welche in der Regel zu Nierenversagen führt. Ausgehend von Hinweisen, dass durch den Einsatz von ACE-Hemmern der Verlauf deutlich verzögert werden kann [6], wurde die Behandlung mit *Ramipril* im Rahmen eines RCT untersucht. Aufgrund der vielversprechenden Behandlung, und da diese bereits im Kindesalter ansetzt, bestanden vielfach Vorbehalte gegenüber einer Randomisierung, und eine sofortige Behandlung wurde vorgezogen. Dem wurde bereits im Studiendesign Rechnung getragen, einerseits durch eine unbalancierte Randomisierung zugunsten von Ramipril, bei der die Ramipril-Gruppe 2-mal so groß war wie die Placebokontrollgruppe, und andererseits durch Einschluss einer dritten, unverblindeten Gruppe (Open-label-Arm). Letztlich beinhaltete der RCT 20 Patienten, während zusätz-

Präv Gesundheitsf <https://doi.org/10.1007/s11553-023-01016-9>  
© Der/die Autor(en) 2023

T. Friede · C. Röver · T. Mathes

## Verknüpfung von randomisierten kontrollierten Studien und Real World Data

### Zusammenfassung

**Hintergrund.** Randomisierte kontrollierte Studien („randomized controlled trials“ [RCT]) sind der Goldstandard für die Evaluation von Interventionen. Allerdings wird oft angeführt, dass diese nur schwer durchzuführen seien und dass sie daher ggf. unter kleinen Fallzahlen liden. Zudem wird häufig kritisiert, dass RCT aus diesem Grund sowie durch (zu) enge Einschlusskriterien und zu starke Standardisierung vielfach nicht den klinischen Routinebedingungen entsprechen. Beides kann zu Einschränkungen in der Aussagekraft von RCT führen.

**Fragestellung.** Der Artikel zeigt auf, wie RCT und Real World Data (RWD)-basierte Studien voneinander profitieren können.

**Methoden.** Es wurde eine selektive Übersicht der Literatur zur Verknüpfung von Daten aus RCT und RWD erstellt.

**Ergebnisse.** Die RCT-Daten und RWD können mit unterschiedlichen Zielen verknüpft werden. Zum einen kann die Verknüpfung dazu dienen, die Effizienz der Auswertung eines RCT zu erhöhen. So können hierarchische Modelle zur Evidenzsynthese RWD nutzen, um die Präzision der RCT-Effektschätzung maßgeblich zu erhöhen. Zum anderen können RWD genutzt werden, falls die Übertragbarkeit von RCT auf die Routineversorgung zweifelhaft ist. Zur Erhöhung der externen Validität können u. a. verschiedene Gewichtungverfahren und Modellierungsmethoden verwendet werden. Umgekehrt können RCT-Daten genutzt werden, um eine systematische Verzerrung in

RWD zu bereinigen. Bei der „comprehensive cohort study“ erfolgt die Durchführung der RCT- und der Kohortenstudie parallel. Sie erlaubt die Einschätzung der externen Validität eines RCT und kann zudem bei einer gemeinsamen Auswertung von RCT und Registern sehr effizient sein.  
**Schlussfolgerungen.** Es bestehen diverse vielsprechende Möglichkeiten, Daten aus RCT und RWD zu verknüpfen. Es erscheint daher wünschenswert, dass Verknüpfungen vermehrt Anwendung finden. Hierbei ist wichtig, dass diese prospektiv geplant werden.

### Schlüsselwörter

Randomisierte kontrollierte Studie · Real World Data · Hierarchische Modelle · „Comprehensive cohort study“ · Externe Validität

## Combining randomized controlled trials and real world data

### Abstract

**Background.** Randomized controlled trials (RCTs) are the gold standard for evaluating interventions. However, they are often considered to be difficult to conduct and may therefore suffer from small case numbers. In addition, it is often claimed that RCTs do not represent clinical routine well. These may lead to limited significance or relevance of the results from RCTs.

**Objectives.** To show how RCTs and real world data (RWD)-based studies can benefit from each other.

**Methods.** This is a selective review of the literature on approaches for linking data from RCTs and RWD.

**Results.** RCTs and RWD can be linked with different aims. First, RWD can be used to increase the efficiency of the evaluation of RCTs. More specifically, hierarchical models for evidence synthesis can be utilized to combine RWD and RCT data to increase the precision of the RCT effect estimate. Second, RWD can be used if the applicability of RCTs results to routine care is doubtful. Here, various stratification methods and modelling methods are available that can increase the external validity of the RCT results. Conversely, information from RCTs can be utilized to adjust for bias in RWD. In the comprehensive cohort study design, the RCT and the cohort

study are carried out in parallel. It allows to assess the external validity of an RCT and can also be very efficient when the RCT and registry are evaluated jointly.

**Conclusions.** There are various promising ways of linking data from RCTs and RWD. Therefore, a more routine joint consideration of RCT and RWD data appears desirable. It is important that this is planned prospectively.

### Keywords

Randomized controlled trials · Real-World clinical trials · Hierarchical models · Comprehensive cohort study · External validity

liche Beobachtungsdaten einerseits aus dem Open-label-Arm (42 Patienten mit Ramipril) und andererseits aus Registerdaten (28 unbehandelte Patienten) zur Verfügung standen. Die schließlich durchgeführte Analyse basierte dann auf einem gemeinsamen hierarchischen Modell, wobei der RCT-Effekt wiederum durch einen Shrinkage-Schätzer quantifiziert werden konnte [7]. Auch in diesem Fall, wie schon oben im CJK-Beispiel, entspricht der Zugewinn an Information in etwa einer Verdoppelung der Fallzahl des RCT [7].

## Überblick über alternative Ansätze

Neben hierarchischen Modellen gibt es weitere Modelle zur kombinierten Analyse von Datenquellen unterschiedlicher Evidenzstärke. Ein prominenter Ansatz ist der sog. „power prior“, welcher in der Regel von einer sequentiellen Analyse z. B. von RWD und RCT-Daten ausgeht, bei welcher dann die Ergebnisse der ersten Analyse die A-priori-Information für die Bayes-Analyse der eigentlichen Studie darstellen. Um der begrenzten Vergleichbarkeit der Daten-

quellen Rechnung zu tragen, werden die RWD explizit heruntergewichtet [9]. Gewisse Analogien bestehen zwischen „power prior“ und hierarchischen Modellen bei Betrachtung zweier Studien [17]. Ein Problem des „power prior“ bleibt die Spezifizierung der Gewichtung der Daten; anders als in einem hierarchischen Modell, bei dem sich die Skalenparameter anschaulich in den gleichen Einheiten wie die eigentlichen Effekte bewegen [19], wirkt sich der Gewichtungsparameter *multiplikativ* auf den Standardfehler aus. Auch eine Inter-

pretation des Gewichtungsparameters als weitere Unbekannte in einem Bayes-Modell (und die damit notwendige Spezifikation einer sinnvollen A-priori-Verteilung) stellt keine einfache Lösung dar.

Generell kann man auch versuchen, externe Evidenz in seiner Gesamtheit durch Befragung von Experten zu quantifizieren („expert elicitation“; [15]). Information aus RWD-Quellen kann hier natürlich einen Teil der Evidenz darstellen und somit Teil des Expertenwissens sein. Grundsätzlich ist dabei aber zu beachten, dass es sich hierbei dann um eine sehr „informelle“ Art der Synthese handelt, und dass implizite Annahmen oder Schlussfolgerungen eventuell im Verborgenen bleiben.

### Nutzung von RWD zur Erhöhung der Übertragbarkeit auf Populationen in der Routineversorgung

Häufig wird angeführt, dass die Ergebnisse aus RCT nicht ohne weiteres übertragbar auf Patienten unter „Real-life“-Bedingungen seien [11]. Verschiedene Faktoren können die externe Validität eines RCT beeinflussen. Hierzu zählen insbesondere Unterschiede in den Populationscharakteristika, des Settings, sowie Unterschiede bzgl. Flexibilität der Erbringung und Anwendung der Intervention (nicht-medikamentöse Verfahren) bzw. Adhärenz (Medikamente; [13]).

Eine eingeschränkte externe Validität sollte nicht *per se* unterstellt werden, sobald Abweichungen zwischen RCT (z. B. Population) und Zielpopulation und/oder Gegebenheiten vorliegen, sondern es sollten begründete Annahmen bestehen (z. B. ein plausibler Mechanismus), dass diese auch zu einer Abweichung des Behandlungseffekts führen können. Das heißt, Unterschiede in Populationscharakteristika, wie Alter, Geschlecht und Komorbidität, sind für die externe Validität der Schätzung von Behandlungseffekten nur von Bedeutung, wenn davon auszugehen ist, dass relevante, im besten Fall belegte Subgruppeneffekte (unterschiedliche Effekte in verschiedene Gruppen, z. B. Altersgruppen) existieren. Beispielsweise wird häufig

kritisiert, dass Probanden in RCT jünger und gesünder sind als Patienten im Versorgungsalltag und aus diesem Grund die Ergebnisse nicht übertragbar seien. Es liegt auf der Hand, dass ältere Menschen häufiger versterben als jüngere Menschen. Dennoch bedeutet diese Abweichung der absoluten Häufigkeiten nicht zwangsläufig, dass sich die relative Schätzung des Behandlungseffekts unterscheidet. Darüber hinaus müssen sich die Studienpopulation und die Zielpopulation bzw. Gegebenheiten in einem relevanten Ausmaß unterscheiden. Gleiches gilt für Unterschiede in der Interventionserbringung, der Adhärenz oder dem Setting. Erst wenn sowohl relevante Abweichung als auch relevante Effektmodifikatoren, also Interaktionen mit dem Behandlungseffekt, vorliegen, führen diese zu einem Unterschied im (relativen) Behandlungseffekt.

Weiterhin sollte bedacht werden, dass die externe Validität immer nur für eine bestimmte Zielpopulation erreicht werden kann, jedoch so gut wie nie allgemeine Generalisierbarkeit. Eine Erhöhung der externen Validität für eine bestimmte Zielpopulation (z. B. ältere Menschen, bestimmte Region) kann ggf. sogar eine geringere Übertragbarkeit für eine andere Zielpopulation (jüngere Menschen, andere Region) implizieren. Aus diesem Grund sollte unabhängig vom Konzept die Zielpopulation stets sehr eindeutig definiert sein.

Falls Anhaltspunkte bestehen, dass die externe Validität für eine bestimmte relevante Zielpopulation eingeschränkt ist, können weitere Studien zur Prüfung des Behandlungseffekts unter Real-life-Bedingungen angezeigt sein. Neben der Durchführung einer neuen Studie unter Routinebedingungen als Ergänzung der Evidenz aus RCT existieren verschiedene Ansätze um die Ergebnisse aus RCT und die Ergebnisse unter „Real-life-Bedingungen“ anzunähern. Bei den im folgenden Abschnitt skizzierten Methoden werden nur eingeschränkt übertragbare RCT mit Hilfe von vorhandener RWD kalibriert, mit dem Ziel, die externe Validität zu erhöhen. Für diese Ansätze ist mindestens ein RCT sowie Informationen über die Verteilung der Merkmale von Interesse (z. B. relevante

Subgruppen) in der Zielpopulation (z. B. Altersverteilung) notwendig. Im Abschnitt „Modellierung der Verzerrung in Evidenzsynthesen von RWD“ werden Methoden aufgezeigt, die einen umgekehrten Ansatz verfolgen. Hier werden potenziell intern nicht valide – jedoch extern valide – RWD-basierte Studien mit Hilfe von Daten aus RCT kalibriert, mit dem Ziel, die interne Validität zu erhöhen. Damit diese Ansätze angewendet werden können, werden ein Set an RWD-basierten Studien und externe Informationen über die Auswirkung von Störgrößen (Verzerrung) benötigt. Der letzte Abschnitt stellt einen Ansatz dar, wie RWD zur Durchführung von RCT beitragen können.

### Stratifikation und Gewichtung von RCT

Sowohl die Stratifikation als auch die Gewichtung zielen darauf ab, die Studienpopulation und die Zielpopulation zu harmonisieren. Die Stratifikation funktioniert gleichermaßen wie eine stratifizierte Stichprobenziehung. Zunächst werden möglichst homogene Schichten definiert (d. h. einzelne Untergruppen, z. B. verschiedene Altersgruppen). Dann wird eine „Stichprobe“ aus dem RCT gezogen. Die Größe der einzelnen Stichprobe wird dabei so gewählt, dass die Größe der einzelnen Schichten proportional zur Größe der Schicht der Zielpopulation ist. Dadurch ist jede Schicht in dem RCT in gleicher Relation wie in der Zielpopulation vertreten. Der Behandlungseffekt wird dann in dieser neuen RCT-Population geschätzt. Wie direkt ersichtlich, ist ein großer Nachteil dieser Methode, dass hier nicht die gesamte RCT-Population genutzt wird. Jeder zusätzliche Effektmodifikator und jede zusätzliche Schicht führt potenziell zu einer Verkleinerung der RCT-Population. Die Methode ist daher insbesondere im Falle von kleinen RCT und/oder vielen Effektmodifikatoren nicht gut geeignet.

Bei der Gewichtung wird zunächst ein Modell gebildet, in dem die Wahrscheinlichkeit an dem RCT teilzunehmen in Abhängigkeit der Probandencharakteristika modelliert wird [24]. Hierfür ist ein aus RCT-Population und RWD-

Population kombinierter Datensatz notwendig, der die relevanten Prädiktoren für die Teilnahme enthält. Die Schätzung erfolgt mittels Regressionsmodellen, wie logistische Regression im Falle von binären Daten oder Cox-Regression im Fall von Überlebenszeiten. Mit Hilfe dieser Modelle wird für jede Beobachtung ein *Propensity Score (PS)*, d. h., die Wahrscheinlichkeit, mit der ein Patient in den RCT eingeschlossen wird, ermittelt. Die RCT-Probanden werden mit der inversen Wahrscheinlichkeit an der Studie teilzunehmen gewichtet. Sowohl die Schichtung als auch die Gewichtung können lediglich Homogenität bzgl. der Population herstellen. Abweichungen aufgrund von anderen Aspekten (z. B. Adhärenz, Setting) können nicht ausgeglichen werden.

### Modellierungsansätze zur Extrapolation der Effekte aus RCT

Im Gegensatz zu den im vorangegangenen Abschnitt aufgezeigten Methoden existieren Ansätze, die nicht nur Unterschieden in der Population, sondern auch anderen unter Real-life-Bedingungen ggf. abweichenden Faktoren (z. B. geringere Adhärenz) Rechnung tragen können. Es finden sich insbesondere in der gesundheitsökonomischen Literatur zahlreiche Beispiele in denen Daten aus RCT mit RWD mittels Modellierung verknüpft werden, um den Effekt unter Routinebedingungen zu schätzen [16]. Im Gegensatz zu den anderen hier im Artikel beschriebenen Ansätzen werden in Modellen nicht nur zwei „Arten“ von Evidenz (aus RWD- und RCT-Studien) verknüpft, sondern es geht eine Vielzahl an verschiedenen Daten aus verschiedenen Quellen (z. B. epidemiologischen und klinischen Studien) in das Modell ein. Das heißt, RWD und Daten aus RCT bilden nur einen Teil der Inputdaten, die in das Modell eingehen.

Zu den verwendeten Modellierungsverfahren zählen v. a. Mehrstadienmodelle und diskrete Ereignissimulationen (vgl. z. B. [21] und [10]). Die möglichen Auswirkungen von in der Routine ggf. abweichenden Patientencharakteristika, Variation der Interventionserbringung oder abweichendem Patientenverhalten

werden in der Regel durch Variation der Input-Parameter im Rahmen von Sensitivitätsanalysen analysiert, um den Einfluss auf die im Modell verglichenen Interventionen zu untersuchen. Die Parameter (z. B. verschiedene hohe Adhärenz) werden dabei über vordefinierte plausible Spannweiten verändert, wie sie aus RWD bekannt sind. Die Ergebnisse sind daher nicht genau eine Schätzung des Behandlungseffekts unter Real-life-Bedingungen – auch wenn dieses theoretisch möglich wäre – sondern eine Vielzahl an Ergebnissen für die einzelnen Parameter und Variation, die die potenziellen Auswirkungen von möglichen Abweichung auf den Behandlungseffekt quantifizieren („deterministic sensitivity analysis“). Da die Schätzung der Input-Parameter mit Unsicherheit verbunden ist, kann in die Analyse zudem die Unsicherheit der Schätzung eingehen (Verteilung um den Punktschätzer) und hierdurch für die verschiedenen Ergebnisse ein Vertrauensbereich geschätzt werden („probabilistic sensitivity analysis“). Insgesamt werden diese Ansätze jedoch bisher wenig genutzt mit dem primären Ziel Real-life-Behandlungseffekte abzuschätzen [16].

### Modellierung der Verzerrung in Evidenzsynthesen von RWD

In den vorangehenden Abschnitten wurden Methoden beschrieben, die darauf abzielen, die externe Validität eines oder mehrerer RCT zu erhöhen. Die in diesem Abschnitt skizzierten Methoden gehen den umgekehrten Weg. Hier werden intern nicht-valide Beobachtungsstudien für ihr potenzielles Verzerrungsrisiko adjustiert. Es finden sich in der Literatur verschiedene Ansätze, um den zusammengefassten Effektschätzer für den Behandlungseffekt aus einer Metaanalyse zu adjustieren [23]. Diese Ansätze wurden ursprünglich für Beobachtungsstudien i. Allg. entwickelt. Sie können somit auch auf RWD-basierte nicht-randomisierte Studien angewendet werden. Alle Methoden haben gemein, dass zunächst das Risiko für Verzerrung für die einzelnen Studien in der Metaanalyse mit einem Bewertungstool für die Studienqualität bewertet werden muss (z. B. mittels

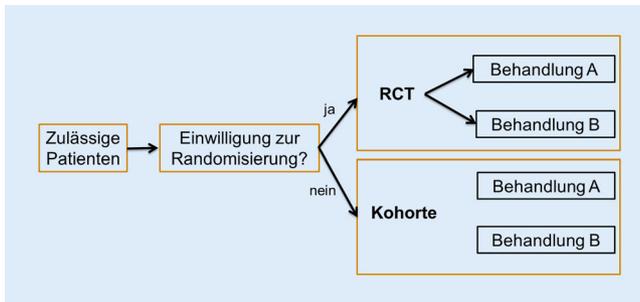
des „risk of bias in non-randomized studies of interventions“ [ROBINS-I]-Tools; [22]). Die Referenz für die Bewertung ist dabei ein intern valider RCT. Anschließend muss der potenzielle Einfluss dieses Verzerrungsrisikos quantifiziert werden, um in den gemeinsamen Schätzer für den Behandlungseffekt mit einfließen zu können.

Turner et al. haben vorgeschlagen, dass das Gesamtrisiko für Verzerrung für jede Studie direkt um einen von Experten geschätzten Faktor bereinigt wird [25]. Der für jede Studie bereinigte Effekt geht dann in die Metaanalyse ein. Der Hauptkritikpunkt an dieser Methode ist, dass diese Einschätzung, auch wenn sie durch mehrere Gutachter vorgenommen wird, sehr subjektiv bleibt.

Beim Vorschlag von Welton et al. beruht die Einschätzung der Verzerrung auf metaepidemiologischen Studien zur Untersuchung der Auswirkung von Verzerrungsquellen (z. B. einer fehlenden Verblindung) auf die Schätzung des Behandlungseffekts [27]. Es werden anhand dieser empirischen Daten A-priori-Annahmen für die mittlere Verzerrung, ihre Heterogenität und Unsicherheit abgeleitet. Diese gehen über ein Bayesianisches hierarchisches Modell (vgl. Abschnitt „Höhere Effizienz in RCT-Auswertung durch RWD“) in die Schätzung des kombinierten Behandlungseffekts mit ein. Ein Nachteil dieses Ansatzes sind die Modellannahmen, dass die Verzerrung für alle Studien innerhalb einer Metaanalyse austauschbar ist, und dass die Schätzungen aus vorherigen Metaanalysen auch für die neue Metaanalyse gelten. Zudem sind für die Schätzung der A-priori-Verteilung möglichst genaue Angaben (z. B. gleiche Population und Indikation) aus metaepidemiologischen Studien notwendig.

### Nutzung von RWD für RCT („pragmatic trials“)

Häufig wird unter RWD ausschließlich Daten aus nicht-randomisierte Studiendesigns verstanden. Allerdings ist dies streng genommen nicht korrekt, da die Randomisierung an sich nicht mit der externen Validität in direktem Zusammenhang steht. Vielmehr ist die einge-



**Abb. 2** ◀ Schematische Darstellung einer Studie im „comprehensive cohort design“. RCT „randomized controlled trial“. (Angepasst nach Schmoor et al. [20])

schränkte externe Validität oftmals ein „selbst gemachtes Leid“, das aufgrund von Restriktionen im Studienprotokoll (z. B. enge Einschlusskriterien, Standardisierung der Intervention) sowie den Studienkontext hervorgerufen wird. Ein vielversprechender Ansatz, um das Beste aus beiden Welten zu vereinen, ist es, für die Durchführung von RCT auf RWD zurückzugreifen [14]. Für die Datensammlung eines RCT wird dabei, soweit wie möglich, auf eine bereits bestehende Datenbank von RWD (z. B. Register) zurückgegriffen. Allerdings bedeutet die Nutzung von RWD nicht automatisch eine Erhöhung der externen Validität. Eine Erhöhung der externen Validität gegenüber einem „konventionellen“ RCT kann nur erreicht werden, wenn die Nutzung von RWD mit einem Pragmatic-trial-Ansatz, d. h. insbesondere einem wenig restriktiven Studienprotokoll (z. B. breite Einschlusskriterien, flexible Interventionserbringung) verbunden wird. Diese Thematik wird in dem Artikel „Pragmatischere randomisierte Studien mit Fokus auf Registerbasierung“ von Lange und Lauterberg [12] dieser Artikelserie tiefergehend betrachtet.

Neben der potenziellen Erhöhung der externen Validität hat die Nutzung von vorbestehenden RWD-Datenbanken den Vorteil, dass durch das Zurückgreifen auf bereits bestehenden Daten der Aufwand für die aktive Datenerfassung reduziert werden kann.

### Das „comprehensive cohort study“ Design

Eine „comprehensive cohort study“ besteht aus zwei Teilen – einem RCT und einer Kohortenstudie. Abgesehen von der Randomisierung sind die Abläufe in den

beiden Teilen ansonsten gleich. Insbesondere gelten die gleichen Ein- und Ausschlusskriterien, auch die Behandlungen und die Nachbeobachtung unterscheiden sich nicht. Eine schematische Darstellung einer „comprehensive cohort study“ ist in **Abb. 2** gezeigt; diese ist der Abb. 1 in Schmoor et al. [20] nachempfunden. Schmoor et al. geben des Weiteren einen Überblick über die Entwicklung der Designs seit den 1980er-Jahren [20].

Die ursprüngliche Idee der „comprehensive cohort study“ war es, dass die externe Validität des RCT durch den Vergleich zur Kohorte (bzgl. Charakteristika der Population, der Prognose und der Behandlungseffekte) eingeordnet werden konnte [20]. Sofern die Ein- und Ausschlusskriterien pragmatisch gewählt wurden, sind so auch Rückschlüsse auf die externe Validität in Bezug auf Routinepopulationen möglich. Die Schätzung des Behandlungseffekts in der Kohorte bedarf des Einsatzes besonderer Auswerteverfahren wie z. B. Propensity Scores. Theoretisch können auch die oben beschriebenen Verfahren der Stratifikation und Gewichtung angewendet werden. Man kann aber auch noch einen Schritt weiter gehen. Die Behandlungseffekte aus den beiden Teilen der „comprehensive cohort study“, nämlich dem RCT und der Kohorte, können durch das oben beschriebene Bayes-Verfahren der Shrinkage-Schätzung auch formal zusammengeführt werden. So erhalten wir z. B. mittels dynamischer Gewichtung (in Abhängigkeit von der Übereinstimmung) einen Behandlungseffekt für den RCT im Lichte der Daten der Kohorte [17].

### Infobox Mehr Informationen zum Thema

- <http://www.jeremy-oakley.staff.shef.ac.uk/shelf/>
- <https://cran.r-project.org/package=bayesmeta>
- <https://methods.cochrane.org/bias/risk-bias-non-randomized-studies-interventions>

### Abschließende Bemerkungen

Wie die obigen Ausführungen gezeigt haben, haben sowohl RCT wie auch RWD ihren Platz. Für die Praxis ist insbesondere deren gemeinsame Betrachtung von Interesse, da so die Vorteile beider Ansätze genutzt werden können. Bei der gemeinsamen Auswertung von RCT und RWD sollten – soweit möglich – Prinzipien, wie sie bei der Auswertung von RCT Anwendung finden, auch hier angewendet werden. Dies bedeutet insbesondere, dass die gemeinsamen Auswertungen von RCT und RWD prospektiv geplant und in einem statistischen Analyseplan (SAP) dokumentiert werden.

### Fazit für die Praxis

- Auch wenn RCT („randomized controlled trials“) manchmal nur schwer durchzuführen sind und häufig nicht Populationen repräsentieren, wie sie in der klinischen Routine tatsächlich vorkommen, so liefern diese doch robuste Evidenz zur Wirksamkeit und Sicherheit von Therapien.
- Auch kleine RCT in Populationen seltener Erkrankungen und pädiatrischen Populationen können wichtige Informationen liefern, wenn sie durch geeignete statistische Verfahren mit Real World Data (RWD) verknüpft werden.
- Verschiedene Verknüpfungsansätze von RWD und RCT können die Übertragbarkeit der Ergebnisse auf die Routineversorgung erhöhen.
- Eine „comprehensive cohort study“ erlaubt die Einschätzung der externen Validität eines RCT, sofern die Ein- und Ausschlusskriterien pragmatisch gewählt wurden, und kann bei

- einer gemeinsamen Auswertung von RCT und Register sehr effizient sein.
- Gemeinsame Auswertungen von RCT und RWD sollten prospektiv geplant und dokumentiert werden.

## Korrespondenzadresse



**Prof. Dr. Tim Friede**  
 Institut für Medizinische  
 Statistik, Universitätsmedizin  
 Göttingen  
 Humboldtallee 32,  
 37073 Göttingen,  
 Deutschland  
 tim.friede@  
 med.uni-goettingen.de

**Funding.** Open Access funding enabled and organized by Projekt DEAL.

## Einhaltung ethischer Richtlinien

**Interessenkonflikt.** T. Friede, C. Röver und T. Mathes geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

**Open Access.** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Borenstein M, Hedges LV, Higgins JPT et al (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 1(2):97–111
- Concato J, Corrigan-Curay J (2022) Real-world evidence—where are we now? *N Engl J Med* 386:1680–1682
- Friede T, Röver C, Wandel S et al (2017) Meta-analysis of few small studies in orphan diseases. *Res Synth Methods* 8(1):79–91
- Friede T, Posch M, Zohar S et al (2018) Recent advances in methodology for clinical trials in small populations: the InSPiRe project. *Orphanet J Rare Dis* 13:186
- Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
- Gross O, Licht C, Anders HJ et al (2012) Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy. *Kidney Int* 81:494–501
- Gross O, Tönshoff B, Weber LT et al (2020) A multicenter, randomized, placebo-controlled, double-blind phase 3 trial with open-arm comparison indicates safety and efficacy of nephroprotective therapy with ramipril in children with Alport's syndrome. *Kidney Int* 97(6):1275–1286
- Held L (2008) Methoden der statistischen Inferenz: Likelihood und Bayes. Spektrum, Heidelberg
- Ibrahim JG, Chen MH (2000) Power prior distributions for regression models. *Stat Sci* 15(1):46–60
- Karnon J, Stahl J, Brennan A et al (2012) Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—4. *Med Decis Making* 32(5):701–711
- Kennedy-Martin T, Curtis S, Faries D et al (2015) A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 16(1):1–14
- Lange S, Lauterberg J (2022) Pragmatische randomisierte Studien mit Fokus auf Registerbasierung. *Präv Gesundheitsf.* <https://doi.org/10.1007/s11553-022-00974-w>
- Loudon K, Tweek S, Sullivan F et al (2015) The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ.* <https://doi.org/10.1136/bmj.h2147>
- Mathes T, Buehn S, Prengel P et al (2018) Registry-based randomized controlled trials merged the strength of randomized controlled trials and observational studies and give rise to more pragmatic trials. *J Clin Epidemiol* 93:120–127
- O'Hagan A, Buck C, Daneshkhan A et al (2006) Uncertain judgements: eliciting experts' probabilities. Wiley, Hoboken
- Panayidou K, Gsteiger S, Egger M et al (2016) GetReal in mathematical modelling: a review of studies predicting drug effectiveness in the real world. *Res Synth Methods* 7(3):264–277
- Röver C, Friede T (2020) Dynamically borrowing strength from another study through shrinkage estimation. *Stat Methods Med Res* 29(1):293–308
- Röver C, Friede T (2021) Bounds for the weight of external data in shrinkage estimation. *Biom J* 65(5):1131–1143
- Röver C, Bender R, Dias S et al (2021) On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res Synth Methods* 12(4):448–474
- Schmoor C, Olschewski M, Schumacher M (1996) Randomized and non-randomized patients in clinical trials: Experience with comprehensive cohort studies. *Stat Med* 15:263–271
- Siebert U, Alagoz O, Bayoumi AM et al (2012) State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force—3. *Med Decis Making* 32(5):690–700
- Sterne JA, Hernán MA, Reeves BC et al (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* <https://doi.org/10.1136/bmj.i4919>
- Stone JC, Glass K, Munn Z et al (2020) Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. *J Clin Epidemiol* 117:36–45
- Stuart EA, Ackerman B, Westreich D (2018) Generalizability of randomized trial results to target populations: design and analysis possibilities. *Res Soc Work Pract* 28(5):532–537
- Turner RM, Spiegelhalter DJ, Smith GC et al (2009) Bias modelling in evidence synthesis. *J R Stat Soc A* 172(1):21–47
- Varges D, Manthey H, Heinemann U et al (2017) Doxycycline in early CJD—a double-blinded randomized phase II and observational study. *J Neurol Neurosurg Psychiatry* 88(2):119–125
- Welton NJ, Ades AE, Carlin JB et al (2009) Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc A* 172(1):119–136