

3 Messen des klinischen Nutzens

Von etwa 300 Studien im Jahr 2011 in Deutschland, die gemäß MPG vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) genehmigt wurden, hatte die Mehrzahl ein lediglich „einarziges Design“, das heißt, es fehlte eine Vergleichsgruppe.^[9] Nach Windeler et al. sollte jegliche nicht preisneutrale Modifikation eines Produktes bei den Klasse-III-Produkten durch RCT evaluiert sein. Dies gilt auch, wenn sich der Nutzen der Modifikation auf nicht-klinische Parameter, wie beispielsweise die Art der Anwendung, bezieht.^[9] Für neue Produkte der Klasse IIb sollten demnach schon bei der Zulassung klinische Daten, allerdings nicht notwendigerweise RCT, vorgelegt werden. Neue MP der Klasse III dagegen bedürfen bereits vor Marktzugang einer klinischen Prüfung. Nach Ansicht des IQWiG sollten gänzlich neue Produkte dabei in RCT evaluiert werden.^[61] Vor dem Hintergrund des GKV-VSG erscheint diese Forderung in ihrem vollen Umfang nicht darstellbar. Die sogenannten Schrittinnovationen sind vom § 137h SGB V mehrheitlich nicht betroffen, zumindest dann nicht, wenn sie kein „neues theoretisch-wissenschaftliches Konzept“^[45] aufweisen, sondern eine Weiterentwicklung bestehender Produkte sind.

3.1 Herausforderungen beim Messen des klinischen Nutzens

Das Messen von Gesundheit, d.h. der damit in Verbindung stehenden klinischen Effekte, stellt eine wesentliche Herausforderung der Nutzenbewertung von Methoden mit MP hoher Klassen dar. Das Messen ist ein Prozess, bei dem einer Entität (gegebenen Größe) der realen Welt numerische Werte (Zahlen) zugeschrieben werden. Die Zuschreibung muss entsprechend klar definierter

Regeln erfolgen. Eine Grundidee bei der Bewertung des klinischen Nutzens ist, dass sich die Intervention bzw. der komparative Vergleich von Interventionen durch ein oder mehrere klinische Effekte beschreiben lässt.

Mit Hilfe des Messens versucht man zu verstehen, was den Patientennutzen kausal beeinflusst. Zusammenhänge zwischen Interventionen und der Gesundheit können so erklärt werden. So könnten die Entscheidungen über die therapeutischen Maßnahmen mit MP unterstützt oder aber einen Vergleich konkurrierender Methoden aufgezeigt werden. Darüber hinaus könnten die Erkenntnisse klinischer Studien zur Entstehung neuer und zur Weiterentwicklung oder Verfeinerung bestehender Interventionen respektive Produkte beitragen. Fortschritt kann ohne die Bewertung und Quantifizierung des klinischen Nutzens nicht stattfinden. Deshalb nimmt das Messen einen so hohen Stellenwert in der klinischen Forschung und im Versorgungsalltag ein.

Die wesentliche Voraussetzung für die vergleichende Bewertung ist, dass dem Nutzen (numerische) Werte zugeordnet werden können. Hinsichtlich der Messung sind sowohl das Konstrukt „Nutzen“ als auch das Konzept „Gesundheit“ problematisch. Skalen und Maßeinheiten, welche allgemein die Gesundheit oder den Nutzen repräsentieren, sind nicht bekannt. Gesundheit kann als Konzept verstanden werden. Es handelt sich um einen theoretischen Begriff, da man davon ausgehen kann, dass weder Dritte noch das Individuum selbst die Gesundheit vollumfänglich direkt beobachten oder messen können. Gesundheit ist ein multidimensionales Konzept, die Operationalisierung erfolgt auf Basis multipler Zielkriterien. Interventionen zielen auf die Verbesserung der Mortalität, Morbidität, Lebensqualität und Zufriedenheit der Patienten.

3.2 Zielkriterien, Zielgrößen oder Nutzendimensionen

Oft bleibt unklar, welche klinischen Zielkriterien, Zielgrößen oder Nutzendimensionen zur Beurteilung positiver oder negativer Effekte herangezogen werden können. Die Ableitung von Zielkriterien bzw. patientenrelevanten Endpunkten ist abhängig von der zugrundeliegenden Definition des Nutzens. Im Methodenbericht des IQWiG wird der Begriff „Nutzen“ als kausal begründeter positiver Effekt einer medizinischen Intervention auf patientenrelevante Endpunkte definiert.^[4] Hier scheint eine Definition gewählt, die sich ausschließlich an den klinischen Effekten orientiert, also einem „medizinischen Nutzenbegriff“. Diese Definition umfasst nur klinisch messbare Effekte, nicht den Nutzen, der beim Patienten entsteht. Bei dieser Definition handelt es sich nicht um den ökonomischen Nutzenbegriff. Im Internet Glossar definiert das IQWiG einen „engen“ und einen „weiten“ Nutzenbegriff. „In seiner engen Bedeutung lehnt er sich an die EbM an und spiegelt den reinen medizinischen Nutzen zur Beurteilung einer Maßnahme wider (= Gesundheitseffekte/Outcomes)“.^[62]

Erst in seiner weiten Bedeutung bezieht sich der Nutzenbegriff des IQWiG auf den beim Patienten entstandenen Nutzen. Der Nutzen wird dann nicht nur

als ein Gesundheitseffekt einer Intervention definiert, „sondern berücksichtigt auch den Wert, den der Patient diesem Effekt zuschreibt (sogenannte Nutzwerte/Utilities)“.^[62] Dieser Begriff lehnt sich scheinbar an die neoklassische Theorie an, d.h. den Wert den der Konsument einer Intervention, Handlung oder Alternative beimisst. Der durch den Experten oder Konsumenten prognostizierte Nutzen muss nicht immer dem tatsächlich eintretenden Nutzen entsprechen. Man spricht vom Erwartungsnutzen, wenn zum Zeitpunkt der Entscheidung aufgrund von Unsicherheit der Nutzen nicht genau quantifiziert werden kann. Offen bleibt, welcher Nutzenbegriff handlungsleitend für die Nutzenbewertung sein wird. Die Bewertung des Patientennutzens ist mit der Berücksichtigung des „engen“ Nutzenbegriffs nicht möglich.

Bei der Bewertung von Gesundheitstechnologien werden heute nicht mehr nur so genannte „objektive“ klinische Maße (z.B. Laborparameter) gemessen („klinische Evidenz“). Auch die subjektive Bewertung von Wirkungen und Nebenwirkungen einer Therapie seitens der Patienten ist neben der klinischen Messung ein Teil der Nutzenbewertung.^[25] Für die Messung des Patientennutzens ist es notwendig, die nicht direkt beobachtbaren theoretischen Begriffe (Konstrukte, theoretische Variablen) in direkt beobachtbare, messbare empirische Entsprechungen (Indikatoren, manifeste Variablen) zu transformieren.

Als Indikatoren für die Variablen werden klinisch relevante und patientenrelevante Effekte (Mortalität, Morbidität, Lebensqualität, Aufwand und Zufriedenheit) zur Analyse der latenten Variablen des „Patientennutzens“ ausgewählt. Wichtiger Bestandteil einer Nutzenbewertung ist die Offenlegung der Gründe für die in einer Konzeptspezifikation vorgenommenen Präzisierungen und Einschränkungen. Konkret bedeutet dies eine Dokumentation der Identifikation von Variablen der Effekte und der Argumentation zur Ableitung der latenten Variablen des Patientennutzens. Es ist notwendig, die als patientenrelevant identifizierten Eigenschaften und deren Herleitung zu dokumentieren.

3.3 Nachweis der Kausalität

Grundsätzlich bezieht sich der Begriff des klinischen Nutzens auf einen positiven klinischen Effekt und wird dem Risiko bzw. Schaden gegenübergestellt.^[50] Diese Betrachtung erfolgt immer unter der Annahme, dass Nutzen und Schaden aus einer spezifischen Intervention resultieren. Bei der Entwicklung innovativer MP ist es wichtig, nachweisen zu können, dass ein neues theoretisch-wissenschaftliches Konzept einen positiven Effekt auf die Gesundheit des Patienten hat und somit als Ursache für einen gesteigerten Patientennutzen angesehen werden kann. Eine Beziehung zwischen Ursache und Wirkung wird auch als Kausalität bezeichnet. Ohne die Annahme der Kausalität macht eine tiefergehende Analyse von Nutzen und Schaden keinen Sinn. Demnach ist die ursächliche Verbindung für die Schlussfolgerungen der Nutzen-

bewertung wichtig^[51], d.h. „die kausale Verbindung von Intervention und Auswirkung“ muss bestmöglich belegt werden.^[25] Die Bewertung des Kausalzusammenhangs ist immer komplex und mit Unsicherheiten verbunden.

Der erhoffte Informationsgewinn aus klinischen Studien ist folglich der Kausalitätsnachweis einer Intervention.^[50, 63] Grundbedingung für die Bestimmung einer Ursache-Wirkungs-Beziehung ist die zeitliche Abfolge der beobachteten Ereignisse.^[64] Eine Voraussetzung für die Identifizierung eines Zusammenhangs zwischen einer Intervention und dem Auftreten eines klinischen Ereignisses ist daher die Abgrenzung der Kausalität.^[65] Die zeitliche Richtung in der Abfolge der Ereignisse ist demzufolge ein essentielles Kriterium für die Kausalitätsbestimmung.^[64] Das bedeutet, dass der Aufbau einer Studie die Ableitung des ursächlichen Zusammenhangs einer Behandlung sowie der gesundheitlichen Auswirkungen ermöglichen muss. Dieser Kausalitätsanspruch bedingt bestimmte Anforderungen an das Design einer Studie. Kausalität kann immer nur dann angenommen werden, wenn ein fester zeitlicher Zusammenhang zwischen einer Ursache und einer zeitlich darauffolgenden Wirkung besteht.

Zur Beurteilung des Studiendesigns sind Faktoren zu beachten, die das Studienergebnis verzerren können.^[64] Für die Planung einer Studie ist es entscheidend, die Kriterien zu kennen, welche die Erkennung einer Kausalität unterstützen. Weiterhin sind Kenntnisse über die wissenschaftlichen Methoden zur Bestimmung eines Kausalzusammenhangs zwischen zwei Ereignissen notwendig. Zur Erkennung von Kausalität und der Bewertung einer Ursache-Wirkungs-Beziehung stehen verschiedene Beurteilungskriterien zur Verfügung.^[64]

1. Den stärksten Hinweis auf eine Ursache-Wirkungs-Beziehung liefert das Experiment am Menschen. Dies gilt als Standardkriterium für klinische Studien.
2. Das zweite Kriterium umfasst die Frage nach der Stärke der Assoziation von Ursache und Wirkung. Um eine starke Assoziation zwischen Ursache und Wirkung nachweisen zu können, ist das Studiendesign entscheidend.
3. Eine Assoziation, die in einer Studie nachgewiesen wurde, sollte zudem in anderen Studien reproduzierbar sein.
4. Das vierte Kriterium ist der temporäre Zusammenhang von Ursache und Wirkung. Jedoch ist die Definition bzw. die Feststellung einer zeitlichen Reihenfolge von Ursache und Wirkung aufgrund der Komplexität nicht immer einfach zu realisieren. In der Theorie folgt im zeitlichen Ablauf die Wirkung auf die Ursache. In der Praxis ist dies gegebenenfalls schwierig zu unterscheiden.^[64]

3.4 Randomisierte kontrollierte Studie

In Studien muss ein Vergleich verschiedener Patienten unter gleichen Ausgangsbedingungen erfolgen, um Auswirkungen möglicher Störgrößen bzw. beeinflussender Faktoren zu minimieren. Das bedeutet, dass der Vergleich von Behandlungseffekten an Probanden strukturell vergleichbarer Gruppen erfolgen sollte.^[50, 61, 63] Dieses Konzept wird in vergleichenden Interventionsstudien angewendet, welche Aussagen über die Effekte einer medizinischen Maßnahme ermöglichen.^[50] Als verlässlichste Interventionsstudie gilt die RCT, welche auch als „Goldstandard“ deklariert wird.^[19, 50, 66–69]

Die randomisierten kontrollierten Studien untersuchen die Wirksamkeit einer Behandlungsmöglichkeit anhand des Vergleiches mit Behandlungsalternativen und können den Nachweis der kausalen Effekte auf die Gesundheit liefern. Oftmals werden hierfür die Standardbehandlung und/oder die Nicht-/Scheinbehandlung zum Vergleich herangezogen. Wird die neue Behandlung mit der aktuellen Standardtherapie verglichen, spricht man von einem aktiven Komparator.^[63, 70]

Bei einer RCT werden Patienten nach dem Zufallsprinzip auf die zu untersuchende Intervention und mindestens eine Kontrollintervention (Vergleichstherapie, keine Behandlung oder Placebo) verteilt. Zudem erfolgt ein analoger Vergleich ihrer Behandlungsergebnisse.^[66, 67] Insbesondere in der Zufallszuteilung (Randomisierung) der Patienten ist der Vorteil der RCT begründet. Bei diesem Vorgehen wird davon ausgegangen, dass sich alle bekannten und unbekannten Patientenmerkmale, welche in einer Studie als Störgrößen wirksam werden könnten, zufällig auf die beiden Gruppen verteilen. Das bedeutet durch die Randomisierung wird eine Strukturgleichheit (z.B. ähnlicher Altersdurchschnitt, Geschlechtsverteilung) angestrebt, welche den Vergleich zweier Patientengruppen in Bezug auf ihre Behandlungsergebnisse erlaubt.^[66] Allerdings kann die Randomisierung einer einzelnen Studie noch keine völlige Strukturgleichheit gewährleisten, eine Restunsicherheit bleibt trotzdem bestehen. Die Vergleichbarkeit der Ergebnisse hängt letztendlich von der tatsächlichen Struktur der einzelnen Gruppen ab.^[63, 70] Ein Vergleich mehrerer randomisierter kontrollierter Studien kann diese Unsicherheit reduzieren.^[67]

RCT weisen dabei einige markante Charakteristika auf. Im Rahmen von RCT wird oftmals von unterschiedlichen Studienarmen oder Gruppen gesprochen. Jede einzelne Behandlung oder Nichtbehandlung, die miteinander verglichen werden soll, bildet jeweils einen Arm oder eine Gruppe der Studie. Das heißt, wenn eine Behandlung mit der Standardbehandlung und mit einer Scheinbehandlung (Placebo) verglichen werden soll, gibt es drei Studienarme/Gruppen.^[70] In den meisten Fällen wird als Studiendesign der RCT allerdings das zweiarmlige Parallelgruppendesign gewählt.^[70]

Ein weiteres Charakteristikum einer RCT ist die Verblindung. Es gibt unterschiedliche Formen der Verblindung: die einfachblinde, doppelblinde, dreifachblinde oder offene Studienform. Dieser Vorgang beschreibt, ob die Behan-

delten und/oder der Prüfarzt sowie der Auswertende in Kenntnis darüber sind, welcher Patient zu welcher Gruppe zugeordnet ist. Die Verblindung dient dem Ausschluss von Verzerrungen der Ergebnisse durch die Kenntnis der Gruppenzugehörigkeit. Prinzipiell wird die höchst mögliche Form der Verblindung empfohlen.^[70] Die hohen Anforderungen an diese Studien begründen sich u.a. dadurch, dass die Effekte auf Endpunkte oftmals eher klein sind und daher ihr Nachweis von Störfaktoren beeinträchtigt werden kann.^[50, 51] Wenn sich hingegen durch Interventionen „dramatische Effekte“ zeigen, können die Studienanforderungen geringer sein.^[67, 71]

Als verlässlichste Interventionsstudie gilt die RCT.^[19, 50, 66-69] Sie ist der einzige Studientyp, der den Kausalitätsnachweis liefern kann. Geringere Evidenzgrade liefern den Nachweis nicht. Insofern wird eine RCT immer ein Grundbaustein einer Nutzenbewertung sein müssen.

Vielfach wird die RCT als Königsweg in der Bewertung der erwünschten und unerwünschten Effekte von medizinischen Interventionen bezeichnet. Bei Fehlen eines solchen Studiendesigns wird der Evidenzgrad einer Studie durch die HTA-Behörden kritisch bewertet.^[72]

Das kontrollierte Studiendesign mit gleichen Versorgungsbedingungen und sorgfältiger Verlaufsbeobachtung für alle Patienten soll quasi-experimentelle Bedingungen gewährleisten (Strukturgleichheit). Die Randomisierung soll ein Confounding verhindern. Dies beschreibt die Verzerrung einer scheinbaren beobachteten Beziehung zwischen zwei Faktoren auf Basis einer dritten Variablen („Confounder“), die mit beiden Faktoren unabhängig verknüpft ist. Dies gilt als entscheidender methodischer Vorteil von randomisierten gegenüber nicht-randomisierten Studien.

Vor dem Hintergrund der Besonderheiten der MP könnte gegebenenfalls auf RCT-Varianten zurückgegriffen werden, wie beispielsweise:

- Registry-based RCT^[73]
- Platform Trial^[74]
- Partially Randomized Patient Preference Trials^[75]
- Zelen's/Präferenz Designs^[76]
- Experten-basierte Designs^[76, 77]

Generell sollten Aussagen innerhalb der Nutzenbewertung hohe Standards erfüllen. Diese sollen den Einfluss von Störfaktoren dezimieren, wie es bei der randomisierten kontrollierten Studie als geeignetes Studiendesign zur Nutzenbewertung der Fall ist.^[50] Wobei auch die randomisierte kontrollierte Studie, genauso wie andere Studiendesigns, nicht frei von methodischen Schwierigkeiten ist (s. Tab. 2).^[78] Auf mögliche Probleme in der Umsetzung von RCT soll im Weiteren genauer eingegangen werden.

Tab. 2 Vor- und Nachteile von randomisierten kontrollierten Studien

Vorteile	Nachteile
<ul style="list-style-type: none"> ■ Aufwand zur Erhebung zahlreicher Merkmale zur Beschreibung der Ausgangssituation kann bei einer Randomisierung drastisch reduziert werden.^[63] ■ Aussagen aus randomisierten Studien sind weniger fehleranfällig als nicht randomisierte Studien.^[63] ■ Randomisierung führt mit hoher Wahrscheinlichkeit zu einer Vergleichbarkeit zwischen Patientengruppen hinsichtlich prognostischer Faktoren.^[79, 80] 	<ul style="list-style-type: none"> ■ Doppelte oder Dreifachverblindung ist oft nicht möglich, z.B. bei Implantaten oder chirurgischen Interventionen.^[81] ■ Bei hoher Drop-Out Rate kann ein zugrundeliegender systematischer Fehler (Selektionsbias) nicht mehr ausgeschlossen werden.^[81] ■ RCT können oft nur kurzfristige und relativ häufige unerwünschte Ereignisse dokumentieren.^[81, 82] ■ Mangelnde Generalisierbarkeit des Studienergebnisses durch häufiges Ausschließen von Patienten mit Begleiterkrankungen (Ein- und Ausschlusskriterien).^[81, 83] ■ Patientenrelevante Endpunkte werden oft unzureichend beantwortet.^[80]

3.5 Probleme bei der Umsetzung von klinischen Studien

Momentan liegen in Deutschland RCT für MP nicht immer vor. Kritiker argumentieren, dass niedrigere Evidenzstufen die Nutzenbewertung sowie die darauf basierenden Erstattungsentscheidungen erschweren.^[5, 38, 84] Die Evidenzgrundlage für die klinische Wirksamkeit (clinical effectiveness) der meisten Methoden mit MP hoher Klassen ist traditionell schwächer als die für Arzneimittel.^[85] Nach Ansicht einiger Kritiker beruht der fehlende Anreiz zur Durchführung einer RCT teilweise auf den regulatorischen Vorgaben. Für ein Produkt, das einem Vorgängerprodukt oder einem vergleichbaren Produkt eines Konkurrenten sehr ähnlich ist, kann für die Zulassung nach aktuellem Recht gegebenenfalls auf die Studienergebnisse der Wettbewerber zurückgegriffen werden. Dieser sogenannte Klassen-Effekt hat negative Auswirkungen auf die Bereitschaft eine neue RCT durchzuführen.^[17, 61] Der Klassen-Effekt ist mit einem Schlüsselkonzept der Bewertung vor dem Inverkehrbringen verbunden: der „substanziellen Äquivalenz“. Dieses Konzept wurde im Rahmen der „Pre-Market Notification“ (PMN) bzw. des „510(k)“-Verfahrens der FDA eingeführt. Es wurde als Möglichkeit zur Vereinfachung und Erleichterung der Marktzulassung in den US-amerikanischen Gesundheitsmarkt geschaffen. Derzeit ist PMN einer der wichtigsten Zulassungskanäle von MP.

Bei der Implementierung einer RCT ergeben sich verschiedene Herausforderungen, die auf den speziellen Merkmalen der MP, den Charakteristika von RCT oder den regulatorischen Vorgaben beruhen.^[17] Auch wenn die randomisierte kontrollierte Studie als Standard angesehen werden muss^[19, 50, 66–69], führen Kritiker einige Probleme an, die deren Umsetzung erschweren:

Rigidität von RCT: Methoden mit MP umfassen eine heterogene Gruppe von Produkten, die mehrheitlich physikalisch wirken und weitreichende Auswirkungen auf den Betroffenen und auf den Anwender der Produkte haben. Das Stu-

diendesign wird im Voraus festgelegt und ist im weiteren Verlauf wenig veränderbar. Dies führt dazu, dass am vorab definierten Protokoll festgehalten werden muss und die Ergebnisse somit zum Teil nur schwer auf die Realität übertragbar sind. Dieser Zustand wird als geringe externe Validität bezeichnet.^[17] Die Rigidität von RCT beschreibt die kaum vorhandene Möglichkeit im Rahmen einer Studie auf Veränderung zu reagieren oder neue Erkenntnisse bei der Durchführung einfließen zu lassen. Eine Anpassung des Studiendesigns ist schwer umsetzbar.

Schnelle Innovationszyklen: Im Zusammenhang mit der Rigidität und Klassen-Effekten steht die Vorstellung des schnellen Wandels von MP. MP unterliegen kurzen Innovationszyklen von im Mittel zwei Jahren.^[9] Die schnellen Innovationszyklen von MP beruhen vor allem auf der stetigen Weiterentwicklung bestehender Produkte. Die Durchführung eines RCT ist zeitintensiv und eher mit einer langen Studiendauer verbunden. Für die Durchführung von RCT birgt dies das Risiko, dass eine Studie noch durchgeführt wird, obwohl schon Neuentwicklungen oder Weiterentwicklungen der eingesetzten MP zur Verfügung stehen. Für ein Medizinproduktehersteller kann es somit finanziell belastend sein, ein RCT durchzuführen, wenn das zu testende MP nicht über die Dauer der Studie auf dem neuesten Stand der Entwicklung verweilt und möglicherweise zum Abschluss der Studie schon technisch überholt ist.^[17] Schrittinnovationen als geringfügige Verbesserungen eines bestehenden Produkts machen einen Großteil der Neuentwicklungen aus.^[31]

Geeignete Auswahl der Outcomes: Der klinische Erfolg kann weit in der Zukunft liegen. Deshalb behelfen sich RCT oftmals mit intermediären Outcomes (Surrogaten). Der kausale Zusammenhang zwischen Surrogaten und klinisch relevanten Endpunkten muss ausreichend dargelegt werden. Die Betrachtung der endgültigen klinischen Ergebnisse erfordert lange Follow-Ups, die sehr zeit- und kostenintensiv sein können.^[17]

Einwände gegenüber der Scheinbehandlung (Placebo): Ethische Einwände gegenüber Nicht-/Scheinbehandlungen entstehen bei MP, die einen bisher nicht behandelbaren und gleichzeitig schweren oder lebensbedrohlichen Krankheitszustand betreffen. Es wird als unethisch erachtet, wenn einem Teil der Studienteilnehmer durch die Randomisierung die Behandlung aberkannt wird. In diesem Fall kann eine vollumfängliche Randomisierung ethisch nicht vertretbar und die Durchführbarkeit einer RCT nicht möglich sein. Zudem werden viele MP, wie beispielsweise Gelenkimplantate, operativ eingesetzt. Hierdurch entstehen z.B. Narben, welche für eine Scheinbehandlung fingiert werden müssten. Der Patient müsste dem Risiko einer Operation ausgesetzt werden, ohne eine tatsächliche Behandlung zu erhalten.

Unterschiedliche Präferenzen und Verblindung: Sowohl die Präferenzen der Anwender als auch der Patienten können zu einer veränderten Wirkung der Behandlung führen. Die Präferenzen wirken sich beispielsweise auf die unvoreingenommene Einschätzung der Wirkung aus und könnten somit ein

verzerrtes Bild abgeben.^[17] Ist eine Studie offen angelegt, findet keine Verblindung statt.

Die Art der Verblindung ist einfach, zweifach als auch dreifach möglich. Die Möglichkeit der Verblindung kann eingeschränkt sein, da der Einsatz oder die Anwendung eines MP oftmals sowohl für den Anwendenden als auch den Behandelten offen ersichtlich ist.^[17, 61] Dieses Problem der doppelten Verblindung kann ebenso zu ethischen und zusätzlich zu praktischen Problempotenzialen führen.^[17, 61]

Es muss jedoch deutlich gemacht werden, dass die Verblindung kein konstituierendes Merkmal der RCT ist (gewinnt aber an Bedeutung bei starken Präferenzen). Eine offene Studie stellt kein Hindernis für die Durchführung von RCT zur Untersuchung von MP dar. Es wird zwar die bestmögliche Verblindung empfohlen^[70], aber kein bestimmtes Niveau der Verblindung zwingend gefordert.^[61] Im Allgemeinen ist die Möglichkeit der Verblindung abhängig vom Umfang bzw. dem Ausmaß der Veränderungen der NUB. Im Bereich der invasiven MP ist beim Vergleich mit der Standardintervention davon auszugehen, dass eine Doppelverblindung umsetzbar ist, da der Patient unter Narkose im Unklaren über das eingesetzte Produkt ist und der Auswerter der Studiendaten ebenfalls verblindet werden kann.^[86]

Lernkurveneffekte und Anwenderqualifikation: Die erfolgreiche Behandlung mit einem MP ist maßgeblich von den Fähigkeiten und Fertigkeiten des Anwenders abhängig.^[17, 61] Das Erfahrungslevel bzw. die Lernkurveneffekte des Benutzers (z.B. Chirurgen, Operationsteam, Krankenhaus) haben potenziell einen Einfluss auf die Behandlungsergebnisse (Outcomes). Durch die wiederholte Anwendung eines Produktes verfeinern die Anwender ihre Fertigkeiten und verbessern ihre Lernkurve.^[87] Die klinischen Ergebnisse hängen somit nicht ausschließlich vom MP selbst, sondern oft von der Ausbildung, Kompetenz und Erfahrung der Endverbraucher bzw. der Anwender ab.^[43, 88] Eine weitere Herausforderung bei der Beurteilung der Wirksamkeit liegt in der Komplexität der Therapie, die als Ganzes durch ein multifaktorielles Zusammenwirken verschiedener Fachdisziplinen (Chirurgie, Anästhesie, Pflege etc.) charakterisiert ist. Das Gelingen einer wirksamen Behandlung hängt damit nicht allein von der NUB oder dem einzelnen Produkt ab, sondern auch von der Qualität der beteiligten Leistungserbringer. Dies umfasst die prä-, peri- und postoperative Medizin.^[19] Auch die Auswertung einer RCT wird hierdurch beeinflusst, da die Wirkeffekte nicht eindeutig zugeordnet werden können.^[17, 61]

Auswirkungen bei unwirksamen Behandlungen: Ein eingesetztes MP, wie z.B. ein Gelenkimplantat, kann nicht ohne schwerwiegende Folgen wieder entfernt werden. Das Ausmaß der Auswirkungen unwirksamer Behandlungen mit einem MP stellt eine Herausforderung hinsichtlich der Rücknahme einer Behandlung dar. Somit kann es unerwünschte Folgen haben, wenn beispielsweise als Vergleichstherapie ein MP verwendet würde, das im weiteren Verlauf aufgrund von ungünstigen Effekten wieder entfernt werden sollte.^[17] Schließlich könn-

te die Endgültigkeit der Behandlungsmethoden (insbesondere bei implantierbaren Produkten) zu erheblichen versunkenen Kosten führen.^[43]

3.6 Evidenzstufen in der Nutzenbewertung

In der klinischen Forschung gibt es verschiedene Typen klinischer Studien, die entsprechend ihrer Beweiskraft in Evidenzklassen eingeteilt sind. Mit Hilfe von Evidenzklassen/Evidenzgraden wird die wissenschaftliche Aussagefähigkeit klinischer Studien beschrieben. Evidenz beschreibt die Offenkundigkeit bzw. Klarheit eines Sachverhalts.^[64] Eine hohe Evidenz zielt auf die eindeutige Dokumentation eines Sachverhalts ab. In der Verfahrensordnung des G-BA werden die Evidenzklassen Ia bis III für Untersuchungsmethoden unterschieden. Behandlungsmethoden werden zudem noch weiter in die Klassen IV und V unterteilt (s. Tab. 3).

Tab. 3 Übersicht der Evidenzstufen nach Untersuchungs- und Behandlungsmethoden (eigene Darstellung nach ^[89])

Evidenzstufen der Unterlagen zu Untersuchungsmethoden	Evidenzstufen der Unterlagen zu Behandlungsmethoden
Ia: Systematische Übersichtsarbeiten von Studien der Evidenzstufe Ib	Ia: Systematische Übersichtsarbeiten von Studien der Evidenzstufe Ib
Ib: Randomisierte kontrollierte Studien	Ib: Randomisierte klinische Studien
IIa: Systematische Übersichtsarbeiten von Studien zur diagnostischen Testgenauigkeit der Evidenzstufe II b	IIa: Systematische Übersichtsarbeiten von Studien der Evidenzstufe II b
IIb: Querschnitts- und Kohortenstudien, aus denen sich alle diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse, positiver und negativer prädiktiver Wert) berechnen lassen.	IIb: Prospektive vergleichende Kohortenstudien
III: Andere Studien, aus denen sich die diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse) berechnen lassen	III: Retrospektive vergleichende Studien
	IV: Fallserien und andere nicht vergleichende Studien
	V: Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte, u.ä.; nicht mit Studien belegte Meinungen anerkannter Expertinnen und Experten, Berichte von Expertenkomitees und Konsensuskonferenzen

Vereinfacht gilt: Je höher die Evidenzklasse einer Studie, desto breiter ist ihre wissenschaftliche Basis. Studien der Klasse Ia haben die höchste Evidenz, Studien der Klasse V die geringste. Systematische Reviews weisen den höchsten Evidenzgrad auf und umfassen Bewertungen aller Studien, die vorher festgelegten Einschlusskriterien genügen (Design der Studie, Studienpopulation, Interventionsart etc.) und durch deren vollständige Erfassung Verzerrungen (Bias) vermieden werden sollen.^[64]

Sowohl EU-Richtlinien zu MP als auch die FDA-Vorschrift erkennen Daten aus randomisierten und nicht-randomisierten, experimentellen und beobachtenden klinischen Studien als gültige wissenschaftliche Quellen der klinischen Wirksamkeit an.^[34, 90, 91] Dieser Nachweis vor dem Inverkehrbringen kann zur Beeinflussung der Bewertung der Erstattungsfähigkeit nach dem Inverkehrbringen von MP verwendet werden. Das NICE erkennt beispielsweise in seinem Pilotbericht den Wert der Berücksichtigung experimenteller als auch beobachtender Daten im HTA-Prozess des Instituts an.^[92] Die Notwendigkeit alle relevanten (randomisierten und nicht-randomisierten) Nachweise zu berücksichtigen und quantitativ zu synthetisieren, wirft eine Reihe von methodischen Fragen, sowohl für die Entscheidungsträger als auch die Analysten, auf. Vor allem stellt sich aber die Frage, wie man von einer fragmentierten, heterogenen und möglicherweise verzerrten Evidenzgrundlage für die Beurteilung der klinischen und wirtschaftlichen Wirksamkeit von MP, angesichts des Mangels an verfügbaren Daten, Gebrauch machen kann.^[17]

3.7 Adaptive Studiendesigns und alternative Studiendesigns

Die besonderen Eigenschaften von MP lassen vermuten, dass die Sammlung von klinischen Daten auf die Post-Launch-Phase verschoben wird, im Gegensatz zu Arzneimitteln, wo dies mehrheitlich in der Pre-Launch-Phase geschieht.^[43] Daher sind in der Medizintechnik klinische Daten nur in begrenztem Umfang verfügbar, insbesondere zum Zeitpunkt der Produkteinführung. Um den diskutierten Problemen und Besonderheiten der Durchführung einer RCT mit MP oder klinischen Studien mit sehr hohem Evidenzgrad zu begegnen, werden international verschiedene Ansätze beschrieben.

Adaptive Studiendesigns

Traditionelle klinische Studiendesigns sind nicht dafür geschaffen, die Komplexität heutiger Entscheidungsprozesse zu erfassen.^[2] Einschätzungen des Verhältnisses von Risiko/Schaden und Nutzen werden für viele Subgruppen von Patienten nicht mehr mit herkömmlichen RCT unterstützt werden können. Daher sind diese Studien zwangsläufig mit Unsicherheiten verbunden, die nur über einen längeren Zeitraum analysiert werden können. Diese Vorgehensweise mag für eine initiale Zulassung nicht praktikabel sein.^[2] RCT

werden nur in wenigen Subgruppen möglich sein. Dies gilt insbesondere für seltene Erkrankungen, für die diese Informationen über Risiko und Schadenspotenziale oft nur auf Basis von realen Daten in einer späteren Phase des Produktlebenszyklus erhoben werden können.^[2]

Adaptive Studiendesigns können helfen, die erforderlichen Daten für jene Patienten(-gruppen) zu generieren, für die ein bestimmtes Maß an Ungewissheit akzeptiert wurde (ohne zunächst eine größere Anzahl von anderen Patienten diesem Risiko auszusetzen).^[2] Ein adaptives Design für eine klinische Studie mit MP ist im Rahmen des FDA-Richtlinienentwurfs zu „Adaptive Designs for Medical Device Clinical Studies“ als klinische Studie definiert, die prospektiv geplante Änderungen basierend auf der Akkumulation von Studiendaten erlaubt, ohne dabei die Integrität und Gültigkeit der Studie zu verletzen.^[20] Adaptive Designs eröffnen damit die Möglichkeit, erste Studienergebnisse zu bestimmten Zeitpunkten während der Durchführung auszuwerten und das Studiendesign entsprechend anzupassen.^[2] Wenn sie richtig umgesetzt werden, können adaptive Designs demnach den Ressourcenbedarf reduzieren und/oder die Chance auf den Studienerfolg erhöhen.^[20] Adaptive Studien bieten die wichtige Chance, den Informationsbedarf der Aufsichts- und Regulierungsbehörden und der Kostenträger in einer Studie zu erfüllen^[93] und die Kontinuität der Phasen bis nach der Zulassung zu operationalisieren.^[94]

Um die Integrität und Validität der (adaptiven) Studie zu erhalten, sollten Änderungen zumeist prospektiv geplant und im Protokoll der klinischen Studie vor Beginn beschrieben werden. Allerdings können Modifikationen der Studie auch nach deren Beginn wissenschaftlich valide sein. Die FDA diskutiert für diesen Fall verschiedene Arten von adaptiven Änderungen am Studiendesign sowie deren Vorteile und Grenzen.^[20] Im einfachsten Fall eines zweistufigen adaptiven Designs wird die Studie abhängig vom Ergebnis der Zwischenauswertung mit Ablehnung oder Beibehaltung der Null-Hypothese beendet oder mit einem zweiten Studienteil fortgesetzt. Für die Planung dieses zweiten Teils können alle Informationen aus dem ersten Studienteil (oder auch Ergebnisse aus mittlerweile abgeschlossenen anderen Studien) verwendet werden. Die Möglichkeiten von Design-Änderungen gehen dabei weit über eine Modifikation des ursprünglich festgelegten Stichprobenumfanges hinaus.^[95] Zu den adaptiven Änderungen zählen unter anderem:

- Änderungen am Studiendesign
 - „Gruppensequentielle Designs“, mit und ohne Neubewertung der Stichprobengröße
 - Fallenlassen/Beendigung eines Behandlungsarms
- Änderungen an der Studiendurchführung
 - Adaption der Stichprobe/Bayesian Stichproben-Adaptation
 - Änderung des Randomisierungsverhältnisses
 - „Adaptive Enrichment“/Adaptive Erweiterung der Patienten(sub-)gruppe

- Adaptive Planung bezogen auf die Gesamtinformation und adaptive Entscheidung über Studienende/Stopp der Studie
- Anpassung des Produkts oder der Endpunkte
- „Nahtlose Studien“ (das Entwicklungs- und Evaluationsprotokoll kann eine Machbarkeitsstudie enthalten, die den nahtlosen Übergang in eine Zulassungsstudie mit einem vorab geplanten Ablauf erlaubt)
- Änderungen der statistischen Hypothesen oder der finalen Analysemethoden^[20]

Aufgrund ihrer außerordentlichen Flexibilität haben sich adaptive Designs in kurzer Zeit in der klinischen Forschung etabliert. Angesichts der weitreichenden Freiheiten hinsichtlich möglicher Design-Modifikationen sind aber Entscheidungshilfen notwendig, um deren Potenzial effektiv nutzen zu können.^[95] Mehrere Faktoren tragen zu der Entscheidung bei, ob ein adaptives Design gewählt wird. Die wichtigste Fragestellung ist dabei, ob ein adaptives Design möglich und vorteilhaft im Vergleich zu einem festen (nicht-adaptiven oder konventionellen) Design ist.^[20]

Wann sind Adaptive Designs angemessen?

Falls Studien schnell Teilnehmer einschreiben, kann unter Umständen nicht genügend Zeit sein, Änderungen am Studiendesign vorzunehmen. Zum Beispiel, wenn viele Teilnehmer schnell rekrutiert werden und diese das finale Follow-Up der Studie nahezu zeitgleich erreichen, ist es schwer bis unmöglich, den Stichprobenumfang anzupassen. In solchen Fällen kann der Hersteller/Auftraggeber eine Verlangsamung der Einschreibung anstreben, um Zeit zu haben, aus den anfallenden Daten zu lernen und vorgeplante Anpassungen vorzunehmen.

Adaptive Designs sind möglicherweise nicht geeignet für sehr komplexe Studien, die mehrere primäre Endpunkte oder mehrere sekundäre Endpunkte analysieren. Studien mit kürzeren Laufzeiten, aber längeren Rekrutierungszeiten, können sich für ein adaptives Design eignen. Studien, in denen die Zeit bis zur Bewertung des primären Endpunkts lang ist, können ebenso von einem adaptiven Design profitieren.

In einem festen (nicht-adaptiven) Design basiert die Fallzahlplanung in der Regel auf angenommenen Werten für einige Parameter. Eine grundlegende Frage ist, wie viel Vertrauen wird in die Wahl dieser Parameter gesetzt? Angenommen, eine Studie sei für einen optimistischen Wirkeffekt geplant, aber der beobachtete Behandlungseffekt beträgt nur 80% des erwarteten, ist aber immer noch klinisch relevant und signifikant. In einem festen Design, das auf den optimistischen Wirkeffekt ausgerichtet ist, könnte die Erfolgswahrscheinlichkeit in Bezug auf den Wirksamkeitsendpunkt kleiner als geplant und damit inakzeptabel sein. In diesem Fall würde das feste Studiendesign,

basierend auf einer zu optimistischen Einschätzung der Effektstärke, wahrscheinlich zu einer gescheiterten Studie für den Sponsor führen. Im Gegensatz dazu könnte ein adaptives Design mit einer geplanten Zwischenanalyse zur Anpassung der Stichprobengröße eine potenziell erfolglose Studie in eine erfolgreiche umwandeln. Ein adaptives Design kann vor diesen Unwägbarkeiten durch das Lernen aus anfallenden Daten während der Studie schützen.^[20]

Wann ist ein Adaptives Design vorteilhaft?

Wenn ein adaptives Design eine realistische Option für eine Studie darstellt, bleibt noch die Frage, ob ein adaptives Design im Gegensatz zu nicht-adaptiven (festen) Designs gewählt werden sollte. Es wird empfohlen, das optimale Studiendesign für die jeweilige Situation zu wählen, egal ob dies ein adaptives oder ein festes (nicht-adaptives) Design ist. Um festzustellen, ob ein adaptives Studiendesign verfolgt werden soll, kann es helfen, im Vorfeld eine Reihe von möglichen realistischen Szenarien des Studienablaufs durchzuspielen. Für jedes Szenario ist die Herausforderung zu beurteilen, wie wahrscheinlich jedes Szenario ist. Daneben müssen für dieses Design die Chance auf Erfolg, die durchschnittliche Größe der Studie und die Eigenschaften der Durchführung (Wahrscheinlichkeit eines Fehlers erster Art und die statistische Power) benannt werden. Für adaptive Designs erfolgt die Berechnung unter Verwendung von analytischen Techniken oder durch Computer-Simulationen.^[20]

Adaptive Pfade mit unterschiedlichen Evidenzgraden

Auf europäischer Ebene hat die EMA erkannt, dass die bestehenden Regulierungspfade nicht ausreichen, um sich mit der Herausforderung „Zugang versus Evidenzgrad“ zufriedenstellend zu befassen.^[96] Neue Modelle der sogenannten „adaptive pathways“ (AP) werden eingesetzt, um die Bedürfnisse der Patienten und die Vorsicht der Aufsichtsbehörden auszubalancieren. Ansätze der adaptiven Lizenzierung bzw. der „adaptiven Pfade“ (AP) basieren auf einer progressiven Steuerung und Reduktion der Unsicherheit bei gleichzeitigem Aufbau von Evidenz.

Zumindest während der ersten Genehmigungsstufe wird die adaptive Lizenzierung voraussichtlich eine Abwägung zwischen einem frühen Zugang für die Patienten und einem erhöhten Maß an akzeptabler Unsicherheit über Vorteile und Risiken beinhalten. Jedoch wird erwartet, dass sich die Unsicherheit mit zunehmendem Maß an Evidenz verringert.^[30] Die „adaptiven Pfade“ ermöglichen einen schnelleren Zugang von Patienten zu innovativen Therapien. Dabei ist jedoch zu beachten, dass in diesem Verfahren sowohl die Patienten, Leistungserbringer, Hersteller und Aufsichtsbehörden bereit sein müssen, höhere Risiken und Unsicherheiten in Bezug auf den Nachweis von Wirksamkeit und Sicherheit in Kauf zu nehmen.^[2, 30] Dabei erkennen sowohl die FDA als auch die EMA an, dass nicht für alle Entscheidungen und Bewertungen

RCT notwendig sind. Ab einem gewissen Grad an erreichter Evidenz würden beispielsweise Anwendungsbeobachtungen ausreichen.^[2, 30] Diese Flexibilität der Evidenz/-generierung wird unter anderem auch durch adaptive Studiendesigns erreicht. Dabei muss im Prozess der adaptiven Pfade ein bestimmtes Sicherheitsniveau eines Produktes initial gewährleistet sein. Im Zeitverlauf besteht die Gelegenheit, weitere Evidenz zu generieren und Risiken abzubauen bzw. einen höheren Nutzen zu belegen.

Ein weiteres Ziel besteht in der Reduktion der Unsicherheit in Bezug auf Endpunkte und das Studiendesign. Eine vorläufige Nutzenanerkennung könnte dabei auf einem (noch nicht vollumfänglich validierten) Surrogatendpunkt basieren. Erst für die endgültige Nutzenbewertung bzw. finale Entscheidung müssten klinische Endpunkte belegt werden. In diesem Ansatz wäre es zudem möglich, zunächst auf Basis einarmiger Studien, Beobachtungsstudien oder Registerdaten zu entscheiden. RCT wären erst für die endgültige Entscheidung nötig. In ähnlicher Weise kann hier dem Problem der „real-world“ Anwendungen begegnet werden. Durch ein adaptives und flexibles Studiendesign würde eine erste Entscheidung auf Basis explorativer Studien vorgenommen, während pragmatische RCT für die „endgültige“ Nutzenbewertung gebraucht würden.^[30]

Die adaptiven Ansätze sind besonders gut geeignet, um sich mit Problemen einer komplexen und fragmentierten Evidenzgrundlage, wie sie in der Regel für Methoden mit MP zur Verfügung stehen, zu befassen. Diese Methoden ermöglichen naturgemäß die Synthese von mehreren Quellen der Evidenz mit heterogenen Designs (z.B. Evidenz aus RCT und Beobachtungsdaten). Studien, die sich auf niedrig-gewichtete empirische Evidenz konzentrieren, bilden die Evidenzbasis.^[97] Darauf folgende Ansätze kombinieren randomisierte und nicht-randomisierte Studien.^[98] Aktuell werden Erhebungen mit einer Kombination verschiedener Methoden vorgeschlagen und eingesetzt, um gleichzeitig interne und externe Quellen für einen Bias zu betrachten. Ein Beispiel findet sich bei einer Evaluation des NICE zur Schwangerenvorsorge.^[99]

Hierarchische Modelle werden als leistungsfähige Plattform empfohlen, um den klinischen Effekt bei inkrementellen Innovationen zu untersuchen.^[100] So zum Beispiel Bayes'sche hierarchische Modelle für geclusterte Daten oder die Durchführung der Netzwerk-Metaanalyse^[101], um den Umgang mit Lernkurveneffekten zu ermöglichen.^[102, 103] Hierarchische Modelle können die Korrelation zwischen dem klinischen Effekt einer Methode und dem Effekt, der mit einer früheren Version des Produkts assoziiert wird, berücksichtigen. Mit anderen Worten, sie erkennen die hierarchische Struktur, die den Produktentwicklungsprozessen von Methoden mit MP zugrunde liegt. Dabei treten manchmal Änderungen an Produkten und der Nutzung von Protokollen während der Durchführung von klinischen Studien zur Darstellung der Sicherheit und Leistungsfähigkeit auf.^[17, 104]

Die Umsetzung flexibler und adaptiver Ansätze benötigt Zeit und Investitionen, insbesondere im Hinblick auf die Generierung personeller Kapazitäten. Allerdings gibt es erste Versuche, die Unsicherheit, die mit der Evidenzgrundlage von MP verbunden ist, zu charakterisieren.^[17, 105] Der Ansatz, RCT und Beobachtungsstudien als Ergänzung in einem Studiendesign zu berücksichtigen, bietet einen fundierten und umfassenden Rahmen für die klinische Bewertung von MP. Dies kann als ein erster Schritt gesehen werden, die klinische Wirksamkeit der MP unter nicht-klinischen Bedingungen zu bewerten.^[106] Der adaptive und flexible Ansatz müsste in die Methodenpapiere der Regulierungsbehörden für die Bewertung vor und nach dem Inverkehrbringen aufgenommen werden. Damit könnten Anreize für die Generierung einer soliden klinischen Evidenzgrundlage für NUB mit MP geschaffen werden.^[17] Ähnlich dem derzeitigen Status Quo wäre im Rahmen der primären Nutzenbewertung eine Nutzenanerkennung für einen begrenzten Zeitraum mit einem anschließenden umfassenden Beleg der Wirksamkeit und Verträglichkeit denkbar. Der Nachweis des Langzeitnutzens könnte über Register bzw. eine „long-term surveillance“ erfolgen.

Alternative Studiendesigns

Alternative Studiendesigns wurden in der Literatur diskutiert, um die oben genannten Herausforderungen der Evidenzgenerierung in Bezug auf MP zu bewältigen.^[17] Nicht-standardisierte flexible Strategien – unter anderem der Einsatz von Scheinbehandlungen, welche die Offenlegung der Studienhypothese verhindern – können verwendet werden, um die Wirkung der Verblindung als mögliche Quelle für einen Bias zu minimieren.^[107] Tracker-Studien bieten das flexibelste Design für eine RCT und ermöglichen die Verwendung von stufenweiser Entwicklung von MP, sobald sie verfügbar werden.^[108] Daneben werden in der Literatur verschiedene Ansätze diskutiert, um Lernkurveneffekte zu neutralisieren (mittels festgelegtem Maß an Erfahrung oder Sicherstellung, dass alle Interventionen durch den gleichen Arzt erbracht werden) oder zu erklären (z.B. mit hierarchischen Modellen).^[102]

Die komplementäre Nutzung von RCT und prospektiven Beobachtungsstudien in umfassenden Kohortenstudien wurde als mögliche Lösung vorgeschlagen, um Fragen im Zusammenhang mit individuellen Behandlungspräferenzen und der Dokumentation von langfristigen klinischen Ergebnissen zu klären.^[109] Discrete-Choice Experimente (DCE) können zur Identifizierung und Bewertung von Eigenschaften, die Einfluss auf die Behandlungspräferenzen von Patienten und Leistungserbringern haben, genutzt werden (z.B. Benutzerfreundlichkeit des Produkts, Nebenwirkungen, Einfluss auf die Lebensqualität etc.).^[110, 111] Aktive Komparatoren, wie die „Regelversorgung“ sind eine Alternative, um ethische Bedenken bei der Verwendung von Scheinverfahren zu überwinden.^[17, 112] Um Fragestellungen in der (Nutzen-)Bewertung von MP zu beantworten, stehen verschiedene Studienansätze zur Verfügung. Welcher

Studientyp zum Einsatz kommt, hängt von verschiedenen Faktoren ab. Dazu zählen:

- die Fragestellung
- finanzielle und zeitliche Ressourcen
- zur Verfügung stehende Daten
- die betrachtete Indikation u.a.

Eine besondere Herausforderung bei der Messung klinischer oder nicht-klinischer Effekte ist der Vergleich verschiedener Patienten bzw. Patientengruppen unter gleichen Ausgangsbedingungen. Zielsetzung ist die Minimierung der Auswirkungen von Störgrößen bzw. beeinflussender Faktoren. Dies ist besonders dann wichtig, wenn vergleichende Interventionsstudien Aussagen über die Effekte einer medizinischen Maßnahme relativ zur Vergleichstherapie ermöglichen sollen.^[50]

Grundsätzlich lassen sich die Studienansätze wie folgt einteilen: Die Medizinische Forschung unterteilt sich in Primär- und Sekundärforschung. Während die Grundlagenforschung sowie Klinische und Epidemiologische Forschung unter die Primärforschung fallen, lassen sich die Metaanalyse und die Reviews in die Sekundärforschung einordnen (s. Abb. 4).^[113]

Registrierungsdaten können als Quelle für realistische Informationen über die „Effektivität“ (effectiveness) von Behandlungen dienen. Die Voraussetzungen dafür sind Register, die komplette, leicht verfügbare Informationen liefern und regelmäßig aktualisiert werden. Die Pilotstudien der „EUnetHTA Joint Action 2“ zum Rapid Relative Effectiveness Assessment greifen auf reale Daten (inkl. Register) zurück.^[114, 115] Diese Pilotstudien sowie die Arbeit im Rahmen des „GetReal Projekts“ der Innovative Medicines Initiative haben vielversprechende Ergebnisse gebracht. Im Hinblick auf die Datensammlung und Analyse-Methoden von realen Registerdaten besteht ein Mangel an Harmonisierung und Standardisierung. Zudem fehlen methodische Richtlinien. Daneben erschweren auch unterschiedliche Anforderungen an die Evidenz und Unterschiede in der Versorgungspraxis (Standard of care) in den verschiedenen Ländern und Rechtssystemen eine EU-weite Sammlung von Registerdaten in standardisierter Form.^[116]

Im Bereich der Pharmakovigilanz erkennen die geltenden Rechtsvorschriften die Möglichkeit an, mit Post-Autorisierungs-Studien zu Sicherheit und Wirksamkeit fortzufahren, in denen die Verwendung von realen Daten besonders relevant ist. Auf der anderen Seite zeigt auch die EU-Medizinprodukt-Richtlinie, welche noch reformiert wird, deutlich in Richtung einer gesteigerten Bedeutung des Follow-Ups von MP über ihren gesamten Lebenszyklus.^[117, 118]

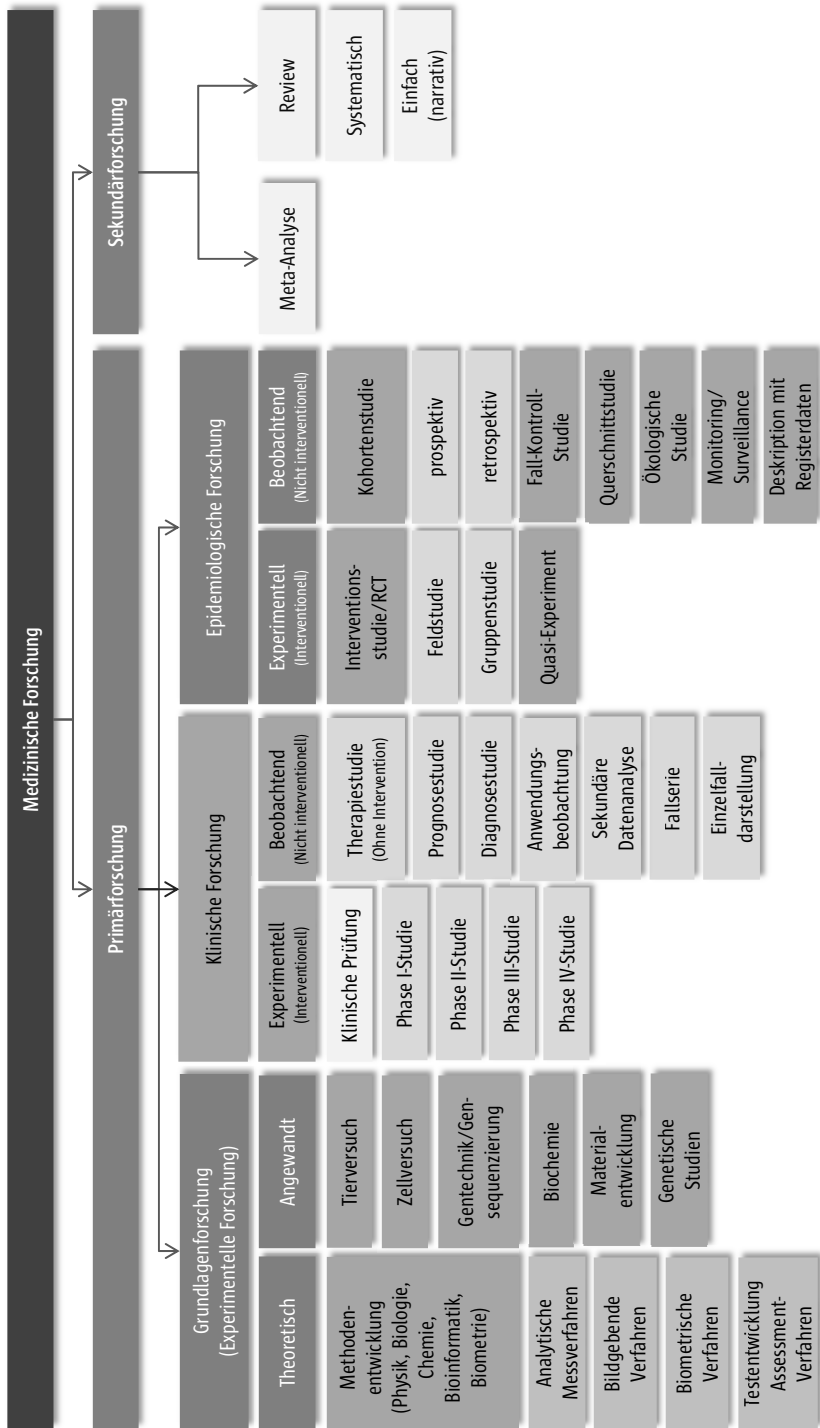


Abb. 4 Studientypen in der medizinischen Forschung (eigene Darstellung nach [113])