

H. Pfaff¹, U.S. Albert², R. Bornemann³, N. Ernstmann¹, J. Gostomzyk⁴, M.G. Gottwik⁵, G. Heller⁶, U. Höhmann⁷, U. Karbach¹, O. Ommen¹, M. Wirtz⁸

Methoden für die organisationsbezogene Versorgungsforschung*

Methods for organisational health services research

Das Deutsche Netzwerk Versorgungsforschung e. V. (DNVF e.V.) hat am 1. Juli 2009 – getragen von den genannten, im DNVF organisierten Fachgesellschaften und Organisationen – ein Memorandum III „Methoden für die Versorgungsforschung“ verabschiedet, das in der Zeitschrift *Gesundheitswesen* 2009; 71: 505–510 und in der *Dtsch Zahnärztl Z* 2010; 65: 360–366 publiziert wurde. Die vorliegende Publikation stellt eine Vertiefung des Memorandums III „Methoden für die Versorgungsforschung“ dar. Im Mittelpunkt stehen die Fragestellungen und Methoden der organisationsbezogenen Versorgungsforschung. Im ersten Schritt werden die zentralen Fragestellungen der organisationalen Versorgungsforschung dargestellt. Im zweiten Schritt wird auf die methodischen Grundlagen und die Mindeststandards in Bezug auf (a) die Auswahl der Untersuchungseinheiten, (b) die Messung der Merkmale und (c) die Auswahl des Untersuchungsdesigns eingegangen. Wir stellen ein Phasenmodell der Durchführung komplexer Interventionsstudien vor, mit dem man eine besonders hohe Studienqualität in der organisationalen Versorgungsforschung erreichen kann. Dieses Modell integriert hierbei verschiedene Methoden der empirischen Sozialforschung und zeigt, in welcher Phase des Forschungsprozesses die einzelnen Methoden ihre spezifische Stärke haben.

Schlüsselwörter: Versorgungsforschung, Methoden, empirische Sozialforschung, Organisationen

On 1 July 2009, the German Network for Health Services Research (Deutsches Netzwerk Versorgungsforschung e.V. [DNVF e.V.]) approved the Memorandum III “Methods for Health Services Research”, supported by the member societies mentioned below and published in the journal *Gesundheitswesen* 2009; 71: 505–510 and in the journal *Dtsch Zahnärztl Z* 2010; 65: 360–366. The focus of this part of the Memorandum III “Methods for health services research” is on the questions and methods of organisational health services research. In a first step, we describe the central questions which are at the core of organisational health services research. In a second step, we describe the methodological standards and requirements with regard to a) sampling, b) measurement and c) research design. We present a phase model for complex intervention trials. This model allows to conduct high quality organisational health services research, to integrate different methods of social research and to show in which phase they are of special importance.

Keywords: health services research, methods, empirical social research, organisation

¹ Zentrum für Versorgungsforschung, c/o IMVR – Institut für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaft der Humanwissenschaftlichen Fakultät und der Medizinischen Fakultät der Universität zu Köln (KöR)

² Klinik für Gynäkologie, gynäkologische Endokrinologie und Onkologie, Universitätsklinikum Giessen und Marburg GmbH

³ Fakultät für Gesundheitswissenschaften, Universität Bielefeld

⁴ ltd. Medizinaldirektor a. D., Augsburg

⁵ Professor em. für Innere Medizin / Kardiologie, Nürnberg

⁶ Wissenschaftliches Institut der AOK (WIdO), Berlin

⁷ Evangelische Fachhochschule Darmstadt

⁸ Institut für Psychologie der Pädagogischen Hochschule Freiburg

* Ein vertiefender Beitrag zum Memorandum III „Methoden für die Versorgungsforschung“, Teil 1 (*Gesundheitswesen* 2009; 71: 505 – 510)

Das Memorandum III wird von folgenden Mitgliedsgesellschaften des Deutschen Netzwerks Versorgungsforschung e.V. getragen:

- Aktionsbündnis Patientensicherheit
- Berufsverband der Fachärzte für Orthopädie und Unfallchirurgie
- Kompetenzzentrum Versorgungsforschung in der Dermatologie
- Deutsche Dermatologische Gesellschaft
- Deutsche Diabetes Gesellschaft
- Deutsche Gesellschaft für Allergologie und klinische Immunologie
- Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin
- Deutsche Gesellschaft für Angiologie
- Deutsche Gesellschaft für Chirurgie
- Deutsche Gesellschaft für Epidemiologie
- Gesellschaft für Gefäßmedizin
- Deutsche Gesellschaft für Gynäkologie und Geburtshilfe

- Deutsche Gesellschaft für Innere Medizin
- Deutsche Gesellschaft für Kardiologie
- Deutsche Gesellschaft für Medizinische Informatik, Biometrie u. Epidemiologie
- Deutsche Gesellschaft für Medizinische Psychologie
- Deutsche Gesellschaft für Medizinische Soziologie
- Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie
- Deutsche Gesellschaft für Prävention u. Rehabilitation v. Herz-Kreislaufkrankungen
- Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde
- Deutsche Gesellschaft für Psychoanalyse, Psychotherapie, Psychosomatik und Tiefenpsychologie
- Deutsche Gesellschaft für Psychosomatische Medizin und ärztliche Psychotherapie
- Deutsche Gesellschaft für Rehabilitationswissenschaften
- Deutsche Gesellschaft für Senologie
- Deutsche Gesellschaft für Sozialmedizin und Prävention
- Deutsche Gesellschaft für Unfallchirurgie
- Deutsche Gesellschaft für Zahn-, Mund- und Kieferheilkunde
- Deutsche Hochdruckliga/ Deutsche Hypertonie Gesellschaft
- Deutsche Krebsgesellschaft
- Deutsche Psychotherapeutenvereinigung
- Deutsche Schlaganfall-Gesellschaft
- Deutscher Verband für Gesundheitswissenschaften und Public Health
- Deutsches Kollegium für Psychosomatische Medizin
- Forschungsverbund Public Health, Sachsen-Sachsen Anhalt
- Gesellschaft f. Arzneimittelanwendungsforschung u. Arzneimitteltelepidemiologie
- Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung
- Wissenschaftliches Institut der Niedergelassenen Hämatologen und Onkologen GmbH

Einleitung

Die methodische Qualität von Versorgungsforschungsstudien wird häufig kritisiert. Einheitliche Standards auf dem Gebiet der Versorgungsforschung fehlen bisher. Das Deutsche Netzwerk Versorgungsforschung e.V. (DNVF e.V.) sah sich deshalb aufgefordert, hier eine Hilfestellung zu geben und hat am 01.07.2009, getragen von den genannten, im DNVF e.V. organisierten Fachgesellschaften und Organisationen, das Memorandum III „Methoden für die Versorgungsforschung“ verabschiedet und in der Zeitschrift Gesundheitswesen 2009; 71: 505–510 und der Dtsch Zahnärztl Z 2010; 65: 360–366 publiziert. Das Memorandum III „Methoden für die Versorgungsforschung“ Teil 1 stellt die methodischen Grundprinzipien und Mindeststandards, die bei der Durchführung und Veröffentlichung von Versorgungsforschungsstudien beachtet werden sollten, in drei Themenbereichen „Epidemiologische Methoden für die Versorgungsforschung“, „Methoden für die organisationsbezogene Versorgungsforschung“ und „Methoden für die Lebensqualitätsforschung“ dar. Diese Themenbereiche werden in separaten Beiträgen vertieft. Das Memorandum wendet sich sowohl an die Versorgungsforscher, die Studien planen, durchführen und veröffentlichen, als auch an Gutachter, die Anträge und Publikationen zu Versorgungsforschungsstudien bewerten. Entsprechend dem Erkenntnisfortschritt in der Versorgungsforschung ist vorgesehen, das Memorandum in sinnvollen Zeitabschnitten regelmäßig zu aktualisieren und einen weiteren Teil herauszubringen. Das Memorandum ist daher als „Work in Progress“ anzusehen. Der vorliegende Beitrag stellt eine Vertiefung zu dem Themenbereich „Methoden der organisationsbezogenen Versorgungsforschung“ dar.

Organisationen wie Kliniken und Arztpraxen stellen wesentliche Grundelemente des Gesundheitswesens in Deutschland dar. Das Netzwerk dieser Organisationen bildet das „Grundgerüst“ des Gesundheitssystems. Die empirische Untersuchung dieser Organisationen und ihrer Netzwerke ist daher eine Kernaufgabe der Versorgungsforschung. Zur Bewältigung dieser Aufgabe kann die Versorgungsforschung auf bereits etablierte Methoden der empirischen Sozial- und Organisationsfor-

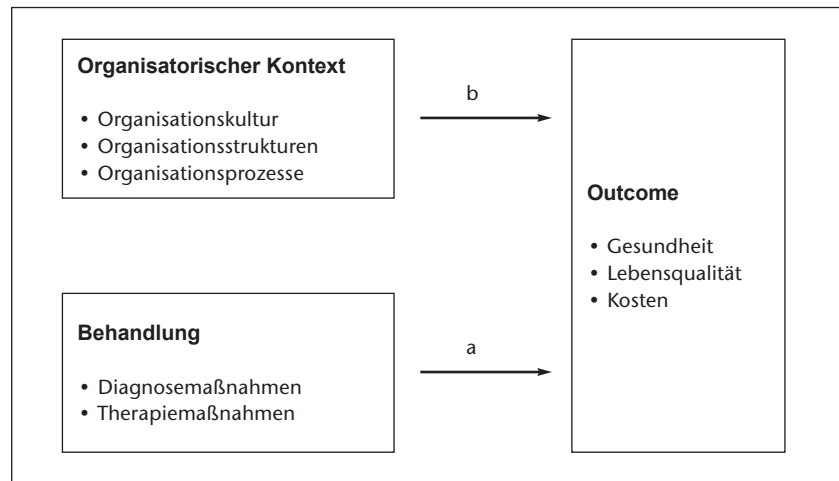


Abbildung 1 Behandlung und organisatorischer Kontext als Determinanten des Versorgungsergebnisses.

schung zurückgreifen. Diese bedürfen einer sach- und fachgerechten Anwendung. In diesem Kapitel werden die wichtigsten methodischen Grundlagen für (a) die Auswahl der Untersuchungseinheiten, (b) die Messung von Merkmalen und (c) die Planung und Darstellung von Studiendesigns in der Organisationsforschung thematisiert. Für diese Themenbereiche werden die Methoden in ihren Grundzügen beschrieben, wichtige Mindeststandards dargestellt und Kriterien für eine möglichst hohe Studienqualität formuliert.

Organisationen als Gegenstand der Versorgungsforschung

Gegenstand der organisationsbezogenen Versorgungsforschung sind alle Organisationen im Gesundheitssystem. Sie werden im Folgenden als Gesundheitsorganisationen bezeichnet. Gesundheitsorganisationen können aufgeteilt werden in Versorgungsorganisationen und Nicht-Versorgungsorganisationen. Unter Versorgungsorganisationen verstehen wir Organisationen, die das primäre Ziel verfolgen, die Kranken- und/oder Gesundheitsversorgung mit professionellen Mitteln durchzuführen. Versorgungsorganisationen erbringen direkte medizinische Versorgungsleistungen. Beispiele dafür sind Krankenhäuser, Arztpraxen, Rehabilitationskliniken, ambulante Pflegedienste und psychoonkologische Dienste.

Als Nicht-Versorgungsorganisationen bezeichnen wir Organisationen des Versorgungssystems, die keine direkten Gesundheitsdienstleistungen erbringen. Sie ermöglichen, steuern oder unterstützen die Erbringung dieser direkten patientenbezogenen Gesundheitsdienstleistungen. Beispiele hierfür sind Ärztekammern, Kassenärztliche Vereinigungen, Krankenkassen und Selbsthilfeorganisationen.

Die Aufgaben der organisationsbezogenen Versorgungsforschung bestehen darin,

- 1) die jeweiligen personellen, sozialen, materiellen und versorgungsbezogenen Strukturen und Prozesse in den Gesundheitsorganisationen zu beschreiben und zu erklären,
- 2) deren Auswirkungen auf Inanspruchnahme, Qualität und Kosten der Kranken- und Gesundheitsversorgung zu untersuchen,
- 3) Konzepte zur Optimierung der Gesundheitsorganisationen (Organisations- und Versorgungsinnovationen) empirie- und theorieorientiert zu entwickeln,
- 4) die Implementierung von Organisations- und Versorgungsinnovationen begleitend zu erforschen und die Qualität der Implementierungstechniken zu überprüfen,
- 5) die Wirkung der Organisations- und Versorgungsinnovationen und der Implementierungstechniken auf personen-, organisations- und gesellschaftsbezogene Parameter (z. B.

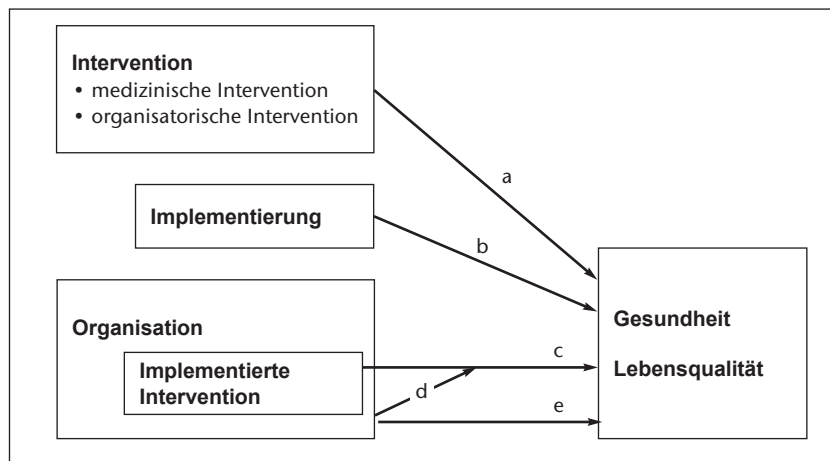


Abbildung 2 Wirkkomponenten der Versorgungssituation.

Mortalität, Lebensqualität, organisationale Leistung, soziale Gerechtigkeit, Kosten-Nutzen) zu erfassen und 6) die Implementierungs- und Programmqualität sowie die Organisations- und Versorgungsinnovationen langfristig weiterzuentwickeln und zu optimieren.

Zur Verfolgung dieser Aufgaben werden je nach Fragestellung Individuen (z. B. Patienten, Ärzte), Gruppen (z. B. Stionsteams), Organisationen (z. B. Kliniken) oder Organisationsnetzwerke (z. B. integriertes Versorgungsnetz) untersucht. Es gibt unter den vielen Aufgabenstellungen, die für die organisationsbezogene Versorgungsforschung kennzeichnend sind, einige von zentraler Bedeutung. Auf diese soll kurz eingegangen werden.

Eine wichtige Aufgabenstellung der organisationalen Versorgungsforschung ist die Untersuchung der Wirkung von Versorgungsorganisationen auf Mortalität und Morbidität der behandelten Patienten. Die Versorgungssituation setzt sich zusammen aus dem konkreten Behandlungsprogramm und dem organisatorischen Kontext, in dessen Rahmen das Behandlungsprogramm zur Anwendung kommt (z. B. Arztpraxis, Klinik). Eine zentrale Frage der organisationsbezogenen Versorgungsforschung ist, ob sich neben dem Behandlungsprogramm (Abb. 1, Pfeil a) auch der organisatorische Kontext der Behandlung auf die Gesundheit und die Lebensqualität der Patienten auswirkt (Abb. 1, Pfeil b).

Eine weitere zentrale Aufgabe der organisationsbezogenen Versorgungsforschung ist die Erfassung der Wirkung der Einführung einfacher und komplexer Interventionen in Versorgungsorganisationen. Bei der Einführung von Versorgungs- und Organisationsinnovationen sind mehrere Wirkkomponenten zu unterscheiden (Abb. 2): (a) die Wirkung der Intervention an sich (Pfeil a), (b) die Wirkung der Implementierung der Intervention (Pfeil b), (c) die Wirkung einer an die Organisationsrealitäten angepassten Intervention (Pfeil c), (d) der moderierende Einfluss der Organisationsmerkmale auf die Wirkung der Intervention im Sinne einer Abschwächung oder Verstärkung dieser Wirkung (Pfeil d) und (e) der direkte Einfluss der Organisation auf die Gesundheit und Lebensqualität der Patienten (Pfeil e).

Evaluationsstudien können somit unterschiedliche Zielsetzungen verfolgen, die z. B. die Identifikation von Veränderungsbedarf, die Entwicklung von Interventionsansätzen, den Nachweis der Effektivität oder Effizienz der Intervention oder die praktische Implementierung betreffen. Für Studien, die die Wirksamkeit von Organisationsinterventionen messen sollen, ist zu beachten, dass die Art des Studiendesigns erheblich davon abhängt, welche Form der Intervention gegeben ist. Die Intervention kann an der Person (z. B. Patient, Arzt) oder an der Organisation (z. B. Abteilung, Klinik) ansetzen. Im ersten Fall sprechen wir von personenbezo-

genen Interventionen, im zweiten Fall von organisationsbezogenen Interventionen. Bei den personenbezogenen Interventionen kann zusätzlich unterschieden werden zwischen verblindbaren und nicht-verblindbaren Interventionen⁹. Bei nicht-verblindbaren Interventionen sind randomisierte kontrollierte Studien (RCT) auf Ebene der Einzelpatienten in Organisationen ggf. kritisch, da z. B. „Verunreinigungen“ (contaminations) zwischen der Kontroll- und der Interventionsgruppe möglich sind. Können solche Verunreinigungen nicht ausgeschlossen werden, gilt als Goldstandard die Durchführung von clusterrandomisierten Studien (cluster randomized trials: CRT)¹⁰. Handelt es sich um organisationsbezogene Interventionen, also um Interventionen, die auf der Ebene der Organisation ansetzen (z. B. Abteilung, OP-Team oder Klinik), ist für Studien mit dem Ziel des Wirksamkeitsnachweises ebenfalls die Clusterrandomisierung der Goldstandard. Der Aufwand an organisationsübergreifender Standardisierung, Koordination und Konsentierung ist bei CRTs wesentlich höher als bei RCTs. Deshalb wird später auf diese Besonderheiten gezielt eingegangen.

Eine dritte wesentliche Aufgabe der organisationsbezogenen Versorgungsforschung besteht darin, die Qualität der Implementierungstechniken und -verfahren durch gezielte Evaluation systematisch zu verbessern. Dies ist Gegenstand der organisationsbezogenen Implementierungsforschung. Sie beschäftigt sich mit der Frage, welche Methoden am besten dafür geeignet sind, Maßnahmen und Programme in einer Versorgungsorganisation (z. B. Arztpraxis, Klinikum) erfolgreich zu implementieren. Auch für diese Form der Evaluation stellt das CRT-Studiendesign den Goldstandard dar. In der Praxis ist diese Form der Evaluation nur mit hohem Aufwand durchzuführen, da eine sehr große Zahl an Kliniken oder Arztpraxen dafür gewonnen werden muss, eine bestimmte, ihnen vorgegebene Intervention in einer ganz bestimmten Art und Weise einzuführen und diese Art der Einführung standardisiert anzuwenden.

Die skizzierten Aufgabenstellungen der organisationalen Versorgungsfors-

⁹ Eine Verblindung auf Ebene der Auswerter sollte grundsätzlich erfolge, um verzerrte Effektivitätsschätzungen zu vermeiden.

¹⁰ CRTs (Randomisierte Elemente: organisationale Einheiten) stellen einen Spezialfall von RCTs (Randomisierte Elemente: z. B. Patienten) dar.

Gegenstand der Versorgungsforschung	Menschlicher Körper	Medizintechnik	Denken, Fühlen, Verhalten und subjektives Befinden der Akteure im Versorgungssystem	Wechselseitiges Verhalten der Akteure (Individuen und Organisationen) im Versorgungssystem
Systemtyp	Menschlicher Organismus	Technisches System	Psychisches System	Soziales System
Datenart	Biomedizinische Daten	Technische Daten	Psychosoziale Daten	Sozioökonomische Daten
Datengewinnung	Methoden der klinischen Forschung und der Biomedizin	Methoden der Ingenieurwissenschaften	Methoden der empirischen Sozialforschung	
	Methoden der Epidemiologie			

Tabelle 1 Methoden der Datengewinnung in der Versorgungsforschung [47].

schung können mit verschiedenen Daten und Methoden bearbeitet werden. Bei den Daten unterscheiden wir biomedizinische, technische, psychosoziale und sozioökonomische Daten (vgl. Tab. 1) und bei den Methoden qualitative und quantitative Methoden der Sozialforschung. Eine umfassende Darstellung der verschiedenen Methoden der Datengewinnung kann im Rahmen dieses Memorandums nicht geleistet werden. Der interessierte Leser sei daher auf die entsprechende Übersichtsliteratur verwiesen [26, 37, 46].

Bei der Planung von Versorgungsforschungsstudien müssen Entscheidungen bezüglich der Stichprobe, der Operationalisierung des Forschungsgegenstands, der Messung der Variablen und des Designs getroffen werden. Für alle drei Bereiche stehen unterschiedliche Methoden mit unterschiedlichen Standards zur Verfügung. Auf diese soll im Folgenden eingegangen werden. Dabei geht es vor allem darum, die Kriterien für eine möglichst hohe Studienqualität zu definieren. Generell gilt dabei in der Versorgungsforschung das Prinzip der Angemessenheit der Methoden. Es gibt nicht eine richtige Methode für alle Fragestellungen, vielmehr gibt es in Abhängigkeit von der Fragestellung mehr oder weniger angemessene Methoden. Eine Methode, die bei einer bestimmten Fragestellung den Goldstandard darstellt, z. B. randomisiert-kontrollierte Studien (RCTs) bei personenbezogenen Effectiveness-Untersuchungen, kann bei einer anderen Fragestellung, z. B. formative Evaluation einer Intervention, unangemessen sein.

Im Folgenden wird auf die Stichprobenziehung, die Messung und das Studiendesign im Rahmen quantitativer Studien der Organisationsforschung eingegangen. Ziel ist es dabei, einen orientierenden Überblick über diese Themen zu geben. Somit wird auf eine detaillierte Beschreibung weiterer organisationsbezogener Methoden wie z. B. Risikoadjustierung und Mehrebenenanalyse verzichtet. Ebenfalls nicht betrachtet werden qualitative Verfahren in der organisationsbezogenen Versorgungsforschung. Dies bedeutet jedoch nicht, dass qualitative Verfahren keinen Platz in der organisationalen Versorgungsforschung haben [37]. Da jedoch für qualitative Ansätze andere Prinzipien, Regeln und Methodenstandards gelten, würde deren Darstellung den Rahmen dieses Teils des Memorandums III sprengen. Für eine Übersicht der relevanten Prinzipien und Regeln sei auf die einschlägige Literatur verwiesen [26, 27, 41, 52].

Standards bei der Auswahl der Untersuchungseinheiten (Stichprobenziehung)

In der Organisationsforschung gibt es zwei Typen von Untersuchungseinheiten: 1) Personen und 2) Sozialsysteme. Bei den Personen handelt es sich in der Regel entweder um Mitarbeiter (z. B. angestellte Ärzte, Pflegekräfte) oder um Kunden der Organisation (z. B. Patienten, einweisende Ärzte). Bei den zu untersuchenden Sozialsystemen kann es sich um Arbeitsgruppen, Stationen, Abteilungen oder ganze Versorgungsorganisationen (z. B. Krankenhaus, Arztpra-

xis) handeln. Nur bei kleinen Grundgesamtheiten (z. B. alle Kliniken eines Bundeslandes) lassen sich alle Elemente im Rahmen einer Totalerhebung einbeziehen. Häufiger wird aus Kostens- und/oder Praktikabilitätsgründen nur ein Teil dieser Einheiten (Stichprobenerhebung) oder im Extremfall nur eine einzige Einheit (Fallanalyse) untersucht.

Die Wahl der Stichprobenart zielt normalerweise darauf ab, Irrtümer im Hinblick auf die Verallgemeinerbarkeit der Ergebnisse zu minimieren, indem sichergestellt wird, dass die für die Messeffekte bedeutsamen Einflussgrößen in der Stichprobe ebenso verteilt sind wie in der Grundgesamtheit. Eine Garantie, dass dies immer erfolgreich gelingt, kann jedoch keine Form der Stichprobenziehung bieten. Ein expliziter, möglichst tabellarischer Vergleich der Verteilung zentraler Merkmale zwischen Stichprobe und Grundgesamtheit sollte in jedem Fall angestrebt werden. Dieser Vergleich ist insbesondere dann nötig, wenn die Vergleichbarkeit der Verteilungen nicht durch die Form der Stichprobengewinnung (i. d. R. Zufallsauswahl) gewährleistet werden kann. Um die angenommene Schätzgüte der gewählten Stichprobe transparent zu machen, sind Margen und Kennziffern für die abwägende Inkaufnahme statistischer Irrtümer unter Berücksichtigung der inhaltlichen Untersuchungsziele, der praktischen Untersuchungsmöglichkeiten sowie ethischer Erwägungen im Hinblick auf den Repräsentationsschluss zu kalkulieren und im Rahmen gängiger statistischer Konventionen anzugeben [z. B. 6–8, 10,

17, 29, 50]. Relevante Kennziffern sind zum Beispiel 1) der Standardfehler der Schätzung, 2) die Effektstärkenschatzung und das Vertrauensintervall, 3) die Prüfverteilungen, 4) die Kenngrößen des Signifikanztests und 5) die Irrtumswahrscheinlichkeiten. Im Rahmen der Planung einer Studie sollte weiterhin sichergestellt werden, dass die Stichprobenumfänge hinreichend groß sind. Dies sollte erfolgen, indem eine solche Stichprobengröße gewählt wird, mit der eine empirisch minimal relevante Effektgröße bei einer β -Fehlerwahrscheinlichkeit von 0.2 als signifikant erwartet werden kann [10, 25]. Wesentlich für die „kluge“ Konstruktion einer Stichprobe ist die Kenntnis und Beschreibung der Grundgesamtheit („target population“), auf die hin verallgemeinert werden soll. Dies ist oftmals nicht gegeben (z. B. alle AIDS-gefährdeten Personen; alle von Versorgungsbrüchen betroffenen Patienten) und muss über unterschiedliche Formen der Hochrechnung (gängig: Verhältnis-schätzung, Differenzschätzung, [lineare] Regressionsschätzung, „capture-recapture Methode“ [vgl. 23, 40]) und theoretische Überlegungen geleistet werden [29]. Grundsätzlich ist zu berücksichtigen, dass die Repräsentativität sowohl auf der Ebene der Individuen innerhalb einer Organisationseinheit als auch für die Organisationseinheiten selbst sichergestellt werden muss. Für die organisationsbezogene Versorgungsforschung sind Zufallsauswahlen und gesteuerte Auswahlen von besonderer Bedeutung.

Zufallsauswahlen

Einfache Wahrscheinlichkeitsauswahlen (equal probability selection method)

Bei Zufallsauswahlen müssen alle Elemente einer bekannten oder angenommenen Grundgesamtheit die gleiche Chance haben, in die Stichprobe einzugehen. Ihre Auswahl erfolgt in einem einzigen Auswahlvorgang aus einem Gebiet oder aus einer „Liste“ anhand statistischer Algorithmen. Die geringsten Verzerrungseffekte bergen dabei Lotterielosverfahren, Zufallszahlentafeln oder das Auswahlverfahren „Tag des Ge-

burtstagsdatums“. Bei systematisierten Zufallsverfahren wie z. B. der Ziehung jedes n 'ten Falls ist sicherzustellen, dass die Einheiten zufällig verteilt in den genutzten Auswahllisten vorliegen. Sind nicht alle Elemente einer Grundgesamtheit direkt zugänglich (z. B. drogenabhängige Jugendliche), so muss die Zufallsstichprobe aus einer verfügbaren Teilmenge (Patienten teilnehmender Kliniken) gezogen werden. Für alle über diese Stichprobe hinausgehenden Repräsentationsschlüssen sind die oben genannten Fehlerparameter nach gängigen Konventionen anzugeben und inhaltliche Begründungen für die Berechtigung von Verallgemeinerungsschlüssen darzulegen [8, 17].

Mehrstufige Zufallsauswahlen

Mehrstufige Auswahlverfahren sind immer dann erforderlich, wenn die Erhebungseinheiten (Pfleger, Ärzte, drogenabhängige Jugendliche) nicht direkt zugänglich sind, sondern nur über eine oder mehrere übergeordnete Einheiten (z. B. Kliniken, Stationen) identifiziert werden können. Fragestellungen der Organisationsforschung erfordern in der Regel, dass die natürliche Clusterung der Einzelpersonen in organisationalen Einheiten durch die Anwendung von Mehrebenenanalysen statistisch explizit berücksichtigt werden muss, um die Validität der statistischen Ergebnisse sicherstellen zu können [39]. Da bei diesen mehrstufigen Zufallsauswahlen nicht mehr jedes Mitglied der Grundgesamtheit automatisch die gleiche Chance hat, in einer Stichprobe vertreten zu sein, haben alle mehrstufigen Stichprobenverfahren Entscheidungen über das Zusammenspiel inhaltlicher und statistischer Überlegungen zur Voraussetzung, die explizit gemacht werden sollten. Folgende Methoden, für die je nach Untersuchungsziel und Untersuchungsmöglichkeiten Variationen und Kombinationen möglich sind, zielen darauf ab, die Irrtümer bezüglich Erkenntnisgewinn und Verallgemeinerungsschlüssen gering zu halten.

Geschichtete (stratifizierte) Auswahlen

Diese Stichprobenformen sind sinnvoll, wenn 1) Annahmen darüber bestehen, welche Determinanten für das zu unter-

suchende Merkmal eine Rolle spielen, 2) deren Wirkung besonders geprüft werden soll oder 3) die Heterogenität des interessierenden Merkmals in der Grundgesamtheit sehr hoch ist und gleichzeitig die Schätzfehler der Aussagen gering gehalten werden sollen. Auf der Grundlage des Vorwissens wird die Grundgesamtheit anhand relevanter Merkmale künstlich in Schichten (Untergruppen) unterteilt (z. B. nach Berufsgruppen oder Krankenhausgröße). Diese Unterteilung ist inhaltlich und statistisch insbesondere im Hinblick auf Schichtungsmerkmale zu beschreiben und zu begründen. Je heterogener die Schichten und je homogener die Elemente der jeweiligen Schichten, desto eher ist eine Präzisionsverbesserung der Schätzung durch eine geschichtete Auswahl erreichbar. Auf der ersten Schichtungsebene sollte die Stichprobengröße möglichst proportional zur Grundgesamtheit gewählt werden (PPS Stichprobe = probability proportional to size) und auf der zweiten und jeder weiteren Schichtungsebene können einfache Zufallsverfahren angewandt werden. Wird eine im Hinblick auf die relevante Grundgesamtheit disproportionale Stichprobengröße gewählt, ist dies aufgrund erheblichen Vorwissens über die Bedeutung der Schichtungskriterien zu begründen. Komplexe Schätzformeln (z. B. [8, 34]) sind dann zur Minimierung der Fehlerintervalle einzusetzen.

Klumpenstichproben (Clusterstichproben)

Als Sonderfall der Schichtungs-konstruktion werden zur Reduktion des Erhebungsaufwandes bereits natürlich vorgruppierte Teilmengen – sog. Klumpen oder Cluster (z. B. Klinik) – einer Grundgesamtheit (z. B. alle Kliniken Deutschlands) als Auswahl-einheiten genutzt. Diese Wahl der Teilelemente ist begründungspflichtig. Es gilt folgende Faustregel zur Minimierung von Verzerrungseffekten durch die Klumpenauswahl (sog. Clustereffekte): Die Fehlerbreite der Schätzungen ist invers zur Anzahl der berücksichtigten Cluster, sie wächst hingegen mit der Größe der Cluster relativ zur Population und der Homogenität der Elemente innerhalb der Cluster sowie der Heterogenität der Elemente zwischen den Clustern. Die Verallgemeinerbarkeit der Aussagen aus Klumpenstichproben steigt – ebenso wie bei ge-

schichteten Stichproben –, wenn die zu untersuchenden Cluster (z. B. Krankenhäuser) zufällig aus der Grundgesamtheit (z. B. aller deutschen Kliniken) ausgewählt wurden und dann innerhalb der Klumpen wiederum Zufallsstichproben der zu untersuchenden Personen (z. B. Krankenschwestern) gezogen werden. Auch hier sind Begründungen und Fehlerschätzungen für die jeweils getroffenen Entscheidungen anzugeben (vgl. hierzu zusammenfassend [23]). Wie bei der mehrstufigen Zufallsauswahl, sollte hier bei der statistischen Datenanalyse die Mehrebenenanalyse zum Einsatz kommen, um die Validität der verwendeten Fehlerschätzer gewährleisten und Interaktionen von organisationalen und individuellen Eigenschaften angemessen modellieren zu können. Für diese Analysen gilt als Orientierung für die angemessene Stichprobengröße: es sollten mindestens 30 organisationale Einheiten und ca. 30 Personen pro Organisationseinheit gezogen werden. Abweichungen nach unten sind dabei eher für die Individuenanzahl pro Einheit zu akzeptieren. Liegt das besondere Interesse an der Bestimmung von Interaktionen zwischen den Ebenen, so sollten mindestens 50 Organisationseinheiten und ca. 20 Individuen pro Einheit in der Stichprobe enthalten sein [33].

Gesteuerte Auswahlen (bewusste, kontrollierte Auswahlen, judgement samples)

Gesteuerte Stichproben stützen sich nicht systematisch auf Zufallsentscheidungen. Aus Praktikabilitätsgründen nutzen sie vielmehr gezielte Kenntnisse sowie berechtigte Annahmen über die Verteilung und Bedeutung von Substrukturen in der Grundgesamtheit. Sie erfüllen unterschiedliche statistische Verallgemeinerungskonventionen und werden zum Teil in qualitativen Forschungsdesigns genutzt.

Quotaverfahren

Bei Quotaverfahren erfolgt die Auswahl der Untersuchungseinheiten anhand von kenntnisreich festgelegten Quoten für unterschiedliche Schichtungen oder Klumpungen der Grundgesamtheit mit dem Ziel, auf diese Art und Weise die zentralen messrelevanten Eigenschaften der Grundgesamtheit abzubilden. Erfahrung mit vorhersagestarken Quotierungen

besteht vor allem in der Bevölkerungs-, Markt- und Meinungsforschung. Die Quotierung kann, je nach Vorwissen, anhand von Einzelmerkmalen (z. B. Geschlecht) oder Merkmalskombinationen (z. B. unter 30-jährige ledige Frauen mit Kleinkind) erfolgen. Es können allerdings nur solche Merkmale quotiert werden, deren Bedeutung für den Untersuchungseffekt bekannt sind. Der Quotenplan ist zu begründen und präzise zu dokumentieren. Eine zu enge Quotierung erhöht allerdings das Risiko der Verfälschung bei der Erhebung (z. B. [24]).

Weitere gesteuerte/gezielte Auswahlverfahren

Bei hoher gesicherter Kenntnis über Merkmalszusammenhänge und deren Bedeutung in der Grundgesamtheit besteht die Möglichkeit, zum tieferen Verständnis der Zusammenhänge proportional oder quasi-zufällig typische Fälle, Extremfälle und/oder Experten zur Analyse herauszugreifen. Die jeweiligen Ein- und Ausschlusskriterien sind in diesem Fall jedoch präzise zu begründen und zu beschreiben. Auf die Methode der gesteuerten Gelegenheitsauswahl (z. B. freiwillige Meldung auf Ausschreibungen) kann in der Versorgungsforschung ebenfalls zurückgegriffen werden. Dies ist dann nötig, wenn es schwierig ist, Untersuchungseinheiten zur Forschungsteilnahme zu bewegen (z. B. Arztpraxen, die an einem Shared-Decision-Making-Experiment oder einer Organisationsentwicklungsmaßnahme teilnehmen sollen). Die Population so ausgewählter Freiwilliger unterliegt einem besonders hohen Verzerrungsrisiko in Bezug auf den Verallgemeinerungsanspruch. Wenn der Zugang zu Organisationen oder Individuen institutionell schwierig ist (z. B. häuslich pflegende Töchter; alkoholgefährdete Beschäftigte in Kliniken), ist man auf sogenannte Schneeballverfahren in verschiedenen Varianten [z. B. 28, 40] angewiesen. Hier werden Kontakte von oder über Zielpersonen an weitere Zielpersonen weitergereicht. Für Dunkelfeldstudien ist es u. U. sinnvoll, dies ohne direkten Kontakt zur Forschergruppe, allein im Kreise der Probanden, zu organisieren. Bei den genannten Auswahlmethoden ist die problematische Verallgemeinerbarkeit gegen den Gewinn innovativer Erkenntnisse abzuwägen.

Fehlertransparenz

Um stichprobenbedingte Fehlerquellen transparent zu machen, sind zusätzlich zur Angabe der oben genannten statistischen Referenzgrößen und der inhaltlichen und ethischen Erwägungen für jede Stichprobe folgende weitere Informationen zu den gegebenenfalls getroffenen Entscheidungen anzugeben und zu begründen [17]:

- Eindeutige Ein- und Ausschlusskriterien für alle Untersuchungseinheiten sowohl auf Ebene der organisationalen Einheiten als auch auf Ebene der Individuen (insbesondere Paarbildungs- und Matchingkriterien).
- Systematische ebenen- und zeitpunktbezogene Non-Response-Analyse bezüglich der Ausschöpfungsquote, der Unit-Non-Response (Ausfall von Erhebungseinheiten) und der Item-Non-Response (z. B. anhand des Flow-Diagramms aus dem Consort-Statement [16]). Hierbei sollten bei der Studienplanung potenzielle Prädiktoren des systematischen Drop-Outs ermittelt und zu Studienbeginn erhoben werden, um eine angemessene Analyse der Missing-Data-Prozesse durchführen zu können [49].
- Erwartete Stichprobenreduktion (Panelmortalität) und -verzerrungen bei Längsschnitt- bzw. Paneluntersuchungen.

Standards für die Messung von Merkmalen und Datenqualität

Die folgenden Ausführungen zu den Standards, die bei der Messung von Organisations- und Personenmerkmalen einzuhalten sind, beziehen sich auf die sozialwissenschaftlichen Datenerhebungsformen Befragung (Skalen und Indizes) und Beobachtung. Auf die Nutzung psychometrischer Skalen im Rahmen der Versorgungsforschung wird im Teil „Methoden der Lebensqualitätsforschung“ des Memorandums III genauer eingegangen. Im Folgenden werden nur Aspekte behandelt, die über die dort behandelten Punkte hinausgehen. Für die Qualität der Messung im Rahmen der dritten Form der sozialwissenschaftlichen Datenanalyse, der Sekundärdatenanalyse [46], sei auf den umfassenden Leitfaden „Gute Praxis Sekundärdaten-

analyse“ [3] verwiesen. Die dort aufgeführten Kriterien besitzen auch für die Sekundärdatenanalyse im Rahmen der Organisationsforschung Gültigkeit.

Erfassung von Merkmalen durch befragungsbasierte Skalen und Indizes

Ausgangspunkt der Messung ist eine implizite oder explizite Theorie über den Gegenstand der Untersuchung. Man unterscheidet in der Psychometrie latente Merkmale und manifeste Indikatoren. Die Messung manifester Größen (z. B. 30-Tagesmortalität) erfolgt über Einzelgrößen. Die Messung von latenten Merkmalen erfolgt in der Regel mittels Skalen. Skalen bestehen aus einer Gruppe von Einzelitems bzw. Einzelmerkmalen, die dazu dienen, theoretisch postulierte latente Merkmalsausprägungen zu schätzen (z. B. Depressivitätsskala, Lebensqualitätsskala). Das Ziel bei der Messung von Konstrukten ist es, manifeste Größen zu erhalten, die den interessierenden Konstruktwert valide und mit einem möglichst geringen Messfehler repräsentieren. Alle quantitativen Messungen müssen testtheoretischen Gütekriterien genügen. Entsprechende Angaben zur Sicherstellung dieser Kriterien und zur Erfüllung von Güteanforderungen müssen berichtet werden. Bei Abweichungen von gängigen Grenzwerten bei den Gütekriterien sollte dies begründet werden.

Allgemein gilt, dass Skalenneuentwicklungen nur dann erfolgen sollten, wenn für die zu messenden Konstrukte nicht bereits getestete Instrumente vorliegen. Die in der Literatur berichteten Gütekriterien (z. B. Objektivität, Reliabilität, Validität) sollten dokumentiert werden. Das Ziel bzw. der Nutzen einer Neuentwicklung sollte begründet werden. Wird eine Neuentwicklung vorgenommen, so kann die Skalenstruktur innerhalb einer Item-/Merkmalsgruppe exploratorisch oder konfirmatorisch analysiert werden. Ausgangspunkt der exploratorischen Analyse ist die Verwendung exploratorischer, dimensionsreduzierender Verfahren wie z. B. die exploratorische Faktorenanalyse. Falls möglich, ist eine Kreuzvalidierung exploratorisch ermittelter Ergebnisse durchzuführen. Ausgangspunkt beim konfirmatorischen Vorgehen ist eine Theorie über die Zusammenhänge der Einzelmerkmale (Indikatoren) mit dem

zu messenden Konstrukt. Das theoriegeleitete Vorgehen erfolgt typischerweise durch den Einsatz der konfirmatorischen Faktorenanalyse oder von Item-Response-Modellen (z. B. Rasch-Modell). Bei Verwendung des konfirmatorischen Vorgehens bietet sich die Anwendung von Strukturgleichungsmodellen zur Bestimmung der Konstrukt- und Kriterienvalidität an.

Abweichend vom Konzept der klassischen Skalenentwicklung [9, 14] können Messwerte auch über Indizes gebildet werden. Als Indizes werden Summenwerte über Items bezeichnet, die (a) nach inhaltlichen Kriterien einen zu erfassenden Merkmalsbereich erschöpfend abbilden und (b) als gewichtete Summe geeignet sind, die interessierende Konstruktausprägung vorherzusagen. Hierfür ist es notwendig, dass die Einzelmessungen als formative oder konstituierende Indikatoren des zu erfassenden Konstrukts aufgefasst werden können. Dies ist dann der Fall, wenn die Einzelindikatoren als kausale Antezedenzen für die Konstruktausprägung fungieren (z. B. Konstrukt: Hotelqualität eines Krankenhauses; Indikatoren: Zufriedenheit mit dem Essen, Zufriedenheit mit dem Zimmer, Zufriedenheit mit den sanitären Anlagen). Für Indizes sind spezielle Gütekriterien zu berichten [22], die sich auf die (a) inhaltliche Validität der Konstruktdefinition und Indikatorspezifikation, (b) Kolinearität der Indikatoren und (c) externe Validität des Index beziehen. Im Gegensatz zu Skalen müssen Indizes insbesondere keine interne Konsistenz aufweisen. Die Gütekriterien für formative Messungen lassen jedoch keine eigentliche empirische Überprüfung der zugrunde liegenden Messtheorie zu, wie dies für Skalen der Fall ist. Als messtheoretisch problematisch erweist sich hier insbesondere, dass keine Basis dafür gegeben ist, den Einfluss von Messfehlergrößen abzuschätzen. Da Indizes in der Regel eine lediglich aufgrund von Expertenwissen zusammengestellte Operationalisierung darstellen, besitzt eine solche Messung eine geringere psychometrische Qualität als eine Messung, die durch eine psychometrisch fundierte und validierte Messskala erfolgt ist. Ebenso gilt, dass die Aussagekraft von theoriegeleitet entwickelten und konfirmatorisch überprüften Skalen als höherwertig einzuschätzen ist als diejenige explorativ ermittelter Skalen.

Qualitätskriterien für Beobachtungsdaten

Bei der systematischen Beobachtung werden in der Regel die Ausprägungsgrade definierter Merkmale durch unabhängige Dritte eingeschätzt (z. B. Einschätzung von Patientenmerkmalen durch Behandler; Beurteilung von Qualitätsindikatoren im Peer-Review-Verfahren; z. B. [55]). Der Beobachter oder Beurteiler kann hierbei technisch als Messinstrument betrachtet werden, sodass sich die Frage der Gütekriterien für Messungen in derselben Weise wie für andere Formen der Operationalisierung stellt. Um die Zuverlässigkeit dieses Messprozesses überprüfen zu können, muss insbesondere das Ausmaß der Übereinstimmung der Urteile verschiedener Beurteiler bestimmt werden. Da das diagnostische Ergebnis im Allgemeinen ausschließlich als Indikator der Merkmalsausprägung der eingeschätzten Objekte weiter verwertet werden kann, ist zu fordern, dass die Beurteilungen weitestgehend unabhängig von der Person des jeweiligen Beurteilers sind. Als Übereinstimmungsindikatoren müssen Koeffizientenausprägungen berichtet werden, die das Ausmaß standardisiert quantifizieren, in dem Beurteiler bei der Beurteilung derselben Personen oder Objekte zuübereinstimmenden oder ähnlichen Einschätzungen gelangen. Die Auswahl der angemessenen Kenngröße wird vor allem durch die Skalendignität bestimmt [58]. Für nominale Datenformate werden in der Regel „Cohens κ “ oder das „gewichtete Cohens κ “ als Maßzahl verwendet. Für graduell abgestufte Einschätzungen mittels Ratingskalen sollte Kendalls W im Falle ordinaler Daten und die Intraklassenkorrelation für intervallskalierte Daten bestimmt werden [57]. Für alle genannten Koeffizienten kann ab dem Wert 7 von einer zufrieden stellenden Übereinstimmung ausgegangen werden. Der Mittelwert mehrerer Urteiler hat in der Regel systematisch bessere psychometrische Eigenschaften als das Urteil eines einzelnen Beurteilers. Deshalb kann ggf. der Mittelwert über mehrere Beurteiler als diagnostische Größe verwendet werden, wenn eine zufrieden stellende Übereinstimmung für einzelne Urteiler nicht erreicht werden kann.

Bezogen auf den gesamten Erhebungsprozess ist darauf zu achten, dass in

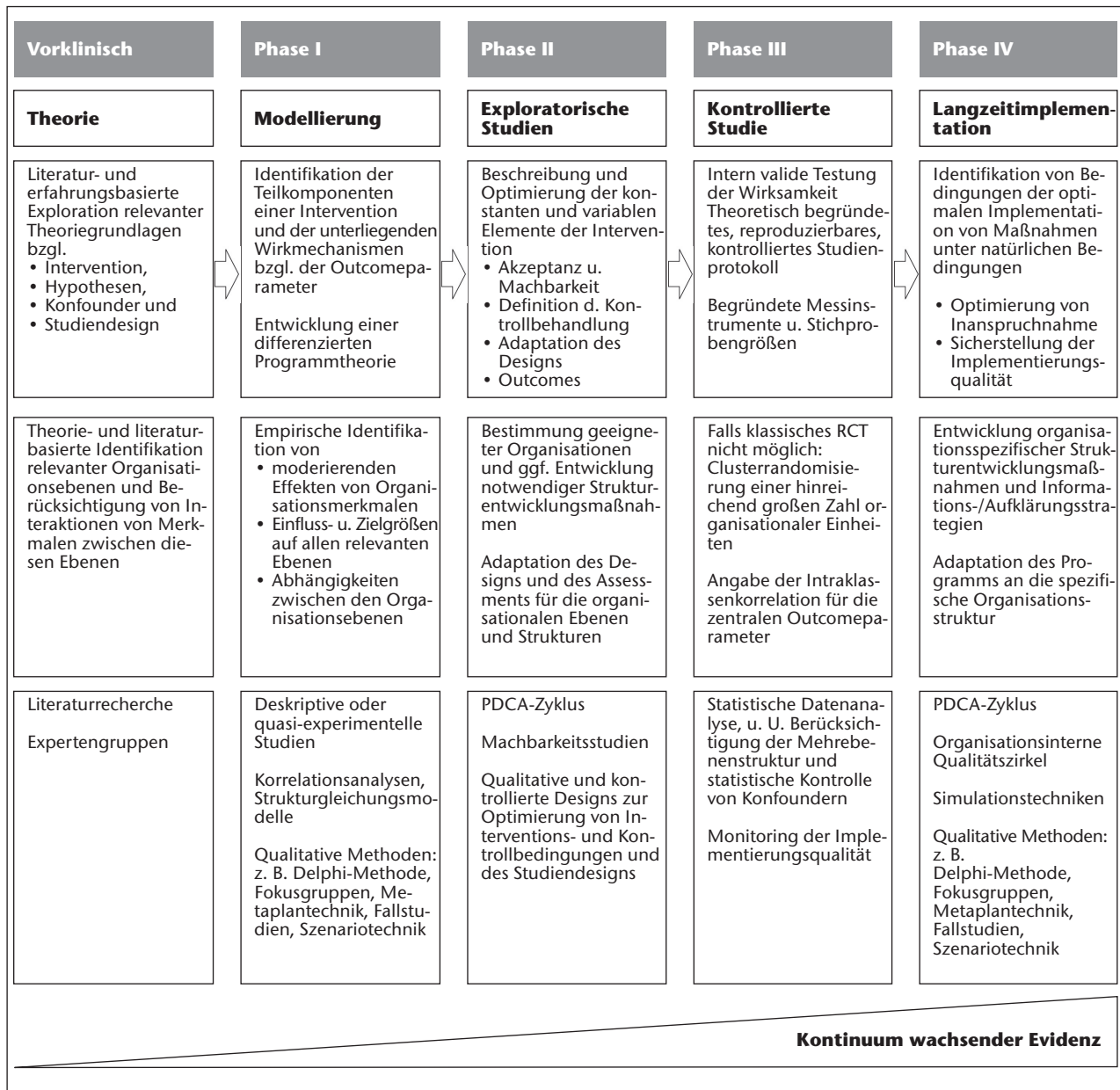


Abbildung 3 Phasenmodell für die Evaluation komplexer Interventionen nach [15].

der Regel die Güte der Beurteilerübereinstimmung vor der Haupterhebung in einer Pilotstudie sichergestellt werden muss. Mängel der Beurteilungsgüte sollten anschließend durch ein systematisches Beurteilertraining identifiziert und beseitigt werden [57]. Hierbei sind Methoden anzuwenden, die z. B. eine eindeutige und intersubjektiv geteilte Definition des zu beurteilenden Merkmals unterstützen, den theoretischen Hintergrund der Beurteiler vereinheitlichen und die notwendigen Informationsverarbeitungsprozesse – z. B. durch Einschränkung des Interpretationsspielraums – minimieren. Zudem sollte eine systematische Analyse möglicher

verzerrender Beurteilungstendenzen (z. B. Halo-Effekt, Tendenz zur Mitte, Tendenz zu den Extremen) erfolgen [31]. Erst nachdem durch ein solches Beurteilertraining eine hinreichende Übereinstimmung sichergestellt wurde, kann die Haupterhebung erfolgen.

Standards und ein Rahmenmodell für Studiendesigns in der organisationsbezogenen Versorgungsforschung

Das Ziel der Organisationswissenschaft im Rahmen der Versorgungsforschung

ist es, Erkenntnisse über den Status Quo zu erhalten und – falls möglich – Maßnahmen zur Verbesserung der Versorgungssituation zu entwickeln und deren Wirksamkeit zu überprüfen. Studien zu komplexen Interventionen, die aus verschiedenen, miteinander verknüpften Teilmaßnahmen bestehen und in natürlichen Settings implementiert werden sollen, sollten dabei sowohl einen Beitrag zur Verbesserung der Programm- als auch der Implementierungstheorie leisten [20]. Das Phasenmodell von Campbell [15] zur Planung und Durchführung von Studien zur Evaluation komplexer Interventionen kann hierbei als Orien-

tierung für eine gute organisationsbezogene Versorgungsforschung empfohlen werden. Dieses Modell berücksichtigt, dass an die Evaluation komplexer Interventionen besondere Anforderungen zu stellen sind, weil die Entwicklung, Identifikation, Dokumentation und Reproduzierbarkeit der Interventionskomponenten evaluiert werden muss. Da die Implementierung einer Intervention in die Versorgungspraxis nur unter Berücksichtigung der natürlichen Rahmenbedingungen erfolgen kann, kommt der empirischen Untersuchung der Generalisierbarkeit und externen Validität der Befunde ein besonderer Stellenwert zu. Eine umfassende Evaluation muss dabei auf qualitativen und quantitativen Forschungsmethoden basieren. Im Folgenden werden die zentralen Fragestellungen und Forschungsziele pro Phase dargestellt und anschließend werden Implikationen für die Methodenstandards abgeleitet [vgl. 43]. Abbildung 3 zeigt den Phasenverlauf im Überblick.

Vorklinische oder theoretische Phase

In dieser Phase wird Wissen generiert und systematisiert, das die Evidenz für die Erreichung der gewünschten Effekte und damit die Wirksamkeit der geplanten Intervention erhöht. Zudem sollte der Kontext analysiert werden, um den Einfluss von Konfounder- und Moderatorvariablen zu erkennen und im weiteren Studienverlauf berücksichtigen zu können. Ziel ist die Entwicklung einer vorläufigen Programmtheorie, die das Vorgehen in den folgenden Studienphasen begründet. In der vorläufigen Programmtheorie sollten insbesondere Stakeholder, Populationsmerkmale, beeinflussbare und nicht beeinflussbare Merkmale, Kontexteinflüsse, Outcomeparameter (Surrogat- und Zielgrößen) und plausible Wirkbeziehungen formuliert werden. Für die Versorgungsforschung muss darauf geachtet werden, dass alle relevanten organisationalen Ebenen (z. B. Ebene 1: Patient, Ebene 2: behandelnder Arzt, Ebene 3: Klinik) berücksichtigt werden. Dabei muss spezifiziert werden (a) welche Merkmale auf den jeweiligen Ebenen bei der Theoriebildung relevant sind, (b) ob Interaktionen zwischen den verschiedenen Ebenen bestehen können und (c) welche kausalen Zusammenhänge zwischen

den Ebenen vermutet werden können. Angenommen, eine Theorie formuliert, dass die Behandlungsmotivation des Patienten bei Aufnahme seine Compliance im Behandlungsverlauf positiv beeinflusst und durch diese der Behandlungserfolg mit determiniert wird, dann muss davon ausgegangen werden, dass dieses Patientenmerkmal durch Merkmale der Versorgungseinheiten systematisch beeinflusst wird. So wäre anzunehmen, dass beispielsweise die Empathie des Arztes oder Strukturmerkmale der Klinik (z. B. Betreuungsverhältnis: Ärzte pro 10 Patienten) die am Patienten erhobenen Merkmale beeinflussen und deren Beziehung untereinander moderieren. Die Versorgungsforschung zeichnet sich gegenüber klassischen klinischen Forschungsansätzen dadurch aus, dass diese strukturellen Abhängigkeiten ein explizites Ziel der Theoriebildung sind. Die darauf aufbauenden empirischen Studien müssen diese Abhängigkeiten methodisch angemessen analysieren. Literaturrecherchen und Expertenbefragungen sind in dieser Phase die wichtigsten methodischen Grundlagen. Es sollte dabei insbesondere auf vorliegende Studien Bezug genommen werden, die verwandte Interventionen und Konstrukte in vergleichbaren Populationen untersucht haben. Für die Organisationsforschung ist die Berücksichtigung von empirischen Studien unter ähnlichen oder auch abweichenden organisatorischen Rahmenbedingungen von besonderer Bedeutung für die Theoriebildung.

Phase I: Definition der Interventionskomponenten (Modelling)

Einzelne Prozesselemente und Wirkzusammenhänge, die gemäß der Programmtheorie von zentraler Bedeutung sind, werden in dieser Phase isoliert untersucht und das Verständnis des Forschungsgegenstands wird empirisch vertieft. Dies kann beispielsweise geschehen, in dem einzelne Interventionskomponenten in Pilotstudien getestet, Einzelfälle intensiv beobachtet und analysiert oder alle an komplexen Versorgungsprozessen beteiligten Behandler befragt werden. Ziel ist es hierbei, empirische Hinweise auf die Gültigkeit elementarer Hypothesen zu erhalten, Optimierungs- und Modifikationsbedarf zu erkennen und somit die Basis für geziel-

te Verbesserungsmaßnahmen zur Verfügung zu stellen. Zudem sollten mögliche Barrieren bei der Umsetzung von Interventionsprogrammen in die Praxis dokumentiert werden.

In komplexen Versorgungsorganisationen muss darauf geachtet werden, dass potentiell moderierende organisationale Rahmenbedingungen systematisch auf ihre Wirkung hin untersucht werden. Kann davon ausgegangen werden, dass unterschiedliche Organisationstypen existieren, in denen die Intervention implementiert werden soll, können beispielsweise Einzelfallanalysen in Institutionen jeden Typs wichtige Hinweise auf zu beachtende Heterogenitäten in den Wirkprozessen liefern. Alternativ können vor allem Extremgruppenvergleiche eingesetzt werden. Zudem ist es wichtig, Einflussgrößen und relevante Outcomeparameter auf allen Ebenen (z. B. Ebene 1: Lebensqualität des Patienten; Ebene 2: stationsspezifische Mitarbeiterzufriedenheit; Ebene 3: Kosten-Effektivität auf Klinikenebene) zu identifizieren und deren Abhängigkeiten empirisch zu prüfen. In dieser Phase können unterschiedliche Forschungsmethoden gewählt werden, die ein breites Methodenspektrum abdecken. Die Angemessenheit der Methode zur Beantwortung der jeweiligen Fragestellung determiniert die Qualität des forschungsmethodischen Vorgehens. Formative methodische Strategien mit dem Ziel einer kontinuierlichen Optimierung stellen dabei einen methodischen Schwerpunkt dar. Hierbei können z. B. qualitative Analysen zusammen mit Fokusgruppen, Dokumentenanalysen, Einzelfallstudien sowie deskriptive, korrelative und quasi-experimentelle Ansätze zum Einsatz kommen. Zur Erhöhung der Evidenz der Gültigkeit theoretisch postulierter multivariater Wirkbeziehungen können insbesondere pfadanalytische Modelle oder Strukturgleichungsmodelle eingesetzt werden [36]. Diese Modelle erlauben die vergleichende Evaluation alternativer Modelle und sind somit geeignet, den Raum möglicher Kausalmodelle einzuschränken. Durch Strukturgleichungsmodelle ist es zudem möglich, die bedeutsamen Konstrukte messtheoretisch besser abzubilden und damit die Validität der statistischen Schätzungen zu erhöhen. Für Strukturgleichungsmodelle müssen lokale und globale Ma-

ße [4] des Modellfits berücksichtigt werden, um die Interpretierbarkeit der Ergebnisse gewährleisten zu können. Wurde eine Klumpenstichprobenziehung oder eine mehrstufige Zufallsauswahl angewandt, so ist die Mehrebenenstruktur bei solchen Analysen zu berücksichtigen [38, 51].

Phase II: Definition der Machbarkeit und des Interventionsdesigns (Exploratory trial)

Die optimale Gestaltung der Intervention und des endgültigen Studiendesigns stehen bei dieser Phase im Mittelpunkt. Dabei müssen sowohl die Akzeptanz als auch die Machbarkeit der Intervention in der Versorgungspraxis sichergestellt werden. Für die Erhöhung der Akzeptanz stehen standardisierte Techniken der Konsensfindung zur Verfügung. Zu den gängigsten Verfahren gehören der Nominale Gruppenprozess, die Delphi-Technik und die formale Konsenskonferenz im Format nach *Vang* [21, 54]. Alle drei Verfahren sind mehrstufig, interaktiv und zielen mit einem kontrollierten Umfrage-Feedback auf Konsens. Die Wahl des einzusetzenden Verfahrens ist u. a. abhängig von der Teilnehmeranzahl [5, 44]: Nominaler-Gruppen-Prozess (bis 20 Teilnehmer), Delphi-Verfahren (> 10 Teilnehmer; [21]), strukturierte Konsensuskonferenz im Format nach *Vang* (> 50 Teilnehmer; [54]). Zur weiteren Verbesserung der Akzeptanz und Machbarkeit können gegebenenfalls unterschiedliche Interventionsvarianten in einer Testphase implementiert und optimiert werden. Zudem muss sichergestellt werden, dass die Intervention in der Praxis tatsächlich angemessen umgesetzt werden kann (z. B. Identifikation von Weiterbildungsbedarf, Schulung bezüglich der Interventionsdurchführung, Bedarf an personellen Ressourcen, organisatorische Rahmenbedingungen). Weiterhin muss geprüft werden, wie eine adäquate Kontrollbedingung gestaltet werden kann. Für Barrieren und Widerstände bei der Implementierung der Intervention und des Studiendesigns müssen Lösungsstrategien entwickelt und geprüft werden (z. B. Rekrutierung und Aufklärung potenzieller Teilnehmer). Es ist zu prüfen, in welchen Organisationen und auf welchen organisationalen Ebenen eine Implementierung der Intervention grundsätz-

lich möglich ist und welche Strukturentwicklungsmaßnahmen ggf. notwendig sind, um die Voraussetzung für die Intervention bzw. die Kontrollbehandlung zu schaffen.

Ein weiterer Schwerpunkt sollte auf die Pilottestung der Outcomeparameter gelegt werden. Die verwendeten Outcomeparameter sollten die zuverlässige und valide Erfassung der Interventionseffekte für alle Betroffenen ermöglichen. Welche Assessmentmethoden aussagekräftige Ergebnisse liefern und in der Praxis akzeptiert werden, muss empirisch geprüft werden. Es muss gewährleistet werden, dass alle für die Gesamtversorgungsstruktur relevanten Effekte erfassbar sind. Für hierarchisch übergeordnete Ebenen im Versorgungsprozess muss dabei insbesondere geprüft werden, ob Merkmale dieser höheren Ebenen direkt auf den entsprechenden Ebenen bestimmt werden sollen oder valider durch Aggregation von Daten niedrigerer Ebene geschätzt werden können. So wird die Empathie des behandelnden Arztes in der Regel am besten durch Mittelung entsprechender Patientenbeurteilungen bestimmt. Im Rahmen der Datenanalyse sollte der Mittelwert dann aber als Arzt und nicht als Patientencharakteristikum analytisch behandelt werden. Die Empfehlungen für die angemessene Methodenauswahl entsprechen im Wesentlichen denjenigen in Phase I. Nützliche Informationen können vor allem durch kontrollierte Designs zur Optimierung von Interventions- und Kontrollbedingungen gewonnen werden.

Phase III: Summative Überprüfung der Intervention (definitive randomised control trial)

Das Interventionsprogramm wird implementiert und im Vergleich zu einem Kontrollprogramm summativ im Hinblick auf seine Effektivität evaluiert. Es sollte eine randomisierte Zuteilung zu der Treatment- und Kontrollgruppe auf Einzelpatienten oder Clusterebene erfolgen, um eine maximale interne Validität zu gewährleisten. Erlauben die praktischen Rahmenbedingungen keine Randomisierung, stellen quasi-experimentelle Designs mit gematchten Stichproben oder mit anschließender Risikoadjustierung eine Alternative dar. Wenn Patienten starke Präferenzen

für eine bestimmte Behandlungsform haben und in der späteren Praxis die Patientenpräferenz für die Auswahl der Behandlungsform mit entscheidend ist, können Designs gewählt werden, die die Patientenpräferenz mit berücksichtigen (preference trial designs; [12, 59]).

In dieser Phase muss der gesamte Studienverlauf transparent dokumentiert werden, sodass eine Replikation ermöglicht wird. Es müssen standardisierte und validierte Assessmentverfahren oder wissenschaftlich akzeptierte Maßzahlen verwendet werden. Es sind ferner angemessene statistische Datenanalysemethoden zu wählen und im statistischen Modell sollten die in der Programmtheorie formulierten Zusammenhänge adäquat abgebildet sein [56]. Darüber hinaus sollte eine Kalkulation der notwendigen Stichprobenumfänge erfolgen, um sicherzustellen, dass klinisch bedeutsame Effekte mit hinreichender Teststärke nachgewiesen werden können. Zudem sollte dokumentiert werden, welche Selektionsprozesse zur Nichtteilnahme von Patienten geführt haben, um Verzerrungen der Wirksamkeitseinschätzungen durch spezifische Charakteristiken der Untersuchungsstichprobe identifizieren und ggf. korrigieren zu können. Aufgrund des begrenzten zeitlichen Horizonts können theoretisch gut belegte Surrogatparameter als Erfolgsmaße verwendet werden [20]. In clusterrandomisierten Studien muss angegeben werden, welcher Varianzanteil in den Outcomeparametern und den Interventionseffekten durch die Zugehörigkeit zu den entsprechenden organisationalen Einheiten erklärt werden kann. Dies sollte durch die Angaben der Intraklassenkorrelation erfolgen [39]. In dieser Phase steht die interne Validität der gewonnenen Befunde im Vordergrund. Die Überprüfung der Effektivität der Intervention erfolgt mittels quantitativer empirischer Methoden. Es gelten die klassischen Stufen der evidenzbasierten Medizin zur Beurteilung der klinischen Evidenz (z. B. [35]):

- Evidenzstufe Ia: Vorliegen von mindestens einer systematischen Übersichtsarbeit (Metaanalyse) auf der Basis methodisch hochwertiger randomisierter und kontrollierter Studien.
- Evidenzstufe Ib: Vorliegen von mindestens einer hochwertigen randomi-

sierten und kontrollierten Studie mit ausreichender Stichprobengröße.

- Evidenzstufe IIa: Vorliegen von mindestens einer hochwertigen kontrollierten Studie (nicht randomisierte Kontrolle).
- Evidenzstufe IIb: Vorliegen von mindestens einer anderen hochwertigen quasiexperimentellen Studie.
- Evidenzstufe III: Vorliegen von mehr als einer hochwertigen nichtexperimentellen Studie (z. B. mehrere hochwertige deskriptive Studien).
- Evidenzstufe IV: Expertenmeinungen, Grundlagenforschung, Konsensuskonferenzen.

Eine formative Begleitung der Implementierung sollte insbesondere zur Sicherstellung der Validität von Effektinterpretationen erfolgen (Monitoring). Neben der Analyse von Programmeffekten mittels zumeist varianz- und mehr-ebenenanalytischer Methoden können Pfadanalysen oder Strukturgleichungsmodelle zur Überprüfung der in der Programmt heorie postulierten Wirkzusammenhänge herangezogen werden.

Phase IV: Unterstützung einer effektiven Interventionsimplementierung (Long term implementation)

Nachdem in Phase III die Wirksamkeit der Intervention nachgewiesen wurde, muss nun in der Phase IV untersucht werden, wie die Intervention in der Versorgungspraxis optimal unter Alltagsbedingungen implementiert werden kann, welche bisher nicht genauer untersuchten organisationalen und Kontextmerkmale die Effektivität beeinflussen und wie die Akzeptanz bei Leistungserbringern und -nutzern optimiert werden kann. Gegebenenfalls müssen Anreizsysteme entwickelt werden, um eine maximale Anzahl von Organisationen und Patienten sowie eine möglichst hohe Treatment- und Behandlungscompliance erreichen zu können. Das Vorhandensein von Kontrollgruppen wird in dieser Phase bei der Bearbeitung dieser Fragen als nachrangig betrachtet, Qualitätsunterschiede bei der Programmdurchführung durch unterschiedliche Leistungserbringer werden im Sinne der externen Validierung als wichtiger erachtet. Die Empfehlungen für die angemessene Methodenauswahl entsprechen denjenigen in Phase I.

Grundsätzlich bedarf die Implementierung dabei des Einsatzes theoriebasierter Veränderungsstrategien, die einen unterschiedlichen Wirkungsgrad aufweisen können [32]. Die Strategien und die ihnen zugrunde liegenden Theorien sind je nach Intervention entsprechend darzulegen. Für Untersuchungen von Implementierungsstrategien sind eingebettete Prozessevaluationen geeignet, die Aussagen zur Qualität der Implementierung erlauben und outcome-relevante Kontextfaktoren identifizieren [45]. Der Einsatz von indikatorgestützten Verfahren, die einerseits organisationsinterne als auch -externe Qualitätsvergleiche von Interventionen zulassen, ist in Versorgungsstudien nutzbar. Die Erkenntnisse der Phase III sind dabei für die Priorisierung und Ermittlung von geeigneten Messgrößen zur Beurteilung der Versorgungsqualität zu berücksichtigen [1, 30]. Bei der Ermittlung von geeigneten Messgrößen der Versorgung (Qualitätsindikatoren) werden kombinierte Verfahren eingesetzt, die neben der Evidenzsynthese strukturierte methodische Konsensusverfahren anwenden, die bereits in Phase II beschrieben wurden [2, 13, 48].

Die Phase IV-Forschung beschäftigt sich jedoch nicht nur mit Fragestellungen der Optimierung und Sicherstellung von Interventionen unter spezifischen Organisations- und Kontextbedingungen, sondern hat auch die Untersuchung von Outcomes zur Qualitätsverbesserung der medizinischen Versorgung der Bevölkerung zur Aufgabe. Unter dem Gesichtspunkt der Langzeitimplementierung ist die Identifikation von Barrieren und die Erhebung des Bedarfs für weitergehende Forschung zu sehen.

Während in Phase III vor allem Surrogatkriterien als Effektivitätsindikatoren untersucht werden, sollte in Phase IV ebenfalls die Veränderung langfristiger Zielgrößen (z. B. kardiale Ereignisse bei Herzinfarktpatienten) systematisch evaluiert werden. Aufgrund der in der Regel zeitlichen Beschränktheit von Phase III-Studien findet die Untersuchung dieser Effekte in Phase IV statt. Die Untersuchung der Auswirkungen der Intervention auf langfristige Zielgrößen muss sich an der Methodik von Phase III-Studien orientieren (i. d. R. kontrollierte experimentelle Studien).

Nutzen und Limitationen des Rahmenmodells für die organisationsbezogene Versorgungsforschung

Obwohl das Modell von *Campbell* [15] als Stufenprozess formuliert ist, sollte berücksichtigt werden, dass häufig erst in späteren Phasen des Prozesses Mängel in der Beantwortung relevanter Fragen aus den Vorphasen deutlich werden. Diese müssen erst zufriedenstellend beantwortet werden, bevor in den nachfolgenden Phasen auf Basis hinreichender Evidenz weiter gearbeitet werden kann. Deswegen weicht der Prozessverlauf in der Regel von der linearen Darstellung in Abbildung 3 ab. Auch der Gesamtprozess ist als zirkulärer Prozess zu sehen: In Phase IV wird in der Regel weiterer Forschungsbedarf erkennbar werden, der es erforderlich macht, alle Phasen des Modells erneut zu durchlaufen, um neue Interventionskonzepte praxisgerecht entwickeln und evaluieren zu können. Nach der Vorstellung von *Campbell* [15] sollte eine umfassende Evaluation komplexer Interventionsmaßnahmen alle fünf Phasen durchlaufen. Dies gilt sowohl bei neu zu entwickelnden Interventionsprogrammen als auch bei der Implementierung bereits entwickelter Programme, da immer von der Notwendigkeit einer Adaptation der Inhalte und des Studiendesigns an die lokalen studienspezifischen Rahmenbedingungen im Versorgungskontext ausgegangen werden muss. In der Forschungspraxis müssen sich Studien aus Gründen der Praktikabilität häufig auf eine dieser Phasen beschränken. Dann ist aber zu fordern, dass angegeben wird, (a) weshalb die empirischen Belege, die nach der Modellvorstellung eigentlich in den vorhergehenden Phasen hätten identifiziert werden müssen, als vorliegend angenommen werden können und (b) wie die jeweilige Studie in einen komplexen Phasenprozess eingebettet ist. Letzteres ist notwendig, um beurteilen zu können, wie die jeweilige Einzelstudie in einem komplexeren Forschungsprozess verankert ist, der die Erreichung der Ziele gewährleistet, die von *Campbell* [15] für ein komplexes empirisches Evaluationsprojekt gefordert werden.

Eine eindeutige Evidenzstufung empirischer Befunde erweist sich für dieses Modell als schwierig. Wie in Abbildung 3 dargestellt wurde, erhöht sich die empirische Evidenz, wenn eine höhere

Phase des insgesamt fünfstufigen Verlaufs erreicht wird – vorausgesetzt, dass die Fragen der vorherigen Phasen methodisch adäquat bearbeitet und die Phasen erfolgreich abgeschlossen wurden. Zusätzlich muss unbedingt berücksichtigt werden, dass innerhalb jeder Phase die Evidenz der gewonnenen Erkenntnis in Abhängigkeit vom methodischen Vorgehen variieren kann. Eine Studie der Phase III besitzt je nach methodischem Vorgehen einen unterschiedlichen Evidenzgrad (s. oben). Für Studien der Phase III müssen also die klassischen Evidenzstufungen mit berichtet werden. Für Studien der ‚Vorklinischen Phase‘ sowie der Phasen I, II und IV können keine phaseninternen Evidenzgrade angegeben werden, da hier die Angemessenheit der Methode bzw. der Methodenkombination in Bezug auf die jeweilige Fragestellung für die Erkenntnisgüte entscheidend ist. Offen bleibt grundsätzlich, wie die Kombination der Evidenzgrade aufgrund der Phaseneinteilung und der Evidenzgrade innerhalb der Phasen gemeinsam bewertet werden kann: Liegt für Phase III ein adäquater Wirksamkeitsnachweis mittels experimenteller Studien vor, so kann ggf. eine hohe Evidenz angenommen werden, auch wenn die vorangehenden Phasen nicht systematisch durchlaufen wurden. Für eine Phase IV-Studie ist jedoch das Vorliegen einer erfolgreichen Phase III-Studie unverzichtbar.

Abschließend soll darauf hingewiesen werden, dass durch das Modell von *Campbell* [15] die Standards der evidenzbasierten Medizin nicht in Frage gestellt werden. Vielmehr liegt der Nutzen des Modells darin, für spezifische Fragen zu sensibilisieren, die entstehen, wenn komplexe Interventionen in ggf. stark variierenden natürlichen Settings zur Anwendung kommen sollen. Eine maximale Evidenz gemäß klassischer Kriterien muss für jede empfohlene Intervention gegeben sein. Dass diese Empfehlungen jedoch in enger Verzahnung mit den praktischen Rahmenbedingungen in ökologisch validen Kontexten (insbesondere Merkmale der jeweiligen organisationalen Einheiten) entwickelt werden müssen, Implementierungsempfehlungen und -standards einer gesonderten empirischen Fundierung bedürfen und jedes komplexe Maßnahmenbündel durch vielfältige Optimie-

rungs- und Adaptationsprozesse begleitet werden muss, wird durch das Modell in besonderer Weise hervorgehoben [19]. Der besondere Nutzen des Modells für die organisationsbezogene Versorgungsforschung besteht darin, dass es eine gut begründete Systematik für komplexe empirische Interventions- und Evaluationsstudien liefert.

Standards für experimentelle und nicht-experimentelle quantitative Studien

Wie bereits in Kapitel 2 dargestellt, ist bei Experimenten in Phase III (Abb. 3) zu unterscheiden, ob es sich um eine Intervention (z. B. Therapie-, Diagnose- oder Psychoedukationsprogramm) handelt, die an einem Individuum ansetzt und den Erfolg an diesem Individuum misst (z. B. Patient, Mitarbeiter), oder ob es sich um eine Intervention handelt, die auf der Ebene eines sozialen Systems ansetzt (z. B. Abteilung, OP-Team, Klinik). Im ersten Fall ist in der Regel die randomisierte kontrollierte Studie (RCT) der Königsweg, im zweiten Fall die cluster-randomisierte Studie (CRT; [10]). Neben der inhaltlichen Entscheidung, welche Form der Randomisierung optimal ist, müssen häufig auch Aspekte der empirischen Realisierbarkeit berücksichtigt werden. In vielen Fällen ist eine Randomisierung von Individuen auch in Versorgungsforschungsstudien möglich. Dies ist vor allem dann der Fall, wenn es sich bei den Untersuchungseinheiten, die zufällig der Experimental- oder der Kontrollgruppe zugeordnet werden, um Individuen handelt und die Experimental-Individuen nicht mit den Kontroll-Individuen in Austausch treten können. Beide Gruppen sollten nichts von der unterschiedlichen Intervention wissen. Die Qualitätsstandards für den Bericht von Studienergebnissen, die auf der Basis von RCTs zustande gekommen sind, wurden in den revidierten CONSORT-Kriterien niedergelegt [11, 42, 60].

Kann zwischen den Mitgliedern der per Randomisierung zustande gekommenen Vergleichsgruppen – z. B. wegen räumlicher Nähe – ein Austausch stattfinden, so ist es möglich, dass aufgrund der Veränderung des Verhaltens der Mitglieder oder der Art der Intervention – wie z. B. schnell austauschbares Treatment (z. B. Informationsbroschüre über

Therapiealternativen) – die interne Validität und damit die Aussagekraft des Experiments erheblich beeinträchtigt wird. Trotz der durch Randomisierung erreichten Strukturgleichheit der Gruppen zu Untersuchungsbeginn können auf diese Weise für die normale Klinikpraxis verzerrte oder nicht verwertbare Erkenntnisse resultieren. Nach *Cook* und *Shadish* [18] kann die interne Validität dadurch beeinträchtigt werden, dass Kliniker und Patienten wahrgenommene Ungerechtigkeiten in der Versorgungspraxis (Verum vs. Placebo) durch anderweitige Maßnahmen (kompensatorischer Ausgleich) oder durch Imitation oder Übernahme der Aktivitäten in der Vergleichsgruppe (Treatment-Diffusion) auszugleichen versuchen. Das wäre zum Beispiel dann der Fall, wenn die Kontroll-Individuen die für sie nicht gedachten Informationsbroschüren bei den Experimental-Individuen entdecken und ebenfalls lesen. Fühlen sich Patienten aufgrund der Zugehörigkeit zur vermeintlich schlechter behandelten Gruppe (z. B. Placebo-Gruppe) benachteiligt oder nehmen Kliniker die Kontrollintervention (Placebo) als suboptimal wahr, so kann dies zu einer Verringerung oder aber auch zu einer Zunahme der Motivation und Compliance auf Seiten der Patienten und Behandler führen. Im ersten Fall wird die Validität der Studienbefunde z. B. durch ‚empörte Demoralisierung‘ beeinträchtigt, im zweiten Fall durch ‚kompensatorischen Wettstreit‘. Können diese Effekte durch die Untersuchungsanordnung nicht ausgeschlossen werden, ist das Ergebnis des Experiments eingeschränkt interpretierbar und nicht allein auf das applizierte Treatment zurückführbar (vgl. [43]). Als wichtigste Sicherungsmaßnahme zur Identifikation dieser potenziellen Beeinträchtigungen der internen Validität sollte stets ein Studien begleitendes Monitoring vorgesehen werden, in dem Hinweise auf solche Prozesse identifiziert und dokumentiert werden können. Ist eine Gefahr der Kontamination der Intervention (z. B. Diätintervention; Schulung eines Teils der Ärzte) gegeben, strebt man oft eine randomisierte Zuweisung von Gruppen von Individuen (z. B. Familie, Ärzteteam) zur Kontroll- und Interventionsgruppe an. In manchen Fällen ist die Randomisierung von Gruppen von Individuen die einzig machbare Form des Experiments (Clus-

terrandomisierung). Das ist meist bei Feldexperimenten, z. B. Gemeindeexperimenten oder Organisationsexperimenten, der Fall [16]. Die Clusterrandomisierung ist der Königsweg des Feldexperiments in der Versorgungsforschung. Die Qualitätskriterien, die an diese Form des Designs zu stellen sind, wurden in den CRT-Kriterien des CONSORT-Statements niedergelegt [16]. Diese werden hier als Basis auch für die Planung und Durchführung von CRTs in Organisationsstudien empfohlen. Die Datenauswertung sollte dann die spezielle, geschachtelte Datenstruktur angemessen berücksichtigen (z. B. Mehrebenenanalyse [39]).

Der Begriff der nicht-experimentellen Studien umfasst eine Reihe von Studiendesigns. In der Regel handelt es sich hierbei um Beobachtungsstudien, die meist in die drei Subtypen 1) Fall-Kontroll-Studie, 2) Querschnittsstudie und 3) Kohortenstudie unterteilt werden. Diese Studientypen werden in organisationsbezogenen Versorgungsstudien vor allem im Rahmen von versorgungsepi-

demologischen Fragestellungen eingesetzt. Als Qualitätsstandards für den Bericht von epidemiologischen Beobachtungsstudien gelten die STROBE-Kriterien (Strengthening The Reporting of Observational Studies in Epidemiology [53]). Die Beachtung dieser Kriterien bei der Planung und Durchführung solcher Formen von nicht-experimentellen Studien erhöht ihre Qualität. Diese Kriterien stellen die Mindeststandards für diese Studienarten in der Versorgungsforschung dar.

Fazit

Im Mittelpunkt dieses Teils des Memorandums III standen die Fragestellungen und Methoden der organisationsbezogenen Versorgungsforschung. Ziel war es, diese in ihren Grundzügen vorzustellen. Es konnte gezeigt werden, welche Mindeststandards in Bezug auf (a) die Auswahl der Untersuchungseinheiten, (b) die Messung von Merkmalen und (c) die Auswahl des Untersuchungsdesigns einzuhalten

sind. Weiterhin konnte ein Rahmenmodell für die Durchführung komplexer Studien vorgestellt werden, das bei Berücksichtigung der Standards der evidenzbasierten Medizin eine besonders hohe Studienqualität in der organisationalen Versorgungsforschung erlaubt. DZZ

Nachdruck aus: Gesundheitswesen 2009; 71: 777–789, © Georg Thieme Verlag KG Stuttgart – New York

Korrespondenzadresse

Prof. Dr. Holger Pfaff
Deutsches Netzwerk Versorgungsforschung e.V.
c/o IMVR – Institut für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaft der Humanwissenschaftlichen Fakultät und der Medizinischen Fakultät der Universität zu Köln (KöR)
Eupener Str. 129
50933 Köln
Tel.: 02 21 / 478 – 9 71 01
E-Mail: holger.pfaff@uk-koeln.de

Literatur

1. Albert U, Koller M, Lorenz W, et al.: Implementierung und Evaluation von Leitlinien auf nationaler Ebene: Entwicklung eines Konzeptes für die Stufe 3-Leitlinie Brustkrebs Früherkennung in Deutschland. *ZaeFQ* 98, 347–359 (2004)
2. Albert U, Schulz K, Kopp I: Methodik. In Albert U, Hrsg: Stufe-3-Leitlinie Brustkrebs-Früherkennung in Deutschland 1. Aktualisierung 2008. Zuckschwerdt Verlag, München 2008, 37–46
3. Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) und Arbeitsgruppe Epidemiologische Methoden. GPS – Gute Praxis Sekundärdatenanalyse: Revision nach grundlegender Überarbeitung. *Das Gesundheitswesen* 70, 54–60 (2008)
4. Bagozzi R, Baumgartner H: The evaluation of structural equation models and hypothesis testing. In Bagozzi R, Hrsg: Principles of marketing research. Blackwell Publishers, Cambridge 1994, 386–422
5. Black N, Murphy M, McKnee M, et al.: Consensus development methods: a review of best practice creating clinical guidelines. *J Health Services Res Policy* 4: 236–248 (1999)
6. Blalock HM: Social statistic. Mc Graw Hill, New York, Toronto, London 1960
7. Blalock HM, Blalock AB: Methodology in social research. Mc Graw Hill, New York, Toronto, London 1985
8. Böltken F: Auswahlverfahren. Eine Einführung für Sozialwissenschaftler. Teubner, Stuttgart 1976
9. Bollen W, Bollen KA: Structural equations with latent variables. John Wiley & Sons, New York 1989
10. Bortz J, Döring N: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (4. Aufl.). Springer, Berlin 2006
11. Boutron I, Moher D, Altman DG, et al.: The CONSORT group extending the CONSORT statement to randomized trials of nonpharmacologic treatment. Explanation and elaboration. *Ann Intern Med* 148, 295–309 (2008)
12. Brewin CR, Bradley C: Patient preferences and randomised controlled trials. *BMJ* 289, 315 (1989)
13. Brook R, Chassin M, Fink A, et al.: A method for the detailed assessment of the appropriateness of medical technologies. *Int J Tech Assess Health Care* 2, 53–63 (1986)
14. Bühner M: Einführung in die Test- und Fragebogenkonstruktion. Pearson, München 2006
15. Campbell M, Fitzpatrick R, Haines A: Framework for the design and evaluation of complex interventions to improve health. *BMJ* 321, 694–696 (2000)
16. Campbell MK, Elbourne DR, Altman DG: CONSORT statement: extension to cluster randomised trials. *BMJ* 328, 702–708 (2004)
17. CONSORT Group. <http://www.consort-statement.org/> [11.07.2009]
18. Cook TD, Shadish WR: Social experiments: some developments over the past fifteen years. *Annual Review of Psychology* 45, 545–580 (1994)
19. 55 Craig P, Dieppe P, Macintyre S, et al.: Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ* 337, 979–983 (2008)
20. Christiansen G: Evaluation – ein Instrument zur Qualitätssicherung in der Gesundheitsförderung. Eine Expertise. Bundeszentrale für gesundheitliche Aufklärung, Köln 2001
21. Delbecq AL, Van de Ven AH, Gustafson DH: Group techniques for program planning: a guide to nominal group and delphi process. Scott, Foresman and Company, Glenview, IL, USA 1975
22. Diamantopoulos A, Winklhofer HM: Index construction with formative indicators: an alternative to scale development. *J Marketing Res* 38, 269–277 (2001)
23. Diekmann A: Empirische Sozialforschung Grundlagen, Methoden, Anwendungen. Rowolt, Hamburg 2008
24. Dorroch H: Meinungsmacherreport. Wie Umfrageergebnisse entstehen.

- Steidl, Göttingen 1994
25. Erdfelder E, Faul F, Buchner A: GPOWER: A general power analysis program. *Behavior research methods. Instruments & Computers* 28, 1–11 (1996)
 26. Flick U, von Kardorff E, Steinke I: *Qualitative Forschung. Ein Handbuch.* Rowohlt Taschenbuch Verlag, Hamburg 2004
 27. Flick U: Design und Prozess qualitativer Forschung. In Flick U, von Kardorff E, Steinke I, Hrsg: *Qualitative Forschung. Ein Handbuch.* Rowohlt Taschenbuch Verlag, Hamburg 2004, 252–265
 28. Gabler S: Schneeballverfahren und verwandte Stichprobendesigns. *ZUMA-Nachrichten* 31, 47–69 (1992)
 29. Gabler S: *Stichproben in der Umfragepraxis.* Westdeutscher Verlag, Opladen 1997
 30. Geraedts M, Selbmann H, Ollenschläger G: Beurteilung der methodischen Qualität klinischer Messgrößen. *ZaeFQ* 96, 91–96 (2002)
 31. Greve W, Wentura D: *Wissenschaftliche Beobachtung. Eine Einführung.* Psychologie Verlags Union, Weinheim 1997
 32. Gross PA, Greenfield S, Cretin S, et al.: Optimal methods for guideline implementation: Conclusions from Leeds Castle meeting. *Med Care* 39, 1185–1192 (2001)
 33. Hox IJ, Maas CJM: Multilevel models for multimethod measurement. In Eid M, Diener E, Hrsg.: *Handbook of multimethod measurement in psychology.* Washington D.C. Am Psycholog Assoc 269–281 (2005)
 34. Kalton G: *Introduction to survey sampling.* Sage University paper, London 1983
 35. Kleespies C, Kaiser T, Sawicki PT: *Begriffe und Methoden der evidenzbasierten Medizin. Ein Glossar.* DiEM – Institut für evidenzbasierte Medizin. Köln 2005; <http://www.diem.de/publikationen.php> [11.07.2009]
 36. Kline RB: *Principles and practice of structural equation modeling.* Guilford, New York 2005
 37. Kühl S, Strodtholz P, Taffertshofer A: *Handbuch Methoden der Organisationsforschung. Quantitative und qualitative Methoden.* VS Verlag für Sozialwissenschaften, Wiesbaden 2009
 38. Langer B, Wetter T: Physician's attitudes towards health telematics – an empirical study. In Hasman A, Blobel B, Dudeck J, et al., Hrsg: *Medical infobahn for europe. Proceedings of MIE2000 and GMDS2000.* IOS Press, Amsterdam 2000, 1106–1111
 39. Langer W: *Mehrebenenanalyse. Eine Einführung für Forschung und Praxis.* VS Verlag für Sozialwissenschaften, Stuttgart 2004
 40. Lee RM: *Doing research on sensitive topics.* Sage, London 1993
 41. Merkens H: Auswahlverfahren, Sampling, Fallkonstruktion. In Flick U, von Kardorff E, Steinke I, Hrsg: *Qualitative Forschung. Ein Handbuch.* rororo, Hamburg 2000, 286–299
 42. Moher D, Jones A, Lepage L: Use of the consort statement and quality of reports of randomized trials a comparative before- and -after evaluation. *JAMA* 285, 1992–1995 (2001)
 43. Morfeld M, Wirtz M: *Methodische Ansätze in der Versorgungsforschung. Das Beispiel Evaluation der Disease-Management-Programme.* Bundesgesundheitsbl – Gesundheitsforsch – Gesundheitsschutz 49, 120–129 (2006)
 44. Murphy M, Black N, Lamping D, et al.: Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 2, 1–88 (1998)
 45. Oakley A, Strange V, Bonell C, et al.: Process evaluation in randomized controlled trials of complex interventions. *BMJ* 332: 413–416 (2006)
 46. Pfaff H, Bentz J: *Qualitative und quantitative Methoden der Datengewinnung.* In Schwartz FW, Badura B, Busse R, et al., Hrsg: *Das Public Health Buch. Gesundheit und Gesundheitswesen.* Urban & Fischer, München und Jena 2003, 419–434
 47. Pfaff H: *Konzept und Methodik der Versorgungsforschung.* *Die Psychiatrie* 2, 101–106 (2005)
 48. Reiter A, Fischer B, Geraedts M, et al.: QUALIFY – a tool for assessing quality indicators. *ZaeFQ*, 101, 683–688 (2007)
 49. Schafer JL, Graham JW: Missing data: our view of the state of the art. *Psychological Methods* 7(2), 147–177 (2002)
 50. Schnell R, Hill PB, Esser E: *Methoden der empirischen Sozialforschung.* Oldenbourg Verlag, München 2008
 51. Skrondal A, Rabe-Hesketh S: *Generalized latent variable modeling: multi-level, longitudinal, and structural equation models.* Chapman & Hall/CRC, Boca Raton 2004
 52. Steinke I: Gütekriterien qualitativer Forschung. In Flick U, von Kardorff E, Steinke I, Hrsg: *Qualitative Forschung. Ein Handbuch.* rororo, Hamburg 2000, 319–331
 53. Vandembroucke JP, von Elm E, Altman DG, et al.: Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Medicine* 4, 1628–1654 (2007)
 54. Vang J: The consensus development conference and the European experience. *Int J Technol Assess* 2, 65–76 (1986)
 55. Vogel H, Lemisz W, Liebeck H, et al.: Zur Bewertung des Gutachterverfahrens für die ambulante Verhaltenstherapie durch die GutachterInnen. *Verhaltenstherapie* 12, 228–231 (2002)
 56. Wegscheider K: *Statistische Verfahren in der Rehabilitationforschung.* In Bengel J, Koch U, Hrsg: *Grundlagen der Rehabilitationswissenschaften.* Springer, Berlin 2000, 395–408
 57. Wirtz M, Caspar F: *Beurteilerübereinstimmung und Beurteilerreliabilität.* Hogrefe-Verlag, Göttingen 2002
 58. Wirtz M: *Methoden zur Bestimmung der Beurteilerübereinstimmung.* In Petermann M, Hrsg.: *Handbuch der Psychologie – Psychologische Diagnostik.* Hogrefe, Göttingen 2006, 369–382
 59. Zelen M: Randomized consent designs for clinical trials. An update. *Statistics Med* 9, 645–656 (1990)
 60. Zwarenstein M, Treweek S, Gagnier JJ, et al.: Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 337, a2390 (2008)