

Volker E. Amelung, Birgit Dembski, Franz-Helmut Gerhards, Daniel Grandt, Rainer Hess, Jutta Wendel-Schrief, Chantal Friebertshäuser, Sabine Barz, Ann-Kathrin Klähn, Martin Göhl

Wissenschaftliche Studien als Katalysator für innovative Versorgungsprojekte

Nominierte und Preisträger des MSD Gesundheitspreises 2011 – 2021 teilen ihre Erfahrungen

Der MSD Gesundheitspreis zeichnet seit 2011 innovative Ansätze aus, die dazu beitragen, die Gesundheitsversorgung in Deutschland zu verbessern. Rund 100 Projekte wurden in den vergangenen zehn Jahren von einer unabhängigen Jury nominiert. Aus dem Pool der zahlreichen Einreichungen ragten diese Versorgungsmodelle heraus, weil sie Herausforderungen und Alltagsprobleme auf neue, kreative Weise anpacken und teilweise immense Potenziale für Versorgungsverbesserungen aufzeigen. Doch das Augenmerk der Juror:innen lag stets auch noch auf einer weiteren Komponente: der wissenschaftlichen Evaluation.

Warum sind wissenschaftliche Studien so wichtig?

Das Out-of-the-Box-Denken und die Hands-on-Mentalität der Initiator:innen sind wertvolle Ressourcen im schwerfälligen deutschen Gesundheitssystem. Es hat sich in der Vergangenheit immer wieder gezeigt, dass der Enthusiasmus, die Überzeugungskraft und der Tatendrang von Projektteams und einzelnen Persönlichkeiten beeindruckende Best-Practice-Beispiele hervorgebracht haben. Begeisterung allein reicht aber nicht aus, um sich langfristig einen Platz in der Regelversorgung zu sichern. Die Versicherten als Finanzierende des Systems haben einen Anspruch darauf, dass die verfügbaren finanziellen und personellen Mittel effektiv und effizient eingesetzt werden [1]. Das Fünfte Sozialgesetzbuch (SGB V) übersetzt diesen Anspruch in das Gebot der Wirtschaftlichkeit. Wirtschaftlichkeit darf hier allerdings nicht auf den Kostenfaktor reduziert werden – vielmehr geht es um einen sinnvollen Ressourcen-Einsatz, der die Aspekte Zweck, Notwendigkeit und Kosten in ein angemessenes Verhältnis bringt. In § 12 Abs. 1 SGB V heißt es: „Die Leistungen müssen ausreichend, zweckmäßig und wirtschaftlich sein; sie dürfen das Maß des Notwendigen nicht überschreiten. Leistungen, die nicht notwendig oder unwirtschaftlich sind, können Versicherte nicht beanspruchen, dürfen die Leistungserbringer nicht bewirken und die Krankenkassen nicht bewilligen.“

Demnach kann der Nachweis der Wirtschaftlichkeit als Eintrittskarte in die Regelversorgung begriffen werden. Für innovative Versorgungsmodelle heißt das: Es muss sichergestellt sein, dass Zeit, Geld

Zusammenfassung

Im Rahmen des MSD Gesundheitspreises wurden in den vergangenen Jahren rund 100 innovative Versorgungsprojekte nominiert. Ein wichtiges Augenmerk der Jury lag dabei stets auf der Frage nach einer wissenschaftlichen Begleitevaluation. Denn nur wenn ein Projekt nachweisen kann, dass es die Versorgung messbar verbessert, hat es Chancen, dauerhaft in die Regelversorgung übernommen zu werden. Im vorliegenden Beitrag teilen nominierte Projekte aus zehn Jahren MSD Gesundheitspreis ihre Erfahrungen rund um die Planung und Durchführung von wissenschaftlichen Studien. Die zehn Empfehlungen, die daraus abgeleitet werden können, bieten wertvolle Hinweise für alle, die bei der Planung eines neuen Versorgungsmodells gerade am Anfang stehen.

Schlüsselwörter

Evaluation, wissenschaftliche Studien, Evaluationsdesign, Methoden der Evaluation, innovative Versorgungsprojekte, MSD Gesundheitspreis, Versorgungsforschung, Verhandlungen mit Kostenträgern, Empfehlungen für Begleitevaluationen

und Energie an den richtigen Stellen für die richtigen Maßnahmen aufgewendet werden. Der Beweis kann nur geführt werden, wenn sich die gute Idee in messbare Größen herunterbrechen lässt. Und das ist Aufgabe, Sinn und Zweck einer wissenschaftlichen Begleitevaluation [2]. Zu ihren Erfahrungen im Bereich der Evaluation wurden im Juli 2021 Projekte befragt, die in den Jahren 2011 bis 2021 zu den Nominierten oder Preisträgern des MSD Gesundheitspreises zählten.¹

Drei Viertel der Befragten gaben an, dass die Evaluation eine sehr große oder eher große Rolle bei der Entwicklung des Versorgungsprojekts gespielt habe. Hinsichtlich des Nutzens der Evaluation fielen die Zahlen sogar noch deutlicher aus: 88 Prozent beurteilten den Nutzen der Evaluation als eher groß oder sehr groß (Abb. 1).

Die Bedeutung der Studien liegt in den Augen der Teilnehmenden auf mehreren Ebenen: Zum einen habe die Durchführung der Evaluation vielfach zu den Bedingungen für eine finanzielle Förderung gehört. Zum anderen liefere sie wichtige Argumente bei Verhandlungen mit Kostenträgern. Zwei Drittel der Befragten waren der Auffassung, dass die Evaluation bei Vertragsverhandlungen mit Partnern, z. B. Kranken-

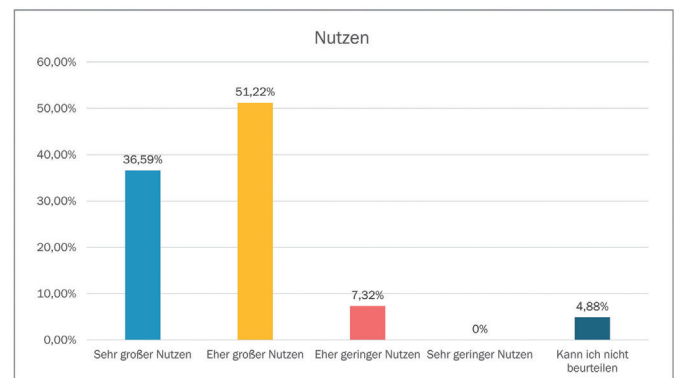


Abbildung 1: Wie beurteilen Sie den Nutzen der Evaluation? (n = 41). Quelle: inav GmbH

¹ Es wurden insgesamt 95 Personen eingeladen, an der Umfrage teilzunehmen, 41 von ihnen füllten den Online-Fragebogen aus. Das entspricht einer Rücklaufquote von 43 Prozent.

kassen, eine sehr große oder eher große Rolle gespielt habe (Abb. 2). Weiterhin könne die Evaluation als Gütesiegel für das Versorgungsmodell dienen. Sie bilde eine wertvolle Grundlage für Folgeprojekte und habe Modellcharakter für ähnliche Projekte in anderen Regionen. Einige Teilnehmende berichteten darüber hinaus, dass die Evaluation es ermöglicht habe, Publikationen in renommierten Journals zu platzieren.

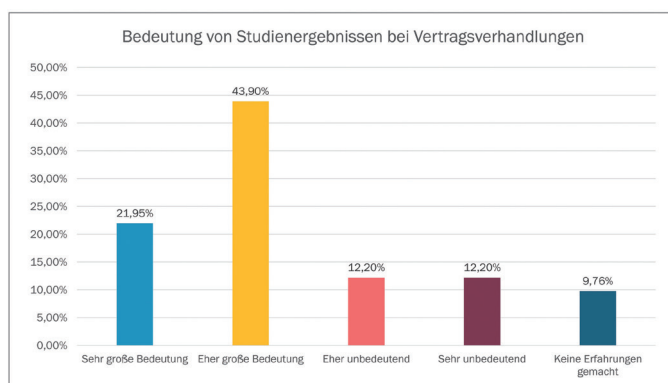


Abbildung 2: Welche Bedeutung hatten Ihre Studienergebnisse bei Vertragsverhandlungen mit Partnern, z. B. Krankenkassen? (n = 41). Quelle: inav GmbH

Welcher Studientyp eignet sich für welche Fragestellung?

Bei wissenschaftlichen Evaluationen im Gesundheitswesen denken viele Menschen zunächst an klinische Studien im Bereich der Medikamentenzulassung. Zahlen, Daten, Fakten rund um Wirkungen und Nebenwirkungen von Arzneimitteln zu erheben – das erscheint vergleichsweise einfach. Wirksamkeitsnachweise von therapeutischen Interventionen in klinischer Forschung erfolgen oftmals unter Idealbedingungen (efficacy). In der Versorgungsforschung geht es dagegen um die Untersuchung der Wirksamkeit unter realen Versorgungsbedingungen mit allen Einflüssen und Wechselwirkungen (effectiveness) [3]. Hier steht die Alltagswirksamkeit der Maßnahmen stärker im Vordergrund.

Was bedeutet es also für die Evaluation, wenn es zum Beispiel um Patientenzufriedenheit, Gesundheitskompetenz, Patienten-Empowerment oder Prozessverbesserungen geht? Oder im Bereich Prävention: Wie können wir messen und damit belegen, dass durch eine Intervention ein Krankheitsereignis nicht eintritt?

Antworten darauf liefert die Versorgungsforschung. Während in der klinischen Forschung vorzugsweise der Typ der randomisierten kontrollierten Studie (engl. Randomized Controlled Trial – RCT) zur Anwendung kommt, hat sich der methodische Werkzeugkasten der Versorgungsforschung in den vergangenen Jahrzehnten erheblich gefüllt und ausdifferenziert [4]. Neben Varianten des RCT wie Stepped Wedge Design oder Cross-Over Design können Interventionsstudien mit quasi-experimentellen Designs (mit oder ohne Kontrollgruppe) die neuen Interventionen prospektiv evaluativ begleiten. Unterstützend können zudem Beobachtungsstudien wie Kohortenstudien, Querschnittsuntersuchungen, Fall-Kontroll-Studien oder Fallstudien zum Einsatz kommen. Jede dieser Methoden hat Vor- und Nachteile (vgl. Tab. 1, Tabelle basierend auf [5]). Die evaluative Kunst besteht darin, zu beurteilen, welches Studiendesign für welche Fragestellung besonders gut geeignet ist.

Dass unter den Teilnehmenden der Umfrage die Evaluationsdesigns der randomisierten kontrollierten Studie (RCT) (44 Prozent) und der nicht-randomisierten kontrollierten Studie (NRCT) (34 Prozent) favorisiert werden (Mehrfachauswahl möglich), liegt sicherlich auch am hohen Bekanntheitsgrad und Ansehen dieser Studientypen. So wird die RCT

weiterhin als Goldstandard in der Wirksamkeitsforschung gesehen [6]. Vor dem Hintergrund, dass wir uns bei Versorgungsprojekten vielfach im Bereich komplexer Interventionen bewegen, muss allerdings bedacht werden, dass der ausschließliche Einsatz von RCTs und NRCTs häufig nicht ausreichend ist.

Komplexe Interventionen bestehen aus mehreren Einzelkomponenten, die sich wechselseitig bedingen und mit den Kontexten interagieren, in denen sie implementiert wurden. Es ist daher schwierig, den Einfluss der Einzelkomponenten am Gesamteffekt und ihre Interaktionen mit dem Setting auszumachen [7]. Außerdem sollte nicht ausschließlich die Wirksamkeit der Intervention im Fokus stehen. Im Rahmen von Prozessevaluationen gilt es, auch die Umsetzungsqualität zu überprüfen [8]. Es bietet sich daher in den meisten Fällen ein Methoden-Mix aus quantitativen und qualitativen Ansätzen an [9]. Hier ist es Aufgabe des evaluierenden Partners, die Projektverantwortlichen hinreichend zu beraten und die Vor- und Nachteile der unterschiedlichen Evaluationsmethoden abzuwägen.

Die Ergebnisse der Befragung zeigen, dass die in den Projekten definierten Endpunkte häufig „weiche“ Outcome-Parameter wie Akzeptanz/Zufriedenheit (68 Prozent), Lebensqualität (54 Prozent) und Gesundheitskompetenz (39 Prozent) oder prozessorientierte Komponenten wie Zugang zu Versorgungsleistungen (29 Prozent) und Qualität der Leistungserbringung (29 Prozent) umfassten (Mehrfachauswahl möglich). Gerade bei der Überprüfung solcher Endpunkte bieten sich Mixed-Methods-Ansätze an, also Evaluationsdesigns, die qualitative Erhebungsmethoden einbeziehen. Die Erfahrungen aus den Projekten bestätigen die Aussagekraft qualitativer Ansätze: So gaben 71 Prozent der Befragten an, sehr gute oder eher gute Erfahrungen mit qualitativen Methoden gemacht zu haben (Abb. 3).

Gute Planung ist die halbe Miete

Das Ziel muss es stets sein, mit der jeweiligen Methodik und den eingesetzten Instrumenten, wie Fragebögen, qualitativen Interviews, medizinisch gemessenen Werten etc., zu belastbaren Daten zu gelangen, die Aussagen über die Wirksamkeit der Intervention zulassen. Die Voraussetzung dafür besteht zum einen in einer umfassenden Methodenkompetenz. Zum anderen ist aber auch ein tiefer Einblick in das Versorgungsetting notwendig, in dem die Intervention angesiedelt ist. Nur wenn die zu messenden Endpunkte sinnvoll gewählt und die Erhebungen so aufgesetzt sind, dass sie sich in der Versorgungsrealität abbilden lassen, ist eine Validität der Ergebnisse gewährleistet.

Deshalb gilt es bei der Konzeption innovativer Versorgungsmodelle, die Evaluation von Beginn an mitzudenken und den Evaluator bereits in die Entwicklungsphase einzubeziehen. Je besser auf der einen Seite das

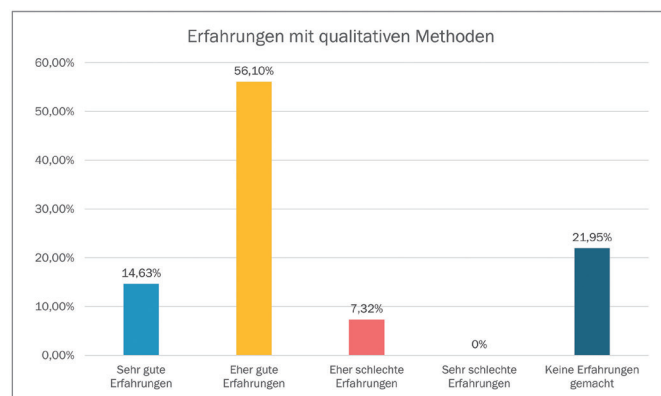


Abbildung 3: Welche Erfahrungen haben Sie mit qualitativen Methoden gemacht? (n = 41). Quelle: inav GmbH

Verständnis der Projektverantwortlichen für die Erfordernisse der Evaluation ist, umso hochwertiger ist erfahrungsgemäß am Ende die Datenlage. Auf der anderen Seite trägt es zu einem reibungsloseren Ablauf der Erhebung bei, wenn die Evaluatoren ein gutes Verständnis des Versorgungsalltags im Projekt haben. Dies gilt besonders vor dem Hintergrund, dass einige Befragte konstatierten, die „Denke“ der Wissenschaft habe sich teilweise nur schwer mit der Versorgungsrealität in Einklang bringen lassen.

In zahlreichen Förderprogrammen, beispielsweise im Innovationsfonds, ist es zudem nicht oder nur mit großem bürokratischem Aufwand möglich, das Evaluationsdesign oder einzelne Teile davon während der Förderlaufzeit zu modifizieren. Umso wichtiger ist es daher, sich frühzeitig mit dem Evaluator auszutauschen und von vornherein sorgfältig zu planen.

Eine befragte Person bringt diesen Zusammenhang treffend auf den Punkt: „Die Evaluation zwingt bei der Planung von Versorgungsprojekten, den Output im Auge zu behalten, und verhindert einen reinen Selbstzweck.“

Typische Hürden bei der Durchführung von Evaluationen

Zu den wichtigsten Lernerfahrungen aus den Projekten gehört, dass die Evaluationsziele häufig schwerer zu erreichen waren als zunächst angenommen, beispielsweise in Bezug auf die Rekrutierung der Teilnehmenden und das Erreichen der Fallzahlen sowie auf den Nachweis der Effekterwartung. 59 Prozent der Befragten gaben an, dass es sehr schwierig oder eher schwierig gewesen sei, die Studienziele zu erreichen (Abb. 4). Mehrere Personen berichteten, dass die Rekrutierung von Patientinnen und Patienten eine sehr anspruchsvolle Aufgabe darstelle, an der man permanent „dranbleiben“ müsse. Darüber hinaus habe die Pandemie-Situation es zusätzlich erschwert, auf die geplanten Fallzahlen zu kommen. Zudem sei vielfach Überzeugungsarbeit nötig, um Arztpraxen oder Zentren für die Beteiligung an Studien zu gewinnen, auch weil die vorgesehene Kompensation vonseiten des Förderers teilweise als zu niedrig angesehen werde.

Daraus lässt sich ableiten, dass es sinnvoll ist, die anvisierte Stichprobe nicht größer zu planen, als dies im Rahmen der statistischen Aussagekraft nötig ist. Zu hohe Planzahlen, die am Ende nicht erreicht werden, führen einerseits zu Demotivation und Frustration bei den Projektpartnern. Andererseits können sie den Projektaufwand erhöhen, weil Anträge auf eine Verlängerung der Förderzeit gestellt oder neue Geldgeber gefunden werden müssen, um die Studie sauber abschließen zu können.

Fazit: Dreieck aus Zeit – Aufwand – Kosten in Einklang bringen

Und was sind die Erfolgsfaktoren für die Planung und Durchführung von Evaluationen? Kurz gesagt: Wem es gelingt, die Faktoren Zeit – Aufwand – Kosten in eine harmonische Relation zueinander zu bringen, ist auf einem guten Weg.

Im Hinblick auf die Studiendauer lagen die Angaben der Umfrage-Teilnehmenden zwischen drei Monaten und fünf Jahren, wobei der Durchschnittszeitraum bei 2,5 Jahren liegt. Auch hier gilt es, sorgsam zu planen. Insbesondere sollten Pufferzeiten bedacht werden für den Fall, dass unvorhergesehene Ereignisse den Projektverlauf verzögern, sei es durch Schwierigkeiten in der Rekrutierung, durch Personalwechsel bei den Projektverantwortlichen, Probleme mit IT-Systemen oder andere Hürden.

Darüber hinaus darf der Aufwand aufseiten der Leistungserbringer nicht unterschätzt werden, wenn es darum geht, Patientinnen und Patienten aufzuklären, einzuschreiben, zu schulen und die Versorgungsdaten zu erheben. Eine leistungsstarke IT, die auf die Bedürfnisse derjenigen zugeschnitten ist, die die Daten erheben, dokumentieren und übermitteln, ist deshalb essenziell, um die Zeitaufwände im Rahmen und die

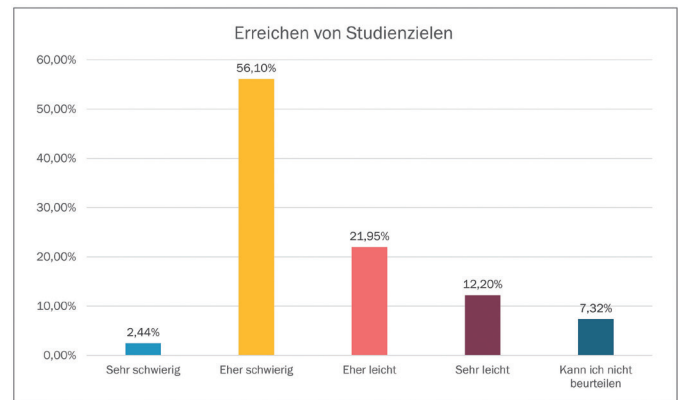


Abbildung 4: Wie schwer war es, Studienziele zu erreichen, beispielsweise in Bezug auf die Rekrutierung, die Erreichung der Fallzahlen, den Nachweis der Effekterwartung? (n = 41). Quelle: inov GmbH

Motivation hochzuhalten. Wenn Ärztinnen und Ärzte oder medizinische Fachangestellte minutenlang vor dem rotierenden Cursor sitzen, weil die Software wieder einmal abgestürzt ist, ist die Frustrationsschwelle schnell erreicht.

Nicht zuletzt ist der Kostenfaktor für die Evaluation zu berücksichtigen. Rund drei Viertel der Befragten empfanden den finanziellen Aufwand für die Studie als eher hoch oder sehr hoch. Gleichzeitig beurteilten aber auch 64 Prozent der Teilnehmenden das Preis-Leistungs-Verhältnis für die Evaluation als eher gut oder sehr gut. Mit anderen Worten: Wer schon einmal an einem Versorgungsprojekt beteiligt war, konnte erleben, wie hoch der Aufwand für eine solide Studie ist. Die Kosten dafür sollten daher schon bei der Beantragung der Fördermittel eingeplant werden, denn – wie eingangs beschrieben – gibt es zu einer guten Evaluation im Gesundheitssystem nun einmal keine Alternative. <<

Zehn Empfehlungen für erfolgreiche wissenschaftliche Begleitevaluationen

Was würden Sie Initiatoren von neuen Versorgungsprojekten raten, die jetzt gerade anfangen, ihre Idee in die Tat umzusetzen? Auf diese Fragen gaben die Teilnehmenden wertvolle Antworten. Die wichtigsten zehn Empfehlungen lauten:

1. Denken Sie die Evaluation bereits in der Konzeption und Planung des Versorgungsprojektes mit.
2. Wählen Sie einen unabhängigen Evaluationspartner mit viel Erfahrung und binden Sie diesen frühzeitig in die Projektplanung ein.
3. Stellen Sie gemeinsam mit dem Evaluator sicher, dass die Ziele des Projektes und die Endpunkte der Evaluation in einem sinnvollen Verhältnis zueinander stehen.
4. Achten Sie darauf, dass das Evaluationsdesign so gestaltet ist, dass es der Intervention gerecht wird (Dauer, Versorgungssetting, Komplexität etc.).
5. Planen Sie die Fallzahlen nicht größer als nötig, unterschätzen Sie den Aufwand für die Rekrutierung von Studienteilnehmenden nicht.
6. Planen Sie bei der Beantragung von Fördermitteln ausreichend Zeit und Geld für die Durchführung der Studie ein.
7. Stellen Sie sicher, dass die im Projekt eingesetzte IT leistungstark ist und den Anforderungen des Projektes entspricht.
8. Sorgen Sie dafür, dass Sie Zwischenergebnisse erhalten: So können Sie die Motivation der Projektpartner heben und ermöglichen ggf. Anpassungen im Versorgungsprojekt.
9. Machen Sie Betroffene zu Beteiligten: Es lohnt sich, die Zielgruppe(n) der Intervention in die Planungsphase einzubeziehen, ihre Bedürfnisse zu hören und ihre Anregungen aufzugreifen.
10. Achten Sie bei der Darstellung der Evaluationsergebnisse auf Transparenz und Nachvollziehbarkeit. Dies erhöht die Akzeptanz bei allen Beteiligten und hilft in weiteren Verhandlungen mit Kostenträgern und anderen Vertragspartnern.

	Übergeordnete Synthese	Interventionsstudien			Beobachtungsstudien			
	Systematischer Review und Metaanalyse [10]	Randomisierte kontrollierte Studie (Randomized Controlled Trial – RCT) [11]	Nicht-randomisierte kontrollierte Studie (Non-RCT) [12]	Quasi-experimentelles Design (nicht-kontrollierte Studie: ohne Kontrollgruppe) [13]	Kohortenstudie [14]	Fall-Kontroll-Studie [15]	Querschnittsdesign [16]	Fallstudie [17]
Ziel	Zusammenfassung und quantitative Analyse der Ergebnisse von Einzelstudien	Kausalität belegen, „Proof of Concept“	Real-World Effekte einer Intervention messen	Kausalität zwischen einer Intervention und einem/ mehreren Outcomes in einer bestimmten Stichprobe belegen	Beobachten der Exponierten und Nicht-Exponierten über einen bestimmten Zeitraum hinsichtlich des Auftretens der Outcomes	Untersuchen, ob eine Exposition mit einem Outcome (Erkrankung) verbunden ist	Beschreiben der Verteilung von Risikofaktoren (Expositionen) oder Krankheiten (Outcomes) in einer Bevölkerung zu einem bestimmten Zeitpunkt	Detailliertes Beschreiben einer indiv. Situation (Fall), ggf. Analyse der spezifischen Probleme
Einteilung und Studienverlauf		- Interventionsstudie - Längsschnitt und Querschnitt - Prospektiv	- Interventionsstudie - Längsschnitt und Querschnitt - Prospektiv	- Prä-Post-Interventionsstudie - Längsschnitt - Prospektiv	- Beobachtungsstudie - Längsschnitt - Prospektiv oder retrospektiv	- Beobachtungsstudie - Längsschnitt - Retrospektiv	- Beobachtungsstudie - Querschnitt	Qualitativ
Beschreibung	„Studie über Studien“ - Zuverlässigste Informationen in der medizinischen Forschung - Einbezogene Studien müssen ähnlich sein bezügl. Auswahl der Teilnehmenden, geprüfte Interventionen (bzw. Forschungsfragen) u. gemessenen Outcomes - Qualitätsprüfung u. statistische Zusammenfassung der Ergebnisse aus verschiedenen Studien	- Zufällige Zuteilung der Teilnehmenden zu den Untersuchungsgruppen - Zuteilung auf Ebene einzelner Personen (indiv. Randomisierung) oder auf Ebene von Bevölkerungsgruppen (Cluster-Randomisierung) „Goldstandard“ der Studientypen	- Einteilung der Probanden in Gruppen auf natürliche Weise, z. B. durch den Wohnort (nicht-randomisierte Zuteilung) - Vergleichbarkeit der Gruppen muss in der Analyse sichergestellt werden	- Oftmals in der Feldforschung eingesetzt - Zwei Typen: - Ein-Gruppen-Prätest-Posttest-Versuchsplan - Interrupted Time Series Design	- Daten werden über den Untersuchungszeitraum hinweg zu unterschiedlichen Zeitpunkten erhoben - Prospektive Kohortenstudie: Die Daten werden nach Beginn der Studie eigens für diese Studie neu erhoben - Retrospektive Kohortenstudie: Datenerhebung fand vor Beginn der Studie statt	- Beobachtungsstudie mit zwei Populationen (erkrankt und gesund): Vergleich des Auftretens der Expositionen - Stellt korrelativen Zusammenhang zwischen Exposition und Outcome her	- Momentaufnahme zu bestimmtem Zeitpunkt (Ergebnis = deskriptive Statistik) - Man unterscheidet zw. Erhebung mit einer oder mehreren Versuchsgruppen (Single-Strata-Approach vs. Cross-Sectional-Approach) - Wird verwendet, um das Ausmaß des Behandlungseffektes zu bestimmen, nicht die Effektivität einer Intervention	Detaillierte qualitative Beschreibung
Für welche Fragestellungen geeignet?	Geeignet für spezifische und klar definierte Fragestellungen –> PICO-Format (problem/patient, intervention, comparison, outcome)	Geeignet für klinische Wirksamkeitsprüfungen von Interventionen	Geeignet für Fragestellungen, bei denen Randomisierung ethisch nicht vertretbar oder wegen geringer Fallzahl nicht möglich ist: Alternative zur RCT	Geeignet für einen Vergleich über die Zeit: Beurteilen zeitlicher Entwicklungen	Geeignet für seltene Expositionen	Gut geeignet bei seltenen Erkrankungen/ Outcomes	Eignet sich, um neue Hypothesen zu entwickeln	Gut geeignet für Beschreibung der natürlichen Situation
Für welche Fragestellungen nicht bzw. wenig geeignet?	Nicht geeignet bei neuen Interventionen: Kann erst durchgeführt werden, wenn bereits mehrere Studien abgeschlossen wurden	Interventionen mit hoher Komplexität, die sich nicht im Experiment abbilden lassen	Wenig geeignet für klinische Wirksamkeitsprüfungen	Nicht geeignet für Fragestellungen, bei denen Kausalität des Interventionseffektes bestätigt werden muss	Nicht geeignet für seltene Outcomes	Nicht geeignet bei selten auftretender Exposition	Nicht geeignet für das „Testen“ der Hypothesen: Kann keine Ursache-Wirkungs-Beziehung aufdecken	Nicht geeignet für das „Testen“ der Hypothesen: Kausale Beziehungen nur sehr eingeschränkt herstellbar
Welche Gruppen werden verglichen?	Wie in eingeschl. Studien, Möglichkeit der Netzwerk-Meta-Analyse bei untersch. Kontrollgruppen	Vergleich von Interventions- und Kontrollgruppe	Vergleich von Interventions- und Kontrollgruppe	Population vor und nach Intervention	Vergleich von Exponierten und Nicht-Exponierten	Vergleich von Fällen und Kontrollen		

Tabelle 1: Übersicht verschiedener Studientypen. Quelle: inav GmbH, basierend auf [5].

Literatur

- [1] Pfaff, H., Abholz, H., Glaeske, G., Icks, A., Klinkhammer-Schalke, M., Nellesen-Martens, G., ... & Stemmer, R. (2011). Versorgungsforschung: unverzichtbar bei Allokationsentscheidungen – eine Stellungnahme. *DMW-Deutsche Medizinische Wochenschrift*, 136(48), 2496-2500.
- [2] Amelung, V. & Wagner, C. (2010). Neue Versorgungsformen und Versorgungsforschung. In *Case Management* (pp. 169-200). Gabler.
- [3] Gräßel, E., Donath, C., Holleeder, A., Drexler, H., Kornhuber, J., Zobel, A. & Kolominsky-Rabas, P. (2015). Versorgungsforschung – evidenzbasiert. Ein Kurzüberblick und Implikationen. *Das Gesundheitswesen*, 77(03), 193-199.
- [4] Geraedts, M., Drösler, S. E., Döbler, K., Eberlein-Gonska, M., Heller, G., Kuske, S., ... & Schrappe, M. (2017). DNVF-Memorandum III „Methoden für die Versorgungsforschung“, Teil 3: Methoden der Qualitäts- und Patientensicherheitsforschung. *Das Gesundheitswesen*, 79(10), e95-e124.
- [5] Pfaff, H., Neugebauer, E., Glaeske, G. & Schrappe, M. (2017). *Lehrbuch Versorgungsforschung. Systematik - Methodik - Anwendung* (2. Auflage). Schattauer GmbH, Verlag für Medizin und Naturwissenschaften.
- [6] Hariton, E. & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology*, 125(13), 1716.
- [7] Mühlhauser, I., Lenz, M. & Meyer, G. (2011). Entwicklung, Bewertung und Synthese von komplexen Interventionen – eine methodische Herausforderung. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 105(10), 751-761.
- [8] Lösel, F. & Heinrichs, N. (2013). Messung von Wirksamkeit und Umsetzungsqualität (Evaluation). In *Entwicklungsförderung & Gewaltprävention* (2): In: Deutsches Forum für Kriminalprävention (2) (pp. 24-26).
- [9] Robert Koch-Institut & Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit (Hrsg.) (2012). *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes*. RKI, Berlin.
- [10] Finckh, A. & Tramèr, M. R. (2008). Primer: strengths and weaknesses of meta-analysis. *Nature Clinical Practice Rheumatology*, 4(3), 146-152.
- [11] Scriven, M. (2008). A summative evaluation of RCT methodology: An alternative approach to causal research. *Journal of multidisciplinary evaluation*, 5(9), 11-24.
- [12] Reeves, B. C., Deeks, J. J., & Higgins, J. P. (2008). 13 Including non-randomized studies. *Cochrane handbook for systematic reviews of interventions*, 1, 391.
- [13] Reichardt, C. S. (2009). Quasi-experimental design. *The SAGE handbook of quantitative methods in psychology*, 46(71), 490-500.
- [14] Grimes, D. A. & Schulz, K. F. (2002). Cohort studies: marching towards outcomes. *The Lancet*, 359(9303), 341-345.
- [15] Schulz, K. F. & Grimes, D. A. (2002). Case-control studies: research in reverse. *The Lancet*, 359(9304), 431-434.
- [16] Wang, X. & Cheng, Z. (2020). Cross-sectional studies: strengths, weaknesses, and recommendations. *Chest*, 158(1), S65-S71.
- [17] Noor, K. B. M. (2008). Case study: A strategic research methodology. *American journal of applied sciences*, 5(11), 1602-1604.

	Übergeordnete Synthese	Interventionsstudien			Beobachtungsstudien			
	Systematischer Review und Metaanalyse	Randomisierte kontrollierte Studie (Randomized Controlled Trial – RCT)	Nicht-randomisierte kontrollierte Studie (Non-RCT)	Quasi-experimentelles Design (nicht-kontrollierte Studie: ohne Kontrollgruppe)	Kohortenstudie	Fall-Kontroll-Studie	Querschnittsdesign	Fallstudie
Gruppenaufteilung	Wie in eingeschlossenen Studien	Randomisierung: Zuteilung zur Kontroll- oder Interventionsgruppe nach Zufallsprinzip	Natürliche Zuteilung zwischen Interventions- und Kontrollgruppe		Aufteilung nach Risikofaktoren zu Beginn der Studie	Aufteilung nach Vorliegen der Erkrankung zu Beginn der Studie		
Vorteile	- Hohe interne Validität durch eingeschlossene RCTs, höhere externe Validität im Vergleich zu einzelnen RCT - Große Stichprobe - Weniger anfällig für Fehler und Verzerrung durch Ausgleich über mehrere Studien - Qualitätsprüfung u. Analyse nach einem festgelegten, methodisch durchdachten Prozess - Statistische Zusammenfassung der Ergebnisse aus versch. Studien	- Hohe interne Validität - Unterschiede zw. den Gruppen hinsichtlich des eintretenden Effekts lassen sich direkt auf die getestete Intervention zurückführen - Gleichmäßige Verteilung von bekannten und nicht bekannten Einflussfaktoren auf alle Gruppen	Hohe externe Validität (vergleicht natürliche Gruppen)	Kontrolle möglicher Beeinträchtigungen der internen Validität: Alle Studienteilnehmenden bekommen die gleiche „Intervention“ unter gleichen Bedingungen	- Zeitliche Abfolge einer Exposition hin zum Outcome kann untersucht werden - Ermöglicht es, die Inzidenz eines Outcomes zu ermitteln - Mehrere Outcomes können gleichzeitig untersucht werden	- Es können mehrere Expositionen gleichzeitig untersucht werden - Im Vergleich zu Kohortenstudien relativ schnell und kostengünstig durchzuführen	Geringer Personen- und Zeitaufwand	Extreme Situationen oder Leistungen können erfasst werden
Nachteile	- Wenn individuelle Studienqualität nicht ausreichend validiert wurde, kann dies zur Überschätzung der Evidenzstärke führen - Qualität hängt von der Qualität der eingeschlossenen Studien ab - Kann erst durchgeführt werden, wenn mehrere Studien vorliegen	- Externe Validität kann durch „künstliches Setting“ des Experiments infrage gestellt werden - Ethische Vorbehalte: Exposition bei risikoreichen Interventionen bzw. Vorenthaltung nützlicher Interventionen - Equipoise: Gleichwertigkeit zw. zwei alternat. med. Interventionen (da noch keine Evidenz vorliegt) als zentr. ethische Voraussetzung	- Interne Validität hängt von Balance der Gruppen ab - Durch feste Zuteilung der Teilnehmenden ist Kontrolle der Störvariablen erschwert: Große Gefahr von „unobserved variable bias“ - Die Gruppen sind nie völlig gleich (äquivalent): Gemessene Unterschiede zwischen den Gruppen können auf schon vorher bestehende Unterschiede zurückgehen	- Interne Validität anfänglich: Gefahr des „unobserved variable bias“ - Ohne Kontrollgruppe kann Kausalität des Interventionseffekts nie bestätigt werden - Geht die Veränderung tatsächlich auf die Intervention zurück? – Ableitung von Schlussfolgerungen nur durch Ausschluss von Alternativerklärungen möglich	- Oft sehr viele Studienteilnehmende notwendig - Erste Ergebnisse erst nach langer Zeit verfügbar (prospektive Kohortenstudie)	- Sehr anfällig für Verzerrungen („Bias“) - Ergebnisse stark abhängig von der Wahl der Kontrollen	- Feststellung von Veränderungen nicht/schwer möglich - Große Gefahr von Fehlschlüssen - Kann keine Ursache-Wirkungs-Beziehung aufdecken	- Oft nur subjektive Daten → begrenzte Generalisierung - Kausale Beziehungen nur sehr eingeschränkt herstellbar
Variante: Stepped Wedge Design Die Intervention wird in allen Clustern durchgeführt, aber zu unterschiedlichen Zeitpunkten. Der Startzeitpunkt für die Intervention wird für jedes Cluster mittels Randomisierung festgelegt. Bis zu diesem Startzeitpunkt befinden sich alle Probanden des jeweiligen Clusters im Kontrollarm.					Variante: Regressions-Diskontinuitätsanalyse Man nutzt eine Diskontinuität oder Unstetigkeit in einer beobachteten Kontrollvariablen, die zu einer fast zufälligen Zuteilung in die Interventions- oder Kontrollgruppe führt.			
					Variante: Cross-Over Design Zwei Heilmittel werden zeitlich versetzt denselben Probanden verabreicht. Es gibt mind. zwei Studienphasen. Phase I: Die Probanden der ersten Gruppe erhalten Therapie A, die Probanden der zweiten Gruppe Therapie B. Die Zuteilung zu den Gruppen erfolgt randomisiert. Phase II: Gruppe 1 erhält Therapie B und Gruppe 2 Therapie A. Anschließend wird die Wirksamkeit der Therapien innerhalb einer Gruppe (Intragruppenvergleich) und zwischen den beiden Gruppen verglichen (Intergruppenvergleich).			

Autor:innen

Prof. Dr. Volker E. Amelung ist Professor für internationale Gesundheitssystemforschung an der Medizinischen Hochschule Hannover sowie Mitglied der Jury für den MSD Gesundheitspreis.

Birgit Dembski ist Mitglied des Vorstands der BAG Selbsthilfe sowie Mitglied der Jury für den MSD Gesundheitspreis.

Franz-Helmut Gerhards ist Chief Digital Officer bei der DAK-Gesundheit und Mitglied der Jury für den MSD Gesundheitspreis.

Prof. Dr. Daniel Grandt ist Chefarzt der Inneren Medizin I am Klinikum Saarbrücken und Mitglied der Jury für den MSD Gesundheitspreis.

Dr. Rainer Hess ist Jurist, ehemaliger unparteiischer Vorsitzender des Gemeinsamen Bundesausschusses und Mitglied der Jury für den MSD Gesundheitspreis.

Dr. Jutta Wendel-Schrief ist Director Market Access bei der MSD Sharp & Dohme GmbH und Mitglied der Jury für den MSD Gesundheitspreis.

Chantal Friebertshäuser ist Geschäftsführerin der MSD Sharp & Dohme GmbH.

Sabine Barz ist Kulturwissenschaftlerin und als Projektmanagerin Versorgungsforschung bei der inav GmbH tätig.

Ann-Kathrin Klähn ist wissenschaftliche Mitarbeiterin der inav GmbH mit dem Schwerpunkt Evaluation innovativer Versorgungsprojekte.

Martin Göhl ist Senior Manager Versorgungsprogramme & Kooperationen bei der MSD Sharp & Dohme GmbH.