



**Deutsches Netzwerk
Versorgungsforschung e.V.**

Entwurf*

Memorandum III:

„Methoden der Versorgungsforschung“

Herausgeber:

Holger Pfaff, Gerd Glaeske, Edmund A.M. Neugebauer, Matthias Schrappe
für den Vorstand des Deutschen Netzwerks Versorgungsforschung e.V.

Köln 2008

1. Auflage

* Stand: 17.10.2008

Gliederung

1. Präambel

Pfaff, H., Glaeske, G., Neugebauer E.A.M., Schrappe, M.

2. Methodische Grundlagen*

Glaeske, G.; Augustin, M., Abholz, H.; Banik, N.; Brüggengjürgen, B.; Hasford, J.; Hoffmann, W.; Kruse, J.; Lange, S.; Schäfer, T.; Schubert, I.; Trampisch, H.-J.; Windeler, J.

3. Methoden der organisationsbezogenen Versorgungsforschung

Pfaff, H.; Albert, U.-S.; Bornemann, R., Ernstmann, N., Gostomzyk, J., Gottwik, M.G.; Heller, G. ; Höhmann, U. Karbach, U.; Ommen, O. Wirtz, M.

4. Methoden der Lebensqualitätsforschung*

Koller, M.; Neugebauer, E.A.M.; Augustin, M.; Büssing, A.; Farin, E.; Klinkhammer-Schalke, M.; Lorenz, W.; Münch, K.; Petersen, C.; Steinbüchel, N.v.; Wieseler, B.

Anhang

Autorenverzeichnis

* in Kooperation mit dem Deutschen Netzwerk Evidenzbasierte Medizin e.V. (DNEbM)

1. Präambel

Das Ziel des vorliegenden Memorandums „Methoden der Versorgungsforschung“ („Memorandum III“) ist es, die methodischen Grundprinzipien und Mindeststandards darzustellen, die bei der Durchführung und Veröffentlichung von Versorgungsforschungsstudien beachtet werden sollten. Dies soll dazu dienen, die methodische Qualität der Versorgungsforschungsstudien zu sichern und für einheitliche Standards auf dem Gebiet der Versorgungsforschung zu sorgen. Dieses Memorandum wendet sich daher sowohl an die Versorgungsforscher, die Studien planen, durchführen und veröffentlichen, als auch an Gutachter, die Anträge und Publikationen zu Versorgungsforschungsstudien bewerten. Das vorliegende Memorandum ergänzt das erste „Memorandum zur Versorgungsforschung in Deutschland“, das im September 2003 auf dem 2. Deutschen Kongress für Versorgungsforschung vorgestellt wurde und das zweite Memorandum „Konzeptionelle, methodische und strukturelle Voraussetzungen der Versorgungsforschung“, das im Jahr 2005 auf dem 4. Deutschen Kongress für Versorgungsforschung verabschiedet wurde. Anders als die bisherigen Memoranden, die programmatisch orientiert waren, widmet sich dieses Memorandum den Grundlagen der Versorgungsforschung und will so zur weiteren Professionalisierung dieses Forschungsfeldes beitragen.

Das vorliegende Memorandum ist das Ergebnis der Zusammenarbeit mehrerer Arbeitsgruppen des Deutschen Netzwerks Versorgungsforschung (DNVF). Das DNVF ist ein Zusammenschluss von derzeit 41 klinischen und gesundheitswissenschaftlichen Fachgesellschaften, die gemeinsam das Ziel verfolgen, die an der Versorgungsforschung im Gesundheitswesen beteiligten Wissenschaftler zu vernetzen, Wissenschaft und Versorgungspraxis zusammenzuführen sowie die Versorgungsforschung insgesamt zu fördern. Die einzelnen Arbeitsgruppen sind interdisziplinär durch Experten aus den Fachgesellschaften für die jeweiligen Themenbereiche besetzt. Jede Fachgesellschaft war aufgefordert, Experten aus ihren Gesellschaften zu benennen. Bei Bedarf wurden für die Erarbeitung der methodischen Prinzipien zusätzlich externe Experten hinzugezogen. Die Einarbeitung der mit * gekennzeichneten Methoden erfolgte in enger Abstimmung mit dem Deutschen Netzwerk Evidenzbasierte Medizin (DNEbM), das korrespondierendes Mitglied des Deutschen Netzwerks ist.

Das Memorandum III versteht sich als „work in progress“. Die hier vorliegende Version repräsentiert somit einen ersten Zwischenstand. Das Memorandum gliedert sich in ein allgemein einführendes Kapitel zu den methodischen Grundlagen der Versorgungsforschung sowie zwei folgenden Kapitel zu zentralen Versorgungsforschungsgebieten. Als zentrale Themengebiete wurden für die erste Auflage die „Methoden der organisationsbezogenen Versor-

gungsforschung“ und die „Methoden der Lebensqualitätsforschung“ ausgewählt. Ein weiteres Themengebiet, die „Methoden der Gesundheitsökonomie“, befindet sich derzeit im Entstehungsprozess. Es ist geplant, bei den folgenden Auflagen des Memorandums III weitere Themenbereiche folgen zu lassen, etwa zu den Themen Krankheitsregister, Patientenzufriedenheit, qualitative Methoden sowie Erhebung und Auswertung von Patienteninformationen.

Die Erarbeitung der methodischen Prinzipien erfolgte in einem mehrstufigen, standardisierten Delphi-Prozess. Zunächst wurden die Beiträge der Arbeitsgruppen-Experten unter Federführung des jeweiligen AG-Sprechers zusammengeführt und innerhalb der Arbeitsgruppen in mehreren Treffen mit anschließenden Delphi-Prozessen konsentiert. Daraufhin wurden die Ergebnisse der Arbeitsgruppen zwischen den Arbeitsgruppen-Sprechern sowie mit dem Deutschen Netzwerk Evidenzbasierte Medizin konsentiert. Anschließend wurde ein Entwurf des Memorandums III im Rahmen einer hierzu einberufenen Klausurtagung mit den Mitgliedern des DNVF im Mai 2008 überarbeitet und abschließend im Oktober 2008 im Rahmen der Mitgliederversammlung anlässlich des 7. Deutschen Kongresses für Versorgungsforschung in Köln von 41 Mitgliedsfachgesellschaften verabschiedet. Koordiniert wurde der gesamte Abstimmungsprozess durch eine Steuerungsgruppe, deren Mitglieder die Herausgeber der vorliegenden Publikation sind.

Köln, Oktober 2008

2. Methodische Grundlagen der Versorgungsforschung

Glaeske, G.; Augustin, M.; Abholz, H.; Banik, N.; Brüggengjürgen, B.; Hasford, J.; Hoffmann, W.; Kruse, J.; Lange, S.; Schäfer, T.; Schubert, I.; Trampisch, H.-J.; Windeler, J.

2.1. Selbstverständnis der Versorgungsforschung

Gesetzliche, politische und ökonomische Eingriffe in das Versorgungssystem modifizieren Versorgungsstrukturen und -prozesse, derzeit in der Regel ohne die Auswirkungen für die Nutzerinnen und Nutzer wissenschaftlich zu überprüfen. Dadurch können häufig weder möglichen Fehlentwicklungen frühzeitig entgegengewirkt noch positive Effekte frühzeitig nachgewiesen und konsolidiert werden. Durch fehlende Evaluation werden potentiell gravierende Auswirkungen sowohl für gesunde als auch erkrankte Menschen auch hinsichtlich der ökonomischen Situation im Gesundheitswesen nicht transparent. Dadurch wird eine Qualitätsentwicklung im Versorgungssystem deutlich behindert. Der Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (früher Sachverständigenrat für die Konzentrierte Aktion im Gesundheitswesen) hat in seinen Gutachten aus den Jahren 1995, 1997, 2001 und zuletzt aus dem Jahr 2007 eine Aktivierung der defizitären Versorgungsforschung in Deutschland angemahnt. Dieser Mahnung hat sich der außerordentliche Deutsche Ärztetag im Februar 2003 angeschlossen. Im Jahr 2005 beschloss der Deutsche Ärztetag ein eigenes Förderprogramm für Versorgungsforschung, das gezielt Themenbereiche adressiert, in denen aktuell Forschungsdefizite bestehen.

Hinsichtlich der Gesundheits- und Krankenversorgung sowie deren Rahmenbedingungen ergeben sich folgende Aufgaben:

- Beschreibung und Analyse der Versorgungssituation („Ist“- und Defizit Analysen),
- hierauf aufbauend Entwicklung von Versorgungskonzepten,
- wissenschaftliche Begleitung der Umsetzung (Implementierung) neuer Versorgungskonzepte,
- Evaluierung neuer und alter Versorgungskonzepte unter realen Bedingungen („medizinischer Alltag“) auf der Ebene der Gesamtbevölkerung oder relevanter Populationen.

Zentrale Fragen und Schwerpunkte der Versorgungsforschung betreffen die Prävalenz, Ursachen und Auswirkungen von Unter-, Über- und Fehlversorgung, die Interaktionen zwischen Diagnostik und Therapie, transsektorale Verläufe und komplexe Interdependenzen in der Versorgung.

Neben der Beschreibung und Analyse von Inanspruchnahme und Versorgung leistet Versorgungsforschung Beiträge zur Erweiterung des Wissensstandes über Effektivität, uner-

wünschten Folgen und gesundheitsökonomische Auswirkungen sowie zur Akzeptanz und Implementierbarkeit therapeutischer, diagnostischer, rehabilitativer und präventiver Verfahren im Rahmen unseres Gesundheitssystems und seiner gesetzlichen Regelungen. Versorgungsforschung liefert Planungsdaten und analysiert Prozesse innerhalb der Gesundheitsversorgung unter Alltagsbedingungen - eines Alltags, der vielfach zutreffend wegen der nahezu unübersehbaren Anzahl von effektmodifizierenden Variablen als „black-box“ charakterisiert wird.

Versorgungsforschung kann in diesem Zusammenhang unterschiedliche Blickrichtungen einnehmen und die Versorgung z.B. aus Sicht

- der Nutzerinnen und Nutzer
- der Anbieter (Ärzte, Apotheker, Krankenhäuser, pharmazeutische Hersteller etc.)
- der Organisatoren (gesetzliche Kranken- oder Pflegekassen (GKV, PKV) oder private Kranken- und Pflegeversicherung (PKV, PPV) oder
- der Politik

analysieren.

Dieses Memorandum benennt die Methoden und Instrumente, mit deren Hilfe Studien in der Versorgungsforschung geplant und durchgeführt sowie deren Güte bewertet werden können.

2.2. Aktuelle Themenfelder der Versorgungsforschung

Zu den derzeit vordringlich bearbeiteten und zu bearbeitenden Fragestellungen und Themenbereichen der Versorgungsforschung gehören:

- Analysen auf der Ebene des Gesundheitssystems, Organisationsforschung. Hierzu gehören Fragen der Organisation der stationären, ambulanten und integrierten Versorgung, z.B. Ermittlung des Versorgungsbedarfes, Implementierungsmöglichkeiten für Disease Management- oder strukturierte Behandlungsprogramme, Analysen von Schnittstellen und deren Auswirkungen (transsektoral, inter- bzw. intraprofessionell z.B. „Professionenmix“), sowie Folgen gesundheitspolitischer Regelungen, z.B. im Rahmen von neuen Vergütungs- oder Entgeltformen wie Pay-for-Performance (P4P) etc.
- Fragen des Zugangs und der Inanspruchnahme aus Sicht der Nutzerinnen und Nutzer unter Berücksichtigung von Präferenzen, psychologischer, ökonomischer, sozialer und kultureller Dimensionen
- Ist-Analyse, Leistungsmessung und Leistungsbewertung im Gesundheitswesen einschließlich gesundheitsökonomischer Aspekte und dem Vergleich der Auswirkungen verschiedener Methoden in der Behandlung, der Diagnostik, von Screening- oder Vorsorgeverfahren, von Rehabilitation, Pflege und Prävention (Entwicklung und Implementierung

ergebnisbezogener Qualitätsindikatoren, Messung von mittel- und langfristigen Effekten etc.)

- Qualitätssicherung und Qualitätsmanagement: z.B. Implementierung und Umsetzung von Leitlinien, Evaluation der Auswirkungen auf die Versorgung, einschließlich gesundheitsökonomischer Auswirkungen sowie von Risiken oder nachteiligen Effekten. Analyse der Faktoren, die eine Umsetzung von Leitlinien in die Versorgungspraxis determinieren sowie deren Akzeptanz und Adhärenz bei Leistungserbringern und -empfängern oder Interaktionen mit anderen Leitlinien, Standards und Empfehlungen etc.
- Monitoring und Prüfung der Wirksamkeit der aus Studien an selektierten Patientengruppen und unter Idealbedingungen gewonnenen Ergebnisse in die allgemeine Versorgungspraxis unter komplexen Anwendungsbedingungen. Identifikation von relevanten Subgruppen innerhalb der Bevölkerung. Identifikation innovationsförderlicher und -behindernder Faktoren
- Patienteninformation, Patientenpräferenz und kommunikative Prozesse als Determinanten für Inanspruchnahme (informed consent, Partizipation bei der Wahl der diagnostischen und therapeutischen Strategie und der Definition des Behandlungsziels, sowohl aus Sicht der Akteure wie aus Sicht der Nutzerinnen und Nutzer)
- Analyse versorgungsbezogener Problematiken bei Angehörigen vulnerabler Gruppen, z.B. Migrantinnen und Migranten, Obdachlose, arme ältere Menschen, Menschen mit niedrigem sozialökonomischen Status, niedrigem Bildungsniveau etc.
- Politikfolgenforschung, z.B. die Untersuchung der Konsequenzen von gesundheitspolitischen Entscheidungen, veränderten Strukturen oder Rahmenbedingungen auf z.B. der Angebotsakzeptanz, Entscheidungen der Ärzte oder veränderte finanzielle Belastungen der Krankenkassen

Versorgungsforschung ist nicht an spezifische Versorgungssektoren oder -schwerpunkte gebunden. Häufig ist gerade die transsektorale Betrachtung von Behandlungsabläufen erforderlich. Diese steht in Deutschland noch immer viel zu wenig im Mittelpunkt, d.h. die sektorale Betrachtung der Versorgung (ambulant, stationär, rehabilitativ) herrscht noch immer vor. Die Konsequenz aus solchen sektororientierten Versorgungsformen sind häufig Ineffizienzen und Belastungen für die Patientinnen und Patienten, die z.B. durch eine defizitäre Kooperation und Kommunikation zwischen den verschiedenen Sektoren zustande kommen (Schnittstellenprobleme). Aus den Ergebnissen der Versorgungsforschung können solche Defizite aufgedeckt und, wenn es erforderlich erscheint, Vorschläge für optimierte Abläufe abgeleitet werden.

2.3. Aufgaben der Versorgungsforschung

Es ist Aufgabe der Versorgungsforschung, den Zusammenhang zwischen den eingesetzten Ressourcen und den erreichten Therapieergebnissen zu analysieren und daraus Aussagen über die Effektivität und Effizienz sowie über Verteilungs- und Zugangsbarrieren abzuleiten. In diesem Zusammenhang müssen auch die konkreten Partizipationsmöglichkeiten von Versicherten und Patienten, deren Präferenzen und Rechte berücksichtigt werden.

Die Versorgungsforschung beschäftigt sich aber nicht nur mit der gesamten Gesundheits- und Krankenversorgung und deren kurz-, mittel- und langfristigen Outcomes, sondern auch mit der Frage nach den bestmöglichen Organisationsformen für die Versorgung. Analysiert wird die Beziehung zwischen Struktur-, Prozess- und Ergebnisqualität und deren Folgen, z.B. bezüglich der Honorierungs- oder Entgeltsysteme, der Bedarfsplanung und –steuerung, auf Unter-, Über- und Fehlversorgung. Untersucht wird der Einfluss neuer Kooperations- oder Professionskonzepte („Professionenmix“ und Delegationsverfahren bei der Behandlung, z.B. mit Blick auf die Pflege- oder Heilberufe). Die Versorgungsforschung kann in diesem Feld dazu beitragen, die Anforderungen und Kooperationsmöglichkeiten der ärztlichen und sonstigen Gesundheitsberufe im Bezug auf arbeitsteilige, qualitäts- und effizienzoptimierende Potenziale zu prüfen und aus den Ergebnissen Anforderungen für die Ausbildung (z.B. im Sinne der Akademisierung) oder für neu zu konzipierende Gesundheitsberufe abzuleiten. Hinzu kommen Fragen der regionalen Charakterisierung der Gesundheitsversorgung. Da allen Bürgerinnen und Bürgern eine gleich gute qualitative Versorgung als Grundrecht zusteht, sind regionale (großräumige oder kleinräumige) Abweichungen von einer erreichbaren Versorgungs- und Zugangsqualität häufig als Defizit zu bewerten. Die Versorgungsforschung kann etwaige Defizite aufdecken, analysieren und mit wissenschaftlicher Evidenz die rationale Basis für möglicherweise notwendige Veränderungen schaffen.

Die Ergebnisse aus der Versorgungsforschung können die Versorgung der Bevölkerung direkt oder indirekt verbessern:

- Sie können bei erkannten Defiziten in der Qualität und Organisation der medizinischen Versorgung Optimierungsprozesse anregen und die Evidenzbasis für deren Ausgestaltung liefern.
- Sie können zur Definition von Anreizen für eine verbesserte Versorgung führen, die z.B. im Rahmen selektiver Verträge, Konzepten der integrierten Versorgung oder Medizinischen Versorgungszentren eingesetzt werden.
- Sie können Anforderungen für bessere Produkte oder Verfahren, für Produkt- und Verfahrensalternativen oder Vorschläge für neue Behandlungswege aufzeigen und dadurch der klinischen Forschung neue Impulse geben.

- Sie können Bewertungen von Behandlungsabläufen innerhalb der Sektoren und unter gemeinsamer Betrachtung aller Sektoren ermöglichen, um Optimierungsprozesse einzuleiten.
- Sie können Hinweise dafür liefern, welche Leistungen in andere Versorgungsformen / -sektoren verlagert werden können und gleichzeitig Parameter zum Monitoring der Qualität aufzeigen.
- Sie können helfen, die ökonomischen Konsequenzen unterschiedlicher medizinischer Versorgungsstrategien abzubilden und damit effizienzoptimierend zu wirken.
- Sie geben Hinweise auf die Wirksamkeit von Methoden in der medizinischen Versorgung unter den Bedingungen unseres Gesundheitssystems („letzte Meile“ der Anwendung).

Insgesamt fördert eine aktive Versorgungsforschung eine Evaluationskultur, die im deutschen Gesundheitswesen noch immer wenig entwickelt ist. Insbesondere in Bereichen, in denen eine Evaluation nicht gesetzlich vorgeschrieben ist, die aber wegen ihrer Versorgungsrelevanz und / oder neuen Organisations- und Anreizstrukturen dringend in Bezug auf die Prozess- und Ergebnisqualität einer Begleitforschung bedürften, sollten angemessene Evaluationsstrategien in Zukunft obligat werden. Versorgungsforschung kann daneben auch aus eigener Initiative weitere und speziellere Studien in Gang setzen, die mit geeigneten Studiendesigns definierte Fragestellungen und Hypothesen untersuchen.

In allen Themenfeldern ist es wichtig, Forschungsfragen und wissenschaftliche Zielsetzung transparent darzustellen und zu kommunizieren. Daneben ist darzulegen, in welchem Wertungskontext und Interessenzusammenhang die Forschung durchgeführt wird.

Versorgungsforschung soll versorgungsrelevant sein und auf den Nachweis eines individuellen Nutzens zielen. Auch wenn dieser nur in Gruppen gemessen werden kann, strebt sie damit einen bevölkerungsbezogenen Erkenntnisgewinn an. Die Vorgehensweise muss nachvollziehbar sein, die Studien müssen mit angemessener Methodik durchgeführt werden.

2.4. Methoden der Versorgungsforschung

Die Versorgungsforschung greift auf Theorien und Methoden der beteiligten Disziplinen (Medizin, deskriptive und analytische Epidemiologie, der Sozialwissenschaften, der Psychologie, der Pharmazie und der Ökonomie) zurück. Neben quantitativen Forschungsmethoden sind qualitative Ansätze relevant.

Einer speziellen Forschungsmethodik der Versorgungsforschung bedarf es nicht. Allerdings können spezifische Fragen der Versorgungsforschung und der Zugang zu speziellen Datenkörpern Anlässe für methodische Weiterentwicklungen sein. Im Kontext der Routinedaten stellen sich beispielsweise methodische Anforderungen an die Stichprobenziehungen und

das Datenlinkage, an Validierungsstrategien der verschiedenen Variablen, an die Operationalisierung von Falldefinitionen (Festlegung der Studienpopulation auf der Basis von Diagnose und / oder Inanspruchnahmedaten), von Inzidenz, Prävalenz oder auch von Qualitätsindikatoren. Ebenso sind Verfahren zur Kontrolle von Confoundern oder zur Risikoadjustierung (z.B. Propensity Scores) darzulegen. Methodischer Entwicklungsbedarf besteht beispielsweise hinsichtlich so genannter „unmeasured confounders“, insbesondere für qualitative Studien.

Auch randomisierte klinische Studien können im Rahmen der Versorgungsforschung von Bedeutung sein, wenn die Forschungsfrage ein experimentelles Design (Standardisierung, Randomisierung) erfordert. Die Art und Weise der Randomisierung bzw. die Wahl von Kontrollgruppen stellt eine zentrale Herausforderung für evaluative Studien in der Versorgungsforschung dar. Hierzu liegen noch wenig verallgemeinerbare Erfahrungen vor. Relevante Beispiele sind überwiegend den sogenannten „pragmatischen randomisierten Studien“ zuzurechnen.

Versorgungsforschung definiert sich über ihre Fragestellungen. Ihre Aufgabe ist es, Forschungsbedarfe zu identifizieren, zu priorisieren und geeignete Studiendesigns zur Beantwortung der Fragen anzuwenden. Eigens erhobene Primärdaten und / oder bereits vorhandene Daten (Sekundärdaten, Registerdaten usw.) werden mit einer angemessenen Methodik ausgewertet. Die Auswahl der Methodik ist dabei abhängig von den Fragestellungen.

Grundlage der Anwendung unterschiedlicher Methoden ist eine detaillierte (Versuchs)Planung. Bei gut geplanten Untersuchungen liegt die gesamte Auswertungsstrategie bereits vor dem Beginn der Datenerhebung fest. Je mehr es gelingt, Vorwissen in die Planung einzubringen, desto weniger müssen vorher nicht geplante Methoden nachträglich in die Auswertung eingeführt werden. Ziel jeder Untersuchung sollte es sein, dass die Ergebnisse so wenig verzerrt wie möglich sind. In Abhängigkeit von der Fragestellung sind dabei unterschiedliche Aspekte besonders bedeutsam: Sollen beispielsweise Zustandsbeschreibungen erfolgen (z.B. Prävalenzen), dann muss auf Repräsentativität geachtet werden (= hohe externe Validität). Geht es um die hypothesengeleitete Prüfung von Kausalbeziehungen, muss vor allem die Fairness von Vergleichen (z.B. gleiche Ausgangsbedingungen, standardisierte Instrumente, valide Erfassung der Determinanten und Endpunkte = hohe interne Validität) sichergestellt sein.

2.4.1. Methodische Kriterien der Versorgungsforschung

Zur Beurteilung der Qualität von Studien aus der Versorgungsforschung werden Kriterien vorgeschlagen, die abgeprüft werden sollten (angelehnt an Donner-Banzhoff et al, 2007):

- Fragestellung
- Studiendesign

- Zielvariablen
- Datenbasis
- Adressat
- Interesse
- Qualitätskriterien
- Trugschlüsse

2.4.1.1. Fragestellung

Die Fragestellungen einer Untersuchung müssen klar erkennbar sein, damit die Wahl der Methodik beurteilt und das Ergebnis richtig interpretiert werden können. Die Fragestellungen in der Versorgungsforschung können unterschiedlichen Kategorien zugeordnet werden:

- Quantifizierung eines Versorgungsproblems als „Ist“-Analyse (z.B. einer Behinderung, einer Erkrankung, von Schnittstellenproblemen zwischen unterschiedlichen Versorgungssektoren usw.).
- Defizitorientierte Fragestellungen werden dann untersucht, wenn der Bezug zu einem Standard hergestellt wird und dieser Standard als ausreichend sicher akzeptiert werden kann.
- Das „Warum“ eines Befundes kann in Studien mit analytischen Fragestellungen untersucht werden. Solche Studien können Ergebnisse aus defizitorientierten Fragestellungen sinnvoll ergänzen, um Aufschluss darüber zu bekommen, welche Barrieren für die Umsetzung sinnvoller Interventionen oder Rahmenbedingungen identifiziert werden können (siehe das Beispiel unter 5.)
- Messung versorgungsbezogener Auswirkungen von Veränderungen im Versorgungssystem (z.B. durch neue Versorgungsformen im Rahmen von Interventionsstudien). Fragestellungen können z.B. Wirksamkeit (efficacy), Sicherheit, Akzeptanz und/oder Kosten sein.

2.4.1.2. Studiendesign

Es wurde schon weiter oben darauf hingewiesen, dass sich das Studiendesign an der Fragestellung zu orientieren hat. Grundsätzlich kommen in der Versorgungsforschung unterschiedliche Studientypen in Frage, zu denen Fall-Berichte oder –Serien, Querschnitts-Studien, Fall-Kontroll-Studien, Kohorten-Studien sowie randomisierte Interventionsstudien gehören. Außerdem sind explorative Studien mit unterschiedlichen Methoden der empirischen Sozialforschung (teilnehmende Beobachtung, Befragungen, Fokusgruppen etc.) möglich.

Alle Studienformen, die im Rahmen der Versorgungsforschung angewendet werden, sollten als konstitutive Merkmale aufweisen

- a) eine differenzierte Studienplanung und
- b) auf jeder Stufe der Durchführung und Auswertung eine adäquate Qualitätssicherung.

Ad a): Die Versuchsplanung sollte neben dem Studiendesign (Studienprotokoll) einschließlich einer Betrachtung zum benötigten Studienumfang und der statistischen Aussagekraft (Stichprobenkalkulation, Powerberechnung) ein so genanntes Operationshandbuch enthalten, das im Detail sämtliche Schritte der Planung, Durchführung, Auswertung der jeweiligen Studie und die jeweiligen Instrumente beschreibt und festlegt. Das Operationshandbuch sollte einen solchen Detaillierungsgrad aufweisen, dass durch eine externe Forschergruppe eine Replikation der beschriebenen Studie möglich würde. Unverzichtbarer Bestandteil bereits des Designs ist eine begründete Stichprobenkalkulation.

Bei der Planung und Durchführung sind die einschlägigen Standards und Empfehlungen der betroffenen Fachgebiete zu berücksichtigen.

Ad b): Die Qualitätssicherung sollte alle Schritte der Datengewinnung, Datenverarbeitung und Auswertung umfassen. Bei multizentrischen Studien von grundsätzlicher gesundheitspolitischer Relevanz sollten möglichst durch eine kleine externe Expertengruppe Unabhängigkeit und höchste wissenschaftliche Standards eingebracht werden.

2.4.1.3. Zielvariablen

In allen Studien zur Versorgungsforschung müssen Zielvariablen, Endpunkte bzw. Entscheidungskriterien weitestgehend a priori definiert, begründet, nachvollziehbar dokumentiert sowie transparent dargestellt werden. Ausnahmen hiervon betreffen bestimmte qualitative Verfahren und hypothesengenerierende Untersuchungen. Diese Kriterien müssen sich unmittelbar an der versorgungsrelevanten Fragestellung orientieren. Es muss dabei berücksichtigt werden, dass Zielvariablen bezüglich der Struktur-, Prozess- und Ergebnisaspekte unterschiedliche Sensitivität und Spezifität aufweisen. Da die Versorgungsforschung in vielen Studien die Perspektive der Nutzerinnen und Nutzer des Gesundheitssystems einnimmt, sind Zielvariablen zur Darstellung patientenrelevanter Outcomes von besonderem Interesse.

2.4.1.4. Datenbasis

Studien zur Versorgungsforschung können unterschiedliche Datenzugänge nutzen.

1. Primärdaten (prolektive Studiendesigns), d.h. es werden speziell für das jeweilige Forschungsvorhaben Daten erhoben. Solche Primärdatenerhebungen sind mit einem deutlich höheren Studienaufwand verbunden, wie z.B. Rekrutierung von Kliniken, Praxen, Apotheken

und Patienten. Die Reichweite dieser Primärerhebungen ist meist regional und institutionell begrenzt.

2. Sekundärdaten (Datenbankstudien, retroaktiv). Typisch ist die Nutzung von routinemäßig erhobenen und gespeicherten Daten, die nicht speziell für Forschungszwecke in der Versorgungsforschung erhoben und zusammengetragen wurden (Sekundärdaten). Mit diesen Sekundärdaten sind die bereits beschriebenen Studienansätze wie Kohorten-Studien, Fall-Kontroll-Studien, Querschnittsanalysen und Einzelfallbeobachtungen möglich. Die Daten können mit und ohne Personenbezug ausgewertet werden. Es sind auch datentechnische Verknüpfungen („linkage“) über ein Versichertenpseudonym möglich. Auf diesem Wege können dann z.B. Therapieverläufe von Personen über verschiedene Zeiträume unter Einschluss aller verfügbaren Daten, auch über Behandlungssektoren hinweg, beschrieben werden. Die gängigen Sekundärdaten der gesetzlichen Krankenkassen („Routinedaten“ zur Abrechnung von erbrachten Leistungen) enthalten z.B. personenbeziehbare Daten über verschiedene medizinische Einrichtungen, verschiedene Arztgruppen, über Arbeitsunfähigkeiten, über Diagnosen im Krankenhaus und in der ambulanten Versorgung, über Arzneimittel-, Heil- und Hilfsmittelverordnungen, über Krankengeldzahlungen usw.. Untersuchungen mit Sekundärdaten sind relativ kostengünstig und zeitsparend, da keine Patientinnen und Patienten gezielt rekrutiert werden müssen. Diese Daten haben Vor- und Nachteile. Der Vorteil liegt vor allem in der relativ einfachen und kostengünstigen Zugänglichkeit sowie in der Datenbreite (sektorübergreifende Erhebung). Ein weiterer wesentlicher Vorteil besteht darin, dass eine definierte Bevölkerungsgruppe erfasst wird (dadurch ist der „Nenner“ bekannt und die Population kann hinsichtlich ihrer Repräsentativität für die Bevölkerung, über die eine Aussage getroffen wird, beschrieben werden) und – ein weiterer wichtiger Aspekt - dass die Behandlungsabläufe nicht durch Untersuchungsbedingungen beeinflusst werden. Damit gehen aber eine Vielzahl regulatorischer, gesetzlicher, struktureller und organisatorischer Einflüsse in die Sekundärdaten ein, die bei einer Interpretation der Ergebnisse erkannt und berücksichtigt werden müssen.

Als Limitationen sind zu nennen, dass keine Angaben über die Selbstmedikation oder über den Arzneimittelverbrauch im stationären Sektor vorliegen. Auch sind keine Angaben von anderen, nicht zu Lasten der jeweiligen Krankenkasse oder Versicherung abrechenbaren Leistungen vorhanden (z.B. Inanspruchnahme von individuellen Gesundheitsleistungen (IGeL)). Bei den Diagnose- und Leistungsdaten müssen Rechtfertigungs- oder Abrechnungsstrategien berücksichtigt werden. Unter Umständen kommt es auf der Basis solcher Daten zu einer Fehleinschätzung der Morbidität. Auch sind Codierfehler denkbar. Die Möglichkeit des Versicherungswechsels von Patienten muss ebenfalls berücksichtigt und bezüglich der Auswirkungen auf das Ergebnis abgeschätzt werden. Übermittlungsprobleme führen zu einem Informationsverlust. In der Regel sind auch keine klinischen Daten oder Daten über

den Personenstatus vorhanden (Gewicht, Raucher- oder Nichtraucherstatus, Daten zum Alkoholkonsum, zur Bewegung oder Ernährung, zur sozialen Schichtzugehörigkeit usw.). Die Vor- und Nachteile von Sekundärdatenanalysen gegenüber Primärdaten müssen für jede Fragestellung präzise abgewogen werden. Es können auch Verknüpfungen von Sekundärdatenanalysen mit nachfolgenden prolektiven Studiendesigns sinnvoll sein, sofern dadurch in konkreten Fragestellungen Mängel oder Unwägbarkeiten von Sekundärdatenanalysen gezielt verringert und mit deren genuinem Vorteil der schnelleren und kostengünstigeren Verfügbarkeit verbunden werden können. Dies kann den Zweck von Validierungen bzw. der Qualitätskontrollen der konkreten Sekundärdatenanalyse erfüllen oder zur Beantwortung weitergehender Fragestellungen geschehen.

Es erscheint zweckmäßig, die Qualität und Vollständigkeit der Daten stichprobenartig zu prüfen auf:

- Korrektheit von Kodierungen (insbesondere ICD-10, DRG),
- Vollständigkeit der Stammdaten in Bezug auf Versicherungsphasen,
- Korrektheit von Stammdaten in Bezug auf Alter und Geschlecht
- Vollständigkeit und Korrektheit von Verschlüsselungen (z.B. ATC / DDD)

Bei den Sekundärdatenanalysen müssen, wie bei anderen Studienansätzen, Studienpläne oder Studienprotokolle erstellt und während der Untersuchung eingehalten werden. Für den Personenbezug eines Behandlungsablaufs ist eine Pseudonymisierung der Daten erforderlich.

Anforderungen an den Umgang mit Sekundärdaten sind in der Leitlinie Gute Praxis Sekundärdatenanalyse und der Leitlinie für Gute Epidemiologische Praxis enthalten (AGENS et al., 2008).

3. Nutzung von Registerdaten (retrolektive Studiendesigns). Der Wert von Krankheits- oder Arzneimittelregistern, die regional bzw. national für bestimmte Krankheiten oder ausgewählte medizinische Interventionen / Arzneimitteltherapien über längere Zeiträume geführt werden, hängt vor allem von der Vollzähligkeit der Rekrutierung, der Vollständigkeit und Validität der gespeicherten Daten sowie der Qualität der Nachverfolgung ab.

2.4.1.5. Adressaten

Vor der Durchführung einer Studie im Rahmen der Versorgungsforschung sollte klar sein, für wen die Ergebnisse der Studie Bedeutung haben sollen bzw. wer diese Ergebnisse in die Versorgungspraxis umsetzen soll und kann. Sprechen die Ergebnisse das System (Makroebene), größere Organisationseinheiten oder Netzwerke im System (Mesoebene) oder Individuen wie Ärzte, Patienten usw. (Mikroebene) an, sind sie in ihren Konsequenzen entsprechend vorzustellen und zu diskutieren.

2.4.1.6. Interessen

Grundsätzlich folgen alle hochwertigen Studien bestimmten Interessen, zumal wenn ihre Durchführung erheblichen Ressourceneinsatz voraussetzt. Dies gilt in der Versorgungsforschung im medizinischen Versorgungssystem insbesondere, da alle Ergebnisse ökonomisch interessierte Anbieter einerseits und Nachfrager (z.B. Kassen) bzw. Patienten andererseits tangieren. So können Ergebnisse aus der Versorgungsforschung Erwartungen von Berufs- oder Anbietergruppen im Hinblick auf ihren Einfluss oder Gewinn ebenso dämpfen wie erweitern, sie können die Nachfragerseite stützen und argumentativ stärken. In diesem Zusammenhang werden die Interessenskategorien „Enthusiasmus“ und „Skepsis“ vorgeschlagen. Die Position des „Enthusiasmus“ wird in diesem Zusammenhang möglicherweise eher von den absatz- und umsatzinteressierten Herstellern oder Anbietern vertreten, die der „Skepsis“ eher von Krankenkassen oder solchen Repräsentanten des Versorgungssystems eingenommen werden, für die eine sichere Risiko-Nutzen-Abschätzung besondere Bedeutung hat. In jeder Untersuchung sind Interessenskonflikte von allen Beteiligten vollständig zu offenbaren und transparent zu dokumentieren.

Zur Vermeidung von publication bias ist die Einführung einer a priori Registrierung von Forschungsvorhaben im Bereich der Versorgungsforschung anzustreben.

2.4.1.7. Qualitätskriterien

Die Qualität und Aussagekraft einer Untersuchung in der Versorgungsforschung wird von inhaltlichen, methodischen, statistischen und praktischen Aspekten determiniert. Eine Einschätzung ist immer nur in Bezug auf die spezifische Fragestellung möglich. In der folgenden Tabelle 2.1 sind für einige Arten der Fragestellung und der Datenbasis wichtige Dimensionen aufgeführt.

Tabelle 2.1 Qualitätskriterien (nach Donner-Banzhoff et al., 2007, modifiziert)

Fragestellung	
<u>Kategorie</u>	<u>Methodische Fragen</u>
Deskriptiv	Ist sicher gestellt, dass die interessierenden Fälle lückenlos erfasst sind (konsequente Rekrutierung)? Qualitätssicherungsmaßnahmen [Audit]? Ist der Nenner, auf den das Versorgungsproblem bezogen wird, ausreichend präzise definiert?
Defizitorientiert	Ist der Standard, an welchem das Versorgungshandeln gemessen wird, durch valide klinische Studien mit relevanten Outcomes belegt?
Analytisch	(Kriterien zur Beurteilung für quantitative Epidemiologie s. GEP) (Kriterien zur Beurteilung qualitativer Gesundheitsforschung s. einschl. Lehrbücher).
Interventionell	Sind der definitiven Studie (entsprechend Phase III) Entwicklungs- und Erprobungsstudien vorangegangen? Hat es Veränderungen im System gegeben, welche die Wirkung der Intervention abgeschwächt haben könnten? Bieten die beteiligten Zentren unabhängig von der Studie bereits eine so hohe Versorgungs-Qualität, so dass nur ein geringer Spielraum zur Verbesserung besteht (Decken-Effekt)? War die Teststärke (Power) bzw. Stichproben-Größe ausreichend?
Datenbasis	
<u>Kategorie</u>	<u>Methodische Fragen</u>
Administrativ / Routine	Wie groß ist die Gefahr zufälliger Missklassifikation? Wie groß ist die Gefahr systematischer Fehler (Bias)? Wie hoch ist der Anteil fehlender Daten? Besteht dadurch die Gefahr eines Bias?
Register	Wie vollständig werden die vorgesehenen Patienten (Probleme, Prozeduren) erfasst?
Studien, bei denen die Fragestellung der VF primär ist	Ist, z.B. durch konsequente Rekrutierung, sichergestellt, dass die Stichprobe der Versorgungsrealität entspricht?
Studien, bei denen die Fragestellung der VF nur sekundär ist	Entsprechen die eingeschlossenen Patienten der Wirklichkeit der Versorgung (externe Validität)?

2.4.1.8. Trugschlüsse

Auch wenn die externe Validität vieler klinischer Studien als unzureichend bewertet wird (z.B. fehlender Einbezug älterer Menschen, multimorbider Patienten, Patienten unterschiedlicher ethnischer Gruppen, versorgungsfremdes Setting), so kann daraus umgekehrt nicht der Schluss gezogen werden, nicht experimentelle Studien der Versorgungsforschung mit hoher externer Validität, aber schwacher interner Validität (schwache Studiendesigns) an deren Stelle zu setzen. Solche Studien sind allenfalls geeignet, zusätzliche Aspekte in der Routineversorgung zu erheben (z.B. im Rahmen von Primärdaten- oder Sekundärdatenanalysen) und daraus Hypothesen für weitere klinische Studien (v.a. als RCTs) zu generieren. Studien aus der Versorgungsforschung dagegen können keinen Beitrag zum Nachweis der therapeutischen Wirksamkeit oder zum Nutzen einer medizinischen Intervention leisten. Die Ergebnis- bzw. Empfehlungssicherheit von Aussagen im Hinblick auf die Wirksamkeit und/oder Verträglichkeit einer therapeutischen oder diagnostischen Intervention ist immer basierend auf den beiden Kriterien interne und externe Validität (neben weiteren Qualitätskriterien) zu beurteilen. Trugschlüsse hinsichtlich der Limitationen eines spezifischen Studiendesigns würden die Versorgungsforschung insgesamt diskreditieren.

2.5. Beispiel der Entwicklung einer an der Fragestellung orientierten Datenbasis

Eine Möglichkeit zur Verbesserung der häufig unzureichenden Datenlage im Bereich der Versorgungsforschung bietet der Aufbau von eigenständigen Datenressourcen. So etablierten sich beispielsweise für den Bereich der Pharmakotherapie in vielen Ländern, wie z. B. den Niederlanden, Großbritannien, USA, Kanada und den skandinavischen Ländern, so genannte Pharmakoepidemiologische Datenbanken (PDb), die den Beteiligten des Gesundheitssystems umfangreiche und repräsentative Daten zur Verfügung stellen. Diese Datenbanken basieren auf routinemäßig erhobenen Verordnungs-, Diagnose- und Leistungsdaten, die nach Pseudonymisierung personenbezogen prospektiv über die Zeit miteinander verknüpft werden können.

Fundierte Analysen zu Über-, Unter- oder Fehlverordnungen und den Behandlungskosten werden in der ambulanten Arzneimitteltherapie durch PDbs erheblich erleichtert.

Ein zentraler Vorteil einer PDb besteht in der Verbesserung des Wissens zur Arzneimittelsicherheit, da sie eine Risiko-Bewertung von Arzneimitteln nach der Markteinführung ermöglicht. Seltene schwere unerwünschte Arzneimittelwirkungen (UAW) eines Medikamentes werden in der Regel erst nach Markteinführung manifest, wenn das Arzneimittel in die breite Anwendung eintritt. Zum Zeitpunkt der Zulassung ist der Kenntnisstand über das pharmako-

logische Profil und insbesondere die Anwendungssicherheit eines Arzneimittels auf Grund der begrenzten Fallzahl und Beobachtungsdauer sowie der häufig geringen Repräsentanz Älterer, Frauen und Kinder in den klinischen Studien (Phase I-III) meist noch beschränkt. Auch für die Bearbeitung von pharmakoökonomischen Fragestellungen stellen PDbs ein geeignetes Instrument dar. Um Therapiealternativen ökonomisch zu bewerten, kann eine Darstellung der Gesamtkosten sinnvoll sein. Diese gehen über die direkten Kosten (medizinische, pflegerische, rehabilitative Leistungen, Arznei-, Heil- und Hilfsmittel) hinaus, die über die Abrechnungsdaten zu bestimmen sind. Anhand einer Versichertenstichprobe aus den PDbs können auch die indirekten Kosten ermittelt werden, die sich aus dem durch die Krankheit verursachten Produktivitätsausfall (Arbeitsunfähigkeit, Frühberentung, Tod) ergeben.

Restriktive Auslegung des Datenschutzes und konkurrierende Partikularinteressen haben bisher in Deutschland den Aufbau zentraler Datensammelstellen und den zeitnahen Zugang zu den wenigen existierenden Datenpools für die Versorgungsforschung erschwert. Die im GKV-Modernisierungsgesetz (GMG) geplante Bildung eines Datenpools von pseudonymisierten Daten der GKV (§ 303 a ff SGB V) stellt einen ersten Schritt für den Aufbau einer PDb dar, insbesondere da ab dem Jahr 2004 neben den stationären auch die ambulanten Diagnosen in den Kassendaten verfügbar sind.

Zusammenfassend können PDbs weder Prävalenzstudien noch a priori geplante analytische epidemiologische Studien und gesundheitsökonomische Nutzen-Kosten-Analysen ersetzen. Sie können aber helfen, das Versorgungsgeschehen auf der Bevölkerungsebene unter verschiedenen Aspekten zu beobachten und frühzeitig Hinweise auf relevante Forschungsbedarfe geben (vgl. Faich, G.A. 1996; Garbe, E. u. Müller-Oerlinghausen, B. 1998; Dietrich, E.S. 2002; Hasford, J. et al. 2004).

2.6. Beispiel der Methodenauswahl in einem beispielhaften Forschungsschwerpunkt

Im Folgenden soll dargelegt werden, welche Zusammenhänge zwischen möglichen Fragestellungen und den für ihre Beantwortung zur Verfügung stehenden Methoden bestehen. Als Beispiel wird die Implementierung von Leitlinien ausgewählt, ein Forschungsbereich, der von hoher praktischer Relevanz ist, wenn man das Spannungsfeld zwischen Zahl der Leitlinien, Aufwand für ihre qualifizierte Erstellung und Verbreitung betrachtet. Exemplarisch seien hierfür folgende Fragestellungen benannt:

- Es kann die Frage gestellt werden, in welchem Umfang Leitlinien, ggf. ihre einzelnen Empfehlungen in der Praxis berücksichtigt und umgesetzt werden. Ergänzend können Defizite und Hindernisse als Gründe für eine zu geringe oder fehlerhafte Umsetzung identifiziert werden.

Hierfür stehen Methoden der empirischen Sozialforschung sowie der Epidemiologie zur Verfügung.

- Es könnten die Auswirkungen von einzelnen Leitlinien und Empfehlungen, aber auch der Gesamtheit des Leitliniensystems auf die Versorgung untersucht werden. Es könnte also z.B. um die Fragen gehen, welche Veränderungen im Verhalten Leitlinien nach sich ziehen, und wie diese sich ggf. bei Ärzten und Patienten auswirken (Compliance), wie weit empfohlenes Verhalten tatsächlich gefördert und ungünstiges Verhalten verhindert werden kann, und inwieweit allgemein - auch indirekte - Auswirkungen auf das Gesundheitssystem bzw. spezifische Versorgungsprozesse vorhanden sind.

Für diese Forschungsfragen steht ebenfalls das breite Methodenspektrum der empirischen Sozialforschung sowie der Epidemiologie zur Verfügung.

- Es könnte die Frage untersucht werden, ob Leitlinien die Versorgung im Sinne gesundheitlich relevanter Outcomes verbessern. Dies wären Fragen nach den medizinischen Effekten von Leitlinien (Nutzen, Risiken) in patientenrelevanten Outcomes, inklusive Lebensqualität und PROs (Patient Reported Outcomes).

Für diese Fragestellungen steht das Methodenspektrum der klinischen Epidemiologie zur Verfügung.

- Es könnten Fragestellungen nach den ökonomischen Konsequenzen von Leitlinienimplementierung bzw. auch ökonomischer Hindernisse gegen eine Implementation untersucht werden.

Für diese Fragen steht das Spektrum gesundheitsökonomischer Forschungsmethodik zur Verfügung.

- Es könnten Fragen der Umsetzbarkeit, z.B. auch des Verständnisses oder des Umgangs mit Leitlinien bearbeitet werden.

Hierfür steht unter anderem das Werkzeug qualitativer Forschung zur Verfügung.

- Es könnten Fragen der Patientenperspektive, z.B. die Entscheidungsrelevanz von Leitlinien oder auch die Berücksichtigung von Patientenpräferenzen bearbeitet werden.

Hierfür steht das Spektrum der empirischen Sozialforschung inklusive qualitativer Forschungsmethoden bereit.

Neben diesen auf existierende Leitlinien fokussierenden Fragen kann auch der Bedarf für neue Leitlinien ermittelt werden. Beispielhaft wäre hier der Bereich regionaler Versorgungsunterschiede zu nennen, die im Sinne von Anzeichen für Fehl-, Unter- oder Überversorgung Anlass für die Erarbeitung von Leitlinien sein könnten.

Auch hierfür steht das Spektrum epidemiologischer und sozialwissenschaftlicher Methoden zur Verfügung.

2.7. Fazit

Die Versorgungsforschung nutzt alle Methoden und Instrumente für die Durchführung und Beantwortung relevanter Fragestellungen zur Darstellung und Analyse der medizinischen Versorgung in einem gegebenen System. Ein solches System ist nicht nur gekennzeichnet durch bestimmte Leistungsangebote im Bereich der ambulanten, stationären, rehabilitativen, präventiven oder pflegerischen Versorgung, sondern auch durch organisatorische, personelle, gesetzliche, finanzielle und kulturelle Rahmenbedingungen sowie von den objektiven und subjektiven Belastungen durch Morbidität und Mortalität in einer Bevölkerung. Aufgabe der Versorgungsforschung ist es, durch die Beschreibung von Zuständen und Trends und die Analyse von Zusammenhängen Fehlentwicklungen aufzudecken bzw. erkennbare Optimierungsprozesse zu fördern (z.B. Professionenmix, neue Anreizstrukturen wie Pay-for-Performance oder Integrierte Versorgung). Dabei sind nutzerrelevante Aspekte so weit wie möglich zu berücksichtigen (Alter, Geschlecht, Genderaspekte, Region, sozioökonomischer Status, Versicherungsstatus etc.; siehe z. B. Ziegenhagen et al., 2005) . Hierfür kann die Versorgungsforschung Primär- und Sekundärdaten nutzen, die methodischen Instrumente der klinischen Forschung sind daher in der Versorgungsforschung ebenso präsent wie die Analyse von Daten, die zu anderen Zwecken zusammengetragen wurden (z.B. Krankenkassendaten). Ziel muss es in jedem Falle sein, eine Methodik zu wählen, mit der eine möglichst hohe interne und externe Validität erreichbar ist und die eine Aussage darüber zulässt, wie sich im Vergleich unterschiedlicher Verfahren und Maßnahmen unter realen Versorgungsbedingungen eine bestmögliche und angemessene Prozess- und Ergebnisqualität für die Nutzerinnen und Nutzer eines Gesundheitssystems herstellen lässt. Auf diesem Weg wird die Versorgungsforschung zur Basis einer im deutschen Gesundheitssystem viel zu lang vermissten Evaluationskultur, die es konsequent weiterzuentwickeln gilt.

2.8. Literatur

- Arbeitsgruppe Erhebung und Nutzen von Sekundärdaten der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP), Arbeitsgruppe für Epidemiologische Methoden der Deutschen Gesellschaft für Epidemiologie (DGEpi), der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) und der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) (2008): GPS – Gute Praxis Sekundärdatenanalyse. Revision nach grundlegender Überarbeitung. Gesundheitswesen 70: 54-60.
- Dietrich, E.S. (2002): Grundlagen der Pharmakoepidemiologie und Pharmaökonomie. Eschborn.
- Donner-Banzhoff, N., Schrappe, M., Lelgemann, M. (2007) Studien zur Versorgungsforschung. Eine Hilfe zur kritischen Rezeption. Z.ärztl.Fortbild. Qual. Gesundh.wes (ZaeFQ) 101:463-471
- Faich, G.A. (1996): US adverse drug reaction surveillance 1989-1994. Pharmacoepidemiological Drug Safety, Vol. 5, S. 393-398.

- Garbe, E. und Müller-Oerlinghausen, B. (1998): Datenbanken im Ausland für die Arzneimittelforschung nach der Zulassung. Ein Modell auch für Deutschland? Deutsches Ärzteblatt, Vol. 49, S. 3138-3143.
- Hasford, J., Schubert, I., Garbe, E., Dietlein, G. und Glaeske, G. (2004): Memorandum zu Nutzen und Notwendigkeit Pharmakoepidemiologischer Datenbanken in Deutschland. GEK-Edition. St. Augustin
- Ziegenhagen, D.J., Glaeske, G., Höer, A. und Gieseler, K. (2004): Arzneimittelversorgung von PKV-Versicherten im Vergleich zur GKV. Gesundheitsökonomie und Qualitätsmanagement, Vol. 9, S. 108-115.

3. Methoden der organisationsbezogenen Versorgungsforschung

Pfaff, H.; Albert, U.-S.; Bornemann, R., Ernstmann, N., Gostomzyk, J., Gottwik, M.G.; Heller, G. ; Höhmann, U. Karbach, U.; Ommen, O. Wirtz, M.

3.1. Einleitung

Organisationen wie Kliniken und Arztpraxen sind die Grundelemente des Gesundheitswesens. 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 Organisationsforschung zurückgreifen. Diese bedürfen einer sach- und fachgerechten Anwendung. Ziel dieses organisationsbezogenen Teils des III. Memorandums zur Versorgungsforschung ist es, diese Methoden in ihren Grundzügen zu beschreiben, wichtige Mindeststandards darzustellen und Kriterien für eine möglichst hohe Studienqualität zu definieren. Es werden in dieser Abhandlung 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.

3.2. 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 sicherzustellen. Versorgungsorganisationen erbringen am Patienten 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 haben meist den Zweck, die Erbringung dieser Gesundheitsdienstleistungen zu ermöglichen oder zu unterstützen. Beispiele hierfür sind Ärztekammern, Kassenärztliche Vereinigungen, Krankenkassen und Selbsthilfeorganisationen.

Die Aufgaben der organisationsbezogenen Versorgungsforschung bestehen darin,

- 1) die sozialen, technischen und versorgungsbezogenen Strukturen und Prozesse in den Gesundheitsorganisationen zu beschreiben und zu erklären,
- 2) deren Auswirkungen auf Zugang, 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 neuer 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. 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 (Patienten, Ärzte), Gruppen (z.B. Stationsteams), Organisationen (z.B. Kliniken) oder Organisationsnetzwerke (z.B. integriertes Versorgungsnetz) untersucht. Es gibt unter den vielen Fragestellungen, die für die organisationsbezogene Versorgungsforschung spezifisch 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 (Abbildung 3.1, Pfeil a) auch der organisatorische Kontext der Behandlung auf die Gesundheit und die Lebensqualität der Patienten auswirkt (Abbildung 3.1; Pfeil b).

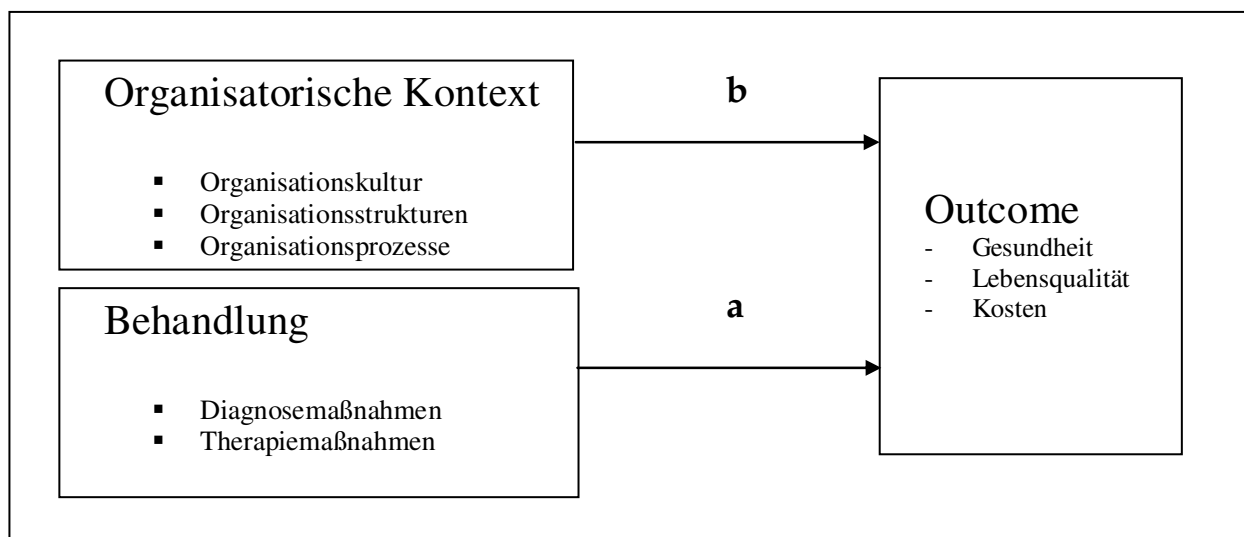


Abbildung 3-1: Behandlung und organisatorischer Kontext als Determinanten des Versorgungsauscomes

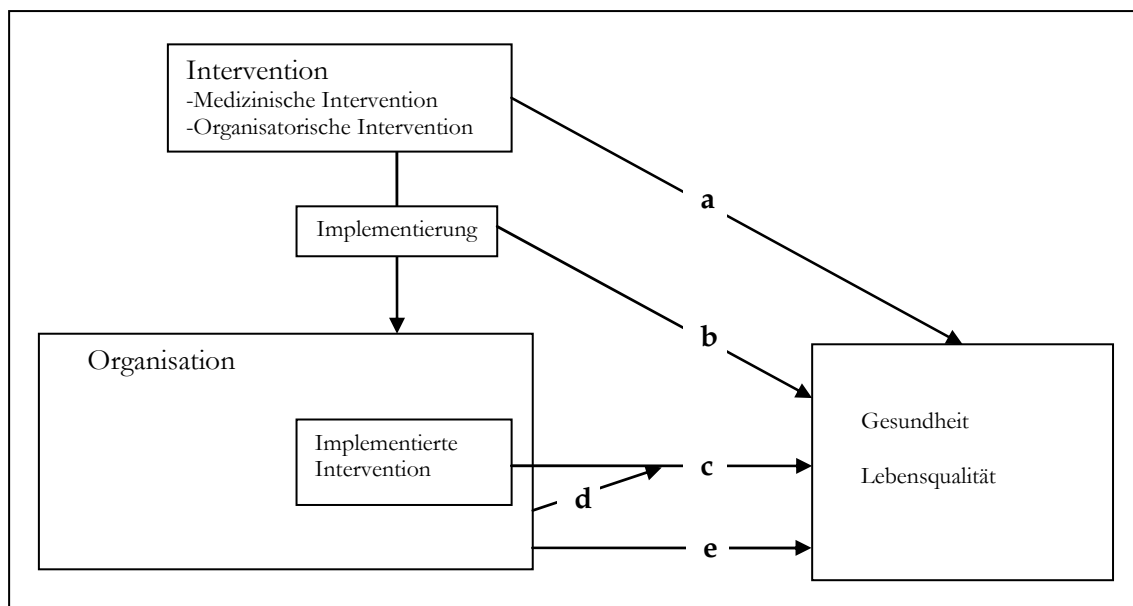


Abbildung 3.2: Wirkkomponenten der Versorgungssituation

Eine weitere zentrale Aufgabe der organisationsbezogenen Versorgungsforschung ist die Erfassung der Wirkung der Art und Weise der Einführung einfacher und komplexer Interventionen in Versorgungsorganisationen. Bei der Einführung von Versorgungs- und Organisationsinnovationen sind mehrere Wirkkomponenten zu unterscheiden. (a) die Wirkung der In-

tervention an sich (Pfeil a in Abbildung 3.2), (b) die Wirkung der Implementierung der Intervention (Pfeil b in Abbildung 3.2), (c) die Wirkung einer an die Organisationsrealitäten angepassten Intervention (Pfeil c in Abbildung 3.2), (d) der moderierende Einfluss der Organisationsmerkmale auf die Wirkung der Intervention im Sinne einer Abschwächung oder Verstärkung dieser Wirkung (Pfeil d in Abbildung 3.2) und (e) der direkte Einfluss der Organisation auf die Gesundheit und Lebensqualität der Patienten (Pfeil e in Abbildung 3.2).

Bei der Evaluation von Organisationsinterventionen 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 personenbezogenen Interventionen, im zweiten Fall von organisationsbezogenen Interventionen. Bei den personenbezogenen Interventionen kann zusätzlich unterschieden werden zwischen verblindbaren und nicht-verblindbaren Interventionen. Bei nicht-verblindbaren Interventionen sind randomisierte kontrollierte Studien (RCT) in Organisationen 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; vgl. Kap. 3.6). Handelt es sich um organisationsbezogene Interventionen, also um Interventionen, die auf der Ebene der Organisation ansetzen (z.B. Abteilung, OP-Team oder Klinik), ist ebenfalls die Clusterrandomisierung der Goldstandard. Der Aufwand an organisationsübergreifender Standardisierung, Koordinierung und Konsentierung ist bei CRTs wesentlich höher als bei RCTs. Deshalb wird im Kapitel 3.6 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 eine Versorgungsorganisation (z.B. Arztpraxis, Klinikum) erfolgreich zu implementieren. Auch für diese Form der Evaluation stellt das CRT-Studiendesign den Goldstandard dar (vgl. Kap. 3.6). 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 müssen, 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 Versorgungsforschung können mit verschiedenen Daten und Methoden bearbeitet werden. Bei den Daten unterscheiden wir biomedizinische, technische, psychosoziale und sozio-ökonomische Daten (vgl. Tabelle 3.1) 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 (Flick, 2000a; Kühl & Strodtholz, 2002; Pfaff & Bentz, 2003).

Tabelle 3.1: Methoden der Datengewinnung in der Versorgungsforschung (nach Pfaff, 2005)

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 (Maschinen)	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			

Bei der Planung von Versorgungsforschungsstudien müssen Entscheidungen bezüglich der Stichprobe, 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 perso-

nenbezogenen Effectiveness-Untersuchungen, kann bei einer anderen Fragestellung, z.B. formative Evaluation einer Intervention, unangemessen sein.

In diesem Teil des Memorandums wird auf die Stichprobenziehung, die Messung und das Studiendesign im Rahmen quantitativer Studien der Organisationsforschung eingegangen. Das Ziel dieses Memorandums ist es, einen orientierenden Überblick über diese Themen zu geben. Somit wird auf eine detaillierte Beschreibung weiterer quantitativer organisationsbezogenen Methoden wie z.B. Risikoadjustierung und Mehrebenenanalyse verzichtet. Dies bedeutet jedoch nicht, dass solche quantitativen, aber auch qualitativen Verfahren in der organisationsbezogenen Versorgungsforschung keinen Platz haben. So gibt es in der Organisationsforschung eine ganze Reihe qualitativer Methoden, die - professionell angewandt - hohe Erkenntnisgewinne bringen können (Kühl & Strodtholz, 2002). Da jedoch für qualitative Ansätze andere Prinzipien, Regeln und Methodenstandards gelten, würde ihre Darstellung den Rahmen dieses Teils des Memorandums III sprengen. Für eine Übersicht der relevanten Prinzipien und Regeln sei auf die einschlägige Literatur verwiesen (Flick, 2004; Flick, 2000a; Flick, 2004; Flick, 2000b; Merrens, 2000; Steinke, 2000).

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

Die Wahl der Art und des Umfangs einer Stichprobe 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 Stichprobe 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 Äquivalenz der Verteilungen nicht durch die Form der Stichprobengewinnung (i.d.R. Zufallauswahl) ge-

währleistet 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. Bortz & Döring 2006; Schnell et al., 2005; Blalock, 1960; Blalock & Blalock, 1985; Böltken, 1976; Gabler, 1997; Consort-Statement Group, 2008c). Relevante Kennziffern sind zum Beispiel 1) der Standardfehler der Schätzung, 2) die Effektstärkenschatzung und das Vertrauensintervall, 3) die Prüfverteilungen, 4) die Signifikanztests und 5) die Irrtumswahrscheinlichkeiten. Im Rahmen der Planung einer Studie sollte weiterhin sichergestellt werden, dass die Stichprobenumfänge hinreichend groß sind. Dies sollte erfolgen, in dem die Stichprobengröße gewählt wird, mit der eine empirisch minimal relevante Effektgröße bei einer α -Fehlerwahrscheinlichkeit von .2 als signifikant erwartet werden kann (Bortz & Döring, 2006; Erdfelder et al., 1996). Wesentlich für die „kluge“ Konstruktion einer Stichprobe ist die Kenntnis und Beschreibung der Grundgesamtheit, der so genannten „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: Verhältnisschätzung, Differenzschätzung, (lineare) Regressionsschätzung, „capture-recapture Methode“ (vgl. Diekmann, 2008, Lee, 1993)) und theoretische Überlegungen geleistet werden (Gabler, 1997). 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.

3.3.1. 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 zu explizierender statistischer Algorithmen. Die geringsten Verzerrungseffekte bergen dabei Lotterielosverfahren, Zufallszahlentafeln oder das Auswahlverfahren „Tag des Geburtstagsdatums“. 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 teilneh-

mender Kliniken) gezogen werden. Für alle über diese Stichprobe hinausgehende Repräsentationsschlüsse sind die oben genannten Fehlerparameter nach gängigen Konventionen anzugeben und inhaltliche Begründungen für die Berechtigung von Verallgemeinerungsschlüssen anzugeben (Böltken, 1976; Consort-Statement Group, 2008b).

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. Deutschland = 1. Ebene, Kliniken = 2. Ebene, Stationen = 3. Ebene) 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 (Langer, 2004). 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 Stichprobeverfahren 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 untersuchende 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 auf Grund erheblichen Vorwissens über die Bedeutung der

Schichtungskriterien zu begründen. Komplexe Schätzformeln (z.B. (Böltken, 1976; Kalton, 1983) sind dann zur Minimierung der Fehlerintervalle einzusetzen.

Klumpenstichproben (Clusterstichproben)

Als Sonderfall der Schichtungskonstruktion werden zur Reduktion des Erhebungsaufwandes bereits natürlich vorgruppierte Teilmengen – sog. Klumpen oder Cluster (z.B. Stationen) - einer Grundgesamtheit (z.B. alle Kliniken Deutschlands) als Auswahleinheiten 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 geschichteten 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 Diekmann, 2008)). 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 (Hox & Maas, 2005).

3.3.2. 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 (Mann/Frau) oder Merkmalskombinationen (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 Fälschung bei der Erhebung (z.B. Dorroch, 1994).

Weitere gesteuerte/gezielte Auswahlverfahren

Bei hoher gesicherter Kenntnis über Merkmalszusammenhänge und deren untersuchungsrelevanter 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 gesteuerte Gelegenheitsauswahlen (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. Gabler, 1992; Lee, 1993)) 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 Erwä-

gungen für jede Stichprobe folgende weitere Informationen zu den gegebenenfalls getroffenen Entscheidungen anzugeben und zu begründen (Consort-Statement Group, 2008a):

- Eindeutige Ein- und Ausschlusskriterien für alle Untersuchungseinheiten (insbesondere Paarbildungs- und Matchingkriterien sowohl auf Ebene der organisationalen Einheiten als auch auf Ebene der Individuen)
- 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, Campbell et al., 2004c).
- Erwartete Stichprobenreduktion bei Längsschnitt- und Paneluntersuchungen.

3.4. 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 Kapitel „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 (Pfaff & Bentz, 2003), sei auf den umfassenden Leitfaden „Gute Praxis Sekundärdatenanalyse“ (Swart et al., 2005) verwiesen. Die dort aufgeführten Kriterien besitzen auch für die Sekundärdatenanalyse im Rahmen der Organisationsforschung Gültigkeit.

3.4.1. 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 Gütekri-

terien genügen (vgl. auch Memorandum III: Methoden der Lebensqualitätsforschung). Entsprechende Angaben zur Sicherstellung dieser Kriterien und zur Erfüllung von Güteanforderungen müssen berichtet werden. Bei Abweichungen von gängigen Grenzwerten sollte bei den Gütekriterien die Beibehaltung der Messvorschrift 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 messende 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 (Bollen, 1989; Bühner, 2006) 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 (Diamantopoulos & Winklhofer, 2001), die sich auf die (i) inhaltliche Validität der Konstruktdefinition und Indikatorspezifikation, (ii) Kolinearität der Indikatoren und (iii) externale 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 auf Expertenwissen zusammengestellte Operationalisierung darstellen, besitzt eine solche Messung eine geringere Evidenz (Evidenzstufe I) als eine psychometrisch fundierte und validierte Messskala. Da die konfirmatorische Analyse zudem theoriegeleitet vorgeht, kann für konfirmatorisch überprüfte Skalen eine höhere Evidenzstufe (Evidenzstufe III) der Operationalisierung angenommen werden als für explorativ ermittelte Skalen (Evidenzstufenleiter II).

Qualitätskriterien für Beobachtungsdaten

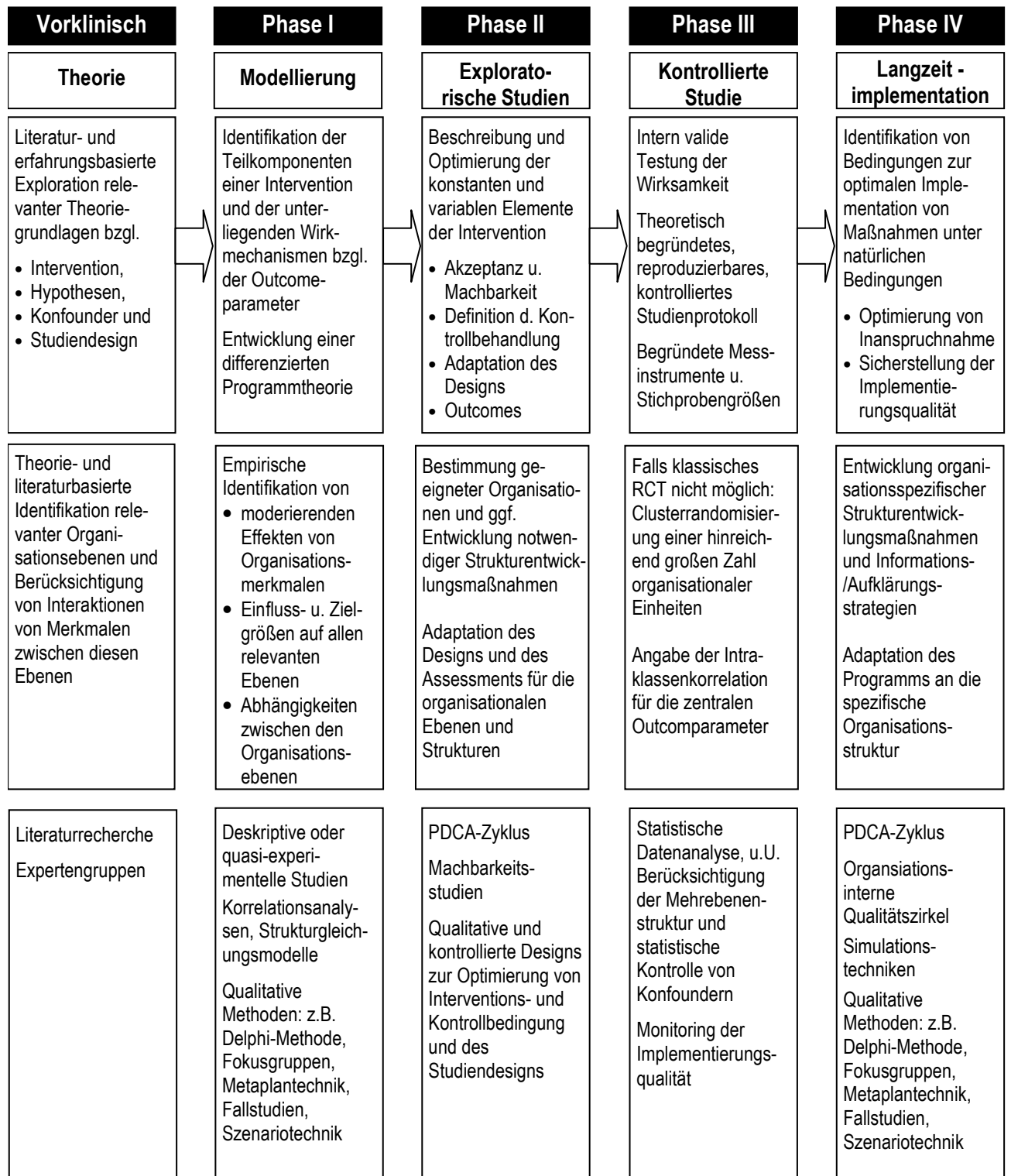
Bei der systematischen Beobachtung werden in der Regel die Ausprägungsgrade definierter Merkmale durch unabhängige Dritte eingeschätzt (z.B. Einschätzung von Patientenmerkmalen durch Behandler; Beurteilung von Qualitätsindikatoren im Peer-Review-Verfahren; z.B. (Vogel et al., 2002). Der Beobachter oder 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, in dem Beurteiler bei der Beurteilung derselben Personen oder Objekte zu übereinstimmenden oder ähnlichen Einschätzungen gelangen, standardisiert quantifizieren. Die Auswahl der angemessenen Kenngröße wird vor allem durch die Skalendignität bestimmt (Wirtz, 2006). 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 (Wirtz & Caspar, 2002). 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 der Mittelwert über mehrere Beurteiler als diagnostische Größe verwendet werden, wenn eine zufrieden stellende Übereinstimmung für einzelne Urteiler nicht erreicht werden kann und für den Mittelwert die Übereinstimmungskriterien erreicht werden.

Bezogen auf den gesamten Erhebungsprozess ist darauf zu achten, dass in der Regel die Güte der Beurteilerübereinstimmung vor der Haupterhebung in einer Pilotstudie sichergestellt werden muss. Mängel der Beurteilungsgüte sollten anschließend durch ein systemati-

ches Beurteilertraining identifiziert und beseitigt werden (Wirtz & Caspar, 2002). 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 (Greve & Wentura, 1997). Erst nachdem durch ein solches Beurteilertraining eine hinreichende Übereinstimmung sichergestellt wurde, kann die Haupterhebung erfolgen.

3.5. Standards und Evidenzstufenleiter für organisationsbezogene Studiendesigns

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 (Christiansen, 2001). Das Phasenmodell von Campbell (Campbell et al., 2000) zur Planung und Durchführung von Studien zur Evaluation komplexer Interventionen kann hierbei als Orientierung für eine gute organisationsbezogene Versorgungsforschung empfohlen werden. Dieses Modell berücksichtigt, dass an die Evaluation komplexer Interventionen besondere Anforderungen zu stellen sind, da 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. Morfeld & Wirtz, 2006). Abbildung 3.3 zeigt den Phasenverlauf im Überblick.



Kontinuum wachsender Evidenz

Abbildung 3.3: Phasenmodell für die Evaluation komplexer Interventionen nach (Campbell et al. 2000)

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 bzw. deren Einzelkomponenten 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 (i) welche Merkmale auf den jeweiligen Ebenen bei der Theoriebildung relevant sind, (ii) ob Interaktionen zwischen den verschiedenen Ebenen bestehen können und (iii) 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, Hinweise auf die Gültigkeit elementarer Hypothesen empirisch zu bekommen, Optimierungs- und Modifikationsbedarf zu erkennen und somit die Basis für gezielte 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 Klinikebene) 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 stehen im Mittelpunkt. 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 (Kline, 2005). 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 Maße (Bagozzi & Baumgartner, 1994) 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 (Langer & Wetter, 2000; Skrondal & Rabe-Hesketh, 2004).

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

Die optimale Gestaltung der Intervention und des endgültigen Studiendesigns stehen bei dieser zweiten 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 (Delbecq et al., 1975; Vang, 1986). 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 (Black et al., 1999; Murphy et al., 1998): Nominaler-Gruppen-Prozess (bis 20 Teilnehmer; Dunham, 1998), Delphi-Verfahren (> 10 Teilnehmer; Delbecq et al., 1975), strukturierte Konsensuskonferenz im Format nach Vang (> 50 Teilnehmer; Vang, 1986). 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 und 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 potentieller Teilnehmer). Es ist zu prüfen, in welchen Organisationen und auf welchen organisationalen Ebenen eine Implementierung der Intervention grundsätzlich 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 Outcome-Parameter gelegt werden. Die verwendeten Outcomeparameter sollten die zuverlässige und valide Erfassung der Interventionswirkungen für alle Betroffenengruppen 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 be-

handelnden 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 Kontrollbedingung gewonnen werden.

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

Das Interventionsprogramm wird implementiert und im Vergleich zu einer Kontrollbehandlung 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 Samples oder mit anschließender Risikoadjustierung eine Alternative dar (vgl. Abschnitt 3.6). 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, (Brewin & Bradley, 1989; Zelen, 1990)).

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 (Wegscheider, 2000). 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 (vgl. Abschnitt 3.3). Aufgrund des begrenzten zeitlichen Horizonts können theoretisch gut belegte Surrogatparameter als Erfolgsmaße verwendet werden (Christiansen, 2001) 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 (Langer, 2004).

Eine formative Begleitung der Implementierung sollte insbesondere zur Sicherstellung der Validität von Effektinterpretationen erfolgen (Monitoring). Neben der Analyse von Programmeffekten mittels zumeist varianz- und mehrbenenanalytischer Methoden können Pfadanalysen oder Strukturgleichungsmodelle zur Überprüfung der in der Programmtheorie postulierten Wirkzusammenhänge herangezogen werden. 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 (Kleespies et al., 2005)

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

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 als nachrangig betrachtet, Qualitätsunterschiede bei der Programmdurchführung durch unterschiedliche Leistungserbringer werden im Sinne der externen Validierung als wichtiger betrachtet. Während in Phase III vor allem Surrogatkriterien als Effektivitätsindikatoren untersucht werden, sollte nun die

Veränderung der langfristigen Zielgrößen (z. B. kardiale Ereignisse bei Herzinfarktpatienten) im Mittelpunkt stehen. Die Empfehlungen für die angemessene Methodenauswahl entsprechen denjenigen in Phase I.

Obwohl das Modell von Campbell (Campbell et al., 2000) 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 zufrieden stellend 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 3.3 ab. Auch der Gesamtprozess ist als zirkulärer Prozess zu sehen: In Phase IV wird in der Regel weiterer Forschungsbedarf notwendig sein, der es erforderlich macht, alle Phasen des Modells erneut zu durchlaufen, um neue Interventionskonzepte praxisgerecht entwickeln und evaluieren zu können. Nach der Modellvorstellung von Campbell (Campbell et al., 2000) 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 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, (i) weshalb die empirischen Belege, die nach der Modellvorstellung eigentlich in den vorhergehenden Phasen hätten identifiziert werden müssen, als vorliegend angenommen werden können und (ii) 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 (Campbell et al., 2000) für ein komplexes empirisches Evaluationsprojekt gefordert werden.

Eine eindeutige Evidenzstufung empirischer Befunde erweist sich für dieses Modell als schwierig. Wie in Abbildung 3.3 dargestellt wurde, erhöht sich die empirische Evidenz, wenn eine höhere Phase des insgesamt fünfstufigen Verlaufs erreicht wird – vorausgesetzt, dass die vorherigen 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: Ob beispielsweise eine experimentelle Studie der Phase III eine niedrigere oder höhere Evidenz hat als eine Phase IV-Studie, die auf einer quasi-experimentellen Überprüfung in Phase III beruht, kann nicht allgemein beantwortet werden.

3.6. Standards für experimentelle und nicht-experimentelle quantitative Studien

Wie bereits in Kapitel 3.5 dargestellt, ist bei Experimenten in Phase III (Abbildung 3.3) 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 clusterrandomisierte Studie (CRT; Bortz & Döring, 2006). 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-Personen in Austausch treten können bzw. beide Gruppen 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 (Moher et al., 2001a)

Kann zwischen den Mitgliedern der per Randomisierung zustande gekommen 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 & Shadish, 1994 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. (Morfeld & Wirtz, 2006). Als wichtigste Sicherungsmaßnahme zur Identifikation dieser potentiellen 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. Das ist meist bei Feldexperimenten, Gemeindenexperimenten oder Organisationsexperimenten der Fall (Campbell et al., 2004b). Diese Form der Randomisierung bezeichnet man als Clusterrandomisierung (CRT: cluster randomized trials). 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 (Campbell et al., 2004a). 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, Langer, 2004). 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.) Querschnittstudie und 3.) Kohortenstudie unterteilt werden. Diese Studientypen werden in organisationsbezogenen Versorgungsstudien vor allem im Rahmen von versorgungsepidemiologischen Fragestellungen eingesetzt. Als Qualitätsstandards für den Bericht von epidemiologischen Beobachtungsstudien gelten die STROBE-Kriterien (Strengthening The Reporting of Observational Studies in Epidemiology, Vandenbroucke et al., 2007). Die Beachtung dieser Kriterien bei der Planung und Durchführung solcher Formen von nicht-experimentellen

Studien erhöht ihre Qualität und stellen die Mindeststandards für diese Studienarten in der Versorgungsforschung dar.

3.7. 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 der Variablen und (c) die Auswahl des Untersuchungsdesigns einzuhalten sind. Weiter konnte dargestellt werden, wie eine besonders hohe Studienqualität in der organisationalen Versorgungsforschung erzielt werden kann.

3.8. Literatur

- Bagozzi, R. & Baumgartner, H. (1994). The evaluation of structural equation models and hypothesis testing. In *Principles of Marketing Research*, (ed. R. Bagozzi), pp. 386-422. Blackwell Publishers: Cambridge.
- Black, N., Murphy, M., McKnee, M., Sanderson, C., Askham, J. & Merta, T. (1999). Consensus development methods: a review of best practice creating clinical guidelines. *Journal of Health Services Research and Policy* 4, 236-248.
- Blalock, H. M. (1960). *Social Statistic*. Mc Graw Hill: New York, Toronto, London.
- Blalock, H. M. & Blalock, A. B. (1985). *Methodology in Social Research*. Mc Graw Hill: New York, Toronto, London.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York.
- Bötker, F. (1976). Auswahlverfahren, eine Einführung für Sozialwissenschaftler. In (Anonymous), Teubner: Stuttgart.
- Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation. In (Anonymous), Springer.
- Brewin, C. R. & Bradley, C. (1989). Patient preferences and randomised controlled trials. *BMJ* 289, 315.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson: München.
- Campbell, M., Fitzpatrick, R. & Haines, A. (2000). Framework for the design and evaluation of complex interventions to improve health. *BMJ* 321, 694-696.
- Campbell, M. K., Elbourne, D. R. & Altman, D. G. (2004). CONSORT statement: extension to cluster randomised trials. *BMJ* 328, 702-708.

- Christiansen, G. (2001). *Evaluation - ein Instrument zur Qualitätssicherung in der Gesundheitsförderung. Eine Expertise*. Bundeszentrale für gesundheitliche Aufklärung: Köln.
- Consort-Statement Group (2008). In (ed. www.Consort-Statement.Org (30.09.2008)).
- Cook, T. D. & Shadish, W. R. (1994). "Social Experiments: Some Developments Over the Past Fifteen Years". *Annual Review of Psychology* 45, 545-580.
- Delbecq, A. L., Van de Ven, A. H. & Gustafson, D. H. (1975). *Group Techniques for Program Planning: A guide to Nominal Group and Delphi Process*. Scott, Foresman and Company: Glenview, IL, USA.
- Diamantopoulos, A. & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research* 38, 269-277.
- Diekmann, A. (2008). *Empirische Sozialforschung Grundlagen, Methoden, Anwendungen*. Rowolt: Hamburg.
- Dorroch, H. (1994). *Meinungsmacherreport. Wie Umfrageergebnisse entstehen*. Steidl: Göttingen.
- Dunham, R. B. (1998). Nominal Group Technique: users' guide. In (Anonymous).
- Erdfelder, E., Faul, F. & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers* 28, 1-11.
- Flick, U. (2000a). Qualitative Forschung. Ein Handbuch. In (ed. U. Flick, E. von Kardorff and I. Steinke), pp. 1-768. rororo: Hamburg.
- Flick, U. (2004). *Qualitative Forschung ein Handbuch*. Rowohlt Taschenbuch Verlag: Hamburg.
- Flick, U. (2000b). Design und Prozess qualitativer Forschung. In *Qualitative Forschung Ein Handbuch*, (ed. U. Flick, E. von Kardorff and I. Steinke), pp. 252-265. rororo: Hamburg.
- Gabler, S. (1992). Schneeballverfahren und verwandte Stichprobendesigns. *ZUMA-Nachrichten* 31, 47-69.
- Gabler, S. (1997). Stichproben in der Umfragepraxis. In (ed. S. Gabler and J. H. P. Hoffmeyer-Zlotnik), pp. 1-246. Westdeutscher Verlag: Opladen.
- Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung*. Psychologie Verlags Union: Weinheim.
- Hox, I. J. & Maas, C. J. M. (2005). Multilevel Models for Multimethod Measurement. In *Handbook of Multimethod Measurement in Psychology*, (ed. M. Eid and E. Diener), pp. 269-281. American Psychological Association: Washington D.C.
- Kalton, G. (1983). *Introduction to Survey Sampling*. Sage UNIVERSITY PAPER: London.
- Kleespies, C., Kaise, T., Sawicki, P. T., für die Arbeitsgruppe Praktische evidenzbasierte Medizin, St. F. H. K. & IeM - Institut für evidenzbasierte Medizin, K. (2005). Begriffe und Methoden der evidenzbasierten Medizin - Ein Glossar. In (ed. 2. S. h. Version

Oktober 2005 and www.di-em.de/publikationen.php), pp. 1-27. DiEM - Institut für evidenzbasierte Medizin, Köln: Köln.

- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. Guilford: New York.
- Kühl, S. & Strodtholz, P. (2002). *Methoden der Organisationsforschung. Ein Handbuch*. Rowohlt.
- Langer, B. & Wetter, T. (2000). Physician's attitudes towards health telematics - an empirical study. In *Medical Infobahn for Europe. Proceedings of MIE2000 and GMDS2000*, (ed. A. Hasman, B. Blobel, J. Dudeck, R. Engelbrecht, G. Gell and H.-U. Prokosch), pp. 1106-1111. IOS Press: Amsterdam.
- Langer, W. (2004). *Mehrebenenanalyse, Eine Einführung für Forschung und Praxis*. VS Verlag für Sozialwissenschaften.
- Lee, R. M. (1993). *Doing Research on Sensitive Topics*. Sage: London.
- Merkens, H. (2000). Auswahlverfahren, Sampling, Fallkonstruktion. In *Qualitative Forschung Ein Handbuch*, (ed. U. Flick, E. von Kardorff and I. Steinke), pp. 286-299. rororo: Hamburg.
- Moher, D., Jones, A. & Lepage, L. (2001). Use of the Consort Statement and Quality of Reports of Randomized Trials
A Comparative Before- and -After Evaluation. *JAMA* 285, 1992-1995.
- Morfeld, M. & Wirtz, M. (2006). Methodische Ansätze in der Versorgungsforschung. Das Beispiel Evaluation der Disease-Management-Programme. *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz* 49, 120-129.
- Murphy, M., Black, N., Lamping, D., McKee, C., Sanderson, C., Askham, J. & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment* 2, 1-88.
- Pfaff, H. & Bentz, J. (2003). Qualitative und quantitative Methoden der Datengewinnung. In *Das Public Health Buch. Gesundheit und Gesundheitswesen*, (ed. F. W. Schwartz, B. Badura, R. Busse, R. Leidl, H. Raspe, J. Siegrist and U. Walter), pp. 419-434. Urban & Fischer: München und Jena.
- Pfaff, H. (2005). Konzept und Methodik der Versorgungsforschung. *Die Psychiatrie* 2, 101-106.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. R. Oldenbourg Verlag: München.
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC: Boca Raton.
- Steinke, I. (2000). Gütekriterien qualitativer Forschung. In *Qualitative Forschung Ein Handbuch*, (ed. U. Flick, E. von Kardorff and I. Steinke), pp. 319-331. rororo: Hamburg.
- Swart, E., Ihle, P., Geyer, S., Grobe, T. & Hofmann, W. (2005). GPS - Gute Praxis Sekundärdatenanalyse. *Das Gesundheitswesen* 67, 416-421.

- Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C. & Mulrow, C. D. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Medicine* 4, 1628-1654.
- Vang, J. (1986). The consensus development conference and the European experience. *International Journal of Technology Assessment* 2, 65-76.
- Vogel, H., Lemisz, W., Liebeck, H. & Palm, W. (2002). Zur Bewertung des Gutachterverfahrens für die ambulante Verhaltenstherapie durch die GutachterInnen. *Verhaltenstherapie* 12, 228-231.
- Wegscheider, K. (2000). Statistische Verfahren in der Rehabilitationforschung. In *Grundlagen der Rehabilitationswissenschaften*, (ed. J. Bengel and U. Koch), Springer: Berlin.
- Wirtz, M. (2006). Methoden zur Bestimmung der Beurteilerübereinstimmung. In *Handbuch der Psychologie - Psychologische Diagnostik*, (ed. M. Petermann), pp. 369-382. Hogrefe: Göttingen.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Hogrefe-Verlag: Göttingen, Bern, Toronto, Seattle 2002.
- Zelen, M. (1990). Randomized consent designs for clinical trials. An update'. *Statistics in Medicine* 9, 645-656.

4. Die Erfassung von Lebensqualität* in der Versorgungsforschung – konzeptuelle, methodische und strukturelle Voraussetzungen

Koller, M.; Neugebauer, E.A.M.; Augustin, M.; Büssing, A.; Farin, E.; Klinkhammer-Schalke, M.; Lorenz, W.; Münch, K.; Petersen, C.; Steinbüchel, N.v.; Wieseler, B.

* Im vorliegenden Memorandum wird Lebensqualität stets in Bezug auf Gesundheit und/oder Krankheit betrachtet. Daher wird auf die Ausdrucksweisen „gesundheitsbezogene Lebensqualität“ (health related quality of life, HRQL) oder „krankheitsbezogene Lebensqualität“ verzichtet und im Sinne der Lesbarkeit einheitlich der Begriff Lebensqualität (Abkürzung LQ) verwendet.

Präambel

Versorgungsforschung ist ein interdisziplinärer Forschungsansatz, der die Prozesse, Ergebnisse und Rahmenbedingungen von Gesundheitsversorgung mit wissenschaftlichen Methoden untersucht.

Lebensqualität wird innerhalb der Medizin, Psychologie und anderen Gesundheitswissenschaften als Maß zur Bewertung von Gesundheitszielen (Outcomes) betrachtet, wobei es von entscheidender Bedeutung ist, dass Patienten selbst Auskunft über ihren Zustand geben. Lebensqualität ist ein mehrdimensionales Konstrukt und umfasst die subjektive Befindlichkeit und das Handlungsvermögen im somatischen, im psychischen und im sozialen Bereich.

Der Schnittpunkt zwischen der Versorgungsforschung und der Lebensqualitätsforschung kann in ihrer Patientenzentrierung gesehen werden.

Beide Disziplinen stellen die Patienten in den Mittelpunkt der wissenschaftlichen Analyse. Bislang ist der Zusammenhang zwischen Versorgungsforschung und Lebensqualität noch nie systematisch dargestellt worden. Eine derartige Analyse ist wegen der inhaltlichen und methodischen Nähe der beiden Forschungsgebiete naheliegend und notwendig.

Der interdisziplinäre Ansatz der Versorgungsforschung mit seinem weiten Aufgabenfeld auf der einen, und die Lebensqualitätsforschung mit ihren Modellen und Methoden auf der anderen Seite, bieten ein hervorragendes Potential für Synergien.

Die Integration von Lebensqualität in die Versorgungsforschung verspricht, die Diskussion über die Ziele der Gesundheitsversorgung und die Werte in unserem Gesundheitssystem in konstruktiver, kritischer und nachhaltiger Weise zu beeinflussen.

4.1. Bedeutung und Eigenheiten des Forschungsgebietes

Die Bedeutung des Konzepts Lebensqualität

Innerhalb der klinischen Forschung hat sich die Erfassung der Lebensqualität (LQ) als patientenrelevantes Outcome durchgesetzt. Namhafte Fachgesellschaften kommentieren den Stellenwert der LQ-Messung in ihren Stellungnahmen zur Methodik klinischer Studien (American Society of Clinical Oncology, 1996). Auch für Arzneimittelzulassung (drug labeling), Nutzenbewertung und gesundheitsökonomische Evaluationen wird die Bedeutung des LQ-Ansatzes anerkannt und in Stellungnahmen diskutiert (FDA & U.S. Department of Health and Human Services, 2006; EMEA, 2005; Chassany et al., 2002; Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2008). Auch ist, von Ausnahmen zu Beginn der klinischen LQ-Forschung abgesehen (Troidl et al., 1979), eine Zunahme von Publikationen klinischer Studien in hochklassigen Zeitschriften zu verzeichnen, in denen LQ als primärer Endpunkt diente (Bauhofer et al., 2007; Deuschl et al., 2008; Wright et al., 2007).

In Deutschland wird in Texten zum Sozialrecht und zur Nutzenbewertung explizit auf den Terminus „Lebensqualität“ Bezug genommen:

§35 b SGB V

"Beim Patienten-Nutzen sollen insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität ... berücksichtigt werden".

G-BA Verfahrensordnung §20, Absatz 2

"Der Nutzen einer Methode ist durch qualitativ angemessene Unterlagen zu belegen. Dies sollen, soweit möglich, Unterlagen der Evidenzstufe 1 mit patientenbezogenen Endpunkten (z. B. Mortalität, Morbidität, Lebensqualität) sein".

Methodenpapier des IQWiG, Kapitel 3.1.1

„Dabei (bei der Nutzenbewertung) werden sowohl die beabsichtigten als auch die unbeabsichtigten Effekte der Interventionen berücksichtigt, die eine Bewertung der Beeinflussung insbesondere folgender patientenrelevanter Zielgrößen zur Feststellung krankheits- und behandlungsbedingter Veränderungen erlauben:

- 1) Mortalität,
- 2) Morbidität (Beschwerden und Komplikationen),
- 3) gesundheitsbezogene Lebensqualität.

Ergänzend können der interventions- und erkrankungsbezogene Aufwand und die Zufriedenheit der Patienten mit der Behandlung berücksichtigt werden. Solche Aspekte kommen allerdings in der Regel nur als sekundäre Zielgrößen in Betracht“ (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2008).

Generell wird die LQ-Messung in der klinischen Forschung oder Routine immer dann von Interesse sein, wenn das subjektive Krankheitserleben im Vordergrund steht und entsprechende, darauf abzielende Behandlungseffekte abgebildet werden sollen. Das kann insbesondere in folgenden klinischen Situationen der Fall sein: chronische Erkrankungen mit ausgeprägtem Leidensdruck, langwierige Genesungsverläufe, palliative Situationen oder massiver akuter gesundheitlicher Einbruch, der in der Erinnerung noch lange Zeit nachwirkt.

Lebensqualität und Versorgungsforschung

Das Spektrum und die Aufgaben der Versorgungsforschung wurden von den Autoren des Arbeitskreises Versorgungsforschung in fünf Punkten zusammengefasst (Arbeitskreis Versorgungsforschung beim Wissenschaftlichen Beirat, 2004). Im Folgenden werden diese fünf Punkte aufgeführt (kursive Darstellungsweise) und der Bezug zum Konzept LQ skizziert.

1. Versorgungsforschung beschreibt die Inputs, Prozesse und Ergebnisse von Kranken- und Gesundheitsversorgung, einschließlich der auf sie einwirkenden Rahmenbedingungen mit quantitativen und qualitativen, deskriptiven, analytischen und evaluativen wissenschaftlichen Methoden.

LQ-Fragebögen stellen ein wichtiges methodisches Instrument für die wissenschaftliche Evaluation dar. Zur LQ-Messung stehen überwiegend quantitative Instrumente zur Verfügung, es ist jedoch auch ein qualitativer Ansatz möglich.

2. Versorgungsforschung versucht, Bedingungsbeziehungen soweit als möglich kausal zu erklären.

LQ ist nicht nur als Endpunkt von Interesse, sondern kann auch als Mediatorvariable andere versorgungsrelevante Endpunkte beeinflussen (z.B. Wiederaufnahme der Arbeit oder Überleben). Insofern ist die Erfassung von LQ geeignet, Bedingungsbeziehungen zu erhellen.

3. Versorgungsforschung soll zur Neuentwicklung theoretisch und empirisch fundierter oder zur Verbesserung vorhandener Versorgungskonzepte beitragen;

4. die Umsetzung dieser Konzepte begleitend oder ex post erforschen und
5. die Wirkungen von Versorgungsstrukturen und -prozessen oder definierten Versorgungskonzepten unter Alltagsbedingungen mit validen Methoden evaluieren.

Validierte LQ-Fragebögen eignen sich zur Evaluation von Versorgungskonzepten. In neuerer Zeit werden diese Instrumente auch zur Diagnostik von LQ- und Versorgungsdefiziten eingesetzt. Insofern lassen die LQ-Perspektive und die LQ-Methodik wichtige und innovative Erkenntnisse zur Verbesserung vorhandener Versorgungskonzepte erwarten.

Letztlich soll sich die Annäherung für beide Gebiete als Win-Win-Situation erweisen. Die Versorgungsforschung bietet einen interdisziplinären Forschungsansatz mit eigenständigen, praktisch relevanten Fragestellungen und Anwendungsfeldern. Die Lebensqualitätsforschung liefert ein ausgereiftes konzeptuelles und methodisches Inventar und kann sich als wichtige methodische Disziplin und als Katalysator für Interventionssysteme einbringen.

4.2. Definition / Begriffsbestimmung

Historische Entwicklung und Rahmenbedingungen

In der Medizin wurde dem Konzept Lebensqualität im Vergleich zu den Sozial- und Geisteswissenschaften relativ spät Aufmerksamkeit geschenkt. In Abgrenzung zu soziologischen und philosophischen Definitionen wurde in der Medizin der Terminus Lebensqualität immer in Beziehung zu Gesundheit und Krankheit konzeptualisiert. So erklärt sich auch die häufige parallele Verwendung von Begriffen wie „gesundheitsbezogene Lebensqualität“ und „subjektive Gesundheit“ (Patrick & Erickson, 1992).

Mehrere Autoren (Siegrist, 1990; Bullinger, 1997) beschreiben drei Phasen der Entwicklung der Lebensqualitätsforschung. Die erste Phase bildete die Diskussion über die Messbarkeit der Lebensqualität in den 70er Jahren. In der zweiten Phase stand die Entwicklung von Messinstrumenten im Vordergrund. In einer dritten Phase ab den 90er Jahren setzt die Anwendung der Messinstrumente in klinischen Studien ein. Eine vierte Phase sollte die gezielte therapeutische Intervention bei Lebensqualitätsdefiziten sein.

Vor allem in dieser ersten Phase war der Begriff Lebensqualität eng mit der Gesundheitsdefinition der World Health Organization verbunden, nach der Gesundheit als „Zustand voll-

kommenen körperlichen, geistigen und sozialen Wohlbefindens und nicht nur die Abwesenheit von Krankheit und Schwäche“ (WHO, 1948) zu verstehen ist. Diese Definition wird von vielen Autoren als idealtypisch oder illusorisch bezeichnet (Koller & Lorenz, 2002). Die Begriffsbestimmung der WHO ist aber insofern wichtig, da hier erstmals sowohl die psychische als auch die soziale Komponente von Gesundheit berücksichtigt wird.

Weitere Gründe für die Entstehung eines patientenzentrierten Ansatzes in der Medizin liegen in einer steigenden Anzahl älterer Menschen in der Bevölkerung und der damit verbundenen Verschiebung des Krankheitsspektrums von akuten zu chronischen Erkrankungen.

Als Bewertungskriterien, z.B. für die Effizienz einer Therapie, wurde lange Zeit nur die Verlängerung des Lebens bzw. die Veränderung der Symptomatik herangezogen. Der Patient selbst wurde nicht nach seinem Erleben von Krankheit und Therapie gefragt. Als Konsequenz der zunehmenden Lebensqualitätsforschung und des damit verbundenen Perspektivenwechsels wurden psychosoziale Aspekte von Krankheitsprozessen mehr betont, und das subjektive Erleben des Krankheitszustands gewann als Bewertungskriterium von Therapien an Bedeutung (Najman & Levine, 1981).

Definition und aktuelle Trends in der Konzeptualisierung von Lebensqualität

Ein grundlegendes Definitionselement von LQ ist, dass sie sich stets auf die subjektive Sicht der Patienten bezieht, d.h. die Betroffenen geben selbst Auskunft über ihre Befindlichkeit. Auf derartige Selbstberichte stützt sich auch die Messung anderer Konstrukte, wie beispielsweise Patientenzufriedenheit oder subjektives Wohlbefinden. Mittlerweile beginnt sich als übergeordnete Kategorie der Terminus PRO (patient reported outcomes) durchzusetzen (Tabelle 4.1). Daher ist es notwendig, begriffliche Klarheit zu schaffen und darzulegen, was in diesem Memorandum unter „Lebensqualität“ zu verstehen ist.

Tabelle 4.1 : Patientenselbstberichte und zugrunde liegende Konzepte. Nach (FDA & U.S.Department of Health and Human Services, 2006)

PRO (patient reported outcome, Patientenselbstberichte): Jede Äußerung von Betroffenen über ihre gesundheitliche Situation und die damit in Zusammenhang stehende medizinische Behandlung		
Begriff	Konzept	Komplexitätsgrad
Symptom	Wahrnehmung eines einzelnen, abgrenzbaren körperlichen oder psychischen Missempfindens	einfach
Körperliche Aktivität	Zusammenspiel motorischer und psychophysiologischer Prozesse zur Bewältigung von Umweltanforderungen oder selbst gesetzter Ziele	mittel
Patientenzufriedenheit	Bewertende Angabe zur medizinischen Behandlung, zu Gesundheitsdienstleistungen und zum Therapieerfolg	mittel / hoch
Patientenpräferenzen	Ordnen verschiedener Gesundheitszustände oder –ziele nach einer Rangreihe (vgl. <i>Utilities</i>)	mittel / hoch
Lebensqualität (LQ)	Mehrdimensionales Konstrukt, das sich auf die gesundheits- bzw. krankheitsbezogenen Komponenten somatisches, psychisches und soziales Befinden/Funktion bezieht (daher auch „gesundheitsbezogene“ oder „krankheitsbezogene“ LQ)	hoch

Innerhalb der Medizin, der Medizinpsychologie und Medizinsoziologie haben viele Autoren Definitionen des Begriffs Lebensqualität formuliert (Neugebauer et al., 1991; Bullinger, 1991; Siegrist, 1990). Die Formulierung

„Lebensqualität innerhalb der medizinischen Wissenschaften umfasst das subjektive Befinden und Handlungsvermögen im körperlichen, im psychischen und im sozialen Bereich“

dürfte die größte gemeinsame Schnittmenge der meisten Definitionsversuche darstellen, die im Laufe der mittlerweile mehr als drei Dekaden dauernden Beschäftigung mit dem Thema Lebensqualität von verschiedenen Autoren und Forschergruppen vorgeschlagen wurden. Im vorliegenden Rahmen wird Lebensqualität stets in Bezug auf Gesundheit (health related

quality of life, HRQL) und/oder Krankheit (krankheitsbezogene Lebensqualität) betrachtet. Die Verwendung des neutralen Begriffs Lebensqualität trägt darüber hinaus der Tatsache Rechnung, dass subjektive Wahrnehmungen und Berichte von vermeintlich gesundheitsbezogenen Symptomen und Empfindungen keineswegs Abbildungen objektivierbarer, medizinischer Gegebenheiten sind, sondern in vielfältiger Weise von psychosozialen Variablen beeinflusst werden (Pennebaker, 1982; Koller et al., 1996).

Aktuelle Trends in der LQ-Forschung, die sowohl für die Präzisierung des LQ-Begriffs als auch für die Versorgungsforschung von Interesse sind, beziehen sich auf die Konzeptualisierung spezifischer Lebensqualitätsdimensionen (z.B. Spiritualität, Kohärenz, Kognition), auf die Integration patienten- und arztseitiger Parameter sowie auf das Phänomen des Response-Shifts.

Die Dimension „Spiritualität“ spielt bei der Bewältigung von chronischen oder lebensbedrohlichen Erkrankungen (Büssing et al., 2007; Büssing et al., 2008) eine große Rolle und ist daher prinzipiell auch in der Versorgungsforschung anwendbar. Unter Spiritualität wird ein Konzept verstanden, das weit über die institutionalisierte Religiosität hinausgeht und die individuelle Suche nach Rückhaltgebenden und Sinnstiftenden Ressourcen berücksichtigt. Ein insbesondere in der Gesundheitspsychologie vielfach untersuchtes, verwandtes Konzept ist die Kohärenz (sense of coherence; Eriksson & Lindström, 2006; Tagay et al., 2006).

Die demographische Veränderung der Gesellschaft und die damit verbundene drastische Zunahme von Demenzerkrankungen mit zunehmendem Alter wirft folgende Frage auf: kann ein so komplexes Konstrukt wie LQ, das über die kognitive Modalität mittels Fragebogen erhoben wird, bei Personen mit kognitiven Defiziten erfasst werden (v.Steinbüchel et al., 2005b)? Interdisziplinäre Arbeiten zeigen Möglichkeiten und Grenzen der Messbarkeit des Konstrukts LQ bei Personen mit kognitiver Einschränkung mit einem MMSE (Mini Mental State Extended) Punktwert unter 12 (Smith et al., 2005; v.Steinbüchel et al., 2005a; Berger et al., 1999).

Für die praktische Umsetzung müssen neben der LQ auch traditionelle, arztbasierte Outcomeparameter und das Kriterium des klinisch relevanten Endpunktes berücksichtigt werden. Diese verschiedenen Messebenen wurden in einem Drei-Komponenten-Modell integriert (Troidl et al., 1998; Lorenz et al., 1999; Koller & Lorenz, 2002) und empirisch überprüft (Klinkhammer-Schalke et al., 2008a; Koller & Lorenz, 2002). Der Aspekt der Integration von patienten- und arztseitigen Urteilen stellt ein wichtiges Ziel einer umfassenden Bewertung medizinischer Interventionen dar (Janse et al., 2004; Klinkhammer-Schalke et al., 2008b).

Response Shift bezeichnet das Phänomen, dass Personen, die eine Veränderung ihres Gesundheitsstatus (z.B. bedingt durch eine Intervention) erleben, unter Umständen ihre internen Beurteilungsstandards und ihre Bewertungen von Gesundheitszuständen verändern (Schwartz et al., 2006). Wenn LQ-Werte zur Bestimmung der Effektivität medizinischer Interventionen herangezogen werden, ist die Berücksichtigung dieses Phänomens von großer Bedeutung.

Neben der Lebensqualität sind in der Versorgungsforschung auch weitere Patienten-relevante Outcome-Konzepte von Bedeutung. Hierzu zählen die Patientenzufriedenheit (van Campen et al., 1995; Hardy et al., 1996), die Patientenpräferenzen (Bennett & Torrance, 1996; Stevens et al., 2005) sowie die patientenseitige Nutzenbewertung möglicher Gesundheitszustände (Froberg & Kane, 1989). Der Nutzenbegriff ist auch aktueller Gegenstand der Diskussion auf der Ebene der Entscheidungsträger in Deutschland (Gesundheitsforschungsrat des BMBF, 2007). Die genannten Konstrukte sind für die Versorgungsforschung bedeutsam, unterscheiden sich aber vom Konzept der Lebensqualität. Aus diesem Grunde werden sie an vorliegender Stelle nicht behandelt, sondern bedürfen einer separaten Betrachtung.

4.3. Methodische Standards: Messung

Messmethode

Besonders prägend für die LQ-Forschung war deren zweite Entwicklungsphase ab den 80er Jahren: über viele Jahre war die Forschung durch methodische Fragestellungen bestimmt und dabei immer eng mit der Fragebogenmethode verbunden.

Wichtige Gründe für die Popularität der Fragebogenmethode sind:

- die recht hohen Erfolgsaussichten bei der Entwicklung von Fragebögen durch die Vorgabe klarer methodischer Regeln;
- die Verfügbarkeit von Messgütekriterien und die damit verbundene Wissenschaftlichkeit und Glaubwürdigkeit und
- der überschaubare logistische Aufwand bei der Integration von Fragebögen in klinische Studien bei hoher Akzeptanz seitens der Patienten.

Dieser Forschungstradition folgend wurde die Entwicklung neuer Fragebögen und deren Validierung, Übersetzung in internationale Sprachen (Bullinger et al., 1993; Koller et al., 2007) oder die Entwicklung unterschiedlicher Fragebogenversionen (Langform, Kurzform,

computeradaptiert) vorangetrieben. Demzufolge ist ein hohes Methodenbewusstsein entstanden, insbesondere durch den Einfluss von Psychologen, Soziologen und Epidemiologen.

Messgüte

Die klassischen Messgütekriterien innerhalb der Psychometrie lauten Reliabilität, Validität und Sensitivität (Tabelle 4.2). Bei moderneren testtheoretischen Ansätzen wie den Item-Response-Modellen (Reeve et al., 2007) ist als weiteres Gütekriterium die Passung zum zugrunde gelegten Messmodell zu berücksichtigen.

Weitere Gütekriterien für LQ-Messinstrumente sind (Albers et al., 2007)

- Objektivität: verschiedene Personen, die die Messungen unabhängig voneinander vornehmen, gelangen zu den gleichen Messergebnissen (Differenzierung nach Durchführungs-, Auswertungs- und Interpretationsobjektivität);
- Normierung: es liegen Messwerte anderer Stichproben (Normstichproben, verschiedene Patientengruppen) vor, die für die Einordnung der individuellen Testergebnisse als Bezugssystem dienen können;
- Ökonomie: Betrachtung der Messung unter wirtschaftlichen Gesichtspunkten, Kosten der Durchführung und der Zeitbedarf der Messung stehen in einem angemessenen Verhältnis zum Erkenntnisgewinn;
- Nützlichkeit: das Messinstrument weist eine gute Anwendbarkeit auf.

Tabelle 4.2: Psychometrische Kriterien der Messgüte eines Fragebogens. Nach Koller, 2007

Kriterium	Inhaltliche Bedeutung	Statistische Verfahren
Reliabilität	<i>Wie genau misst der Fragebogen?</i>	Cronbachs Alpha Split-half Reliabilität Test-Retest Reliabilität
Validität	<i>Misst der Fragebogen das, was er messen soll?</i> inhaltliche Validität: Berücksichtigung aller wichtigen Inhalte Konstruktvalidität: Beziehungen zu verwandten Konstrukten oder Außenkriterien Augenscheinvalidität: „gesunder Menschenverstand“	Gruppenvergleiche Zusammenhänge t-Test Mann-Whitney Varianzanalyse Korrelation Regressionsanalysen
Sensitivität	<i>Kann der Fragebogen Veränderungen erfassen?</i>	vorher-nachher-Vergleiche abhängiger t-Test Wilcoxon

In neuerer Zeit rückt auch das Gütekriterium Interpretierbarkeit in den Vordergrund (FDA & U.S. Department of Health and Human Services, 2006). Es gibt verschiedene Methoden und Ansätze zu deren Bestimmung (Crosby et al., 2003): Ankerbasierte Ansätze, die die Veränderung des LQ-Wertes mit einem anderen klinisch relevanten Maß (z.B. direkte Veränderungseinschätzung, Referenzwerte) in Beziehung setzen, verteilungsbasierte Ansätze, die auf der Analyse statistischer Verteilungsmerkmale (z.B. Effektstärken) basieren, die Bestimmung des gerade noch bedeutsamen Mittelwertsunterschieds (minimal clinically important difference) oder die Bestimmung von cut-off-Werten (zufriedenstellende – nicht zufriedenstellende LQ).

Der Interpretierbarkeit von LQ kommt innerhalb der Versorgungsforschung eine besonders große Bedeutung zu, weil Versorgungsprozesse und -ergebnisse durch eine Vielzahl von Parametern abgebildet werden können. Ein klar interpretierbares LQ-Ergebnis hilft, den Stellenwert der LQ im Rahmen von anderen Messgrößen einzuordnen.

Messinstrumente zur Lebensqualität und deren Klassifikation

LQ-Fragebögen erfassen stets die subjektive Sicht der Patienten. Diese subjektive Perspektive spiegelt die Lebenssituation mit deren spezifischen Erlebnis- und Bewertungsinhalten wider. Daher gibt es nicht, wie aus praktischer Sicht wünschenswert wäre, einen allgemein gültigen Fragebogen zur Erfassung der Lebensqualität, sondern eine Vielzahl derartiger Instrumente. Eine genaue Schätzung über die Anzahl lässt sich schwer angeben, aber man kann von mehreren hundert Fragebögen ausgehen, die Lebensqualität und verwandte Konstrukte bei den unterschiedlichsten Krankheitsbildern erfassen (Schumacher et al., 2003; Bowling, 2001; McDowell & Newell, 1987). Tabelle 4.3 zeigt ein etabliertes Klassifikationssystem von LQ-Fragebögen auf der Basis ihrer inhaltlichen Schwerpunktbildung.

Tabelle 4.3: Klassifikation von LQ-Fragebögen

Fragebogentyp	Inhaltlicher Schwerpunkt	Beispiel
Generisch	multidimensionales Konstrukt der LQ krankheitsübergreifende Aspekte der LQ	SF-36
Krankheitsspezifisch	multidimensionales Konstrukt der LQ Krankheitsspezifische Symptome und Funktionseinschränkungen	EORTC QLQ-C30 POLO-Chart
Präferenzmaß	Rangreihe verschiedener Gesundheitszustände	Euroqol
Individualisiertes Maß	Patientenspezifische Gewichtung verschiedener Dimensionen der LQ	SeiQol

Ein spezifischer LQ-Bogen für die Versorgungsforschung liegt nicht vor. Die Entwicklung eines derartigen Bogens erscheint entbehrlich, da eine Fülle bewährter Instrumente vorhanden sind. LQ-Konzepte und –Messmethoden sind keine Frage des Studientyps (z.B. Studien zur Versorgungsforschung), sondern müssen das Beschwerdebild abdecken, das bezüglich einer bestimmten Patientenpopulation und/oder Forschungsfrage relevant ist. Eine Auswahl von LQ-Fragebögen, die in der Versorgungsforschung Anwendung finden können, zeigt Tabelle 4.4 im Überblick.

Tabelle 4.4 : Wichtige Fragebogeninstrumente zur Lebensqualität. Nach Bullinger & Ravens-Sieberer, 2001

Verfahren (Autoren)	Ursprungsland	Typ	Kennzeichen	Durchführungszeit	Dt. Normwerte
DISABKIDS (Petersen et al., 2005)	Internat.	chronisch-generisch (für Kinder mit chronischen Erkrankungen)	Alter: 4-16 6/37 + Module 1/6 Skalen Selbst-/ Fremdbeurteilung	5-20 min	ja
EuroQoL European Quality of Life Instrument (EURO-QOL) (Kind et al., 1998)	UK	Präferenzmaß	5 Items 1 Skala Selbstbeurteilung	5 min	
EORTC QLQ C30 (Aaronson et al., 1993)	UK	Krankheitsspezifisch (Krebs)	30 Items 15 Scores Selbstbeurteilung	10 min	ja

FACT (Cella et al., 1993)	USA	Krankheitsspezifisch (Krebs)	38 Items 8 Scores Selbstbeurteilung	10 min	
KIDSCREEN (Ravens-Sieberer et al., 2005)	Internat.	generisch (für Kinder)	Alter: 8-18 10/27/52 Items 1/6/10 Skalen Selbst-/ Fremdbeurteilung	10-25 min	ja
KINDL (Ravens-Sieberer & Bullinger, 1998)	D	generisch (für Kinder)	Alter: 4-16 24 Items + Module 6 Skalen, Selbst-/ Fremdbeurteilung	15 min	
Nottingham Health Profile (NHP) (Kohlmann et al., 1997)	UK	generisch	38 Items 6 Skalen Selbstbeurteilung	15 min	
PLC (Siegrist et al., 1996)	D	generisch	40 Items 6 Skalen Selbstbeurteilung	15 min	ja
PoloChart (Pirente et al., 2002)	D	Validiertes Kombinationsinstrument für Traumapatienten	Enthält GOS,VAS, Prestatus. EURO-QoL, SF 36 traumaspez.Modul-TOP mit 10 Dimensionen, 57 Items, 2 Ergänzungsfragen zu Körperbild und Zufriedenheit	35 min	ja
SF-36 Health Survey (Ware & Sherbourne, 1992)	USA	generisch	36 Items 8 Skalen Selbst-/ Fremdrating	15 min	ja
Sickness-Impact- Profile (SIP) (Bergner et al., 1981)	USA	generisch	136 Items 12 Skalen Selbst-/ Fremdbeurteilung	30 min	
World Health Organization Quality of Life Assessment Instrument WHO-QOL-100 (The WHOQOL Group, 1993)	Internat.	generisch	100 Items 6 Skalen Selbstbeurteilung	36 min	

Auswahl eines Fragebogens für ein Projekt

Verschiedentlich sind Algorithmen zur Auswahl von LQ-Fragebögen publiziert worden (Osoba et al., 1991), die in der Forschungspraxis allerdings kaum Verwendung zu finden scheinen.

Bei der Auswahl eines geeigneten Messinstruments für ein Projekt innerhalb der Versorgungsforschung sind die psychometrischen Kennwerte eines Bogens ein wenig selektives Kriterium. Dies deshalb, weil praktisch alle heute etablierten Bögen einen mehrjährigen, standardisierten Entwicklungsprozess durchlaufen haben und über zufriedenstellende bis sehr gute psychometrische Kennwerte verfügen. Allerdings unterscheiden sich Fragebögen erheblich in ihrer inhaltlichen Ausrichtung, d.h. der relativen Gewichtung der physischen, psychischen oder sozialen Dimension sowie der Betonung von negativen oder positiven Erlebnisinhalten (Koller, 2007). Der Forscher ist daher gut beraten, verschiedene Fragebögen aufmerksam zu lesen und in inhaltlicher Hinsicht zu vergleichen. Die Entscheidung für einen Bogen soll sich primär an der Forschungsfrage orientieren: deckt der Bogen wirklich die Inhalte ab, die im Rahmen des zu untersuchenden Versorgungsproblems relevant sind?

Man soll begründen, warum man sich für einen bestimmten Bogen entschieden hat. Der lapidare Hinweis, der Bogen wurde auch in anderen Studien verwendet, genügt als alleinige Begründung nicht, es sei denn, man benötigt eine Vergleichsgruppe im Sinne einer historischen Kontrolle. Pilotstudien und Fokusgruppen sind hilfreiche methodische Mittel, um eine empirische Basis für die Entscheidung zu haben, wenn zwei oder mehrere Bögen für ein Projekt geeignet erscheinen (Götz et al., 2007).

Qualitative Methoden

In der LQ-Forschung dominiert der quantitative methodische Zugang. Daneben gibt es aber auch qualitative Methoden wie Fokusgruppen (Lehoux et al., 2006), kognitive Interviews (Willis, 2005) oder qualitative Einzelinterviews (Ciechanowski & Katon, 2006). Diese stellen eigenständige Methoden dar, können aber auch in Kombination mit quantitativen Verfahren eingesetzt werden. Auch wenn qualitative Verfahren bislang weniger verbreitet sind, können sie für bestimmte Themen der Versorgungsforschung einen relevanten methodischen Zugang darstellen.

4.4. Methodische Standards: Design

Versorgungsforschung versteht sich als „grundlagen- und anwendungsorientiertes fachübergreifendes Forschungsgebiet“, das „Versorgungsstrukturen, -prozesse und –konzepte unter Alltagsbedingungen mit validen Methoden evaluiert“ (Arbeitskreis Versorgungsforschung beim Wissenschaftlichen Beirat, 2004). Diese breite Definition bringt es mit sich, dass das Methodenarsenal der Versorgungsforschung eine Vielzahl von Studientypen und Studiendesigns beinhaltet: experimentelle Grundlagenforschung, klinisch-epidemiologische Forschung, quasi-experimentelle epidemiologische Studien, Metaanalysen, systematische Reviews, und Implementierungsstudien (Arbeitskreis Versorgungsforschung beim Wissenschaftlichen Beirat, 2004). Versorgungsforschung greift auf verschiedene Datenquellen (GKV-Routinedaten, Registerdaten, Befragungen etc.) zu, die im Rahmen unterschiedlicher Studiendesigns (prospektiv, retrospektiv, RCT, Kohortenstudien, Korrelationsstudien etc.) erhoben werden.

LQ-Messung benötigt prospektive Studiendesigns. Die Messzeitpunkte sind im Studienprotokoll festzulegen. Bei Longitudinal-Studien ist die Erhebung von Baseline LQ-Werten wünschenswert (Efficace et al., 2007), da dadurch Aussagen über die Veränderungen der LQ im Zuge des Versorgungsprozesses gemacht werden können. Darüber hinaus kann anhand der Baseline-Daten in kontrollierten Studien die Vergleichbarkeit der Gruppen zu Beobachtungsbeginn geprüft werden.

Bei der vergleichenden Untersuchung von Interventionen sind offene Studien ggf. von eingeschränkter Validität, weil Angaben zur LQ aufgrund ihrer Natur subjektiv sind (FDA & U.S. Department of Health and Human Services, 2006; Chassany et al., 2002). Die Verzerrung der Ergebnisse subjektiver Endpunkte in offenen Studien ist empirisch belegt (Wood et al., 2008). Wenn möglich, sollten Studien zum Vergleich von Interventionen deshalb verblindet durchgeführt werden. Ist eine Verblindung nicht möglich, sollten andere Anstrengungen zur Verzerrungsminimierung oder -einschätzung unternommen werden.

Besondere methodische Umsicht ist bei der Festlegung des primären Endpunkts gefordert. Es ist derjenige Messzeitpunkt zu wählen, der klinisch relevant ist (zu diesem Zeitpunkt sollte eine Erholung oder Veränderung nach medizinischen Vorgaben eingetreten sein). Bei der Verwendung von LQ-Instrumenten mit mehreren Domänen ist die Domäne, die den primären Endpunkt bildet, prospektiv festzulegen. Wird der primäre Endpunkt von mehreren Domänen gebildet, sind Probleme des multiplen Testens adäquat zu berücksichtigen.

LQ ist ein Outcome-Kriterium, das prinzipiell in allen prospektiven Studientypen Verwendung finden kann, die deren valide Erhebung erlauben. Der jeweils geeignete Studientyp hängt von der Fragestellung ab, die die Studie beantworten soll. Tabelle 4.5 zeigt beispielhaft Fragestellungen der Lebensqualitätsforschung und die für die Beantwortung der Fragen geeigneten Studientypen.

Tabelle 4.5: Forschungsfrage und geeigneter Studientyp (Beispiele)

Fragestellung	Studientyp	Anmerkung
Zusammenhang zwischen natürlicher Krankheitsprogression und LQ	Prospektive Kohortenstudie	z.B. auch zur Unterstützung der Interpretation von LQ Daten (anchor-based approach)
Erhebung der LQ in einer definierten Population	Querschnittsstudie	z.B. zur Ermittlung von Normdaten
Vergleich der Effekte von Interventionen auf die LQ	Randomisierte kontrollierte Studie	bei sehr großen Unterschieden zwischen Effekten alternativer Interventionen können ggf. auch nicht-randomisierte Studien hinreichend sichere Ergebnisse liefern (Glasziou et al., 2007)
Einfluss potenzieller Effektmotivatoren (z.B. Alter, Schweregrad der Erkrankung) auf die LQ	Querschnittsstudie oder prospektive Kohortenstudie	u. a. als Hilfestellung bei der Beurteilung der externen Validität von Studien zur Untersuchung von Interventionseffekten (Sugano DS & McElwee NE, 1996)

4.5. Schlüsselkriterien für die Messung von Lebensqualität bei Studien zur Versorgungsforschung

Als Fazit unserer konzeptuellen und methodischen Ausführungen lassen sich folgende Schlüsselkriterien festlegen, die qualitativ hochwertige Studien zur Versorgungsforschung, in denen Lebensqualität als zentrales Outcome-Kriterium gemessen wird, auszeichnen.

Messinstrument: Inhalte/Dimensionen, Messgüte und Auswahl

- Standardgütekriterien müssen erfüllt sein (Reliabilität, Validität, Sensitivität).
- Auswahl des Messinstruments muss begründbar sein. Insbesondere müssen Form und Inhalt des Fragebogens (Dimensionen der LQ) von Relevanz für die wissenschaftliche Fragestellung und dem Alter der zu Untersuchenden angemessen sein. Auch sollte berücksichtigt werden, ob die Interpretierbarkeit der gewonnenen Daten, z. B. durch die Verfügbarkeit von Referenzdaten, cut-off Werten oder anderer Kriterien, gegeben ist.

- Fragebögen, die nicht primär in Deutschland entwickelt wurden, müssen eine standardisierte Übersetzung und kulturell angemessene Adaptierung durchlaufen haben.

Studiendesign und Studienpopulation

- Das Studiendesign muss zur Beantwortung der wissenschaftlichen Fragestellung geeignet sein.
- Es ist eine ausreichend große Stichprobe mit biometrischer Begründung der Stichprobengröße vorzusehen.
- Die Festlegung der Stichprobencharakteristika muss mit der wissenschaftlichen Fragestellung korrespondieren; die Stichprobe muss repräsentativ für die Studienpopulation sein, für die die Generalisierung der Befunde gedacht ist.

Durchführung der Studie

- Lebensqualität ist durch Selbstberichte zu erfassen; Proxy-Berichte (z.B. Angehörige oder Pflegende) sind nur in begründeten Ausnahmefällen zulässig.
- Mess-Situation muss standardisiert sein, um eine valide LQ-Messung zu gewährleisten.
- Zeitpunkt der LQ-Erhebung muss beschrieben und begründet sein; insbesondere ist auf den Messzeitpunkt für das primäre LQ-Kriterium zu achten.

Auswertung / Interpretation / Berichterstattung

- Die Objektivität von Auswertung und Interpretation ist sicherzustellen; insbesondere sind Auswertungsstrategien und Interpretationskriterien (z. B. aktuelle Referenzwerte) prospektiv festzulegen.
- Auf eine allgemein-intuitiv verständliche Darstellung der LQ-Ergebnisse auf Ebene der Stichprobe und Individualpatienten (Profil) ist zu achten.
- Interdisziplinäre Bewertung und Begutachtung (bei der Beurteilung der Studien-/Messergebnisse) ist wünschenswert.
- Berichte und Publikationen zu LQ-Studien sollten sowohl die generellen Standards für wissenschaftliche Publikationen erfüllen – CONSORT (Moher et al., 2001b), STROBE, TREND, QUOROM, MOOSE (siehe zusammenfassend www.consort-statement.org) - als auch LQ-spezifische Publikationsanforderungen berücksichtigen, z. B. ERIQA-Checkliste (Chassany et al., 2002).

Anwendung und praktische Implikationen

- Alle Studien zur Versorgungsforschung sind bereits bei Planung, Auswertung und Interpretation so anzulegen, dass anwendungsrelevante Aussagen zur LQ getroffen werden können.

- Bei Studien mit explizitem Anwendungscharakter muss das LQ-Messinstrument innerhalb des Untersuchungssettings implementiert werden; dabei ist die interdisziplinäre Zusammenarbeit von Methodikern, Nutzern und Therapeuten unverzichtbar.
- Bei Studien mit diagnostischem/therapeutischem Charakter sind Standards für die therapeutische Intervention und Kriterien für die Umsetzung der Intervention (z.B. cut-off Werte) zu spezifizieren.

4.6. Zusammenfassung / Schlussfolgerung

LQ deckt Inhalte des subjektiven Erlebens und individuellen Verhaltens im Schnittpunkt medizinisch-relevanter und sozial-ökonomischer Belange ab.

LQ gibt Auskunft darüber, welche Aspekte der Versorgung beim Patienten „angekommen“ sind. LQ ist primär das, was Patienten empfinden, dann aber auch das, worüber Patienten sprechen, und was in ganz erheblichem Maße die gesellschaftliche Akzeptanz von Versorgungsangeboten und -prozessen bestimmt.

Daher ist LQ als ein zentraler Endpunkt innerhalb der Versorgungsforschung anzusehen.

Es gibt eine Vielzahl etablierter LQ-Bögen, die für viele Forschungsfragen tauglich sind und deren Messgüte in umfangreichen empirischen Studien nachgewiesen wurde. Ein spezifischer LQ-Bogen für die Versorgungsforschung ist nicht bekannt. Die Notwendigkeit zur Entwicklung eines solchen spezifischen Bogens wird nicht gesehen.

Die Entscheidung für einen LQ-Bogen sollte begründet werden und mit der Forschungsfrage und der Patientenpopulation kompatibel sein. Eine Pilotstudie zur Auswahl eines geeigneten Messinstruments ist wünschenswert.

Die Praxis der LQ-Messung in der Versorgungsforschung soll die allgemein gültigen Kriterien für gute Studiendesigns und -durchführung berücksichtigen: prospektives Studiendesign, definierte Messzeitpunkte, Machbarkeit, Vermeidung von lost cases bei Mehrfachmessungen.

Die Interpretierbarkeit von LQ-Ergebnissen ist von größter Bedeutung für versorgungsrelevante Schlussfolgerungen. Es sollen a priori Kriterien für akzeptable/inakzeptable LQ-Werte festgelegt werden (z.B. cut-offs, Referenzwerte). Diese Kriterien sollen nicht nur durch statistische Kennwerte begründet sein, sondern müssen auch der Forschungsfrage und den Interventionsmöglichkeiten gerecht werden.

Von der Arbeitsgruppe wurden Schlüsselkriterien für die Beurteilung der methodischen Güte von Studien in der Versorgungsforschung festgelegt. Diese Schlüsselkriterien beziehen sich auf die fünf Bereiche Messinstrument, Studiendesign/Studienpopulation, Durchführung/Stichprobe, Auswertung/Interpretation, Anwendung/praktische Implikation. Hochwertige Studien werden in allen diesen Bereichen die postulierten Schlüsselkriterien erfüllen.

4.7. Literatur

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J. et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *J.Natl.Cancer Inst.*, 85, 365-376.
- Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J. (2007). *Methodik der empirischen Forschung*. Wiesbaden: Gabler.
- American Society of Clinical Oncology (1996). Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J.Clin.Oncol.*, 14, 671-679.
- Arbeitskreis Versorgungsforschung beim Wissenschaftlichen Beirat. (2004). Definition und Abrenzung der Versorgungsforschung. http://www.versorgungsforschung.nrw.de/content/e54/e104/e259/object260/Def_Versorgungsforsch_2004_BAEK.pdf.
Ref Type: Generic
- Bagozzi, R. & Baumgartner, H. (1994). The evaluation of structural equation models and hypothesis testing. In R.Bagozzi (Ed.), *Principles of Marketing Research* (pp. 386-422). Cambridge: Blackwell Publishers.
- Bauhofer, A., Plaul, U., Torossian, A., Koller, M., Stinner, B., Celik, I. et al. (2007). Perioperative prophylaxis with granulocyte colony-stimulating factor (G-CSF) in high-risk colorectal cancer patients for an improved recovery: A randomized, controlled trial. *Surgery*, 141, 501-510.
- Bennett, K. J. & Torrance, G. W. (1996). Measuring Health State Preferences and Utilities: Rating Scale, Time Trade-Off, and Standard Gamble Techniques. In B.Spilker (Ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials* (pp. 253-265). Philadelphia.
- Berger, E., Leven, F., Pirente, N., Bouillon, B., & Neugebauer, E. (1999). Quality of life after traumatic brain injury: A systematic review of the literature. *Restor.Neurol.Neurosci.*, 14, 93-102.
- Bergner, M., Bobbitt, S. A., Carter, W. B., & Gilson, B. S. (1981). The Sickness Impact Profile: Development and final revision of a health status measure. *Med.Care*, 19, 780-805.

- Black, N., Murphy, M., McKnee, M., Sanderson, C., Askham, J., & Merta, T. (1999). Consensus development methods: a review of best practice creating clinical guidelines. *Journal of Health Services Research and Policy*, 4, 236-248.
- Blalock, H. M. (1960). *Social Statistics*. New York, Toronto, London: Mc Graw Hill.
- Blalock, H. M. & Blalock, A. B. (1985). *Methodology in Social Research*. New York, Toronto, London: Mc Graw Hill.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York.
- Bötker, F. (1976). Auswahlverfahren, eine Einführung für Sozialwissenschaftler. In (Stuttgart: Teubner).
- Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation. In (Springer).
- Bowling, A. (2001). *Measuring disease. A review of disease-specific quality of life measurement scales*. (2 ed.) Buckingham: Open University Press.
- Brewin, C. R. & Bradley, C. (1989). Patient preferences and randomised controlled trials. *BMJ*, 289, 315.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. (2.Auflage ed.) München: Pearson.
- Bullinger, M. & Ravens-Sieberer, U. (2001). Diagnostik der Lebensqualität. In R.D.Stieglitz, U. Baumann, & M. Freyberger (Eds.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie* (pp. 246-257). Berlin: Thieme Verlag.
- Bullinger, M. (1991). Quality of life: definition, conceptualization and implications - a methodologist's view. *Theor.Surg.*, 6, 143-148.
- Bullinger, M. (1997). Gesundheitsbezogene Lebensqualität und subjektive Gesundheit. *Psychother.Psychosom.Med.Psychol.*, 47, 76-91.
- Bullinger, M., Anderson, R., Cella, D., & Aaronson, N. (1993). Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual.Life Res.*, 2, 451-459.
- Büssing, A., Ostermann, T., & Koenig, H. (2007). Relevance of religion and spirituality in German patients with chronic diseases. *Int.J.Psychiatry Med.*, 37, 39-57.
- Büssing, A., Ostermann, T., & Matthiessen, P. F. (2008). Adaptive Coping und Spiritualität als Ressource bei Krebspatienten. *PRÄVENTION - Zeitschrift für Gesundheitsförderung*, 31, 51-53.
- Campbell, M., Fitzpatrick, R., & Haines, A. (2000). Framework for the design and evaluation of complex interventions to improve health. *BMJ*, 321, 694-696.
- Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004a). CONSORT statement: extension to cluster randomised trials. *BMJ*, 328, 702-708.
- Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004b). CONSORT statement: extension to cluster randomised trials. *BMJ*, 328, 702-708.
- Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004c). CONSORT statement: extension to cluster randomised trials. *BMJ*, 328, 702-708.

- Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Lloyd, S., Linn, E. et al. (1993). The functional assessment of cancer therapy (FACT) scale: Development and validation of the general measure. *J.Clin.Oncol.*, 11, 570-579.
- Chassany, O., Sagnier, P., Marquis, P., Fullerton, S., Aaronson, N., & for the ERIQA group (2002). Patient-reported outcomes: the example of health-related quality of life – a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. *Drug Information Journal*, 36, 209-238.
- Christiansen, G. (2001). *Evaluation - ein Instrument zur Qualitätssicherung in der Gesundheitsförderung. Eine Expertise*. Köln: Bundeszentrale für gesundheitliche Aufklärung.
- Ciechanowski, P. & Katon, W. J. (2006). The interpersonal experience of health care through the eyes of patients with diabetes. *Soc.Sci.Med.*, 63, 3067-3079.
- Consort-Statement Group. (30-9-2008a). www.Consort-Statement.Org (30.09.2008).
Ref Type: Internet Communication
- Consort-Statement Group. (30-9-2008b). www.Consort-Statement.Org (30.09.2008).
Ref Type: Internet Communication
- Consort-Statement Group. (30-9-2008c). www.Consort-Statement.Org (30.09.2008).
Ref Type: Internet Communication
- Cook, T. D. & Shadish, W. R. (1994). "Social Experiments: Some Developments Over the Past Fifteen Years". *Annual Review of Psychology*, 45, 545-580.
- Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of care. *J.Clin.Epidemiol.*, 56, 395-407.
- Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group Techniques for Program Planning: A guide to Nominal Group and Delphi Process*. Glenview,IL, USA: Scott, Foresman and Company.
- Deuschl, G., Schade-Brittinger, C., Krack, P., Volkmann, J., Schafer, H., Botzel, K. et al. (2008). A randomized trial of deep-brain stimulation for Parkinson's disease. *N.Engl.J.Med.*, 355, 896-908.
- Diamantopoulos, A. & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, 38, 269-277.
- Diekmann, A. (2008). *Empirische Sozialforschung Grundlagen, Methoden, Anwendungen*. (14. Auflage ed.) Hamburg: Rowolt.
- Dorroch, H. (1994). *Meinungsmacherreport. Wie Umfrageergebnisse entstehen*. Göttingen: Steidl.
- Dunham, R. B. (1998). Nominal Group Technique: users' guide.
Ref Type: Unpublished Work
- Efficace, F., Osoba, D., Gotay, C., Sprangers, M., Coens, C., & Bottomley, A. (2007). Has the quality of health-related quality of life reporting in cancer clinical trials improved over time? Towards bridging the gap with clinical decision making. *Ann.Oncol.*, 18, 775-781.

- EMA. (2005). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. www.emea.europa.eu/pdfs/human/ewp/13939104en.pdf.
Ref Type: Generic
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Eriksson, M. & Lindström, B. (2006). Antonovsky's sense of coherence scale and the relation with health: a systematic review. *Journal of Epidemiology & Community Health*, 60, 381.
- FDA & U.S.Department of Health and Human Services. (2006). Guidance for Industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. <http://www.fda.gov/CDER/GUIDANCE/5460dft.pdf>.
Ref Type: Generic
- Flick, U. (2000a). Qualitative Forschung. Ein Handbuch. Flick, Uwe, von Kardorff, Ernst, and Steinke, Ines. 1-768. Hamburg, rororo. rowohlt's enzyklopädie. Burghard, König.
Ref Type: Serial (Book, Monograph)
- Flick, U. (2004). *Qualitative Forschung ein Handbuch*. Hamburg: Rowohlt Taschenbuch Verlag.
- Flick, U. (2000b). Design und Prozess qualitativer Forschung. In U.Flick, E. von Kardorff, & I. Steinke (Eds.), *Qualitative Forschung Ein Handbuch* (pp. 252-265). Hamburg: rororo.
- Froberg, D. G. & Kane, R. L. (1989). Methodology for measuring health-state preferences I: Measurement strategies. *J.Clin.Epidemiol.*, 42, 345-354.
- Gabler, S. (1992). Schneeballverfahren und verwandte Stichprobendesigns. *ZUMA-Nachrichten*, 31, 47-69.
- Gabler, S. (1997). Stichproben in der Umfragepraxis. Gabler, S. and Hoffmeyer-Zlotnik, J. H. P. 1-246. Opladen, Westdeutscher Verlag.
Ref Type: Serial (Book, Monograph)
- Gesundheitsforschungsrat des BMBF. (2007). Diskussionsforum zur Nutzenbewertung im Gesundheitswesen. Begriffsdefinition und Einführung. Dokumentation des 1. gemeinsamen Workshops von GFR und IQWiG am 4. September 2007 in Berlin. www.gesundheitsforschung-bmbf.de/de/1191.php.
Ref Type: Generic
- Glasziou, P. P., Chalmers, I., Rawlins, M., & McCulloch, P. (2007). When are randomised trials unnecessary? Picking signal from noise. *BMJ*, 334, 349-351.
- Götz, S., Lunz, B., Koller, M., Görse, R., & Ortmann, O. (2007). Evaluation der Lebensqualität bei adjuvanter endokriner Therapie des Mammakarzinoms: Pilotstudie zur Auswahl eines geeigneten Messinstruments. *Senologie - Zeitschrift für Mammadiagnostik und -therapie*, DOI: 10.1055/s-2007-982941.
- Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung*. Weinheim: Psychologie Verlags Union.
- Hardy, G. E., West, M. A., & Hill, F. (1996). Components and predictors of patient satisfaction. *Brit J Health Psychol*, 1, 65-85.

- Hox, I. J. & Maas, C. J. M. (2005). Multilevel Models for Multimethod Measurement. In M. Eid & E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 269-281). Washington D.C.: American Psychological Association.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2008). Allgemeine Methoden: Version 3.0 vom 27.05.2008.
www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf. [Zugriff am 06.06.2008].
 Ref Type: Generic
- Janse, A. J., Gemke, R. J. B. J., Uiterwaal, C. S. P. M., van der Tweel, I., Kimpfen, J. L. L., & Sinnema, G. (2004). Quality of life: patients and doctors don't always agree: a meta-analysis. *J.Clin.Epidemiol.*, 57, 653-661.
- Kalton, G. (1983). *Introduction to Survey Sampling*. London: Sage UNIVERSITY PAPER.
- Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ*, 316, 736-741.
- Kleespies, C., Kaise, T., Sawicki, P. T., für die Arbeitsgruppe Praktische evidenzbasierte Medizin, St. F. H. K., & IeM - Institut für evidenzbasierte Medizin, K. (2005). *Begriffe und Methoden der evidenzbasierten Medizin - Ein Glossar* Köln: DieM - Institut für evidenzbasierte Medizin, Köln.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: Guilford.
- Klinkhammer-Schalke, M., Koller, M., Wyatt JC, Steinger, B., Ehret, C., Ernst, B. et al. (2008a). Quality of life diagnosis and therapy as complex intervention for improvement of health in breast cancer patients: delineating the conceptual, methodological, and logistic requirements (modeling). *Langenbeck's Arch.Surg.*, 393, 1-12.
- Klinkhammer-Schalke, M., Koller, M., Ehret, C., Steinger, B., Ernst, B., Wyatt, J. C. et al. (2008b). Implementing a system of quality-of-life diagnosis and therapy for breast cancer patients: results of an exploratory trial as a prerequisite for a subsequent RCT. *Br.J.Cancer*, 99, 415-422.
- Kohlmann, T., Bullinger, M., & Raspe, H. (1997). Die deutsche Form des Nottingham Health Profiles – Skalenstruktur, Reliabilität und Validität. *Zeitschrift für Sozial- und Präventivmedizin*, 42, 175-185.
- Koller, M. (2007). Outcome und Lebensqualität. In Jauch K-W, Mutschler W, & Wichmann MW (Eds.), *Chirurgie Basisweiterbildung* (pp. 583-589). Heidelberg: Springer Medizin Verlag.
- Koller, M., Aaronson, N. K., Blazeby, J., Bottomley, A., Dewolf, L., Fayers, P. et al. (2007). Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *Eur.J.Cancer*, 43, 1810-1820.
- Koller, M., Kussmann, J., Lorenz, W., Jenkins, M., Voss, M., Arens, E. et al. (1996). Symptom reporting in cancer patients: The role of negative affect and experienced social stigma. *Cancer*, 77, 983-995.
- Koller, M. & Lorenz, W. (2002). Quality of life: a deconstruction for clinicians. *J.R.Soc.Med.*, 95, 481-488.

- Kühl, S. & Strodtholz, P. (2002). *Methoden der Organisationsforschung. Ein Handbuch*. Rowohlt.
- Langer, B. & Wetter, T. (2000). Physician's attitudes towards health telematics - an empirical study. In A.Hasman, B. Blobel, J. Dudeck, R. Engelbrecht, G. Gell, & H.-U. Prokosch (Eds.), *Medical Infobahn for Europe. Proceedings of MIE2000 and GMDS2000* (pp. 1106-1111). Amsterdam: IOS Press.
- Langer, W. (2004). *Mehrebenenanalyse, Eine Einführung für Forschung und Praxis*. VS Verlag für Sozialwissenschaften.
- Lee, R. M. (1993). *Doing Research on Sensitive Topics*. London: Sage.
- Lehoux, P., Poland, B., & Daudelin, G. (2006). Focus group research and "the patient's view". *Soc.Sci.Med.*, 63, 2091-2104.
- Lorenz, W., Troidl, H., Solomkin, J. S., Nies, C., Sitter, H., Koller, M. et al. (1999). Second step: Testing - Outcome measurements. *World J.Surg.*, 23, 768-780.
- McDowell, I. & Newell, C. (1987). *Measuring health: A guide to rating scales and questionnaires*. Oxford: Oxford Univ. Press.
- McGee, H. M., O'Boyle, C. A., Hickey, A., O'Malley, K., & Joyce, C. R. B. (1991). Assessing the quality of life of the individual: the SEIQoL with a healthy and a gastroenterology unit population. *Psychol.Med.*, 21, 749-759.
- Merkens, H. (2000). Auswahlverfahren, Sampling, Fallkonstruktion. In U.Flick, E. von Kardorff, & I. Steinke (Eds.), *Qualitative Forschung Ein Handbuch* (pp. 286-299). Hamburg: rororo.
- Moher, D., Jones, A., & Lepage, L. (2001a). Use of the Consort Statement and Quality of Reports of Randomized Trials
A Comparative Before- and -After Evaluation. *JAMA*, 285, 1992-1995.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001b). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 357, 1191-1194.
- Morfeld, M. & Wirtz, M. (2006). Methodische Ansätze in der Versorgungsforschung. Das Beispiel Evaluation der Disease-Management-Programme. *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz*, 49, 120-129.
- Murphy, M., Black, N., Lamping, D., McKee, C., Sanderson, C., Askham, J. et al. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*, 2, 1-88.
- Najman, J. M. & Levine, S. (1981). Evaluating the Impact of Medical Care and Technology on Quality of Life: A Review and Critique. *Social Science & Medicine* 15, 107-115.
Ref Type: Journal (Full)
- Neugebauer, E., Troidl, H., Wood-Dauphinée, S., Bullinger, M., & Eypasch, E. (1991). Meran consensus conference on quality-of-life assessment in surgery. Part I. *Theor.Surg.*, 6, 123-165.

- Osoba, D., Aaronson, N. K., & Till, J. E. (1991). A practical guide for selecting quality-of-life measures in clinical trials and practice. In D. Osoba (Ed.), *Effect of cancer on quality of life* (pp. 90-104). Boca Radon: CRC.
- Patrick, D. L. & Erickson, P. (1992). Health Status and Health Policy, Pharmacoeconomics in Clinical Trials. In (pp. 337-346). New York, Philadelphia: Oxford University Press.
- Pennebaker, J. W. (1982). *The psychology of physical symptoms*. New York: Springer.
- Petersen, C., Schmidt, S., Power, M., Bullinger, M., Bullinger, C. A., & DISABKIDS Group (2005). Development and pilot-testing of a health-related quality of life chronic generic module for children and adolescents with chronic health conditions: A European perspective. *Qual.Life Res.*, 14, 1065-1077.
- Pfaff, H. & Bentz, J. (2003). Qualitative und quantitative Methoden der Datengewinnung. In F.W.Schwartz, B. Badura, R. Busse, R. Leidl, H. Raspe, J. Siegrist, & U. Walter (Eds.), *Das Public Health Buch. Gesundheit und Gesundheitswesen* (2. ed., pp. 419-434). München und Jena: Urban & Fischer.
- Pirente, N., Bouillon, B., Schäfer, B., Raum, M., Helling, H. J., Berger, E. et al. (2002). Systematic development of a scale for determination of health-related quality of life in multiple trauma patients. The Polytrauma Outcome (POLO) Chart. *Unfallchirurg*, 105, 413-422.
- Ravens-Sieberer, U. & Bullinger, M. (1998). Assessing health related quality of life in chronically ill children with the German KINDL: First psychometric and content analytical results. *Qual.Life Res.*, 7, 399-407.
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W. et al. (2005). The KIDSCREEN-52 Quality of life measure for children and adolescents: Development and first results from a European survey. *Expert Review of Pharmacoeconomics & Outcomes Research*, 5, 353-364.
- Reeve, B. B., Hays, R.D., Bjorner, J. B., Cook, K. F., Crane, P. K. et al. (2007). Psychometric evaluation and calibration of health-related quality of life banks. Plans for the Patient-Reported Outcome Measurement Information System (PROMIS). *Med.Care*, 45, 22-31.
- Schnell, R., Hill, P. B., & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. München: R. Oldenbourg Verlag.
- Schumacher, J., Klaiberg, A., & Brähler, E. (2003). *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden*. Göttingen: Hogrefe.
- Schwartz, C. E., Bode, R., Repucci, N., Becker, J., Sprangers, M. A. G., & Fayers, P. M. (2006). The clinical significance of adaptation to changing health: A meta-analysis of response shift. *Qual.Life Res.*, 15, 1533-1550.
- Siegrist, J. (1990). Grundannahmen und gegenwärtige Entwicklungsperspektiven einer gesundheitsbezogenen Lebensqualitätsforschung. In P.Schölmerich & G. Thews (Eds.), *"Lebensqualität" als Bewertungskriterium in der Medizin* (pp. 59-66). Stuttgart: Fischer.
- Siegrist, J., Broer, M., & Junge, A. (1996). *PLC - Profil der Lebensqualität chronisch Kranker*. Göttingen: Beltz Test.

- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton: Chapman & Hall/CRC.
- Smith, S. C., Lamping, D. L., and Banerjee, S. (2005). Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol. Assess.*, 9, 1-93.
- Steinke, I. (2000). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *Qualitative Forschung Ein Handbuch* (pp. 319-331). Hamburg: rororo.
- Stevens, K. J., Brazier, J. E., McKenna, S. P., Doward, L. C., and Cork, M. J. (2005). The development of a preference-based measure of health in children with atopic dermatitis. *Br J Dermatol*, 373-377.
- Sugano DS & McElwee NE (1996). An epidemiologic perspective. In Spilker B (Ed.), *Quality of Life in Clinical Trials*. (pp. 555-561). Philadelphia: Lippincott-Raven Publishers.
- Swart, E., Ihle, P., Geyer, S., Grobe, T., & Hofmann, W. (2005). GPS - Gute Praxis Sekundärdatenanalyse. *Das Gesundheitswesen*, 67, 416-421.
- Tagay, S., Erim, Y., Brähler, E., & Senf, W. (2006). Religiosity and sense of coherence - Protective factors of mental health and well-being? *Z Med Psychol*, 15, 165-171.
- The WHOQOL Group (1993). *The Development of the WHO Quality of Life Assessment Instrument (The WHOQOL)*. Paris: IPSEN Foundation Press.
- Troidl, H., Menge, K. H., Lorenz, W., Vestweber, K. H., Barth, H., & Hamelmann, H. (1979). Quality of life and stomach replacement. In C. Herfarth & P. M. Schlag (Eds.), *Gastric Cancer* (pp. 312-317). Berlin: Springer.
- Troidl, H., Wechsler, A. S., & McKneally, M. F. (1998). How to choose a relevant endpoint. In H. Troidl, M. F. McKneally, D. S. Mulder, A. S. Wechsler, B. McPeck, & W. O. Spitzer (Eds.), *Surgical research. Basic principles and clinical practice* (3 ed., pp. 303-319). New York: Springer.
- v. Steinbüchel, N., Petersen, N., Bullinger, M., & The QOLIBRI Group (2005a). Assessment of health-related quality of life in persons after traumatic brain injury – development of the QOLIBRI, a specific measure. *Acta Neurochir*, 43-49.
- v. Steinbüchel, N., Richter, S., & Morawetz, C. (2005b). Assessment of subjective health and health-related quality of life in persons with acquired or degenerative brain injury. *Current Opinion in Neurology*, 18, 681-691.
- van Campen, C., Sixma, H., Friele, R. D., Kerssens, J. J., & et al (1995). Quality of care and patient satisfaction: A review of measuring instruments. *Med Care Res Rev*, 52, 109-133.
- Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., & Mulrow, C. D. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Medicine*, 4, 1628-1654.
- Vang, J. (1986). The consensus development conference and the European experience. *International Journal of Technology Assessment*, 2, 65-76.

- Vogel, H., Lemisz, W., Liebeck, H., & Palm, W. (2002). Zur Bewertung des Gutachterverfahrens für die ambulante Verhaltenstherapie durch die GutachterInnen. *Verhaltenstherapie*, 12, 228-231.
- Ware, J. E. & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Med.Care*, 30, 473.
- Wegscheider, K. (2000). Statistische Verfahren in der Rehabilitationforschung. In J.Bengel & U. Koch (Eds.), *Grundlagen der Rehabilitationswissenschaften* (Berlin: Springer.
- Willis, G. B. (2005). *Cognitive interviewing. A tool for improving questionnaire designs*. Thousand Oaks, Calif.: Sage Publ.
- Wirtz, M. (2006). Methoden zur Bestimmung der Beurteilerübereinstimmung. In M.Petermann (Ed.), *Handbuch der Psychologie - Psychologische Diagnostik* (pp. 369-382). Göttingen: Hogrefe.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen, Bern, Toronto,Seattle 2002: Hogrefe-Verlag.
- Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G. et al. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*, 336, 601-605.
- Wright, J. R., Ung, Y. C., Julian, J. A., Pritchard, K. I., Whelan, T. J., Smith, C. et al. (2007). Randomized, double-blind, placebo-controlled trial of erythropoietin in non-small-cell lung cancer with disease-related anemia. *J.Clin.Oncol.*, 25, 1027-1032.
- Zelen, M. (1990). Randomized consent designs for clinical trials. An update'. *Statistics in Medicine*, 9, 645-656.

ANHANG

Autorenverzeichnis

Abholz, Heinz-Harald	Universitätsklinikum Düsseldorf
Albert, Ute-Susann	Universitätsklinikum Gießen und Marburg
Augustin, Matthias	Universitätsklinikum Hamburg-Eppendorf
Banik, Norbert	GlaxoSmithKline GmbH & Co. KG, München
Bornemann, Reinhard	Universität Bielefeld
Brüggenjürgen, B.	Charité, Berlin
Büssing, Arndt	Universität Witten/Herdecke
Ernstmann, Nicole	Uniklinik Köln
Farin, Erik	Universitätsklinikum Freiburg
Glaeske, Gerd	Universität Bremen
Gostomzyk, Johannes	Augsburg
Gottwik, Martin G.	Nürnberg
Hasford, J.;	Universitätsklinikum München
Heller, Günther	Wissenschaftliches Institut der AOK, Bonn
Höhmnn, Ulrike	Evangelische Fachhochschule Darmstadt
Hoffmann, Wolfgang	Institut für Community Medicine, Greifswald
Karbach, Ute	Uniklinik Köln
Klinkhammer-Schalke, Monika	Universitätsklinikum Regensburg
Koller, Michael	Universitätsklinikum Regensburg
Kruse, Johannes	Universitätsklinikum Düsseldorf
Lange, Stefan	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
Lorenz, Wilfried	Universitätsklinikum Regensburg
Münch, Karsten	Bremen
Neugebauer, Edmund	Universität Witten/Herdecke

Ommen, Oliver	Uniklinik Köln
Petersen, Corinna	Universitätsklinikum Hamburg-Eppendorf
Pfaff, Holger	Uniklinik Köln
Schäfer, Torsten	Universität Lübeck
Schubert, Ingrid	Uniklinik Köln
Trampisch, Hans-Joachim	Ruhr-Universität Bochum
von Steinbüchel, Nicole	Universität Göttingen
Wieseler, Beate	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln
Windeler, Jürgen	Medizinischer Dienst der Spitzenverbände, Essen
Wirtz, Markus	Pädagogische Hochschule Freiburg