

Persönliche PDF-Datei für

Mit den besten Grüßen vom Georg Thieme Verlag

www.thieme.de

Dieser elektronische Sonderdruck ist nur für die Nutzung zu nicht-kommerziellen, persönlichen Zwecken bestimmt (z. B. im Rahmen des fachlichen Austauschs mit einzelnen Kollegen und zur Verwendung auf der privaten Homepage des Autors). Diese PDF-Datei ist nicht für die Einstellung in Repositorien vorgesehen, dies gilt auch für soziale und wissenschaftliche Netzwerke und Plattformen.

Verlag und Copyright:

Georg Thieme Verlag KG
Postfach 30 11 20
70451 Stuttgart
ISSN

Alle Rechte liegen beim
Verlag



Studiendesigns zur Wirkungsevaluation in Prävention und Gesundheitsförderung – Die Bedeutung interner und externer Validität für die Evidenzbasierung in Lebenswelten

Study Designs for Effect Evaluation in Disease Prevention and Health Promotion – Meaning of Internal and External Validity in Intervention Studies

Autoren

Wolfgang Bödeker¹, Susanne Moebus²

Institute

- 1 EPICURUS, Wirkungsanalysen, Essen
- 2 Institut für Medizinische Informatik, Biometrie und Epidemiologie, Zentrum für Urbane Epidemiologie, Universitätsklinikum Essen, Essen

Schlüsselwörter

Gesundheitsförderung, Prävention, Lebenswelten, Validität, Evaluation, Evidenz

Key words

Health promotion, prevention, settings, validity, evaluation, evidence

online publiziert 11.4.2019

Bibliografie

Gesundheitswesen 2020; 82: e147–e157

DOI 10.1055/a-0832-2220

ISSN 0941-3790

© 2020. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Korrespondenzadresse

Prof. Dr. Susanne Moebus
Institut für Medizinische Informatik, Biometrie und Epidemiologie (IMIBE)
Universitätsklinikum Essen
Hufelandstraße 55
45122 Essen
Susanne.Moebus@uk-essen.de

ZUSAMMENFASSUNG

Hintergrund Die Gesundheitsförderung und Prävention in Deutschland haben aufgrund des Präventionsgesetzes an Bedeutung gewonnen. Durch das Gesetz erfahren die sogenannten Lebenswelten (settings) eine besondere Beachtung. Bisher gibt es allerdings weder ein einheitliches Verständnis über

diesen Ansatz, noch sind dessen Potenziale empirisch hergeleitet. Es bedarf daher einer Auseinandersetzung mit geeigneten Evaluations- und Studienkonzepten.

Ziele und Methoden Ziele dieser Arbeit sind, die Bedeutung der internen und externen Validität von Interventionsstudien in Gesundheitsförderung und Prävention zu untersuchen. Wir geben einen Überblick zu verschiedenen Studiendesigns im Hinblick auf die erreichbare Verzerrungskontrolle und stellen Anwendungsbeispiele für die Evaluation von Setting-Projekten vor.

Ergebnisse Lebenswelt bezogene Interventionen sind oft durch eine Vielzahl von Maßnahmen, Akteuren und Interventionskontexten gekennzeichnet. Der Evidenz basierten Medizin entlehene Analysemethoden werden in Deutschland für die Gesundheitsförderung kritisch gesehen. Solche Studien würden zwar eine hohe interne Validität aufweisen, die Übertragbarkeit der Ergebnisse auf „die Wirklichkeit“ sei aber gering. Dagegen ist die Übertragbarkeit von Studienergebnissen für Setting-Projekte keine größere Herausforderung als für andere Forschungsfelder. Sie ist nicht durch unterschiedliche Kontexte, sondern durch unterschiedliche Wirkungsbeziehungen begrenzt. Durch einen Wirkungsnachweis soll ein beobachteter Effekt kausal auf eine Intervention zurückgeführt werden. Die erkenntnistheoretischen Anforderungen unterscheiden sich hierbei nicht zwischen Studien, die auf interne oder externe Validität angelegt sind. Die internationale Diskussion fokussiert auf die Weiterentwicklung (quasi)-experimenteller Studiendesigns. Anwendungsbeispiele für Setting-Projekte vorwiegend aus Deutschland verdeutlichen, dass diese Verfahren vereinzelt bereits in der Evaluationspraxis angewendet werden. **Schlussfolgerung** Eine Herausforderung für die Gesundheitsförderung und Prävention in Lebenswelten besteht darin, die heterogene Evidenzbasis systematisch zu erfassen. Derzeit liegt eine Vielzahl von Erkenntnissen in einer Vielzahl von Interventionsfeldern vor. Diese Erkenntnisse müssen zusammengestellt und dahingehend analysiert werden, ob und im welchem Umfang weitere Evaluationen mit welchen Methoden zu initiieren sind.

ABSTRACT

Background Health promotion and disease prevention in Germany have become more important as a result of the Prevention Act that gives special attention to the setting. So far, however, there is neither a common understanding of the terms of this approach, nor are its potentials empirically derived. Therefore, a discussion about suitable evaluation and study concepts is required.

Aims and methods The aims of this study were to address the meaning of internal and external validity of intervention studies in health promotion and disease prevention. We provide an overview of the achievable bias control for different study designs and provide examples for the evaluation of setting-based approaches.

Results Interventions in settings are often characterized by a multitude of measures, actors and intervention contexts. Methods of analysis from evidence-based medicine are viewed critically for health promotion and disease prevention in Germany. Such studies are considered to provide a high degree of internal validity, but the extrapolation of the results to “reality”

is viewed as low. In contrast, the extrapolation of study results is not more of a challenge for setting projects than for any other research area. It is not limited by different contexts, but rather by different causal relationships. Impact assessment aims at causally attributing an observed outcome to the intervention. Thus, the epistemological requirements do not differ between studies that are designed for internal or external validity. The international discussion focuses on the refinement of (quasi)-experimental study designs. Examples for the evaluation of setting projects, mainly from Germany, illustrate that those alternative methods have already been used in evaluation practice.

Conclusion A challenge for health promotion and disease prevention in settings is to systematically assess the different needs for evidence-basing. At present, there is a wealth of findings in a large number of intervention fields. These findings must be compiled and analyzed to determine whether and to what extent further evaluations need to be initiated and by which methods.

Einleitung

Gesundheitsförderung und Prävention in Deutschland haben aufgrund des Präventionsgesetzes weiter an Bedeutung gewonnen. Durch das Gesetz werden insbesondere der gesetzlichen Krankenversicherung (GKV) weitreichende neue Aufgaben zugewiesen, darunter die Mitwirkung bei dem Aufbau gesundheitsförderlicher Strukturen [1]. Eine besondere Betonung erfahren die sogenannten Lebenswelten (Setting-Ansatz), also „... für die Gesundheit bedeutsame, abgrenzbare soziale Systeme insbesondere des Wohnens, des Lernens, des Studierens, der medizinischen und pflegerischen Versorgung sowie der Freizeitgestaltung einschließlich des Sports“. (§20 a (1) Sozialgesetzbuch V). Die gesetzlichen Krankenkassen haben zusätzlich zu eigenen Leistungen zur Gesundheitsförderung und Prävention in Lebenswelten Unterstützungsleistungen zu finanzieren, die der GKV-Spitzenverband bei der Bundeszentrale für gesundheitliche Aufklärung (BzgA) jährlich beauftragt. Hierzu zählt die Entwicklung „... der Art und der Qualität krankenkassenübergreifender Leistungen, deren Implementierung und deren wissenschaftlicher Evaluation“ (SGB V, §20 a (3)). Im Jahr 2016 wurden für Maßnahmen des Setting-Ansatzes ca. 115 Mio. Euro [2] ausgegeben, wovon ca. 30 Mio. Euro auf Unterstützungsleistungen durch die BzgA entfielen. Eine grundsätzliche Neuausrichtung der durch medizinische Leistungen geprägten GKV-Präventionsausgaben ist damit noch nicht erkennbar [3]. Das Präventionsgesetz legt jedoch erstmals eine finanzielle Grundlage, Strukturen für eine systematische Identifikation von Evidenz basierten Maßnahmen der Gesundheitsförderung und Prävention in Lebenswelten zu schaffen und diese Maßnahmen dauerhaft in die Praxis zu überführen.

Der Setting-Ansatz in der Gesundheitsförderung und Prävention wurde vor nunmehr 30 Jahren durch die Ottawa-Charta der WHO verbreitet und gilt als politisch erfolgreich. Die Prävention in Lebenswelten als aufsuchende Interventionsstrategie scheint a priori besonders geeignet, sozial benachteiligte Zielgruppen zu erreichen und Sozillage bedingte schlechtere Gesundheitschancen zu adressieren. Entgegen seiner Plausibilität und politischen Durch-

dringung steht es um die methodische Weiterentwicklung des Settings-Ansatzes nicht gut. So kann bis heute nicht von einem einheitlichen Verständnis der Inhalte, Begriffe und Potentiale gesprochen werden. Das Konzept gilt weiterhin als unscharf oder gar widersprüchlich, schwer von anderen Konzepten abgrenzbar, und die damit verknüpfte Erfolgserwartung ist wesentlich Theorie gestützt ohne bisher aus systematischen Evaluationen abgeleitet worden zu sein [4, 5]. Auch das Präventionsgesetz hat offenbar nicht zu einer Vereinheitlichung beigetragen. Laut o.g. Definition des Sozialgesetzbuches wäre unter Lebenswelt ein Typus von Systemen zu verstehen, wie etwa Kitas oder Schulen, wohingegen die Krankenkassen den Begriff zur Bezeichnung von einzelnen Elementen dieser Typen verwenden (jede Kita, jede Schule). Entsprechend werden von der GKV für das Jahr 2016 Maßnahmen in mehr als 35.000 Settings ausgewiesen, die aber nur in 15 verschiedenen System-Typen erbracht wurden. Auf Grundschulen und Kindergärten/Kita entfallen bereits 60 % der dokumentierten Maßnahmen [2]. Das unterschiedliche Verständnis von Lebenswelten hat unmittelbare Auswirkungen für die Entwicklung und Evaluation von Setting bezogenen Maßnahmen. Es wäre nämlich zunächst zu klären, ob z. B. die Evidenz für die Wirksamkeit von Bewegungsprogrammen in Kita beurteilt werden soll, oder der Effekt einer entsprechenden Maßnahme in einer bestimmten Kita.

Darüber hinaus werden die Ziele der Evaluation von Setting-Maßnahmen derzeit unterschiedlich ausgelegt. So heißt es im Leitfaden Prävention des GKV-Spitzenverbandes [6], dass die Wirksamkeit der Interventionen insbesondere im Hinblick auf eine Verstärkung von Präventionshandeln beurteilt werden soll. Dagegen wird von den Krankenkassen häufig die alleinige Erfassung der Teilnehmerzufriedenheit oder Zugänglichkeit einer Maßnahme als Evaluation akzeptiert [2]. Diese Unterscheidung ist relevant für das methodische Vorgehen in der Evaluation und Evidenzbasierung, denn die Untersuchung, ob eine Maßnahme wirkt oder ob eine wirksame Maßnahme verbreitet angewendet wird, muss durch verschie-

dene Akteure im Evaluationsprozess und mit verschiedenen Verfahren erfolgen. Es ist also festzulegen, auf welcher Ebene welche Evaluation ansetzen soll, und wie die Ergebnisse systematisch erfasst und zusammengeführt werden können.

Die zunehmend geforderte Evaluation und Evidenzbasierung von Interventionen in der Gesundheitsförderung und Prävention wird vielfach als schwierig gesehen. Insbesondere die Lebenswelt bezogenen Interventionen sind oft gekennzeichnet durch eine Vielzahl von Komponenten, Akteuren und Interventionskontexten [7]. Aus dieser Vielfalt wird gefolgert, dass das übliche Vorgehen bei der Evidenzbasierung in Gesundheitsförderung und Prävention nicht angemessen ist [5, 8, 9], da diese der klinischen Medizin entlehnten Verfahren auf experimentelle, d. h. randomisiert, kontrollierte Studien (RCT) fokussieren. Solche Studien würden zwar eine hohe interne Validität aufweisen, also unter den Studienbedingungen den Zusammenhang zwischen Intervention und Effekt beschreiben, die Übertragbarkeit der Ergebnisse auf die „Wirklichkeit“ sei aber gering. Daraus wird gefolgert, dass in der Evaluation und Evidenzbasierung Ansätze verfolgt werden müssten, die auch sogenannte schwächere Studiendesigns einbeziehen. Die internationale Diskussion fokussiert dagegen bereits seit langem auf Evaluationsmethoden in Gesundheitsförderung und Prävention, die ähnlich belastbare Aussagen ermöglichen wie experimentelle Studien [10]. Ziele dieser speziellen Studientypen und Studiendesigns sind die Berücksichtigung der Besonderheiten komplexer Interventionen sowie Aussagen mit hoher externer Validität zu erreichen. In Deutschland ist inzwischen ebenfalls ein zunehmendes Interesse an solchen Studiendesigns erkennbar [11].

Vor diesem Hintergrund soll die Bedeutung der internen und externen Validität von Interventionsstudien in Gesundheitsförderung und Prävention aufgezeigt sowie geeignete Studiendesigns für die verschiedenen Evaluationsebenen von Setting-Maßnahmen vorgestellt werden. Zur Sicherstellung des Einklangs von Präventionszielen, wissenschaftlichem Kenntnisstand und Präventionspraxis wird ein Vorschlag für einen rückgekoppelten Forschungsprozesses entwickelt.

Die Bedeutung von Übertragbarkeit in Interventionsstudien

Interne und externe Validität

„Validität“ ist ein Begriff der allgemeinen Testtheorie und bezeichnet die Eigenschaft eines Messinstruments oder -verfahrens das zu Messende gültig wiederzugeben. Bei empirischen Studien hat sich das Verständnis von Validität differenziert und es wird von interner Validität in Bezug auf definierte Studienpopulationen und -bedingungen gesprochen. Dagegen soll externe Validität ausdrücken, ob der in einer Studie beobachtete Interventionseffekt auch in „Wirklichkeit“, also außerhalb der definierten Studienbedingungen, zutreffend ist. Von hoher interner Validität eines Studienergebnisses wird ausgegangen, wenn ein unter Beobachtung stehender Effekt möglichst eindeutig einer Intervention zugeordnet werden kann, andere Einflüsse also ausgeschlossen sind. Eine hohe externe Validität soll dagegen dann vorliegen, wenn sich der in einer Studie ermittelte Interventionseffekt auch auf andere Populationen und andere Interventionsbedingungen übertragen lässt.

Während für die Beurteilung der internen Validität eine Vielzahl von Kriterien der Studienplanung und -auswertung entwickelt wurden, werden zur Beurteilung der externen Validität zusätzliche Kriterien wie die Reichweite der Ergebnisse, die dauerhafte Etablierung und die Kosten von Interventionen herangezogen [12].

Häufig werden interne und externe Validität in einem Konkurrenzverhältnis gesehen [13]. Je strikter die kontrollierenden Bedingungen, die eine hohe interne Validität begünstigen, desto geringer sei die Übertragbarkeit auf die in einer anderen „Wirklichkeit“ lebenden Zielpopulation. Diese Sichtweise ist durchaus strittig. So wird auch argumentiert, dass interne Validität ein Kriterium einer Studie sei, während es sich bei externer Validität um ein Kriterium handle, das eine Übertragungssituation beschreibt. Für bestimmte Situationen wäre die Übertragbarkeit gut und für andere wäre sie nicht gegeben. Die Konzepte würden daher auch nicht in Konkurrenz stehen [14, 15].

Dass Übertragbarkeit auch von dem Interventionsmodell abhängt, zeigt das Beispiel „Komm mit in das gesunde Boot“ [16, 17]. Hierbei handelt es sich um eine aufwendige Cluster randomisierte Interventionsstudie zur Förderung der physischen Aktivität und gesunden Ernährung bei Kindern in Kindergärten. Die Randomisierung erfolgte unter Berücksichtigung des sozio-ökonomischen Status und der Region. Die Intervention war systemtheoretisch und modellhaft gestützt, mit Eltern und Lehrern weiterentwickelt und angepasst. Die Studie weist damit nach gängiger Meinung eine hohe externe Validität auf, da sie unter Praxisbedingungen durchgeführt wurde. Da sich aber die Randomisierung nur auf Kindergärten erstreckte, die sich aktiv um eine Teilnahme beworben hatten, könnten Effekte in anderen Kindergärten anders ausfallen. Die Übertragbarkeit der Ergebnisse auf nicht interessierte Kindergärten wäre möglicherweise also gering. Allerdings könnte ein besonderes Teilnahmeinteresse der Kindergärten auch als angemessene und Theorie adäquate Voraussetzung einer angestrebten Routineintervention aufgefasst werden. Eine Einschränkung der externen Validität bestünde dann nicht, weil nicht interessierte Kindergärten nicht zur Zielpopulation der Intervention zählten. Sprich, die Zielpopulation nicht interessierter Kindergärten benötigt ein anderes Interventionsmodell.

Darüber hinaus ist die Übertragbarkeit von Ergebnissen aus Interventionsstudien von einer Untersuchungs- auf die Zielpopulation respektive von einer Studienwirklichkeit auf eine „allgemeine Wirklichkeit“ grundsätzlich nicht gegeben. Erstens ist das Erkenntnisziel von Interventionsstudien nämlich stets zukunftsbezogen. Im Vordergrund steht nicht, ob die Intervention zum Zeitpunkt der Untersuchung wirksam war, sondern ob sie zu einem späteren Zeitpunkt wirksam wäre, nämlich dann, wenn die Intervention routinemäßig eingeführt ist. Zweitens gibt es keine „allgemeine Wirklichkeit“, die alle Interventionskontexte und Personen im Setting beschreibt. Der Geltungsanspruch empirischer Studien ergibt sich daher aus der heuristischen Regel, dass von einer Übertragbarkeit der Ergebnisse solange ausgegangen werden sollte, solange nicht begründet Einschränkungen der Aussagekraft angenommen werden müssen. Solche Einschränkungen ergeben sich aber nicht bereits durch Unterschiede in den Populationen oder der Kontexte. Die Übertragbarkeit wäre dann gering, wenn begründet erwartet werden muss, dass die Wirkungsbeziehungen zwischen Interventionen und Effekten für Subgruppen oder in Kontexten differenziell sind. Voraussetzung für eine solche Beurteilung ist zunächst die

genaue Kenntnis der Interventionskontexte und -bedingungen, über die in Publikationen allerdings oft nicht ausreichend berichtet wird. Daher wurden verschiedentlich Kriterien und Instrumente zur Charakterisierung der Übertragbarkeit vorgeschlagen [12, 18–21]

Efficacy- und Effectiveness-Studien

Eine konfliktfreie Beziehung zwischen interner und externer Validität ergibt sich, wenn die Begriffe spezifischen Untersuchungsmotiven und damit entsprechenden Studientypen zugeschrieben werden, nämlich den „efficacy“- und „effectiveness“-Studien. Diese wohl auf Archie Cochrane zurückgehenden Studienbezeichnungen [10] eröffnen eine bedeutsame Unterscheidungsmöglichkeit. Interne und externe Validität sind dann nicht einfach charakterisierende Merkmale derselben Studie, sondern verschiedener Studien mit eigenen Studienzielen. Externe Validität folgt aus effectiveness-Studien, deren Ziel es ist, Interventionen unter Bedingungen zu untersuchen, die den antizipierten Routinebedingungen entsprechen oder nahe kommen.

Die Unterscheidung zwischen „efficacy“- und „effectiveness“-Studien führte in der angelsächsischen Literatur schon seit den 80iger Jahren zu einer regen Diskussion über deren jeweiligen Bedeutung für Interventionen in Prävention und Gesundheitsförderung. Eine Rezeption dieser Diskussion findet in Deutschland offenbar kaum statt, es existiert entsprechend auch keine deutschsprachige Übersetzung der Begriffe. Nach Flay [10] untersuchen efficacy-Studien, ob eine Wirkung einer Intervention unter optimalen Studienbedingungen eintritt. Efficacy-Studien weisen hiernach folgende Merkmale auf: (i) die zu untersuchende Interventionen sind genau spezifiziert und werden standardisiert angewendet; (ii) Interventionen werden einheitlich in standardisierten Kontexten angewendet; (iii) es werden Studienteilnehmer einbezogen, die die Interventionen verstehen, akzeptieren und an der Studie wie geplant und vorgesehen teilnehmen. Eine effectiveness-Studie liegt dagegen dann vor, wenn von den obigen Merkmalen abgewichen wird, um die Wirkung unter „real-world conditions“ zu untersuchen [10]. So könnte etwa auf einen standardisierten Kontext bewusst verzichtet werden, um die Intervention in möglichst variablen Kontexten zu erproben.

Flay [10] weist darauf hin, dass der Durchführung einer effectiveness-Studie 2 verschiedene Motive zugrunde liegen können. Zum einen ginge es – bei bereits erfolgreich durchgeführter efficacy-Studie – um die weitergehende Untersuchung, ob die Effekte auch unter Praxisbedingungen auftreten. Es wird also untersucht, in wieweit die Standardbedingungen der efficacy-Studie variiert werden können, ohne die Wirksamkeit zu verlieren. Zum anderen können effectiveness-Studien auch zur Evaluation bereits operativer Interventionen verwendet werden, ohne dass bereits vorher die Wirksamkeit in efficacy-Studien gezeigt wurde. Allgemeinen sei ein sequentieller Ablauf der Forschung anzustreben, um im politischen Diskurs den Bedarf von Maßnahmen zweifelsfrei belegen zu können oder weiteren Forschungsbedarf zu signalisieren [22].

Sequentieller Ablauf in der Interventionsforschung?

Efficacy- und Effectiveness-Studien stehen nach der oben beschriebenen Sichtweise in einer untersuchungslogischen Abfolge. Sehr deutlich wird dies durch ein Phasenmodell, das von Greenwald &

Cullen [23] in einer einflussreichen Veröffentlichung eingeführt wurde. Zur Sicherung des Wissenstransfers in die Präventionspraxis halten die Autoren fünf Phasen im Prozess der Präventionsforschung für erforderlich. Efficacy- und Effectiveness-Studien werden dabei in diesem Prozess an verschiedenen Stellen verortet (siehe Textkasten). International wird das Phasenmodell in vielen Feldern der Präventionsforschung rezipiert und dient als Schema etwa für die Entwicklung von Vorschlägen für methodische Standards von efficacy- und effectiveness-Studien [22] oder auch für die Diskussion über den mangelhaften Transfer von Präventionswissen in Präventionshandeln (vgl. [24]).

Phasenmodell in der Präventionsforschung nach Greenwald & Cullen (1985)[23]

- Phase 1: Hypothesen über mögliche gesundheitsförderliche Intervention entwickeln. Ausgangspunkt sind verfügbare Kenntnisse über Gesundheitsprobleme, Risikofaktoren, unvorteilhafte Settings, betroffene Personen und Bedingungen. Der Ausgangspunkt kann auch normativ sein.
- Phase 2: Methoden für die Intervention entwickeln. Es wird betrachtet, welche Risikofaktoren/Setting adressiert werden müssen, welche Ansätze verfügbar sind, welche Interventionsbedingungen relevant sind. Es werden ggf. Machbarkeitsstudien, Akzeptanzstudien oder Pilotstudien durchgeführt. Indikatoren für die Intervention als auch Effekte werden entwickelt.
- Phase 3: Efficacy-Studie (Original: controlled intervention trials). Aufgrund der Vorarbeiten wird eine Interventionsstudie durchgeführt. Das Studiendesign sollte auf die verlässliche Ermittlung eines Zusammenhangs ausgerichtet sein, also prinzipiell durch Auswahl günstiger Bedingungen wie homogene Gruppen, guter Zugang, hohe Compliance. Durch geeignete Kontrollgruppen ggf. mit Randomisierung soll ein vorhandener Zusammenhang möglichst zwingend erkannt werden. Die Studienpopulation muss nicht für die spätere Zielgruppe repräsentativ sein.
- Phase 4: Effectiveness-Studie (Original: defined population studies). Die in Phase 3 erfolgreich getestete Intervention wird unter variierten Bedingungen untersucht. Insbesondere wird eine größere Population einbezogen, die der späteren Zielgruppe ähnlicher ist. Die Studien zielen also auf Subgruppenanalysen (für wen wirkt die Intervention) und Kontextanalysen (unter welchen Bedingungen wirkt die Intervention). Zudem sollen in dieser Phase Implementationshürden für eine routinemäßige Anwendung und ggf. weiterer Forschungsbedarf identifiziert werden.
- Phase 5: Umsetzung (Original: demonstration and implementation). Die Intervention wird großflächig, routinemäßig in der eigentlichen Zielgruppe angewendet. Die Erfahrungen führen zu der Etablierung der Intervention als ständiges Programm, das einem kontinuierlichen Monitoring unterzogen wird.

Ob die Interventionsforschung in der Gesundheitsförderung und Prävention stets einem sequentiellen Ablauf entsprechen sollte, ist strittig. Auch Greenwald & Cullen [23] waren der Auffassung, dass Studien der Phase 3 und Phase 4 gegebenenfalls verschmelzen können, sahen in dem strukturellen Ablauf des Prozesses aber eine Absicherung gegen eine geringe Validität von Ergebnissen in Einzelstudien. Nach Glasgow et al. [24] hat eine zu strikte Anwendung dieses Prinzips dazu geführt, dass efficacy-Studien zu Lasten der effectiveness-Studien durchgeführt wurden. Dies habe auch zu Empfehlungen von Interventionen geführt, die unter Praxisbedingungen geringe Erfolgsaussichten hätten. Gerade in der Gesundheitsförderung könne es angemessen sein, anstatt zunächst in efficacy-Studien vereinfachte Interventionen und Interventionsbedingungen zu testen, gleich die relevanten Aspekte einer Intervention zu identifizieren, die für große Populationen unter realistischen Teilnahmebedingungen erforderlich seien. Sie schlagen vor, die Phasen zu verzahnen und moderierende Faktoren in beiden Studientypen gleichermaßen zu berücksichtigen. Zudem sollten Kriterien für die externe Validität in Publikations- und Durchführungsstandards wie z. B. CONSORT [25] berücksichtigt werden. Der sequenzielle Ablauf der Interventionsforschung wird allerdings oft für Studien der Phase 5, also den „Dissemination- and Implementation-Studies“, die in der Regel am Ende eines Evaluationsprozesses stehen, nicht eingehalten [26]. Auch für die Interventionsforschung in der Gesundheitssystem- und Politikanalyse wird eine Abfolge von efficacy- und effectiveness-Studien zwar grundsätzlich für sinnvoll gehalten. Systemänderungen können aber oft nicht auf gestaltbare Interventionsbedingungen beschränkt werden, sodass es keine Möglichkeit für vereinfachte Studien gemäß Phase 3 gibt [27].

Wirkungsnachweis und Studiendesigns

Wie gezeigt, lassen die unterschiedlichen Motive und Ziele im Forschungsprozess sinnvoll zwischen efficacy- und effectiveness-Studien unterscheiden. An verschiedenen Stellen des Forschungsprozesses geht es in beiden Studientypen um die Beurteilung des Zusammenhangs zwischen einer Intervention und einem beobachteten Effekt. Die verzerrungsarme Attribuierung eines Effekts zu einer Intervention muss auch dort angestrebt werden, wo in einer effectiveness-Studie eine hohe externe Validität, also Ähnlichkeit der Studienwirklichkeit zu einer Lebenswirklichkeit, im Vordergrund steht. Im Folgenden sollen daher zunächst die generellen Anforderungen für einen Wirkungsnachweis skizziert und daran anschließend Studiendesigns vorgestellt werden, die für einen Wirkungsnachweis in der Gesundheitsförderung und Prävention geeignet sind und in der Praxis angewendet werden können.

Generelle Anforderungen eines Wirkungsnachweises

Die Ergebnisse einer Intervention lassen sich grundsätzlich nach dem Aussagetypus „Es wurde dies getan und jenes beobachtet“ beschreiben. Ein Wirkungsnachweis, also eine Aussage über den Zusammenhang zwischen der Intervention und einem Effekt, erfordert dagegen Aussagen des Typs: „Es wurde dies getan und deshalb ist jenes geschehen“. Einem Wirkungsnachweis liegt also der Anspruch zugrunde, einen beobachteten Effekt ursächlich auf eine Intervention als kausalen Einfluss zurückzuführen. Dieser Anspruch

muss gegen die Kritik verteidigt werden, dass der Effekt durch andere Einflüsse hervorgerufen wurde oder gar nicht existiert. Kernproblem eines Wirkungsnachweises ist damit die kausale Attribuierung eines beobachteten Effekts zu einer vorangegangenen Intervention. Die damit verbundenen erkenntnistheoretischen Anforderungen unterscheiden sich bei efficacy- und effectiveness-Studien nicht, die Studientypen lassen sich daher auch nicht auf bestimmte Studiendesigns festlegen. So wird eine randomisierte kontrollierte Studie (RCT) zwar oft als typisches Design einer efficacy-Studie gesehen, ist aber auch bei effectiveness-Studien das bevorzugte Studiendesign.

Eine Intervention kann sowohl in einer einzelnen, gut abgrenzbaren Maßnahme bestehen, aber auch ein Bündel von Maßnahmen als Programm oder Politik umfassen. Eine Intervention muss daher nicht zwingend aus benennbaren Einzelmaßnahmen und sich hieraus ergebenden Wirkungsketten bestehen. Sie muss aber abgrenzbar von externen Maßnahmen und Entwicklungen sein, da ansonsten Effekte nicht eindeutig der Intervention zugeschrieben werden können [7]. Die Effekte solcher externen Einflüsse können durch das Untersuchungsdesign kontrolliert werden. Die randomisierte Zuordnung zur Interventions- oder Kontrollgruppe führt zur Vergleichbarkeit beider Gruppen, da sie sich theoretisch nur durch die Intervention unterscheiden. Vor diesem kausal homogenen Hintergrund kann dann ein Effektunterschied auf den Einfluss des einzigen bestehenden Unterschieds, nämlich auf die Intervention, zurückgeführt werden. Randomisierung verhindert somit, dass ein Einflussfaktor die Zuordnung zu den Untersuchungsgruppen beeinflusst haben könnte und damit die Intervention-Effekt-Beziehung nur verzerrt beobachtet werden kann. Hierin besteht der Grund, weshalb eine RCT als Goldstandard einer unverzerrten Studie gilt.

Der Goldstandard ist allerdings sowohl bei efficacy- als auch bei effectiveness-Studien nicht immer erreichbar, da sich ethische, politische oder finanzielle Gründe als Hindernisse erweisen [28]. Gerade bei der Evaluation von weitreichenden politischen Interventionen und in der Implementationsforschung sind solche Hindernisse eine Herausforderung [29]. Eine Vielzahl von Studiendesigns und statistischen Verfahren sind daher als Alternativen zu RCTs entwickelt worden. Alternative Studiendesigns gelten dann als besonders geeignet, wenn sie mögliche Verzerrungen nahezu ähnlich gut kontrollieren wie es mit RCT möglich wäre. Ein besonders verzerrungsarmes Design würde wie eine RCT sowohl die beobachteten als auch die unbeobachteten Faktoren kontrollieren. Das Ausmaß der Verzerrungskontrolle wird daher auch als ein Kriterium für die Hierarchisierung von Studiendesigns gesehen. Verzerrungsarme, randomisierte und nicht randomisierte Studien werden inzwischen auch verbreitet zur Evaluation von politischen Maßnahmen mit Auswirkungen auf große Bevölkerungsgruppen in der Gesundheitsförderung, in Public Health oder der Entwicklungszusammenarbeit angewendet, also in Feldern, die als schwierig zu evaluieren gelten und in denen reine efficacy-Studien begrenzte Aussagekraft hätten [29, 30].

Studiendesigns zur Evaluation von Interventionen

Im Allgemeinen wird zwischen experimentellen, quasi-experimentellen und nicht-experimentellen Studien unterschieden. Die Bezeichnungen der Studiendesigns und Analysestrategien sind aller-

dings nicht einheitlich [31]. Während einige Autoren unter quasi-experimentellen Studien alle Studienansätze mit Kontrollgruppe verstehen, zählen andere hierzu nur Studien, bei denen die Zuordnung zur Interventionsgruppe nicht durch die Untersucher vorgenommen werden kann [32]. Es findet sich sogar ein noch engeres Verständnis, wonach als quasi-experimentell nur Studien bezeichnet werden, die die Kontrolle auch unbeobachteter konfundierender Faktoren („as-if-random“) [33] ermöglichen. Gelegentlich werden allerdings auch Cluster-RCTs zu den quasi-experimentellen Studien gezählt [33] wie auch sogenannte „natural experiments“, worunter manche Autoren wiederum lediglich Studien über natürliche Intervention, wie Erdbeben, verstehen [34, 35].

Die ► **Tab. 1** stellt die Charakteristika und Methoden für (I) experimentelle Studien, (II) quasi-experimentelle Studien und (III) nicht experimentelle Studien zusammen und gibt Anwendungsbeispiele aus Gesundheitsförderung und Public Health. Unter experimentellen Studien werden in den empirischen Wissenschaften prototypisch die RCT sowie die Cluster-RCT verstanden. Durch die in der Verfügung des Untersuchers stehende Randomisierung von Individuen oder Gruppen von Individuen wird theoretisch eine vollständige strukturelle Vergleichbarkeit der Interventions- und Kontrollgruppe erreicht. Dies kann allerdings durch eine Reihe von Gründen gefährdet sein. Probanden könnten z. B. einer Randomisierung widersprechen oder nach erfolgter Randomisierung sich nicht oder nicht durchgehend an der Studie beteiligen. Die Vorteile einer experimentellen Studie wie der RCT hängen daher von etlichen Voraussetzungen ab, die Verzerrungsarmut muss ggf. nachträglich durch besondere Analysetechniken erreicht werden. Bei Cluster-RCT muss zudem eine ausreichende Anzahl an Gruppen zur Verfügung stehen, um tatsächliche Strukturgleichheit erwarten zu können. Cluster-RCT gelten aber als experimentelles Studiendesign, mit dem die Realität von Interventionen in der Zielpopulation gut abgebildet und damit eine hohe externe Validität erreicht werden kann. Die kontrollierten Bedingungen erlauben zudem, die Übertragbarkeit auf andere Setting abzuschätzen.

Quasi-experimentellen Studien (► **Tab. 1.**) kennzeichnen sich durch den Verzicht oder die Unmöglichkeit einer zufälligen Zuordnung von Probanden zu einer Kontrollgruppe. Als besonders geeignete Studiendesigns gelten hier die Regression-Diskontinuität-Analyse (RD), instrumentelle Variablen (IV), Difference-in-Difference (DID)-Modelle und unterbrochene Zeitreihen (ITS). Im Prinzip wird durch Besonderheiten des Studiendesign oder der Analysestrategie eine randomisierte Allokation ersetzt (RD, IV) oder angenähert (DID, ITS). Von einem Ersatz der Randomisierung kann gesprochen werden, wenn das Design eine vollständige Kontrolle konfundierender Faktoren ermöglicht, also sowohl der beobachteten als auch der unbeobachteten.

Bei der Regressions-Diskontinuität-Analyse (RD) handelt es sich um ein in der Ökonomie und Gesundheitssystemanalyse entwickeltes Verfahren mit zunehmender Verbreitung in der Public Health Forschung und Evaluation [36]. Das Verfahren nützt den Umstand, dass Kenngrößen, die für einen Einschluss etwa in Präventionsprogramme verwendet werden (z. B. Einkommen, Gewicht, Bildung) in der Regel unscharf gemessen werden. Individuen, die knapp oberhalb eines Schwellenwertes liegen und damit möglicherweise nicht mehr berechtigt sind, an einer Intervention teilzunehmen, dürften sich daher kaum von Individuen unterscheiden, die gera-

de noch einbezogen wurden. Wird nun ein enges Intervall um einen Schwellenwert gelegt und die Interventionsgruppe in diesem Intervall den Nichtteilnehmern aus diesem Intervall gegenüber gestellt, so entspricht dies quasi einer Randomisierung. RD wird bevorzugt angewendet, wenn eine Randomisierung ethisch problematisch wäre. Im Beispiel aus ► **Tab. 1** wäre die Durchführung einer RCT damit verbunden, eine Gruppe von Schülerinnen von der Schuldgeldbefreiung entgegen bestehenden Rechts auszuschließen. Dem Vorteil, hinsichtlich der Verzerrungskontrolle einem RCT vergleichbar zu sein, steht der Nachteil gegenüber, dass eine größere Anzahl von Probanden, nämlich die außerhalb des Intervalls nicht in die Studien einbezogen wird.

Auch das Verfahren der instrumentellen Variablen (IV) hat seinen Ursprung in der ökonomischen Literatur [37]. Eine Randomisierung der Untersuchungsgruppen kann hier durch die Berücksichtigung einer zusätzlichen, außerhalb der Intervention stehenden Variablen erreicht werden. Vermöge der instrumentellen Variablen wird die Intervention zu einem Mediator und es kann unter bestimmten Voraussetzungen der Effekt der Intervention unverzerrt geschätzt werden. IV ermöglicht wie bei einem RCT die Kontrolle beobachteter und unbeobachteter konfundierender Faktoren und wird daher verbreitet und in sehr verschiedenen Fragestellungen eingesetzt. Die Herausforderung besteht bei diesem Verfahren darin, eine Variable zu finden, die zwar mit der Intervention deutlich assoziiert ist, aber ansonsten von dem unter Beobachtung stehenden Effekt unabhängig ist. Die Voraussetzungen, an denen die Verzerrungsarmut geknüpft sind, gelten daher als teilweise erheblich einschränkend [38].

Das Differenzen-in-Differenzen (DID) Design bildet einen Übergang von Studien mit vollständiger Kontrolle konfundierender Faktoren zu Studien, mit denen lediglich die beobachteten Einflussfaktoren kontrolliert werden können [39]. Die Untersuchungsgruppen werden im Zeitverlauf hinsichtlich der interessierenden Effekte beobachtet. Durch die Betrachtung von Effekt-Differenzen vor und nach der Intervention bei den einzelnen Untersuchungspersonen werden implizit alle zeitkonstanten Merkmale – auch unbeobachtete – dieser Personen kontrolliert. Der Unterschied der Differenzen in der Intervention und Kontrollgruppe kann dann nur noch durch die Intervention selbst und zeitvariante Einflussfaktoren erklärt werden. Letztere müssten durch geeignete Analysetechniken zusätzlich kontrolliert werden.

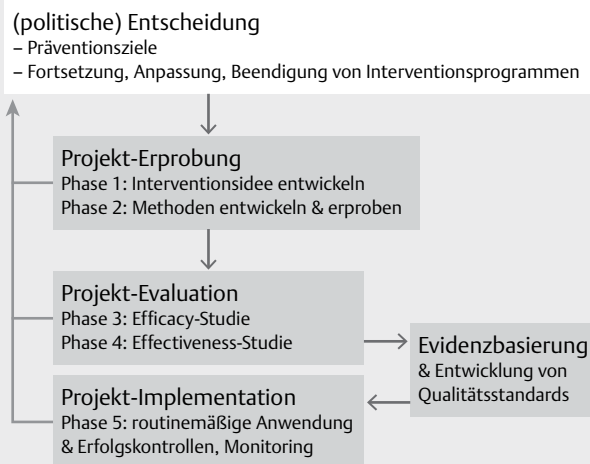
Als quasi-experimentelles Design wird auch die Analyse einer unterbrochenen Zeitreihe (interrupted time series, ITS) verstanden. Obwohl hier keine explizite Kontrollgruppe gebildet wird, besteht der quasi-experimentelle Ansatz darin, für den interessierenden Effekt Daten zu wählen, die längsschnittlich oder wiederholt erhoben sind und somit einen Vergleich über einen längeren Zeitraum vor und nach der Intervention ermöglichen. Die Interventionspopulation vor der Intervention dient daher als eigene Kontrollgruppe. Die Beobachtung des Effekts über einen längeren Zeitraum vor der Intervention lässt annehmen, dass sich der ermittelte Trend ohne Intervention fortgesetzt hätte. Lässt sich eine abrupte Änderung in der Höhe des Effekts oder im Verlauf des Effektes nach der Intervention sehen, so gilt dies als ein deutlicher Hinweis auf die Wirkung der Intervention [40]. Das Studiendesign wird oft verwendet, wenn eine Intervention nachträglich evaluiert werden soll und hierfür auf bestehende Routinedaten zurückgegriffen werden kann. ITS kann auch als nicht experimentelles Studiendesign auf-

► **Tab. 1** Experimentelle, quasi-experimentelle und nicht-experimentelle Studiendesign und Ausmaß der Verzerrungskontrolle.

Studiendesign / Methode	Beschreibung	Verzerrungskontrolle	Beispiel aus Prävention & Gesundheitsförderung
(I) Experimentelle Studien	Zuordnung der Untersuchungspersonen zu Interventions- und Kontrollgruppe erfolgt durch Untersucher	Das Untersuchungsdesign sorgt für eine – bis auf die Intervention - vollständige Vergleichbarkeit der Untersuchungsgruppen	
Randomisiert-Kontrollierte Studie (RCT)	In RCT werden Individuen zufällig auf Interventions- und Kontrollgruppe verteilt. Randomisierung sorgt (ideell) für eine gleichmäßige Verteilung von konfundierenden Faktoren über die Untersuchungsgruppen. RCT können durch weitere Maßnahmen, wie Verblindung von Probanden und/oder Untersucher, ergänzt werden und dadurch weitere Verzerrungsquellen ausgeschlossen werden.	Vollständige Kontrolle beobachteter und unbeobachteter konfundierender Faktoren	Schaller et al. [44] berichten über eine einseitig verblindete RCT mit Patienten aus Reha-Einrichtungen. Als Intervention dient das Programm „movement coach“, dass der üblichen stationären Reha als Kontrolle gegenübergestellt wird. Die Intervention ist auf mehrere Ebenen angelegt (Gespräch/Beratung, Telefon- und Internet Nachbetreuung). Es handelt sich um ein Beispiel einer efficacy-Studie, da sie unter klinischen Bedingungen mit spezifischen Ein- und Ausschlusskriterien durchgeführt wurde. Gleichzeitig kann aber von einer hohen externen Validität ausgegangen werden, da sich die Maßnahmen im Reha-Routinebetrieb (als Zielpopulation) nicht von Studienbedingungen unterscheiden würden.
Cluster RCT	Anstelle von Individuen werden Gruppen zufällig einer Interventions- und Kontrollgruppe zugeordnet. Cluster RCT gelten als besonders angemessenes Design, wenn die Intervention auf Gruppen abzielt (z. B. gesundheitsförderliche Schule) oder Individuen besser auf Gruppenebene angesprochen werden können (z. B. Gewaltprävention bei Schülern)	Vollständige Kontrolle beobachteter und unbeobachteter konfundierender Faktoren	Bühler et al. [45] beschreiben die Cluster-randomisierte Evaluation der gesundheitlichen Aufklärung zur Alkoholprävention bei Jugendlichen „Na-Toll“. Als Intervention fungieren Jugendliche, die bei zufällig ausgewählten Freizeitevents (z. B. Schwimmbad) mit anderen Jugendlichen über Alkohol-Themen sprechen. Peers werden durch Anzeigen und Aushänge an Universitäten gesucht und erhalten einen 3-tägigen Trainingskurs. Es handelt sich also um eine effectiveness-Studie, da Interventionsorte und -anlässe der Lebensrealität von Jugendlichen entsprechen. Die Intervention (trainierte peers) ist nicht anders als unter Routinebedingungen vorgesehen.
(II) Quasi-Experimentelle Studien	Alle kontrollierten Studien ohne Randomisierung	Durch Studiendesign oder durch Analysestrategien wird die Vergleichbarkeit der Untersuchungsgruppen hergestellt	
Regressions-Diskontinuitäts-Analyse (engl.: regression discontinuity)	Anstelle einer Randomisierung wird die Zuordnung zu Interventions- und Kontrollgruppe mithilfe eines Schwellenwertes einer die Intervention charakterisierenden Variable vorgenommen. Individuen knapp oberhalb werden mit denen knapp unterhalb dieses Schwellenwertes verglichen. Es kann angenommen werden, dass die Probanden damit auch hinsichtlich anderer Merkmale vergleichbar sind und alle potentiellen Störvariablen in einem schmalen Band um den Schwellenwert gleichverteilt sind.	Gilt als ein dem RCT ideell identisches Studiendesign. Ist dem RCT dann überlegen, wenn Randomisierung z. B. aus ethischen Gründen nicht erfolgen kann. Kontrolle beobachteter und unbeobachteter konfundierender Faktoren.	Behrmann et al. [46] untersuchen, ob eine längere Schuldauer bei Mädchen zu einer Abnahme der nachfolgend erlebten sexuellen Gewalt führt. Als Schwellenwert verwenden sie ein bestimmtes Lebensalter zum Zeitpunkt der Schulgeldbefreiung in Uganda und Malawi und vergleichen Schülerinnen, die gerade noch die Befreiung erhielten mit denen, die gerade nicht mehr befreit wurden. Es handelt sich um eine effectiveness-Studie, deren Aussagekraft für Schülerinnen außerhalb der untersuchten Altersintervalle möglicherweise eingeschränkt ist.
Instrumentelle Variablen (IV) (engl.: instrumental variables)	Die Randomisierung wird durch die Verwendung einer mit der Intervention assoziierten Variablen ersetzt. Dieses „Instrument“ darf nicht direkt mit dem Effekt assoziiert sein. Dann kann durch eine Zweischritt-Regression der nicht konfundierte Einfluss der Intervention durch die instrumentelle Variable isoliert werden.	Gilt ebenfalls als ein dem RCT ideell identisches Studiendesign. Kontrolle beobachteter und unbeobachteter konfundierender Faktoren.	Basu et al. [29] untersuchen, ob die Teilnahme an einem staatlichen Programm zur finanziellen Unterstützung von Lebensmittelkäufen den Konsum zuckerhaltiger Getränke und damit die Entstehung von Adipositas fördert. Als instrumentelle Variable verwenden sie unterschiedliche Zugangshürden in Staaten der USA. Im ersten Analyseschritt schätzen sie den Zusammenhang dieser Hürden und der Teilnahme an dem Programm und im zweiten Schritt den zwischen der Teilnahme und Zuckerkonsum. Da die Zugangshürden ihrerseits nicht mit dem Konsum direkt assoziiert sind, kann die Wirkung des Programms ohne Störvariablen bestimmt werden.
Differenzen in Differenzen Design (DID) (engl.: difference-in-difference)	Auch als Prä-Post-Design mit Kontrollgruppe oder „non equivalent control group design“ bezeichnet. Betrachtet werden die Differenzen der Prä-Post-Unterschiede in der Interventions- und Kontrollgruppe. Unter der Annahme, dass sich Interventions- und Kontrollgruppe ohne Intervention vergleichbar entwickelt hätten, drückt die Differenz den Einfluss nur der Intervention aus.	Durch die Betrachtung der Differenzen werden alle – auch unbeobachtete – Einflussfaktoren, die sich im Zeitverlauf nicht ändern, implizit kontrolliert. Zeitvariante konfundierende Faktoren werden indes durch das Design nicht kontrolliert und müssen in der Analyse etwa durch Regressions-Adjustment berücksichtigt werden.	Friedrichs et al. [47] evaluieren den ökonomischen Nutzen von Präventionsmaßnahmen. GKV-Versicherte, die an den Bonusprogrammen ihrer Krankenkassen teilnehmen werden mit nicht teilnehmenden Versicherten verglichen. Die unterschiedliche Entwicklung der Leistungsausgaben in der Interventions- und Kontrollgruppe sind Ausdruck der Teilnahme an den präventiven Bonusprogrammen.

► Tab. 1 Fortsetzung.

Studiendesign / Methode	Beschreibung	Verzerrungskontrolle	Beispiel aus Prävention & Gesundheitsförderung
Unterbrochene Zeitreihe (ITS) (engl.: interrupted time series)	Die Untersuchungspopulation wird im Zeitverlauf für interessierende Effekte beobachtet. Es wird angenommen, dass die Intervention sich durch eine Änderung in der Höhe des Effektes oder in einem anderen zeitlichen Verlauf sichtbar wird. Die Untersuchungsgruppe fungiert daher als eigene Kontrollgruppe. Für die Auswertung kann z. B. eine segmentierte Regressionsanalyse für die Zeitpunkte vor und nach der Intervention durchgeführt werden. Es ist daher erforderlich, eine ausreichende Anzahl an Daten im Zeitintervall zu berücksichtigen.	Durch die Betrachtung eines längeren Zeitintervalls ist eine Trendwende leichter erkennbar. Da aber lediglich beobachtete konfundierende Faktoren kontrolliert werden können, ist das Design verzerrungsanfällig.	Sargent et al. [48] untersuchen die Effekte der öffentlichen Rauchverbots in Deutschland. Anhand von Daten einer Krankenkasse wurde die Hospitalisierung für kardiovaskuläre Ereignisse im Zeitraum vor und nach den gesetzlichen Regelungen mithilfe einer ITS beurteilt.
(III) Nicht-Experimentelle Studien	Keine Kontrollgruppe	Keine Verzerrungskontrolle	
Vorher-nachher-Vergleich (engl.: pre-post-design)	Es wird auf eine Kontrollgruppe verzichtet und die Untersuchungspopulation lediglich vor und nach der Intervention verglichen.	Keine Kontrolle von konfundierenden Einflüssen möglich. Der Studientyp geht mit einer geringen Validität einher.	Sickinger et al. [49] beschreiben die Evaluation des Programms „Azubi gscheit und fit“. 51 männliche Auszubildende nahmen an einer Befragung zu Beginn und nach Abschluss des Programms teil. Eine Kontrollgruppe wurde nicht berücksichtigt. Die Studie weist eine geringe Validität auf, es kann weder ausgeschlossen werden, dass die Effekte durch andere Faktoren als das Programm ausgelöst wurden, noch ob überhaupt Effekte eingetreten sind, da eine starke Selbstselektion der Probanden vorlag.



► Abb. 1 Phasen der Präventionsforschung als politischer Rückkopplungszyklus.

gefasst werden. Es können zwar die realen Interventionsbedingungen untersucht werden, allerdings muss wegen der geringen Eingriffsmöglichkeiten auf den Datengenerierungsprozess mit geringer Validität gerechnet werden.

Als Beispiel eines nicht experimentellen Studiendesign ist in ► Tab. 1 noch der einfache Vorher-Nachher-Vergleich aufgeführt. Typischerweise werden hierfür lediglich einmal vor und nach der Intervention Daten in der Interventionsgruppe erhoben. Im Vergleich zu den DID oder ITS eröffnet sich also keine Möglichkeit Einflussfaktoren jenseits der Intervention und weitere Selektionsmechanismen zu kontrollieren. Dieses Studiendesign, das sui-generis nur Ergebnisse mit geringer Validität erzeugen kann, erfreut sich wegen seiner Einfachheit allerdings großer Verbreitung insbesondere auch bei Evaluationen in der Gesundheitsförderung.

Ergänzend zu den in ► Tab. 1 erwähnten Studiendesigns können Analysestrategien zur Verzerrungskontrolle eingesetzt werden. Hierbei handelt es sich um Standard-Analysetechniken, die verbreitet in allen Studiendesigns verwendet werden wie etwa das Adjustieren in Regressionsmodellen, die Stratifizierung von Analysen, das Matching mit oder ohne Propensity Scores. Mit diesen Verfahren können nur beobachtete konfundierende Faktoren kontrolliert werden. In Querschnittsstudien eingesetzt folgt hieraus eine geringe Validität. In Kombination mit einem quasi-experimentellen Studiendesign können sie die Validität erhöhen, da zusätzlich mögliches „Residual-Confounding“ kontrolliert wird.

Systematische Identifikation von Evaluationsbedarfen für die Gesundheitsförderung und Prävention in Lebenswelten

Maßnahmen in Gesundheitsförderung und Prävention zielen auf die Entwicklung gesundheitsförderlicher Strukturen und Kompetenzen sowie auf den Erhalt und die Verbesserung des gesundheit-

lichen Befindens und der Verringerung des Krankheitsgeschehens. Dass der Setting-Ansatz hierfür geeignet ist, wird gesundheitspolitisch meist normativ begründet. Inzwischen wird aber zunehmend eine nach wissenschaftlichen Maßstäben ermittelte Evidenz dafür erwartet, dass diese Ziele mit dem Setting-Ansatz auch erreicht werden. Obwohl wie gezeigt auch für komplexe Evaluationsbedingungen geeignete experimentelle und nicht experimentelle Verfahren vorliegen, können die methodischen Anforderungen einer Wirkungsevaluation nicht in jedem Handlungsfeld im gleichen Umfang erfüllt werden.

Maßnahmen der Gesundheitsförderung und Prävention in Lebenswelten in Deutschland werden oftmals infolge sich im Zeitverlauf ändernder politischer Gestaltungsanreize in der Praxis entwickelt. Entsprechend ist die Evidenzlage derzeit sehr heterogen. Vor dem Hintergrund des obigen Phasenmodells wäre in manchen Settings bei ausreichender Evidenzbasis vordringlich die routinemäßige Implementation der erfolgreichen Maßnahmen erforderlich. In vielen Settings müsste dagegen die Durchführung von efficacy- oder effectiveness-Studien im Fokus stehen, bevor systematische Reviews zur Beurteilung der Evidenz der Maßnahmen in diesem Setting durchgeführt werden könnten. In wiederum anderen Settings schließlich werden vielversprechende Interventionsansätze fehlen, sodass zunächst eine Projekterprobung durchzuführen wäre. Eine Herausforderung besteht mithin darin, die Evaluationsreife eines Settings zu beurteilen, indem Analysebedarfe systematisch erhoben werden und eine Rückkopplung von Evidenzwissen in der Gesundheitsförderung und Prävention sicherzustellen.

Ein solcher Rückkopplungsprozess unter Nutzung des Phasenmodells von Greenwald & Cullen ist in ► **Abb. 1** dargestellt. Ausgangspunkt sind politische Entscheidungen, die in Gesetzen, Programmatiken, Gesundheits- und Präventionszielen festgelegt sein können. Hierbei ist nicht relevant, ob diese Entscheidungen normativ sind oder ihrerseits bereits auf Wissen vorangegangener Interventionsstudien gestützt werden. Die Umsetzung der Ziele beginnt mit einer Projekterprobung, die der Entwicklung von operativen Interventionsansätzen sowie geeigneten Methoden dient. Der Projekterprobung folgt die Projektevaluation, die sowohl efficacy- und effectiveness-Studien als auch die Überprüfung der Prozesse und Strukturen umfasst. Die Evidenzbasierung und Ableitung von Qualitätsstandards setzt eine Anzahl von evaluierten Projekten voraus, geht also über den Einzelfall einer Intervention hinaus. Kennzeichen der Evidenzbasierung sind vorab festgelegte Beurteilungsregeln, mit denen sowohl die Einbeziehung allen verfügbaren Wissens wie auch die möglichst verzerrungsarme Beurteilung dieses Wissens sichergestellt werden soll. Der Evidenzbasierung müsste nicht zwingend eine Abfolge von efficacy- zu effectiveness-Studien vorausgehen (s. o.). Insbesondere bei Interventionen in Settings würde selbst eine vereinfachte efficacy-Studie viele Merkmale einer effectiveness-Studie besitzen, da die Kontexte nicht als konstant vorausgesetzt werden können, sondern sich teilweise erst aus der Intervention ergeben [41]. Aufgabe der Evidenzbasierung ist es, einen möglichst vollständigen Überblick über die Ergebnisse von Interventionsstudien in einem Handlungsfeld zu geben. Die Ergebnissynopse würde dann nach Interventionstypen und Effekten differenzieren und Evidenzlücken sichtbar machen. Als Ergebnis der Evidenzbasierung können schließlich aus den Merkmalen erfolgreicher Interventionen Qualitätsstandards für die

nachfolgende, großflächige Implementation der evidenzbasierten Intervention abgeleitet werden. Dies stellt einen eigenständigen Schritt dar, weil es für die erfolgreiche Umsetzung in die Routine nicht ausreichend ist, dass Interventionen wirksam und Evidenzbasiert sind, sondern es muss auch die Aussicht bestehen, dass relevante Teile der Bevölkerung/Settings erreicht werden. Die routinemäßige Anwendung in Settings unterliegt schließlich einem kontinuierlichen Monitoring, mit dem sichergestellt werden soll, ob die Intervention gemäß den Qualitätsstandards in der Praxis umgesetzt werden kann oder ob die Praxisbedingungen weiteren Forschungsbedarf verdeutlichen. Idealerweise sollte durch die großflächige Anwendung der Intervention das Erreichen der ursprünglich avisierten Präventionsziele unterstützt werden. Die Beurteilung der Zielerreichung schließt daher den Kreislauf und eröffnet ggf. die Diskussion über bessere Interventionsstrategien. Dieser Prozess verfügt über verschiedene Rückkopplungsschleifen, sodass stets von einem sequentiellen Ablauf über alle Stufen begründet abgewichen werden kann.

Übergeordnete Zielsetzung des Rückkopplungsprozess wäre die Sicherstellung des Einklangs von Präventionszielen, wissenschaftlichem Kenntnisstand und Präventionspraxis. Die aufgrund des Präventionsgesetzes bereit stehenden finanziellen Mittel lassen die Organisation eines solchen Rückkopplungsprozesses in Deutschland realisierbar erscheinen und könnten seine Transparenz zudem im Rahmen der Präventionsberichterstattung [42] sichern.

FAZIT

Das Präventionsgesetz bietet die Chance für eine systematische Entwicklung, Evaluation und Evidenzbasierung von Maßnahmen der Gesundheitsförderung und Prävention in Lebenswelten in Deutschland. Der derzeitige Kenntnisstand über die Verbreitung und Wirksamkeit von Maßnahmen ist in den Handlungsfeldern heterogen. Vordringlich ist daher eine systematische Bestandsaufnahme, in welchen Settings welche Evaluationen vorliegen, welche Aussagekraft ihnen zukommt und welche Erkenntnislücken mit Ergebnissen aus welchen Studientypen zu schließen sind.

Anspruch einer Wirkungsevaluation von Maßnahmen der Gesundheitsförderung und Prävention in Lebenswelten muss eine verzerrungsarme Attribuierung eines Effektes zu einer Intervention sein. Hierfür stehen experimentelle und nicht experimentelle Verfahren zur Verfügung, die zu Ergebnissen mit hoher Validität führen. Auch in schwierigen Handlungsfeldern müsste die Evidenzbasierung daher nicht auf Ergebnisse aus schwächeren Studiendesigns zurückgreifen.

Die Praxis der Gesundheitsförderung und Prävention in Lebenswelten ließe sich entsprechend mittelfristig an folgenden Fragen ausrichten:

- Welche Maßnahmen gelten als Evidenzbasiert und werden diese in der Praxis umgesetzt?
- Welche Maßnahmen werden in der Praxis durchgeführt und gelten diese als Evidenzbasiert?

- Welche Maßnahmen gelten als nicht Evidenz basiert und werden diese in der Praxis unterlassen?

Die Beantwortung dieser Fragen würde aufzeigen, in welchem Maße Evidenz basierte Maßnahmen die Präventionspraxis durchdrungen haben. Der damit offengelegte Bedarf an Implementierung und De-Implementierung dürfte allerdings nicht automatisch zu entsprechendem politischen Handeln führen. Einerseits ist aus der Implementations- und Wissenstranslationsforschung bekannt, dass Umsetzungsdefizite nicht ausschließlich aus einem mangelhaften Wissenstransfer folgen, sondern dass z. B. Akteurs-Interessen, Meinungsführerschaften oder Akzeptanz bei Zielgruppen weitere Barrieren darstellen [43]. Andererseits ist Wissen um die Evidenz von Maßnahmen nicht die einzige Grundlage für Entscheidungen in der Gesundheitsförderung und Prävention. So wären etwa Kostenerwägungen, geringe Verbreitungserwartung oder auch eine Priorisierung von politisch bevorzugten Maßnahmen weitere Rationale. Konzepte für die Identifikation von Umsetzungs-Barrieren für Settingprojekte zu entwickeln und Ansätze für die Integration von Evidenzwissen in komplexen Multi-Akteursprozessen zu finden, sind daher weitere Aufgaben der Präventionsforschung.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

- [1] Meierjürgen R, Becker S, Warnke A. Die Entwicklung der Präventionsgesetzgebung in Deutschland. *Prävent Gesundheitsförderung* 2016; 11: 206–213
- [2] GKV Spitzenverband. *Präventionsbericht 2017. Leistungen der gesetzlichen Krankenversicherung: Primärprävention und Gesundheitsförderung Berichtsjahr. 2016.* https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/praevention__selbsthilfe__beratung/praevention/praeventionsbericht/2017_GKV_MDS_Praeventionsbericht.pdf Letzter Zugriff 20.11.2018
- [3] Bödeker W, Moebus S. Normen- und Anreizkonflikte für die gesetzlichen Krankenkassen in Gesundheitsförderung und Prävention. *Gesundheitswesen* 2015; 77: 397–404
- [4] Engelmann F, Halkow A. Der Setting-Ansatz in der Gesundheitsförderung: Genealogie, Konzeption, Praxis, Evidenzbasierung. WZB Discussion Paper 2008
- [5] Dadaczynski K, Baumgarten K, Hartmann T. Settingbasierte Gesundheitsförderung und Prävention: Kritische Würdigung und Herausforderungen an die Weiterentwicklung eines prominenten Ansatzes. *Prävent Gesundheitsförderung* 2016; 11: 214–221
- [6] GKV-Spitzenverband. *Leitfaden Prävention.o.j.* https://www.gkv-spitzenverband.de/krankenversicherung/praevention_selbsthilfe__beratung/praevention_und_bgf/leitfaden_praevention/leitfaden_praevention.jsp Letzter Zugriff 20.11.2018
- [7] Bödeker W. Wirkungen und Wirkungsnachweis bei komplexen Interventionen. In: RKI, Hrsg. *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme?* Berlin: 2012: 33–42. http://www.gbe-bund.de/pdf/evaluation_praevention.pdf Letzter Zugriff 20.11.18
- [8] Bödeker W. Evidenzbasierung in Gesundheitsförderung und Prävention – Der Wunsch nach Legitimation und das Problem der Nachweisstrenge. *Prävent Extra* 2007 3/2007. <https://docplayer.org/60660468-Evidenzbasierung-in-gesundheitsfoerderung-und-praevention-der-wunsch-nach-legitimation-und-das-problem-der-nachweisstrenge.html> Letzter Zugriff 20.11.2018
- [9] Elkeles T, Broesskamp-Stone U. Evidenzbasierte Gesundheitsförderung. BZgA Leitbegriffe Gesundheitsförderung 2015. <https://www.leitbegriffe.bzga.de/alphabetisches-verzeichnis/evidenzbasierte-gesundheitsfoerderung/> letzter Zugriff 20.11.2018
- [10] Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med* 1986; 15: 451–474
- [11] Gerhardus A, Rehfuess E, Zeeb H. Evidenzbasierte Verhältnisprävention und Gesundheitsförderung: Welche Studiendesigns brauchen wir? *Z Für Evidenz Fortbild Qual Im Gesundheitswesen* 2015; 109: 40–45
- [12] Green LW, Glasgow RE. Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology. *Eval Health Prof* 2006; 29: 126–153
- [13] Geldsetzer P, Fawzi W. Quasi-experimental study designs series—paper 2: complementary approaches to advancing global health knowledge. *J Clin Epidemiol* 2017; 89: 12–16
- [14] Windeler J. Externe Validität. *Z Für Evidenz Fortbild Qual Im Gesundheitswesen* 2008; 102: 253–259
- [15] Windeler J. There is no reverse hierarchy. (letter commenting: *J Clin Epidemiol* 2015; 68(11): 1251–1260) *J Clin Epidemiol* 2016; 75: 127–128
- [16] De Bock F, Fischer JE, Hoffmann K et al. A participatory parent-focused intervention promoting physical activity in preschools: design of a cluster-randomized trial. *BMC Public Health* 2010; 10: 49
- [17] De Bock F, Breitenstein L, Fischer JE. Positive impact of a pre-school-based nutritional intervention on children’s fruit and vegetable intake: results of a cluster-randomized trial. *Public Health Nutr* 2012; 15: 466–475
- [18] Shadish W, Cook T, Campbell D. *Experimental and Quasi-Experimental Designs.* Belmont: Wadsworth CENGAGE Learning 2002
- [19] Burchett HED, Blanchard L, Kneale D et al. Assessing the applicability of public health intervention evaluations from one setting to another: a methodological study of the usability and usefulness of assessment tools and frameworks. *Health Res Policy Syst* 2018; 16. doi:10.1186/s12961-018-0364-3
- [20] Schloemer T, Schröder-Bäck P. Criteria for evaluating transferability of health interventions: a systematic review and thematic synthesis. *Implement Sci* 2018; 13. doi:10.1186/s13012-018-0751-8
- [21] Pfadenhauer LM, Gerhardus A, Mozygemba K et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implement Sci* 2017; 12. doi:10.1186/s13012-017-0552-5
- [22] Flay BR, Biglan A, Boruch RF et al. Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prev Sci* 2005; 6: 151–175
- [23] Greenwald P, Cullen J. The new emphasis in cancer control. *J Natl Cancer Inst* 1985; 74: 543–551
- [24] Glasgow RE, Lichtenstein E, Marcus AC. Why don’t we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health* 2003; 93: 1261–1267

- [25] The CONSORT Statement. 2010. <http://www.consort-statement.org/> Letzter Zugriff 20.11.2018
- [26] Brown CH, Curran G, Palinkas LA et al. An Overview of Research and Evaluation Designs for Dissemination and Implementation. *Annu Rev Public Health* 2017; 38: 1–22
- [27] Bärnighausen T, Röttingen J-A, Rockers P et al. Quasi-experimental study designs series—paper 1: introduction: two historical lineages. *J Clin Epidemiol* 2017; 89: 4–11
- [28] Bothwell LE, Greene JA, Podolsky SH et al. Assessing the gold standard – lessons from the history of RCTs. *N Engl J Med* 2016; 374: 2175–2181
- [29] Basu S, Meghani A, Siddiqi A. Evaluating the health impact of large-scale public policy changes: Classical and novel approaches. *Annu Rev Public Health* 2017; 38: 351–370
- [30] Cameron DB, Mishra A, Brown AN. The growth of impact evaluation for international development: How much have we learned? *J Dev Eff* 2016; 8: 1–21
- [31] Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series – paper 5: a checklist for classifying studies evaluating the effects on health interventions – a taxonomy without labels. *J Clin Epidemiol* 2017; 89: 30–42
- [32] Waddington H, Aloe AM, Becker BJ et al. Quasi-experimental study designs series – paper 6: risk of bias assessment. *J Clin Epidemiol* 2017; 89: 43–52
- [33] Bärnighausen T, Tugwell P, Röttingen J-A et al. Quasi-experimental study designs series – paper 4: uses and value. *J Clin Epidemiol* 2017; 89: 21–29
- [34] Craig P, Katikireddi SV, Leyland A et al. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. *Annu Rev Public Health* 2017; 38: 39–56
- [35] Dawson A, Sim J. The nature and ethics of natural experiments. *Journal of Medical Ethics* 2015; 10: 848–853
- [36] Bor J, Moscoe E, Mutevedzi P et al. Regression Discontinuity Designs in Epidemiology: Causal Inference Without Randomized Trials. *Epidemiology* 2014; 25: 729–737
- [37] Martens EP, Pestman WR, de Boer A et al. Instrumental Variables: Application and Limitations. *Epidemiology* 2006; 17: 260–267
- [38] Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29: 722–729
- [39] Wing C, Simon K, Bello-Gomez R. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. *Annu Rev Public Health* 2018; 39: 453–469
- [40] Bernal LJ, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol* 2016; dyw098
- [41] Wildner M. Gesundheit im Kontext. *Gesundheitswesen* 2018; 80: 509–510
- [42] Starke D, Arnold L, Fertmann R et al. Methodische Herausforderungen der Präventionsberichterstattung. *Gesundheitswesen* 2018; 80: 732–740
- [43] Vollmar HC, Santos S, de Jong A et al. Wie gelangt Wissen in die Versorgung?: Implementierungsforschung und Wissenszirkulation. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 2017; 60: 1139–1146
- [44] Schaller A, Petrowski K, Pfoertner T-K et al. Effectiveness of a theory-based multicomponent intervention (Movement Coaching) on the promotion of total and domain-specific physical activity: a randomised controlled trial in low back pain patients. *BMC Musculoskelet Disord* 2017; 18. doi:10.1186/s12891-017-1788-6
- [45] Bühler A, Thurl J, Strüber E et al. Cluster-randomized trial of a German leisure-based alcohol peer education measure. *Health Promot Int* 2016; 31: 385–395
- [46] Behrman JA, Peterman A, Palermo T. Does Keeping Adolescent Girls in School Protect Against Sexual Violence? Quasi-Experimental Evidence From East and Southern Africa. *J Adolesc Health* 2017; 60: 184–190
- [47] Friedrichs M, Friedel H, Bödeker W. Teilnehmerstruktur und ökonomischer Nutzen präventiver Bonusprogramme in der betrieblichen Krankenversicherung. *Gesundheitswesen* 2009; 71: 623–627
- [48] Sargent JD, Demidenko E, Malenka DJ et al. Smoking restrictions and hospitalization for acute coronary events in Germany. *Clin Res Cardiol* 2012; 101: 227–235
- [49] Sickinger S, Carlsohn A, Schleicher K et al. Gesundheitsförderung bei Auszubildenden: Evaluation eines betrieblichen Gesundheitsförderungsprojekts mit den Themen Ernährung und Bewegung. *Prävent Gesundheitsförderung* 2017, doi:10.1007/s11553-017-0628-8