

DNVF-Memorandum III „Methoden für die Versorgungsforschung“, Teil 3: Methoden der Qualitäts- und Patientensicherheitsforschung

Memorandum III, Part 3: Quality of Care and Patient Safety Research Methods



Autoren

Max Geraedts¹, Saskia E. Drösler², Klaus Döbler³, Maria Eberlein-Gonska⁴, Günther Heller⁵, Silke Kuske⁶, Tanja Manser^{7*}, Brigitte Sens⁸, Jürgen Stausberg⁹, Matthias Schrappe^{10 #}

Institute

- 1 Institut für Versorgungsforschung und Klinische Epidemiologie, Fachbereich Medizin, Philipps-Universität Marburg
- 2 Kompetenzzentrum Routinedaten im Gesundheitswesen, Hochschule Niederrhein, Krefeld
- 3 Kompetenzzentrum Qualitätssicherung und Qualitätsmanagement, MDK Baden-Württemberg, Stuttgart
- 4 Zentralbereich Qualitäts- und Medizinisches Risikomanagement, Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden
- 5 Fachbereich Medizin, Philipps-Universität Marburg
- 6 Fliedner Fachhochschule, Düsseldorf
- 7 Institut für Patientensicherheit, Rheinische Friedrich-Wilhelms-Universität Bonn
- 8 Zentrum für Qualität und Management im Gesundheitswesen (ZQ) der Ärztekammer Niedersachsen, Hannover
- 9 Essen
- 10 Institut für Gesundheitsökonomie und Klinische Epidemiologie der Universität zu Köln

Schlüsselwörter

Gesundheitsversorgungsforschung, Versorgungsqualität, Patientensicherheit, Risikoadjustierung, Evaluation komplexer Mehrfachinterventionen

Key words

health services research, quality of health care, patient safety, risk adjustment, evaluation of complex multicomponent interventions

Bibliografie

DOI <https://doi.org/10.1055/s-0043-112431>
Gesundheitswesen 2017; 79: e95–e124
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 0941-3790

Korrespondenzadresse

Prof. Dr. med. Max Geraedts, M.San.
Institut für Versorgungsforschung und Klinische Epidemiologie
Fachbereich Medizin, Philipps-Universität Marburg
Karl-von-Frisch-Straße 4
35043 Marburg
geraedts@uni-marburg.de

ZUSAMMENFASSUNG

Das Deutsche Netzwerk Versorgungsforschung e.V. (DNVF) fördert seit Jahren die methodische Qualität von Versorgungsforschungsstudien auf der Basis von Memoranden und anderen Initiativen. Die Qualitäts- und Patientensicherheitsforschung (QPSF) gilt als Kerngebiet der Gesundheitsversorgungsforschung. Das vorliegende Memorandum erläutert wesentliche etablierte Fragestellungen und Methoden der QPSF. Vor dem Hintergrund der besonderen gesundheitspolitischen Bedeutung des Themas werden Methoden der Messgrößenentwicklung und -prüfung, die Risikoadjustierung, Methoden zur Erhebung von Patientensicherheitsdaten, Instrumente zur Analyse sicherheitsrelevanter Ereignisse und Methoden zur Evaluation der meist multiplen und komplexen QPSF-Interventionen behandelt. Zudem werden vordringliche Forschungsthemen benannt.

ABSTRACT

The German Network for Health Services Research [Deutsches Netzwerk Versorgungsforschung e.V. (DNVF)] fosters the methodological quality of health services research studies by memoranda and other initiatives. Quality of care and patient safety research (QCPSR) form core areas of health services

* Unter Mitarbeit von Hammer A, Huckels-Baumgart S, Mc Dermot F, Monaca C
Am Ende des Beitrags wird die Autorenschaft der einzelnen Kapitel benannt

research. The present memorandum explicates principal QCPSR questions and methods. Based on the issues' particular relevance for health policy, the memorandum exemplifies methods for developing and testing indicators, risk adjustment techniques,

methods for collecting patient safety data, tools to analyse patient safety incidents and methods for evaluating often complex and multicomponent QCPS interventions. Furthermore, we point out urgent research topics.

Einführung

Die Qualitäts- und Patientensicherheitsforschung (QPSF) geht von der zentralen Frage aus, mit welcher Qualität und wie sicher Patienten[‡] behandelt werden. Sie untersucht den Zusammenhang von Rahmenbedingungen, Strukturen und Prozessen mit Ergebnissen für Personen und Personengruppen in der Gesundheitsversorgung. Unterschiede auf staatlicher, regionaler, institutioneller und professioneller Ebene werden in Bezug auf diese Zusammenhänge beschrieben. Für qualitätsrelevante Interventionen entwickelt und evaluiert die QPSF Methoden, Werkzeuge und Konzepte. Sie unterstützt und bewertet sowohl die Einbindung qualitätsrelevanter Aspekte in die Gestaltung des Gesundheitssystems, z. B. bei der öffentlichen Qualitätsberichterstattung (Public Reporting) oder bei leistungsorientierten Vergütungssystemen (Pay-for-Performance, P4P), als auch die Umsetzung qualitätsrelevanter Maßnahmen in einzelnen Einrichtungen oder in Einrichtungsverbänden im Sinne eines Qualitätsmanagements. Im Mittelpunkt stehen Interventionen, die durch Multiplizität sowie hochgradige Komplexität gekennzeichnet sind (so genannte komplexe Mehrfach-Interventionen, complex multicomponent interventions, CMCI [1]). Inhaltlich besteht eine Nähe zur „Improvement Science“ [1, 2], einem übergeordneten Begriff für Konzepte zur Veränderung von Prozessen, Verfahren und Verhalten auf professioneller und organisatorischer Ebene.

Qualität wird allgemein als „Grad, in dem ein Satz inhärenter Merkmale eines Objekts Anforderungen erfüllt“ definiert (DIN EN ISO 9000:2015, Nr. 3.6.2). Sowohl hinsichtlich der Auswahl der relevanten Merkmale als auch hinsichtlich der Spezifikation der Anforderungen sind Aushandlungsprozesse zwischen den Interessengruppen notwendig. In diesem Sinne kommt der Qualitätsforschung innerhalb der (Gesundheits-)Versorgungsforschung, die durch die „doppelte Komplexität“ von Intervention und Kontext gekennzeichnet ist (s. u.), eine Sonderrolle zu, denn diese Aushandlungsprozesse mit den daraus folgenden Zielvorstellungen geben den Kontextbedingungen eine große Bedeutung. Weiterhin zeigt schon die tägliche Erfahrung, dass Interventionen zur Verbesserung der Qualität (z. B. die Etablierung einer Prozessinnovation wie Händedesinfektion) einen unmittelbaren Einfluss auf den Kontext und auch auf die Intervention selbst zur Folge haben, so wie es auch im überarbeiteten „Throughput-Modell“ dargestellt wird (vgl. Schrappe und Pfaff 2016) [3]. Gleiches gilt für die Bestimmung von Qualität (Qualitätsmessung), die selbst als Intervention gelten muss (z. B. die Etablierung von Qualitätsindikatoren) und sofort das Umfeld und auch andere gleichzeitig wirksame Interventionen zur Qualitätsverbesserung beeinflusst.

Qualitäts- und Patientensicherheitsforschung muss deshalb oftmals komplex konzipiert werden. Sowohl für allgemein an der QPSF Interessierte als auch für Nachwuchsforscher/innen im Bereich der

QPSF kann es daher zunächst schwer sein, wissenschaftliche Erkenntnisse dahingehend zu beurteilen, ob die angewandten Methoden dem aktuellen Stand der Forschung entsprechen. Hierbei soll das vorliegende Methodenmemorandum des Deutschen Netzwerks für Versorgungsforschung (DNVF) eine Hilfestellung bieten.

Neben dieser allgemeinen Begründung für ein Memorandum zu den Methoden der QPSF werden nachfolgend spezielle Gründe für dessen Notwendigkeit angeführt:

- 1) Qualität und Sicherheit der Gesundheitsversorgung werden sowohl von Patienten als auch von denjenigen, die sie behandeln oder das Versorgungssystem steuern, oftmals als herausragende Ziele eines Gesundheitssystems genannt. Qualität und Sicherheit sind auch das ultimative Ziel der Gesundheitsversorgungsforschung. Das Verhindern von Fehlversorgung aufgrund nicht qualitätsadäquater Leistungserbringung gilt schon immer als vordringlich („primum non nocere“). Um Qualität und Patientensicherheit erreichen und gewährleisten zu können, muss zunächst geklärt werden, was denn „hohe Qualität“ bzw. „hohe Patientensicherheit“ im Einzelfall, aber auch auf der Systemebene bzw. für die verschiedenen Akteure eigentlich bedeutet. Dies mag im Fall des normativen Begriffs der Patientensicherheit (Abwesenheit von unerwünschten Ereignissen) etwas leichter erscheinen als beim neutralen Begriff der Qualität, aber in beiden Fällen ist eine Zielorientierung unabdingbar. Verhaltensökonomische Erkenntnisse und z. B. die Forschungsergebnisse zu leistungsorientierten Vergütungsmodellen in den USA lassen klar erkennen, dass eine Möglichkeit darin besteht, allgemeingültige Sollwerte für die Gesundheitsversorgung zu definieren. Andere Möglichkeiten gehen von stufenweisen Verbesserungsprozessen oder vom Konzept der relativen Verbesserung im Vergleich zum Ausgangswert aus. Solche Zielvorstellungen oder Sollwerte existieren für das deutsche Gesundheitssystem jedoch nur für einzelne Teilbereiche der Versorgung, z. B. bei der so genannten externen, vergleichenden Qualitätssicherung im Akutsektor. Darüber hinaus hat die Diskussion über konzeptionelle Grundvorstellungen, an denen sich die Entwicklung von Sollvorstellungen orientieren kann, kaum begonnen (vgl. 3-dimensionaler Rahmen bei Schrappe 2014 S. 68ff [4]). Da in Deutschland zwar Gesundheitsziele definiert wurden, aber bislang keine Versorgungsziele benannt sind, die auch Struktur- und Prozessaspekte mit einbeziehen, kann ebenso wenig die Frage beantwortet werden, welches Niveau denn überhaupt ein erreichbares Ziel darstellt und mit welchen Methoden diese Frage zu beantworten wäre.
- 2) Ausgehend von Sollwerten und Zielvorstellungen kann evaluiert werden, ob die gegenwärtige medizinische Versorgung (der Ist-Wert) den Erwartungen bzw. Anforderungen entspricht – möglichst für alle Bevölkerungsgruppen und ggf. unter Berücksichtigung begründeter regionaler Unterschiede.

[‡]Obwohl aus Gründen der Lesbarkeit im Text die männliche Form gewählt wurde, beziehen sich alle Angaben auf Angehörige beider Geschlechter.

Dazu sind differenziert einzusetzende Messverfahren notwendig, die von wissenschaftlichen Erhebungen im kleineren Maßstab über klinisch-epidemiologische Methoden bis zu expliziten Qualitäts- und Patientensicherheitsindikatoren reichen. Die Entwicklung der Messverfahren hat sich in den letzten Jahren zwar methodisch verbessert, sie bedürfen aber der Weiterentwicklung, z. B. hinsichtlich der Einbeziehung all jener Faktoren, die eine Vergleichbarkeit der betrachteten Stichproben oder Populationen garantieren. Hier liegt ein wichtiges Arbeitsfeld der Versorgungsforschung, denn der Einsatz dieser Messverfahren macht eine besonders intensive Theoriebildung in der Planung und der Interpretation der Ergebnisse notwendig – es ist immer das gesamte System im Auge zu behalten (s. [5, 6]). Ebenso sind zum einen Methoden weiter zu entwickeln, die sich je nach Wahl der Messmethode mit der Frage der optimalen Auswahl von Daten für die Beschreibung der Ist-Situation widmen: welche Datenquellen sind angemessen hinsichtlich Datenqualität, Reliabilität und Validität, Stichprobenrepräsentativität und Aussagekraft bei gleichzeitiger Ressourcensparsamkeit? Zum anderen müssen die Problematik sektorenübergreifender Versorgungsprozesse, die Erhebung von Langzeitergebnissen und die Einbindung von Patient Reported Outcomes (PRO) angegangen werden, um nur einige Beispiele zu nennen.

- 3) An die Beschreibung schließt sich die Analyse der Ursachen für Abweichungen zwischen Soll- und Istwerten an. Sind diese Abweichungen und deren Gründe allgemein gültig oder nur bei bestimmten Subpopulationen zu finden? Halten bivariat entdeckte Unterschiede oder Korrelationen auch dann stand, wenn potenzielle Kovariaten oder Störfaktoren (Confounder) in multivariaten Analyseverfahren berücksichtigt werden (Risikoadjustierung)? Welche Methoden sind dazu geeignet, die Relevanz von Analyseergebnissen, die an großen Stichproben erzielt wurden, für den Versorgungsalltag zu beurteilen? Gibt es Hinweise, dass es im Verlauf der Messung zu paradoxen Wirkungen, Ausweichbewegungen und Abwägungsprozessen gekommen ist, die die Ergebnisse verfälschen?
- 4) Schließlich geht es um die Auswahl und Beurteilung von Methoden, anhand derer neue Versorgungsmöglichkeiten erprobt und implementiert werden, die zu einer verbesserten Qualität und Patientensicherheit beitragen sollen. Wie lassen sich hochwertige Beobachtungs- und Interventionsstudien durchführen? Der nach dem klassischen Ansatz der biomedizinischen Forschung definierte Goldstandard der randomisierten, Placebo-kontrollierten, verblindeten Studie ist unter den oft komplexen Alltagsbedingungen der QPSF und bei den hier vorherrschenden komplexen Interventionen kaum möglich. Daher müssen Alternativen mit hoher Aussagekraft beschrieben werden. Wie lassen sich Kausalbeziehungen in der QPSF anderweitig untermauern? Oder aber, wie lassen sich die Ergebnisse experimenteller Studien auf die Versorgungsroutine übertragen? Und was muss bei der Implementierung und Evaluation neuer Modelle im Versorgungsalltag methodisch beachtet werden?

Die Reihe der Fragen, die sich im Rahmen von QPSF-Projekten stellt, ließe sich weit fortsetzen. In diesem Memorandum sollen jedoch primär die nach Auffassung der Verfasser wichtigsten For-

schungsfelder und methodischen Fragen angesprochen werden, wodurch das Memorandum dazu beitragen soll, dass in der QPSF methodische Standards vermehrt Beachtung finden.

Das Memorandum richtet sich v. a. an Nachwuchsforscher/innen, die sich mit den methodischen Standards der QPSF vertraut machen wollen und an Nutzer von QPSF-Ergebnissen, die deren Aussagekraft kritisch reflektieren möchten. Hierzu wird auf die Grundlagenliteratur derjenigen wissenschaftlichen Disziplinen verwiesen, derer sich die QPSF typischerweise bedient. Überall dort, wo bereits umfassende Anleitungen, Memoranden oder methodische Standards existieren, werden diese zitiert und nicht nochmals angeführt. Dies gilt insbesondere für einschlägige Begrifflichkeiten im Kontext Qualität, die bereits seit vielen Jahren durch die Arbeitsgruppe Qualitätsmanagement in der Medizin der GMDS publiziert und fortlaufend überarbeitet werden [7]. Passende Definitionen finden sich auch im österreichischen Gesundheitsqualitätsgesetz [8].

Das von mehreren Autoren erstellte vorliegende Memorandum „Methoden der QPSF behandelt im Einzelnen folgende Punkte:

- Einführung: Aktualität des Forschungsgebietes
- Methoden der Entwicklung und Prüfung von Messgrößen
- Methoden der Risikoadjustierung
- Spezifische Methoden und Instrumente zur Erhebung und Analyse von patientensicherheitsrelevanten Ereignissen mit exemplarischer Darstellung von Studienergebnissen
- Methoden zur Evaluation von Interventionen
- Abschluss: exemplarische Forschungsagenda

Zur Aktualität des Forschungsgebietes

Qualität und Patientensicherheit haben in den letzten Jahren auf allen Ebenen der Gesundheitsversorgung einen noch nie dagewesenen Stellenwert erreicht:

- a. Auf der Ebene der Gesundheitspolitik dominiert der Qualitätsaspekt die aktuelle Gesetzgebung, indem Qualitätsindikatoren erstmals für die Krankenhausvergütung eingesetzt und als planungsrelevante Größen für die Krankenhausplanung verwendet werden, indem Qualitätszu- und -abschläge für Leistungen (P4P) und als „Qualitätsverträge“ bezeichnete Selektivverträge mit Krankenhäusern eingeführt werden. Die im – 2016 in Kraft getretenen – Krankenhausstrukturgesetz (KHSG, SGB V §§ 135a ff.) genannten Attribute „qualitativ hochwertige, patienten- und bedarfsgerechte Versorgung“ unterstreichen die politische Zielsetzung. Somit wird „Qualität“ mehr denn je zum Leitkriterium der Gesundheitsversorgung [4], wobei Patientensicherheit durchgängig als immanente Qualitätsanforderung verstanden wird. Zuvor hatte bereits das Thema Patientensicherheit eine stärkere Akzentuierung durch die Aktualisierung und sektorenübergreifende Ausrichtung der QM-Richtlinien des Gemeinsamen Bundesausschusses (G-BA), auch der zum Qualitätsmanagement im Krankenhaus, erfahren (KQM-RL G-BA, 2005, aktualisiert: QM-RL 2015). Diese enthält die Forderung nach strategisch ausgerichtetem, durchgängigem Risikomanagement als integralem Bestandteil des Qualitätsmanagements, der verbindlichen Einführung von Fehlermeldesystemen (Critical Incident Reporting Systems, CIRS) und einrichtungsübergreifenden Berichts- und Lernsystemen. Mindestmen-

genregelungen, Strukturvorgaben und Elemente der Prozessqualität (u. a. für Organzentren und hochspezialisierte Leistungsbereiche), Zweitmeinungsverfahren sowie der Ausschluss wirtschaftlich begründeter Fallzahlsteigerungen sind weitere Beispiele zur Qualitätsorientierung durch G-BA-Richtlinienkompetenz, ebenso die Berücksichtigung von Qualität und Patientensicherheit in für Patienten verständlichen Qualitätsberichten. Aspekte der Patientensicherheit finden sich zu Behandlungs- und Aufklärungsfehlern (bis hin zur Beweislastumkehr) auch im 2013 verabschiedeten Patientenrechtegesetz. Die kürzlich in Kraft getretene einheitliche Qualitätsmanagement-Richtlinie des G-BA für Ärzte, Zahnärzte, medizinische Versorgungszentren und Krankenhäuser fordert von allen Leistungserbringern ein integriertes Qualitätsmanagement, in dem die genannten Komponenten um konkrete Checklisten (z. B. im OP), Notfall- und Hygienemanagement, Arzneimitteltherapiesicherheit und Sturzprophylaxe zu erweitern sind (QM-RL). Zudem manifestiert sich in der Errichtung der Institute IQWiG und IQTIG der politische Wille, die Patientenversorgung in einem solidarisch finanzierten Gesundheitssystem stärker evidenz- und qualitätsorientiert auszugestalten.

- b. Auf der Ebene der Fachgesellschaften und Verbände hat sich – nicht zuletzt vor dem Hintergrund der enorm veränderten Rahmenbedingungen – eine intensive Auseinandersetzung mit Qualitäts- und Patientensicherheitsfragen entwickelt. Die seit vielen Jahren gesetzlich verpflichtende Qualitätssicherung ausgewählter Leistungsbereiche mit Qualitätsvergleichen bis hin zum so genannten „strukturierten Dialog“ mit „auffälligen“ Einrichtungen ist mit entsprechenden Strukturen fest etabliert. Im Jahr 2015 wurden Daten zu 351 Qualitätsindikatoren aus 25 stationären Leistungsbereichen, insgesamt 3,2 Millionen Datensätze aus 1 538 Krankenhäusern bearbeitet. Die bundesweiten Ergebnisse werden aufbereitet und kommentiert (zunächst durch BQS, dann AQUA, jetzt IQTIG). Die Bestrebungen, auch sektorenübergreifende Verfahren flächendeckend umzusetzen, werden forciert. Auf den entsprechenden Diskussionsforen und Plattformen (z. B. G-BA Qualitätssicherungskonferenz) zeigt sich, in welcher thematischen Breite der medizinischen Versorgungsfelder und in welcher methodischen Tiefe das Thema „Qualitätsorientierung“ präsentiert und diskutiert wird. Flankiert wird dieser Anspruch auch durch Krankenhaus- bzw. Arztpraxis-spezifische Zertifizierungsverfahren (KTQ, QEP, DIN EN 15224) und die zunehmend verbreitete, auf Qualitäts- und Risikomanagement zielende Auslegung der DIN EN ISO 9001:2015 für Gesundheitseinrichtungen. Aktivitäten der Fachgesellschaften, z. B. der GQMG, des Aktionsbündnis Patientensicherheit, wie auch fachgesellschaftsübergreifende Kampagnen wie die „Aktion Saubere Hände“ oder „Gemeinsam klug entscheiden“ sind darauf gerichtet, das Thema Qualität und Patientensicherheit voranzutreiben.
- c. In den Einrichtungen der Gesundheitsversorgung, insbesondere den Krankenhäusern, werden Qualitäts- und Risikomanagement zunehmend konsequenter umgesetzt. Die jüngste Erhebung des Institutes für Patientensicherheit (IfPS) unter 572 Krankenhäusern zeigt, dass die Instrumente des klinischen Risikomanagements zu 68 % (CIRS) bis zu 90 % (Nutzung von Befragungen)

eingesetzt werden. Hinsichtlich einer proaktiv ausgerichteten Risikostrategie besteht jedoch bei mind. der Hälfte der Krankenhäuser noch erheblicher Nachholbedarf [9]. Allerdings werden Qualität und Qualitätsdarlegung nach außen (Transparenz) von zukunftsorientiert aufgestellten Einrichtungen vermehrt als künftig bedeutsamer Wettbewerbsvorteil und damit Treiber für den wirtschaftlichen Erfolg beurteilt [10]. Diese sehen in der Anwendung internationaler Standards für ein zertifiziertes Qualitäts- und Risikomanagement eine Herausforderung darin, Qualität und Sicherheit auf allen Ebenen der Organisation zu gewährleisten (DIN EN ISO 9001:2015, Joint Commission International). Freiwillige Aktivitäten, wie z. B. die Initiative Qualitätsmedizin (IQM) einiger großer Krankenhausträger, setzen darüber hinaus auf die Nutzung von Routinedaten (German Inpatient Quality Indicators) und einrichtungsübergreifender Peer Reviews zur Messung und Verbesserung der medizinischen Ergebnisqualität [11]. Das freiwillige Zertifizierungsverfahren der Deutschen Krebsgesellschaft (DKG), mit dem Leitlinienwissen über die Abbildung von QI, Audits (s. u.) und strukturierte Dialoge in den Versorgungsalltag eingebracht wird, wird in über 1000 organkrebspezifischen Netzwerken in 4 Ländern genutzt. Auch in der ambulanten onkologischen Versorgung existiert ein Indikatorensystem, das der systematischen Qualitätsförderung dient (s. www.winho.de). Für Patienten und Öffentlichkeit besteht seit der verpflichtenden Veröffentlichung bestimmter Qualitätsindikatoren in den seit 2005 zu erstellenden Qualitätsberichten der Krankenhäuser die Möglichkeit, sich über Versorgungsqualität zu informieren. Darauf aufsetzend ist eine Vielzahl von Internetportalen entstanden, in denen die entsprechenden Informationen einrichtungsbezogen abgerufen werden können (z. B. Weisse Liste, TK-Klinikführer, AOK-Navigator, vgl. www.g-ba.de). Die Problematik von Krankenhausinfektionen wird durch das Nationale Referenzzentrum für Surveillance von nosokomialen Infektionen bearbeitet, an der sich Krankenhäuser freiwillig beteiligen (www.nrz-hygiene.de). Ein Nationaler Aktionsplan Arzneimitteltherapiesicherheit soll bundesweit einheitliche Medikationspläne für Patienten realisieren, wobei die Probleme der IT-Vernetzung bei sektoral getrennten Versorgungswelten noch zu überwinden sind [12, 13].

In Anbetracht der Vielzahl an aktuellen Aktivitäten zur Qualitäts- und Patientensicherheitsförderung muss aus Sicht der Versorgungsforschung kritisch angemerkt werden, dass die meisten Aktivitäten nicht hinreichend wissenschaftlich evaluiert wurden. Hierzu gehören u. a. die gesetzlichen, dokumentationsintensiven Qualitätssicherungsmaßnahmen oder auch die Zertifizierung von Organzentren, die hinsichtlich Aufwand und Nutzen bzw. Evidenz zur Wirksamkeit noch nicht ausreichend systematisch untersucht wurden. Der nunmehr politisch eingeschlagene Weg, Qualitätsaspekte zur Grundlage von Vergütung und Versorgungsplanung heranzuziehen, hat in Deutschland bislang keine wissenschaftliche Grundlage. Internationale Studien legen aufgrund kurzfristiger und/oder nicht einheitlich gerichteter Effekte nahe, sehr differenziert an diese Thematik heranzugehen [4].

Implizit streifen einerseits die in Deutschland favorisierten, noch sehr traditionell angelegten Qualitätssicherungsansätze zwar Fragen zur Über-/Unter-/Fehlversorgung und zur Priorisierung/Ratio-

nierung. Andererseits sind aber Empfehlungen zur Fokussierung auf populationsorientierte, auf vernetzte Regionen bezogene und sektorenübergreifende Qualitätsindikatoren sowie zur stärkeren Einbindung von Messgrößen zu Ergebnissen und Erfahrungen, die von Patienten berichtet werden (patient reported outcome measures, PROM, patient reported experience measures, PREM), noch zu wenig berücksichtigt [4].

Insgesamt fehlt der Vielzahl an Initiativen, die unter dem Begriff der „Qualitätsoffensive“ zusammengefasst werden, weiterhin ein umfassendes Rahmenkonzept, das die Zielrichtung und somit auch eine Bewertungsmöglichkeit der Relevanz und des Einflusses der einzelnen Maßnahmen zu beschreiben hilft. International sind solche „conceptual frameworks“ bereits vor Jahren entwickelt worden, so z. B. im 2001 erschienenen Buch „Crossing The Quality Chasm“ des Institute of Medicine (IOM) oder durch die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) [14, 15].

Methoden der Messgrößenentwicklung und -güteprüfung

Einführung

Da es sich bei Qualität um ein latentes Konstrukt handelt, ist Qualität nur durch die Formulierung externer Anforderungen in explizite Messwerte zu transferieren. In der neuen DIN EN ISO 9000:2015 (s. Nr. 3.11.1) wird daher nicht mehr der Begriff der „Qualitätsmessung“, sondern der Begriff der „Bestimmung“ („determination“) verwendet („Tätigkeit zur Ermittlung eines oder mehrerer Merkmale und ihrer Merkmalswerte“). Dieser Umstand weist dem Erkenntnisprozess, dem Erkenntnisinteresse und den theoretischen Modellannahmen eine hohe Bedeutung zu. Man kann z. B. die Krankenhaussterblichkeit direkt messen („3 % Sterblichkeit“), damit ist jedoch noch keine Aussage darüber getroffen, ob dieser Messwert die Qualität der stationären Versorgung wiedergibt (oder über Kontextfaktoren wie die Existenz von Pflegeheimen oder Hospizeinrichtungen in der Umgebung Auskunft gibt).

Zusätzlich ist zu beachten, dass der Messvorgang selbst als komplexe Intervention anzusehen ist: durch die Implementierung von Messinstrumenten (z. B. Indikatoren) wird nicht nur die Intervention, sondern auch der Kontext verändert (z. B. Beobachtung der Händedesinfektion bei entsprechenden Implementierungsstudien, die auf das Verhalten einwirkt). Es ist auch zu bedenken, dass das Ergebnis einer Intervention vom verwendeten Studiendesign determiniert werden kann: so kann z. B. der ausschließliche Einsatz einer randomisierten Kontrollgruppenbildung die Folge haben, dass v. a. Wissen zu isolierten, biomedizinisch dominierten Interventionen zur Verfügung steht, weil komplexe Interventionen u. a. schlechter im randomisierten Design zu untersuchen sind und oftmals einen hohen Finanzbedarf haben.

In den letzten 20 Jahren war es gerade in Deutschland immer wieder notwendig, gegenüber einem rein linearen Indikatorenverständnis (direkte „Qualitätsmessung“ durch Indikatoren) neben der Evaluationsfunktion auch auf die Monitoring-Funktion von Indikatoren zu verweisen („An indicator is a quantitative measure that can be used to monitor and evaluate the quality...“ [16]). Indikatoren, wie sie bspw. bei der externen Qualitätssicherung genutzt werden, die durch eine hohe Sen-

sitivität (unter Verzicht auf hohe Anforderungen an die Spezifität) gekennzeichnet sein sollten, sind daher Gegenstand zahlreicher Missverständnisse, v. a. da sie im klinischen Bereich oft mit diagnostischen Markern (hohe Spezifität bzw. hoher positiver Vorhersagewert PPW) verwechselt werden. Bei der Interpretation von Qualitätsergebnissen ist daher zu beachten, ob ein Indikator vornehmlich zum Monitoring oder zur Messung/Bestimmung der Qualität eingesetzt wurde.

Durch die Komplexität der Messvorgänge, die im Rahmen der Versorgungsforschung nun sehr viel mehr in den Vordergrund gerückt ist (obwohl sie im Alltag immer relevant und in der aktiven Qualitätsarbeit immer gegenwärtig war), ist der Zeitpunkt gekommen, sich dem Thema der Qualitätsmessung oder -bestimmung neu zu nähern und grundsätzlich darzustellen, welche Methoden zur Verfügung stehen. Auch das einfache Vorhersagemodell, wie es dem Indikatorkonzept zugrunde liegt, bedarf einer Erweiterung. Häufig praktiziert, aber in Deutschland in der Qualitätsdiskussion vernachlässigt, sind multivariate Messmethoden, die auf klinisch-epidemiologischen Falldefinitionen beruhen, wie sie im Infektionskontroll-Bereich genutzt werden. Eine besondere Rolle bei der Erweiterung der methodischen Diskussion hat dabei die Patientensicherheitsforschung gespielt, die sich dem Problem der Informations-Asymmetrie zugewandt hat: hier sind sog. „Unknown Unknowns“, also nicht bekannte Ereignisse, bei denen die Tatsache ihres Fehlens nicht bekannt ist („blinde Flecken“), in den Mittelpunkt gerückt sind. Diese sog. generierenden Verfahren (z. B. CIRS) spielen praktisch eine große Rolle, sind theoretisch jedoch noch mangelhaft systematisiert.

In der Versorgungsforschung kann man vor diesem Hintergrund von 4 Gruppen von Messmethoden ausgehen, die jeweils andere Vorgehensweisen und unterschiedliche statistische Implikationen aufweisen:

- Komplexe wissenschaftliche Evaluationen mit schrittweisem Vorgehen (z. B. Theoriebildung) entsprechend dem Vorgehen des Medical Research Council [17];
- Klinisch-epidemiologische Falldefinitionen mit multivariaten Modellen (z. B. Infektionskontrolle);
- Indikatoren mit Monitoring-Funktion;
- Generierende Verfahren zur Identifizierung von „Unknown Unknowns“ (z. B. CIRS).

Unterschiedliche Arten von Qualitätsindikatoren

Quantitative Größen zur Messung von Qualität und Patientensicherheit gesundheitsbezogener Leistungen und Versorgungssituationen werden als Indikatoren bzw. Qualitätsindikatoren bezeichnet. Gemäß der Definition der GMDS können Indikatoren auch als qualitätsbezogene Kennzahlen („Qualitätskennzahlen“) bezeichnet werden [7, 18]. Der Qualitätsindikator ist definitionsgemäß ein Werkzeug („ein quantitatives Maß, welches zum Monitoring und zur Bewertung der Qualität wichtiger Leitungs-, Management-, klinischer und unterstützender Funktionen genutzt werden kann, die sich auf das Behandlungsergebnis beim Patienten auswirken“ [7]). Somit liegt die primäre Zielsetzung darin, Aufmerksamkeit auf potenzielle Problem-bereiche zu lenken, diese zu analysieren und ggf. Verbesserungsmaßnahmen einzuleiten. Darüber hinaus sollen Indikatoren der QPSF Transparenz über bestimmte Aspekte der Versorgungsqualität herstellen (Qualitätsdarlegung) [18, 19]. Der Begriff Indikator soll im Folgenden gleichermaßen in Bezug auf Qualität im Allgemeinen als auch auf Patientensicherheit im Speziellen bezogen werden.

Der Qualitätsindikator soll Qualität messbar machen. Demgemäß folgt die Methodik über die Definition „der Qualität“, also des Anforderungsprofils aus einem Bündel festgelegter (je nach Ebene der Qualitätsbestimmung idealerweise gesellschaftlich bzw. unter den Interessengruppen konsentierter oder auch einrichtungsinterner Qualitätsziele), üblicherweise vorausgesetzter (Patientenerwartungen, Hygiene, Risikobeherrschung) und verpflichtender (Gesetze, Richtlinien) Anforderungen an einen bestimmten Bereich der Gesundheitsversorgung. Aus der Kaskade

Qualität definieren = Anforderungsprofil

Qualität messen = Qualitätsindikator

Qualität bewerten und vergleichen = Referenzwert/-bereich

leitet sich die Methodik der Indikatorenentwicklung ab, wozu die Joint Commission on Accreditation of Healthcare Organizations (JCAHO) ein Formblatt „Indicator Development Form“ entwickelt hat [übersetzt in 18].

Basierend auf den methodischen Grundlagen der JCAHO unterscheidet man ‚aggregate data indicators‘ (stetige oder diskrete Variablen, sog. ‚rate-based indicators‘) und ‚sentinel event indicators‘ (Einzelereignisse).

Qualitätsindikatoren in Form von Zahlenverhältnissen (Proportionen/Raten) messen das Auftreten bestimmter (kritischer oder erwünschter) Ereignisse in einer Gruppe von Fällen bzw. Personen pro Zeiteinheit. Die Anzahl der Personen/Fälle in der Gruppe, die dem zu beurteilenden Ereignis ausgesetzt waren, bildet den Nenner, die Anzahl Personen mit einem Ereignis den Zähler der Rate. Ein kritisches Ereignis für einen ratenbasierten Qualitätsindikator ist bspw. die postoperative Sepsis. Hier wird der Nenner aus der Anzahl der operierten Patienten ermittelt; die Teilmenge der Personen aus dem Nenner mit postoperativem Auftreten einer Sepsis entspricht dem Zähler. Viele Qualitätsindikatoren zur Patientensicherheit, bspw. mit sicherheitsrelevanten Komplikationen oder dem Versterben als kritischem Ereignis werden so gebildet, ebenso qualitativ erwünschte Raten wie der Anteil als „gesund“ entlassener Frühgeborener aus der neonatologischen Intensivversorgung. Die Ausprägung stetiger Variablen (Körpertemperatur, Beatmungsdauer, door-to-needle time) wird in der jeweiligen Einheit angegeben.

Bei seltenen und besonders schwerwiegenden Ereignissen ist die Bildung einer Rate weder methodisch sinnvoll noch inhaltlich vertretbar. Bei solchen Ereignissen (seltene Massivkomplikationen, unerwartete Todesfälle) ist eine unverzügliche Überprüfung des Versorgungsgeschehens erforderlich. Entsprechende Messgrößen werden daher als Sentinel-Event-Indikatoren bezeichnet, da diese immer eine eingehende Analyse des Versorgungsgeschehens nach sich ziehen.

Bei der Prüfung auf Richtigkeit und Evidenzbasierung der Versorgung stehen weniger Ergebnisse als vielmehr Eigenschaften der Prozesse im Vordergrund. Bei den entsprechenden Qualitätsindikatoren wird selten gezählt, sondern vielmehr gemessen, z. B. die Zeitdauer vom Entschluss zum Notfallkaiserschnitt bis zur Geburt des Kindes oder die Menge eines Kontrastmittels bei einer radiologisch überwachten Maßnahme. Das Ergebnis für einen Qualitätsindikator wird dann über Mittelung der Werte für alle Personen in einer definierten Gruppe gebildet. Dies findet sich sowohl bei Indikatoren zu Ergebnissen, die von Patienten berichtet werden (PROM = Patient Reported Outcome Measures) wie der Lebensqualität, da hier Scores eine hohe Bedeutung haben, als auch bei der

Surveillance von nosokomialen Infektionen. Bei PROM können zur besseren Interpretierbarkeit alternativ zur Mittelwertbestimmung Anteile an Patienten einer Einrichtung ermittelt werden, deren Ergebnis oberhalb eines definierten Zielwerts liegt. Erfahrungen mit dem Einsatz von PROM im Bereich der Qualitäts- und Patientensicherheitsbestimmung liegen u. a. aus der Schulterchirurgie vor [20].

Eine Sonderrolle nehmen standardisierte Verhältniszahlen ein, bei denen der Anzahl von eingetretenen Ereignissen – z. B. eines ratenbasierten Indikators – eine Anzahl erwarteter Ereignisse gegenübergestellt wird. Der Quotient aus der Anzahl eingetretener Ereignisse und der Anzahl erwarteter Ereignisse bildet dann ein dimensionsloses Verhältnis, das üblicherweise 1 beträgt, sofern die Anzahlen beobachteter und erwarteter Ereignisse übereinstimmen. Ein Beispiel ist die Hospital Standardized Mortality Ratio (HSMR) bei der der Quotient noch mit 100 multipliziert wird. Im Unterschied zu einem ratenbasierten Indikator sind bei der standardisierten Verhältniszahl die Beobachtungseinheiten im Zähler nicht Bestandteil des Nenners. Die zu erwartende Anzahl kritischer Ereignisse im Nenner einer Beobachtungseinheit (z. B. Krankenhaus oder Fachabteilung), die den beobachteten Ereignissen gegenübergestellt wird, lässt sich an Hand von Informationen – wie z. B. der Altersverteilung der betroffenen und aller beobachteten Fälle – über die gesamte Untersuchungspopulation (aus allen Beobachtungseinheiten) schätzen.

Weiterhin lassen sich mehrere Qualitätsindikatoren zu größeren Messgrößen, den so genannten „Composite Measures“ zusammenfassen (z. B. der so genannte ‚Qualitätsindex der Frühgeborenenversorgung‘ in der Qualitätssicherung Neonatologie). Dies ist insbesondere dann relevant, wenn die Qualitätsbewertung nicht den Ausgangspunkt für eine Schwachstellenanalyse bildet, sondern gesundheitspolitischen Anliegen dient, wie z. B. der Transparenz über die Versorgungsqualität oder der qualitätsorientierten Vergütung [21]. Das methodische Vorgehen zur Generierung von Composites ist sorgfältig abzuwägen [22]. Zur besseren Interpretierbarkeit sollten ggf. Ergebnisse aus Composite Measures als Profil berichtet, d. h. die unterschiedlichen Qualitätsdimensionen auch einzeln angegeben werden [23].

Neben der Unterscheidung über die Art ihrer Berechnung lassen sich Qualitätsindikatoren nach den Ebenen von Qualität nach Donabedian differenzieren. Indikatoren, die sich mit Ressourcen der Gesundheitsversorgung wie der Personalausstattung auseinandersetzen, werden dann der Strukturqualität, Indikatoren, die sich mit der Performance der Leistungserbringung auseinandersetzen, der Prozessqualität zugeordnet. Indikatoren der Ergebnisqualität befassen sich mit den dem medizinischen (und pflegerischen) Handeln zuschreibbaren Veränderungen des Gesundheitszustandes der Patienten einschließlich der von diesen Veränderungen ausgehenden Wirkungen, also auch patientenrelevanten Gesichtspunkten wie Überleben, Lebensqualität oder sozialer Wiedereingliederung.

In aktuellen Diskussionen zu Über-, Unter oder Fehlversorgung z. B. im Rahmen der Initiative „Gemeinsam Klug Entscheiden“ (choosing wisely) ist der Begriff der Indikationsqualität von zunehmender Bedeutung [24]. Entsprechende Indikatoren befassen sich mit der Angemessenheit von Maßnahmen bei gegebener Erkrankungssituation [25].

Entwicklung von Qualitätsindikatoren

Die Entwicklung von Qualitätsindikatoren im Sinne einer Identifikation und Definition möglicher Kandidaten ist nur gering formalisiert.

Ein direkter Zugang eröffnet sich bei Anwendung des PDCA-Zyklus oder der Verfügbarkeit von Leitlinien. Im Qualitätskreis nach Deming (Plan-Do-Check-Act, PDCA) wird mit Qualitätsindikatoren in der Phase „Check“ die Erreichung von Qualitätszielen gemessen, die in der Phase „Plan“ definiert und der Phase „Do“ pilotierend erprobt wurden. Wird in einer Einrichtung als Qualitätsziel die Verbesserung der Patientensicherheit über eine Senkung der Häufigkeit des postoperativen Eintretens einer Sepsis postuliert, kann die Wirksamkeit qualitätsverbessernder Maßnahmen über einen entsprechenden Indikator geprüft werden. Leitlinien beinhalten evidenzbasierte Empfehlungen zu spezifischen Gesundheitsproblemen. Die Empfehlungen umfassen gleichermaßen Anforderungen an Strukturen wie an Prozesse und Ergebnisse (Ziele) [26]. Hierzu gelingt oftmals die Ableitung von Qualitätsindikatoren, die die Umsetzung der Empfehlungen im Sinne eines Erreichungsgrades messen.

Häufig besteht jedoch der Wunsch, Indikatoren für bestimmte Versorgungsbereiche oder Versorgungssituationen unabhängig von Leitlinien oder konkreten, einrichtungsbezogenen Projekten zu definieren. Dies findet sich v. a. im Umfeld gesetzgeberischer Einflussnahme. Die Entwicklung von Indikatoren kann sich dann insbesondere auf die Vorschläge von Experten oder eine systematische Sichtung der Literatur stützen. Als Ausgangspunkt für Indikatoren kommen z. B. Endpunkte infrage, die in klinischen Studien zur Bewertung der Wirksamkeit therapeutischer Verfahren eingesetzt wurden. Insbesondere zur Patientensicherheit können über Spontanberichtssysteme wie dem CIRS Anregungen für mögliche Qualitätsindikatoren aufgenommen werden. Nach Festlegung eines interessierenden Versorgungsbereiches oder einer interessierenden Versorgungssituation können die in der EN 15224:2012 „Dienstleistungen in der Gesundheitsversorgung – Qualitätsmanagementsysteme – Anforderungen nach EN ISO 9001:2008“ beschriebenen 11 Qualitätsmerkmale als Leitschnur zur Identifikation möglicher Qualitätsindikatoren verwendet werden. So spricht auch die EN 15224:2014 davon, dass „Qualitätsindikatoren ... für jedes/jede zutreffende/s Qualitätsmerkmal/-anforderung berücksichtigt werden“. Die erwähnten 11 Qualitätsmerkmale umfassen folgende Aspekte der Gesundheitsversorgung:

- angemessene, richtige Versorgung
- Verfügbarkeit und Erreichbarkeit von Dienstleistungen der Gesundheitsversorgung
- Kontinuität der Versorgung
- Wirksamkeit der Maßnahmen
- Effizienz
- Gleichheit bei der Versorgung
- Evidenz- und Wissensbasierung der Versorgung
- Patientenorientierung der Versorgung
- Mitwirkung und aktive Einbeziehung des Patienten
- Patientensicherheit
- Rechtzeitigkeit und Zugänglichkeit der Gesundheitsversorgung

Diese Qualitätsmerkmale können also systematisch hinsichtlich möglicher Qualitätsindikatoren analysiert werden. Selbstverständlich gehört zu jeder Indikatorenentwicklung die Sichtung des nationalen und internationalen Standes, um Informationen zu anderenorts bereits eingeführten oder aber auch verworfenen Messgrößen zu recherchieren (z. B. in der US-amerikanischen Datenbank des „National Quality Measures Clearinghouse“).

Bei der Indikatorenentwicklung sollte von vornherein beachtet werden, dass für eine erfolgreiche Einführung von Qualitätsindikatoren nicht zuletzt Verantwortlichkeit und Aufwand zur Erfassung der erforderlichen Daten eine herausragende Rolle spielen. Hierbei empfiehlt sich, Verantwortlichkeiten zur Datenerfassung dort anzusiedeln, wo auch Vorteile entstehen, z. B. durch die Unterstützung der Unternehmenssteuerung. Vorteilhaft ist, wenn durch Nutzung vorhandener Daten kein zusätzlicher Erfassungsaufwand entsteht. Hierzu stehen z. B. Daten aus dem Abrechnungskontext, sogenannte administrative Daten oder Routinedaten, zur Verfügung. Deren Nutzbarkeit wie auch deren Qualität / Güte sollte vor Verwendung abgeklärt werden.

Das primäre Ziel der Entwicklung von Qualitätsindikatoren ist jedoch – unabhängig von etwaigen methodischen Unschärfen – deren Nutzung: „Der Qualitätsindikator ist (nur) ein quantitatives Maß, das Aufmerksamkeit auf potenzielle Problembereiche lenken kann, die dann einer konstruktiv-kritischen, intensiven Überprüfung innerhalb einer Organisation bedürfen“ (nach [7]).

Strukturierte Darlegung von Qualitätsindikatoren

Vorschläge für Qualitätsindikatoren sollen strukturiert beschrieben werden. Dies unterstützt bereits bei der Entwicklung die Reflexion über ihre Eignung und bereitet durch eine eindeutige Definition die Evaluation vor. Zu dieser Beschreibung gehören folgende Aspekte:

- Allgemeine Angaben wie eine eindeutige Benennung, eine Begründung für die Wahl des Indikators, ggf. die Nennung eines Qualitätsziels sowie eine Zuordnung zu den Qualitätsebenen nach Donabedian und den Qualitätsmerkmalen der EN 15224:2014.
- Die Berechnungsvorschrift für den Indikator. Bei ratenbasierten Indikatoren umfasst dies die Definition des betrachteten Kollektivs sowie des interessierenden Ereignisses. Zudem sollten Anhaltswerte für die erwarteten Ergebnisse sowie Ziel- oder Referenzwerte angegeben werden.
- Dokumentationsrelevante Aspekte wie Patienten- und Behandlungsmerkmale, die zur Ermittlung des Indikators erforderlich sind.
- Faktoren, die die vom Indikator erfasste Situation beeinflussen. Hierbei kann es sich um patientenseitige Faktoren sowie strukturelle und prozesseitige Faktoren der Leistungserbringung handeln.
- Falls zutreffend, Hinweise auf die Verwendung in einem Berichtswesen, z. B. zu den Zielgruppen, zu der Berichtsfrequenz, zum Berichtszeitraum oder zur Art der Darstellung.
- Literaturangaben zur Evidenz und – falls vorhanden – zu Anwendungserfahrungen des Indikatorvorschlags.

Anwendungsbereiche von Qualitätsindikatoren

Qualitätsindikatoren können für verschiedene Zielsetzungen genutzt werden, bspw. die organisationsinterne Qualitätssteuerung, für Qualitätsvergleiche zwischen Einrichtungen („Leistungserbringervergleich“), für regionale Analysen („area indicators“), für weitere Analysen auf der Systemebene (bspw. Analyse von Änderungen des Einsatzes bestimmter Behandlungsverfahren im Zeitverlauf oder Analysen von Versorgungsketten) oder zur Hypothesenbildung für Fragen der klinischen oder Versorgungsforschung [18, 19, 27].

Innerhalb dieser Zielsetzungen liefern sie die Informationsgrundlage für unterschiedliche Instrumente der Qualitätsdarlegung oder -verbesserung. So können die Indikatoren mit dem Ziel einer einrichtungsbezogenen Qualitätssicherung für ein anonymes oder offenes Benchmarking, für Public Reporting oder eine leistungsorientierte Vergütung (z. B. P4P) genutzt werden [18, 28].

Abhängig von den jeweiligen Zielsetzungen und Instrumenten müssen an die Messgrößen unterschiedliche Anforderungen gestellt werden. Je einschneidender die Konsequenzen des Einsatzes von Qualitätsindikatoren sind, desto höhere Anforderungen müssen an die methodische Güte gestellt werden [19].

Güteprüfung von Qualitätsindikatoren

Die Güteprüfung oder Evaluation möglicher Qualitätsindikatoren befasst sich im Idealfall mit dem Ausmaß, in dem diese die Erreichung vorab definierter Ziele unterstützen. Die Prüfung der methodischen Güte anhand von Gütekriterien [18, 19] geschieht v. a. bei der Entwicklung und der Evaluation von Qualitätsindikatoren [29–31]. Zu beachten ist jedoch, dass die Güte von Qualitätsindikatoren auch kontextabhängig sein kann.

Im Prozess der Entwicklung von Qualitätsindikatoren werden sie genutzt, um für ein bestimmtes Thema eine Auswahl aus potenziell infrage kommenden Indikatoren vorzunehmen, um verschiedene Optionen für eine Operationalisierung gegeneinander abzuwägen oder um Weiterentwicklungsbedarf zu identifizieren.

Die Analyse eines Indikators mithilfe von Gütekriterien sowohl im Rahmen der Entwicklung als auch im Rahmen eines Routineeinsatzes oder einer Evaluation liefert ein Profil seiner methodischen Eigenschaften und macht somit seine Stärken und Schwächen sichtbar. Mithilfe dieses methodischen Profils kann die Eignung eines Indikators für eine bestimmte Zielsetzung oder eine bestimmte Anwendungsform bewertet werden. Das mithilfe von Gütekriterien ermittelte methodische Profil eines Indikators ist somit nicht primär „gut“ oder „schlecht“, sondern für bestimmte Anwendungsformen besser oder schlechter geeignet [29, 32]. Beim Einsatz in einem PDCA-Zyklus kann dies die Eignung zur Verbesserung der Patientensicherheit sein, bei der Verwendung von Indikatoren als Einstieg in ein Peer Review eine hohe Trefferquote von Auffälligkeiten in der Aktendurchsicht, bei der öffentlichen Darlegung von Ergebnissen in Patientenportalen eine Auswahl der Leistungserbringer durch Patienten entsprechend ihrer Präferenzen oder bei der qualitätsorientierten Vergütung eine „gerechte“ (positive oder negative) Sanktionierung. Ganz allgemein lässt sich fordern, dass durch Qualitätsindikatoren aufgedeckte Unterschiede zwischen Leistungserbringern, Regionen, Versorgungskonzepten oder anderen interessierenden Einheiten weitestgehend auf die Beeinflussung durch die jeweiligen Leistungserbringer bzw. auf die interessierenden Faktoren zurückzuführen sein sollten. Andere Faktoren sollten das Messergebnis des Qualitätsindikators nicht verzerren. Ggf. sind derartige Faktoren durch eine entsprechende Korrektur in der Berechnung der Qualitätsindikatoren auszugleichen, die dann als Risikoadjustierung (s. u.) bezeichnet wird. Eine aus Zielen abgeleitete Evaluation von Qualitätsindikatoren stellt (noch) die große Ausnahme dar. In der Regel befassen sich die vorliegenden Evaluationskonzepte mit Gütekriterien, die nur mittelbar mit der Zielerreichung verbunden sind.

Gütekriterien für Qualitätsindikatoren

Zur Evaluation von Qualitätsindikatoren liegen Konzepte und ausgearbeitete Verfahrensempfehlungen vor, die sich allgemein dahingehend unterscheiden lassen, ob sie Zugänge über strukturierte Konsensverfahren mit Einbindung relevanter Expertengruppen oder aber quantitative, Daten-gestützte Verfahren nutzen. Beide Zugänge haben ihre eigenen Stärken, sodass sich eine Kombination empfiehlt. Die zur Güteprüfung verwendeten Kriterien lassen sich zumeist dem in den 70er Jahren des 20. Jahrhunderts geprägten Akronym RUMBA zuordnen {zitiert nach [18]}

- Qualitätsindikatoren sollen wesentliche Versorgungsaspekte erfassen (relevant),
- sie sollten für Leistungserbringer und Patienten verständlich sein (understandable),
- ihre Berechnung sollte mit hoher Reliabilität und Validität möglich sein (measurable),
- sie sollten durch Veränderungen des Verhaltens beeinflussbar (behaviourable) und
- mit vertretbarem Aufwand erreichbar sein (achievable).

Für Deutschland wurden 2001 in einem Konsenspapier erstmals formal Gütekriterien für Qualitätsindikatoren formuliert und den Kategorien Relevanz und Messeigenschaften zugeordnet [18].

Folgende weitere Konzepte liefern systematisch entwickelte oder auf nationaler Ebene verwendete Zusammenstellungen von methodischen Gütekriterien für Messgrößen der Qualitäts- und Patientensicherheitsforschung:

- National Health Service, Großbritannien [27]
 - Kernkriterien: Bedeutung des Versorgungsaspekts, Validität im Sinne der Konstruktvalidität, Praktikabilität, Verständlichkeit, Handlungsrelevanz.
- Agency for Healthcare Research and Quality (AHRQ), USA
Es wird ein Rahmenkonzept vorgestellt, das Gütekriterien von 5 nationalen US-amerikanischen Organisationen zusammenfasst [33].
 - Kernkriterien: Bedeutung des Versorgungsaspekts und Evidenzbasierung, wissenschaftliche Angemessenheit der Messeigenschaften, Praktikabilität, Handlungsrelevanz.
- Konzepte des RAND-Instituts

In Publikationen zu Gütekriterien wird vielfach auf die „RAND/UCLA-Appropriateness Methode“ verwiesen. Dies bedarf einer Erläuterung. Die RAND/UCLA-Appropriateness Methode wurde in den 1980er Jahren entwickelt, um die Angemessenheit der Indikationsstellung medizinischer Leistungen zu bewerten [34]. Die Methode wurde kontinuierlich fortentwickelt und 2001 nochmals explizit beschrieben [35]. Auch in dieser Publikation geht es um die Entwicklung von Kriterien, mit denen die Indikationsstellung medizinischer Prozeduren bewertet werden kann.

Eine Methodik zur Entwicklung von Indikatorensets und zu Gütekriterien zur Auswahl von Indikatoren wird in anderen Dokumenten von RAND dargestellt: Ende der 1990er Jahre hat das RAND-Institut im Auftrag verschiedener Institutionen (u. a. AHRQ) zu 5 übergeordneten Themen Indikatorensets entwickelt [36, 37]. Im Rahmen dieser Entwicklung hat RAND eine Bewertung von Qualitätsindikatoren durch ein Panel vorgenommen, die an das Bewertungsverfahren der Appropriateness-Methode angelehnt ist.

Als Bewertungskriterien werden Validität (validity, definiert mithilfe von 4 Subkriterien) und Machbarkeit (feasibility) verwendet.

- Methodenpapier des AQUA-Instituts
Das AQUA-Institut hat im Rahmen seines Auftrags für die gesetzliche Qualitätssicherung nach §137 SGB V Gütekriterien sowohl für die Indikatorenentwicklung (Auswahl aus einer Liste potenzieller Qualitätsindikatoren) als auch zur Bewertung von Indikatoren im Routinebetrieb bezüglich ihrer Eignung für eine einrichtungsbezogene öffentliche Berichterstattung genutzt [31, 38].
 - Kernkriterien: Relevanz, Klarheit / Verständlichkeit, Praktikabilität, Erhebungsaufwand, Risikoadjustierung, Messeigenschaften und Eignung für die öffentliche Berichterstattung.
- QUALIFY-Instrument
Das QUALIFY-Instrument wurde auf der Grundlage einer systematischen Recherche und Analyse methodischer Gütekriterien entwickelt [29]. Es liefert einen umfassenden Überblick über bewertungsrelevante Gütekriterien und kann aufgrund seines generischen Ansatzes sowohl im Rahmen der Indikatorenentwicklung als auch für eine Evaluation im Einsatz befindlicher Indikatoren genutzt werden. Es umfasst 20 Gütekriterien in den Kategorien Relevanz, Wissenschaftlichkeit und Praktikabilität. Mithilfe von 14 Gütekriterien dieses Instruments wurden 2007 die Indikatoren ausgewählt, die für das damals in Deutschland erstmalig verpflichtende Public Reporting von Krankenhäusern genutzt wurden [31]. Gütekriterien des QUALIFY-Instruments wurden auch für die Entwicklung von Qualitätsindikatoren für die nationalen Versorgungsleitlinien verwendet [32].
- AIRE-Instrument
Das niederländische AIRE-Instrument orientiert sich am Konzept des AGREE-Instruments zur Bewertung von Leitlinien und legt einen besonderen Fokus auf das Vorgehen bei der Entwicklung der Indikatoren. Die Operationalisierung der Gütekriterien wird detailliert beschrieben [39, 40].

Folgende Kategorien von Gütekriterien werden in praktisch allen genannten Publikationen berücksichtigt und können daher als grundlegend angesehen werden:

- Bedeutung des betrachteten Versorgungsaspekts
- Evidenzbasierung
- Messeigenschaften (primär Validität und Reliabilität)
- Praktikabilität (primär Verständlichkeit und Umsetzbarkeit)
- Handlungsrelevanz

Operationalisierung methodischer Gütekriterien

Die genannten „Sets“ methodischer Gütekriterien beschreiben unterschiedlich detailliert die Vorgehensweise für eine Bewertung und nutzen unterschiedliche Modelle zur Operationalisierung. In einigen Publikationen werden nur rudimentär Hinweise zum Vorgehen bei der Bewertung gegeben [33], während für andere Konzepte Operationalisierungen im Detail dargestellt sind [29, 31, 40, 41]. Die vorgeschlagenen Vorgehensweisen unterscheiden sich jedoch erheblich. Daher kann auch für die Operationalisierung derzeit kein Goldstandard identifiziert werden.

Bestimmte Gütekriterien sind für die Bewertung von Prozess- und Ergebnisindikatoren unterschiedlich relevant. So hat die Evi-

denzbasierung für Prozessindikatoren einen herausgehobenen Stellenwert, während die Güte der Risikoadjustierung v. a. bei Ergebnisindikatoren eine wichtige Rolle spielt [19, 29].

Abhängig von der jeweiligen Operationalisierung unterscheiden sich auch die Zeitpunkte, zu denen eine Bewertung stattfinden sollte. Während die Bewertung von Kriterien wie „Bedeutung für das Versorgungssystem“ und „Evidenzbasierung“ bereits früh im Rahmen der Indikatorenentwicklung vorgenommen werden kann, wird eine Bewertung der Messeigenschaften und der Praktikabilität zu diesem Zeitpunkt immer nur vorläufigen Charakter haben. Eine weitergehende Bewertung dieser Kriterien erfordert eine Datenbasis aus einem größer angelegten Probebetrieb oder einem Routinebetrieb.

Aktuell wurde gefordert, für Deutschland ein standardisiertes Set von Gütekriterien zu entwickeln und multiprofessionell zu konsentieren [19]. Solange eine solche Einigung noch nicht vorliegt, sollte für jede Güteprüfung verlangt werden, folgende Aspekte zu dokumentieren:

- transparente Beschreibung der Gütekriterien
- Begründung für die Auswahl der verwendeten Gütekriterien
- Nutzung spezifischer Kriterien, die klar umschriebene Eigenschaften der Messgrößen erfassen. „Generalkriterien“ liefern kein scharfes Profil einer Messgröße.
- nachvollziehbare Beschreibung des Vorgehens bei der Bewertung (Operationalisierung)
- Reproduzierbarkeit der Bewertung
- Validierung des Konzepts

Methodische Bewertung einzelner Messgrößen vs. Bewertung von Indikatorensets

Die Bewertung methodischer Gütekriterien in der Literatur bezieht sich weitgehend auf die Bewertung der Eigenschaften einzelner Messgrößen.

Da jedoch in aller Regel Indikatoren nicht isoliert eingesetzt werden, sondern für ein bestimmtes Thema mehrere Indikatoren („Sets“) zum Einsatz kommen, erscheint es von besonderer Bedeutung, auch das „Konstrukt der Qualität“, das mithilfe dieser Sets abgebildet werden soll, bezüglich seiner methodischen Eigenschaften (insbesondere der Validität) zu beurteilen [7]. Auch bei Indikatorensets kann die Güteprüfung über eine Einschätzung durch Experten oder datengestützt erfolgen. Expertenbasiert lässt sich die Abdeckung der Kategorien nach Donabedian oder der Qualitätsmerkmale der EN 15224:2014 prüfen. Daten-gestützt wäre z. B. der Anteil eingeschlossener Beobachtungseinheiten des interessierenden Kollektivs oder die interne Konsistenz des Indikatorensets zu belegen. Bei ersterem ist möglichst Vollzähligkeit anzustreben, um Fehlanreize zu einem selektiven Einsatz qualitätsrelevanter Maßnahmen zu vermeiden. Weist man der grundsätzlichen Haltung einer Organisation eine wichtige Bedeutung für die Qualität der Leistungserbringung zu, wäre eine hohe interne Konsistenz sinnvoll, um die verschiedenen Aspekte dieser Haltung zu adressieren.

Letztlich besteht ein erheblicher Forschungsbedarf zur Frage, wie Kriterien für eine Bewertung von Indikatorensets identifiziert und operationalisiert werden können und welche methodischen Kriterien bei der Indexbildung auf der Basis einzelner Qualitätsindikatoren beachtet werden müssen [42].

Die Erhebung und Beurteilung von Qualitätsindikatoren ist nicht nur Bestandteil des internen Qualitätsmanagements einer Institution, sondern hat v. a. bei der vergleichenden Bewertung verschiede-

dener Institutionen eine hohe Bedeutung. Damit derartige Qualitätsindikatoren-basierte Vergleiche auch sachgerecht sind, muss das individuelle Risiko eines Behandlungsfalls, ein qualitätsrelevantes Ereignis zu erleiden (und damit im Zähler eines ratenbasierten Indikators erfasst zu werden), berücksichtigt werden. Das dazu erforderliche Verfahren, die Risikoadjustierung, wird inzwischen für die gesetzlich verpflichtende Qualitätssicherung gefordert und im nachfolgenden Abschnitt besprochen. Im Kontext der qualitätsorientierten Vergütung ist davon auszugehen, dass sich v. a. komplexere Risikoadjustierungsmodelle zunehmend etablieren werden.

Methoden der Risikoadjustierung

Im Kontext der Qualitätsmessung und Qualitätssicherung stellt die Risikoadjustierung ein Verfahren dar, das die Qualitätsmessungen unterschiedlicher Einrichtungen vergleichbar machen soll. Auch wenn unterschiedlich erkrankte Patientenpopulationen in den analysierten Einrichtungen vorliegen, soll ein fairer Vergleich der Versorgungsqualität ermöglicht werden [43]. Risikoadjustierungsverfahren werden insbesondere (aber nicht ausschließlich) bei der Analyse von Indikatoren der Ergebnisqualität angewandt. Dagegen kann bei der Erhebung von Struktur- oder Prozessqualität dann auf eine Risikoadjustierung verzichtet werden, wenn die zugrundeliegenden Empfehlungen für ausnahmslos alle Patienten gelten. Es ist zu überdenken ggf. doch eine Risikoadjustierung durchzuführen, wenn populationsabhängige Unterschiede in der Schwierigkeit der Implementierung bestehen. So kann, wie auch in Fußnote 6 thematisiert, gefordert werden, eine Risikoadjustierung nach sozialer Schicht durchzuführen. Umgekehrt kann argumentiert werden, dass Einrichtungen unterschiedlich gut mit höheren Risiken sozialer Randlagen umgehen können bzw. diese besser positiv beeinflussen können. Auch würde durch eine Risikoadjustierung nach sozialer Lage normativ akzeptiert, dass soziale Randlagen schlechtere Behandlungsergebnisse erhalten, wenn dies ggf. die Mehrheit oder ein großer Anteil der Einrichtungen tut. Daher wird in der gesetzlichen stationären Qualitätssicherung in Deutschland keine Risikoadjustierung nach sozialer Schicht durchgeführt.

Im Folgenden wird eine möglichst allgemeinverständliche Einführung zur Bedeutung, der Zielsetzung wie auch der Anwendung und praktischen Umsetzung der Risikoadjustierung gegeben. Weitergehende statistische Erläuterungen und Anmerkungen werden abschnittsweise eingerückt in kleinerem Schrifttyp ergänzt.

Bedeutung der Risikoadjustierung in der gesetzlichen Qualitätssicherung

Das Thema Risikoadjustierung hat im politischen Umfeld der gesetzlichen Qualitätssicherung der letzten Jahre zunehmend an Bedeutung gewonnen:

- Mit dem 2014 verabschiedeten GKV-Finanzstruktur- und Qualitäts-Weiterentwicklungsgesetz (GKV-FQWG) wurde das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTiG) gegründet, das nach dem reformierten § 137a SGB V insbesondere für die Messung und Darstellung der Versorgungsqualität möglichst sektorenübergreifend

abgestimmte risikoadjustierte Qualitätsindikatoren und Instrumente entwickeln soll. Bisher sind laut Qualitätsreport 2015 des IQTiG von den 351 Indikatoren 91 risikoadjustiert.

- Darüber hinaus sollen, gemäß dem 2015 verabschiedeten Gesetz zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz – KHSG) Qualitätsindikatoren auch für die Krankenhausplanung und für eine qualitätsorientierte Vergütung genutzt werden.

Dabei stellen die Funktionalitäten der Qualitätstransparenz, der planungsrelevanten Indikatoren, wie auch der qualitätsorientierten Vergütung höhere Anforderungen an eine faire und damit risikoadjustierte Analyse, weil die Qualitätsanalyse nicht mehr ausschließlich in einem abgeschlossenen und geschützten Rahmen stattfindet, sondern in einem erheblich größeren Ausmaß unmittelbare Konsequenzen nach sich zieht. Vor diesem Hintergrund scheint es verständlich, dass zunehmend justiziable Qualitätsanalysen gefordert werden. Justizierbar bedeutet, dass die Qualitätsindikatoren auch einer antizipierten gerichtlichen Auseinandersetzung Stand halten sollten. Diese Forderung beinhaltet eine tragfähige Risikoadjustierung der Qualitätsindikatoren.

Einordnung der Risikoadjustierung in ein allgemeines Messmodell

Grundsätzlich wird in der Messtheorie oder Messtechnik bei einer Messung davon ausgegangen, dass ein gemessener Wert sich aus einem wahren Wert und einem Messfehler zusammensetzt [44].

- Messwert = wahrer Wert + Messfehler
Dabei können die Messfehler wiederum unterteilt werden in zufällige und systematische Messfehler.
- Messfehler = systematische Fehler + zufälliger Fehler
 - zufällige Messfehler werden mit höheren Fallzahlen geringer und sind nicht gerichtet. Sie weisen demnach bei einer (unendlich) großen Fallzahl einen Mittelwert von 0 auf. Die Größe der zufälligen Messfehler lässt sich mit Verfahren der schließenden oder Inferenzstatistik errechnen bzw. schätzen.
 - systematische Messfehler sind im Prinzip durch feststellbare Ursachen bedingt und einseitig gerichtet. D. h. auch bei Zunahme der Fallzahlen werden systematische Fehler nicht kleiner.

Die Risikoadjustierung ist demnach ein Versuch, systematische Fehler auszugleichen, die durch unterschiedlich erkrankte Populationen in den betrachteten Einrichtungen entstehen. Probleme der Daten,validität' wären ebenfalls systematischen Fehlern zuzuordnen. Es sei allerdings erwähnt, dass auch zufällige Fehler bei der Erhebung von Qualitätsindikatoren eine erhebliche, wenn nicht sogar die größere Rolle bei der Qualitätsmessung spielen [45–49]. Zufällige Fehler werden durch die Risikoadjustierung nicht behoben. Ein großer zufälliger Messfehler kann jedoch auch die Änderungssensitivität eines Qualitätsindikators beeinträchtigen – also die Empfindlichkeit, mit der ein Indikator auf Änderungen der Qualität mit einer Änderung der Ausprägung des Indikators reagiert [23]. Dies äußert sich typischerweise darin, dass die Ergebnisse einer Qualitätsmessung zwischen verschiedenen Messintervallen stark schwanken.

Grundsätzliches Vorgehen bei der Risikoadjustierung

Ganz grundsätzlich wird bei einer Risikoadjustierung versucht, eine heterogene Risikostruktur zwischen Gruppen auszugleichen¹:

- So kann die einschränkende Definition des Nenners eines Qualitätsindikators, wie etwa die Beschränkung auf „reif geborene Einlinge in regelrechter Schädellage“ beim traditionellen Qualitätsindikator zur Sectorate [50], als eine erste Maßnahme zur Risikoadjustierung angesehen werden. Der Nachteil einer Nutzung nur einer einzigen homogenen Gruppe von Patienten ist offensichtlich: Die Analysen beziehen sich nur auf die jeweils definierte Patientengruppe. Über die Behandlungsqualität von Patienten, die nicht in den Nenner dieses Indikators fallen, kann dagegen streng genommen keine Aussage gemacht werden².
- Dementsprechend können mehrere Gruppen von Patienten mit jeweils vergleichbarem Risiko gebildet und deren Ergebnisse für jede Gruppe oder Schicht verglichen werden. In diesem Fall wird von einer geschichteten Analyse oder einer Stratifizierung gesprochen [48]³.
- Ausgehend von mehreren risikogleichen Gruppen können aber auch zusammengesetzte Maße berechnet werden. Hierzu stehen unterschiedliche statistische Verfahren zur Verfügung. Während traditionelle Verfahren typischerweise auf die Zusammenfassung der Ergebnisse weniger Risikogruppen begrenzt sind⁴, werden mittlerweile in der überwiegenden Anzahl der Risikoadjustierungsmodelle Regressionsverfahren angewendet, weil diese eine sehr flexible Modellierung

¹Insgesamt sollte Wert darauf gelegt werden, dass die Risikoadjustierungsvariablen, die zu einer Homogenisierung des Risikos genutzt werden, im Ergebnis auch zu homogen(er)en Risikogruppen führen. Andernfalls kann dies auch in einer schlechteren Vergleichbarkeit der Analyseergebnisse nach einer Risikoadjustierung resultieren, auch wenn sich Maße des Modellfits verbessern [51, 52]. Nicholl (2007) adressiert das Problem, dass auch innerhalb einer durch Risikofaktoren definierten Gruppe das resultierende Risiko ungleich sein kann und nennt dies „constant risk fallacy“. Er schlägt Analysen von Interaktionstermen vor, um ggf. inhomogene Risiken zu identifizieren. Dabei wird allerdings übersehen, dass diese Inkonsistenz durch unterschiedliche Behandlungsqualität der Einrichtungen in einem unvollständigen Prognosemodell bedingt sein könnte, wie auch die Tatsache, dass die Interpretation derartiger Analysen bei einer Vielzahl von potenziellen Risikofaktoren praktisch sehr schnell unmöglich wird. Um die Frage eines „constant risk“ bzw. einer homogenen Risikogruppe empirisch prüfen zu können, müssten vielmehr alternative und bessere Messmethoden des Risikos, quasi eine Art von Goldstandard, vorliegen. Da dies in aller Regel nicht der Fall ist, muss auf indirekte Analyseverfahren bei der Evaluation von möglichen Risikofaktoren oder aber zusätzliche Informationen, wie etwa Expertenmeinungen über Kodiergewohnheiten, und Annahmen bei der Entwicklung von Risikoadjustierungsmodellen zurückgegriffen werden.

²Allerdings könnte ggf. angenommen werden, dass die Ergebnisse für diese Patientengruppe auch Aussagen über die Behandlungsqualität für andere Patientengruppen ermöglichen, wie es traditionell für ein Tracerkonzept üblich ist [53].

³Einen Sonderfall bilden sogenannte additive Scores, wie etwa der CRB-65-Score im Leistungsbereich ambulant erworbene Pneumonie. Hier wird basierend auf den Risikovariablen Desorientierung (C=confusion), Atemfrequenz (R=respiratory rate), niedriger Blutdruck (B=blood pressure) und Alter (>=65) ein Summenscore gebildet, mit dem das Risiko für die einzelnen Summenscores ermittelt und anschließend für die Subgruppen ausgewertet wird [48].

⁴Als ein sehr bekanntes traditionelles Verfahren wäre bspw. der Mantel-Haenszel Schätzer zu nennen [54].

und die Abbildung komplexer Risikokonstellationen ermöglichen⁵.

Vorgehen bei der Entwicklung eines komplexeren Risikoadjustierungsmodells

Zu Beginn sollte klar definiert sein, auf welchen Endpunkt eine Analyse abzielt und welche Grundgesamtheit (zumeist Patienten oder Behandlungsfälle) dabei betrachtet werden soll. Anschließend sollten mögliche Risikoadjustierungsfaktoren (z. B. über systematische Literaturrecherchen) identifiziert werden. Dabei ist zu beachten, dass nicht alle Prognosefaktoren gleichzeitig Risikoadjustierungsfaktoren im Sinne einer Risikoadjustierung sind. Vielmehr sind nur Patienteneigenschaften zu berücksichtigen, die von der behandelnden Einrichtung nicht beeinflussbar sind. Prozessvariablen, wie etwa Art der Operation (z. B. Endoprothese vs. Osteosynthese bei der Therapie einer proximalen Femurfraktur) sind demnach primär ausgeschlossen [55, 56]⁶.

Demnach unterscheidet sich ein Risikoadjustierungsmodell von einem Prognosemodell dadurch, dass es nicht alle relevanten Prognosefaktoren enthält und demnach ein unvollständiges Prognosemodell darstellt.

Wird bspw. als Analyseendpunkt die Sterblichkeit betrachtet, kann typischerweise davon ausgegangen werden, dass die problemlos verfügbaren demografischen Variablen Alter und Geschlecht vergleichsweise stark mit dem Endpunkt assoziiert sind, sodass (zusammen mit diversen Begleiterkrankungen) eine gute bis sehr gute Modellgüte resultiert.

Wird dagegen ein Risikoadjustierungsmodell für einen Qualitätsindikator „perioperative Wundinfektion“ berechnet, ist davon auszugehen, dass der Zusammenhang mit dem Alter eher schwach ausgeprägt ist⁷ und sich zusätzlich eher schwache Assoziationen

⁵Im Bereich der Medizin/Epidemiologie sind logistische Regressionen am weitesten verbreitet. Allerdings wären grundsätzlich auch andere Link-funktionen, wie etwa Probit-Regressionen denkbar, wie sie im Bereich der Ökonometrie gerne genutzt werden. Der Vollständigkeit halber soll erwähnt werden, dass regressionsbasierte Verfahren auch den Effekt einer metrischen Einflussvariable (etwa des Alters) auf ein Zielergebnis direkt schätzen können, sodass ein funktionaler Zusammenhang (wie etwa die Zunahme eines Sterberisikos um 1 % pro Lebensjahr) postuliert wird, ohne dass eine homogene Risikogruppe im engeren Sinne definiert wird.

⁶Andererseits werden gelegentlich unterschiedliche Prozeduren als Proxyvariablen für die Schwere einer Erkrankung verwendet. Eine erweiterte Kolonresektion im Vergleich zu einer einfachen Kolonresektion könnte demnach als ein Proxy für ein fortgeschrittenes Kolonkarzinom angesehen werden [57]. Daneben wird die Zugehörigkeit zu einer sozialen Schicht gerne als Risikofaktor in Qualitätsanalysen genutzt [z. B. www.wales.nhs.uk/sitesplus/922/page/50311]. Es mag Einigkeit darüber bestehen, dass die soziale Schicht an sich durch den Leistungserbringer nicht oder nur in Ausnahmen beeinflusst werden kann. Allerdings bestehen die erklärenden Faktoren der sozialen Schicht aus verschiedenen Komponenten: So kann erwartet werden, dass primär inhärente schichtspezifische Erkrankungsrisiken durch Einrichtungen nicht beeinflusst werden können. Dies gilt allerdings nicht für schichtspezifische Kommunikationsbarrieren, die wiederum das Behandlungsergebnis beeinflussen können. Daher wird für Qualitätsvergleiche von leistungserbringenden Einrichtungen mitunter durchaus gefordert, keine Risikoadjustierung nach der sozialen Lage durchzuführen [55]. Die beschriebenen Zusammenhänge weisen darauf hin, dass hier noch weiterer Forschungsbedarf besteht.

⁷In dem Sinne schwach, dass nur ein geringer Teil des Auftretens von Wundinfektionen erklärbar ist.

mit der ein oder anderen Begleiterkrankung, wie etwa Diabetes, zeigen. Darüber hinaus sind Einflussvariablen, die sinnvolle Hygienemaßnahmen abbilden, sehr stark mit dem Auftreten von Wundinfektionen assoziiert. Diese Maßnahmen stellen nun aber gerade diejenigen Prozessvariablen und sinnvollen Qualitätsmaßnahmen der Einrichtung dar, für die nicht adjustiert werden sollte, weil andernfalls Qualitätsunterschiede nicht mehr sichtbar wären.

In der Konsequenz sind die üblichen Fitmaße von Prognosemodellen eher von untergeordneter Bedeutung⁸ bzw. sollten im Kontext des aktuellen Analyseendpunktes und der verwendeten Grundgesamtheit betrachtet werden. Idealerweise liegen bereits veröffentlichte Risikoadjustierungsmodelle für vergleichbare Endpunkte und Grundgesamtheiten vor, sodass eingeschätzt werden kann, ob alle relevanten und bekannten Risikofaktoren im Risikoadjustierungsmodell präsent sind und der Modellfit den bisherigen Standards genügt⁹.

Sind alle relevanten Risikofaktoren identifiziert und sinnvoll erhoben, wird deren Einfluss im Rahmen statistischer Modellierungen in einem iterativen Verfahren geprüft. Im Rahmen dieser Modellentwicklung werden ggf. einzelne Variablen umkodiert und typischerweise nur diejenigen Risikoadjustierungsvariablen im Modell belassen, die einen ausreichend sicheren Zusammenhang mit dem untersuchten Endpunkt aufweisen¹⁰.

⁸ Dies wird an dieser Stelle betont, da mitunter Auffassungen verbreitet sind, dass Risikoadjustierungsmodelle nur dann genutzt werden sollen, wenn sie einen bestimmten Modellfit, wie z. B. eine Fläche unter der ROC Kurve von 0,75 erreichen. Heller & Schnell [51] erläutern am Beispiel von 2 Einrichtungen mit gleicher Qualität und gleich stark erkranktem Patientengut, dass die Adjustierung nach einem inhomogen gemessenen Risikofaktor zu fälschlich unterschiedlichen risikoadjustierten Ergebnissen führt, wenn dieser Risikofaktor mit dem Outcome assoziiert ist.

⁹ Bspw. führten Heller & Misselwitz (2008) [58] eine vergleichende Studie zur Prognose von Patienten mit akutem Schlaganfall durch. Endpunkt der Analyse war die Sterblichkeit während des Krankenhausaufenthaltes. Datengrundlagen waren stationäre Abrechnungsdaten von AOK-Patienten und Daten der Arbeitsgemeinschaft Schlaganfall in Hessen. Es zeigte sich, dass der neurologische Status und die Vigilanz bei Aufnahme in den Abrechnungsdaten nur sehr unvollständig abgebildet waren, was zu einem merklich schlechteren Modellfit in den Abrechnungsdaten führte. Die Autoren sahen dies als einen Hinweis darauf, dass in den Abrechnungsdaten relevante Risikofaktoren fehlten und daher die Risikoadjustierung auf Basis von Abrechnungsdaten für Patienten mit akutem Schlaganfall für den Endpunkt Krankenhaussterblichkeit mutmaßlich nur eingeschränkt möglich ist.

¹⁰ Eine mögliche und vergleichsweise typische Vorgehensweise wäre:

- Aufspalten der verfügbaren Daten in einen Entwicklungs- und einen Validierungsdatensatz.
- Prüfung, ob metrische Einflussvariablen in der ursprünglichen Metrik in das Modell aufgenommen werden können, bzw. ob eine Umkodierung dieser Variablen sinnvoll ist.
- Schätzen eines vollständigen Modells mit Entwicklungsdatensatz.
- Exklusion von Risikoadjustierungsvariablen ohne einen signifikanten Einfluss.
- Prüfung auf Multikollinearität, falls vorhanden ggf. Umkodieren oder Ausschluss problematischer Risikoadjustierungsvariablen im Entwicklungsdatensatz.
- Ermittlung der Modellgüte im Entwicklungsdatensatz.
- Evaluation des Modells und der Modellgüte mit Evaluationsdatensatz, ggf. mit erneuter Überarbeitung, z. B. Ausschluss von Risikoadjustierungsvariablen, die nur im Entwicklungsdatensatz sicheren (signifikanten) Einfluss ausüben (vgl. [56] oder [59]).

Besonderheiten von Risikoadjustierungsverfahren unter Nutzung von Routinedaten

Wodurch unterscheidet sich das Vorgehen, wenn sogenannte Routinedaten (oder allgemein andere sekundäre Datenquellen) für die Entwicklung von Risikoadjustierungsmodellen genutzt werden können? Einen Überblick über eventuell brauchbare Routinedaten, deren Nutzungsmöglichkeiten und –probleme findet sich z. B. bei Mansky & Nimptsch [60] sowie Swart et al. [61]. Der Unterschied besteht im Wesentlichen darin, dass einerseits geprüft werden sollte, ob sich relevante Risikofaktoren mit den Routinedaten ausreichend abbilden lassen und andererseits mit dem bestehenden Routinedatensatz bereits (mitunter umfangreiche) Angaben zum Krankheitsgeschehen und -verlauf vorliegen. Somit können weitere explorative Analysen durchgeführt werden, um weitere relevante Risikofaktoren zu identifizieren und (falls sich dies als sinnvoll herausstellt) in das Risikoadjustierungsmodell einzubinden. So stellte sich bei einer ersten Analyse mit § 21 Abrechnungsdaten im Rahmen der Weiterentwicklung für den Leistungsbereich Dekubitus heraus, dass eine starke Assoziation zwischen der Diagnose Stuhlinkontinenz und Dekubitus bestand [62]. In vorherigen Analysen [63] war dies nicht als Risikofaktor beschrieben worden. Im Rahmen der Entwicklung eines Risikoadjustierungsmodells wurde daher diskutiert, diesen ICD-Kode in eine Risikoadjustierung mit aufzunehmen.

Neben dem Strang einer Identifikation von relevanten Risikoadjustierungsvariablen über Literaturrecherche und Expertenwissen kann daher in einem Routinedatensatz überprüft werden, inwieweit statistische Assoziationen zwischen potenziellen Risikoadjustierungsvariablen (z. B. ICD-Kodes) und dem betrachteten Endpunkt existieren. Dabei sollte geprüft werden, ob sich

- der Kode aufgrund inhaltlicher Überlegungen als Risikoadjustierungsvariable eignet,
- die Stärke und Richtung des Zusammenhangs mit dem Endpunkt inhaltlich plausibel ist,
- der hinter der Kodierung stehende Risikofaktor mit Blick auf die Kodiergewohnheiten und -richtlinien, aber auch aufgrund der beobachteten empirischen Ergebnisse, plausibel und valide abbildet [64, 65].

Fazit – Risikoadjustierung

Zusammengefasst stellt die Risikoadjustierung einen Versuch dar, im Setting von Beobachtungsstudien systematische Verzerrungen auszugleichen, die durch unterschiedlich erkrankte oder vorbelastete Patientenpopulationen in den betrachteten Einrichtungen entstehen.

Aktuell werden als statistische Verfahren ganz überwiegend Regressionsverfahren genutzt, weil diese eine flexible Modellierung und umfassende Risikoadjustierung erlauben [48, 56, 59]. Neben konventionellen Regressionsverfahren kommen international mittlerweile zunehmend Mehrebenenmodelle zum Einsatz. Diese berücksichtigen die Tatsache, dass die Patienten innerhalb einer Einrichtung ähnlicher sein können als zwischen unterschiedlichen Einrichtungen, bzw. den Umstand, dass ggf. nicht alle Einflüsse in einem Risikoadjustierungsmodell berücksichtigt werden können. So können bspw. „zufällige“ Konstellationen im Rahmen von sogenannten random effects Modellen in bayesianischen Ansätzen in die Modellbildung aufgenommen werden, um so auch zufälligen Fehlern zu begegnen [66, 67].

Die aktuell in der externen Qualitätssicherung in Deutschland verwendeten Risikoadjustierungsmodelle nutzen aber nur vorbestehende patientenbezogene Risikofaktoren als Risikoadjustierungsvariablen, sodass diese Risikoadjustierungsmodelle üblicherweise unvollständige Prognosemodelle darstellen [65]. Mindestanforderungen an die statistischen Modelle im Sinne von Gütekriterien des Modellfits sind daher von untergeordneter Bedeutung. Vielmehr sollte aus inhaltlicher Sicht geklärt werden, ob alle relevanten Risikofaktoren im Modell enthalten sind und sinnvoll abgebildet werden.

Das zentrale Problem bei der Identifikation homogener Risikogruppen im Rahmen der Entwicklung von Risikoadjustierungsmodellen stellt die Identifikation aller für eine Adjustierung relevanten Risikofaktoren wie auch die Sicherstellung deren valider Abbildung dar. Dies gilt insbesondere auch für eine Nutzung von Routinedaten, wo zusätzlich explorative Analysen zur Identifikation von möglichen sinnvollen Risikoadjustierungsvariablen durchgeführt werden können [64, 65].

Jeder Endpunkt benötigt eine eigene Risikoadjustierung, die auf die aktuelle Datenbasis angepasst werden muss. Eine allgemeine Nutzung von publizierten Scores (wie z. B. Charlson et al. [68], Elixhauser et al. [69]) oder auch nur eine Beschränkung der darin genutzten Risikoadjustierungsvariablen ersetzt nicht die Entwicklung eines Risikoadjustierungsmodells für einen gegebenen qualitätsrelevanten Endpunkt [70].

Der nachfolgende Abschnitt behandelt das Thema Patientensicherheit, in dessen Kontext die zuvor besprochenen Inhalte wie bspw. Methodik der Qualitätsindikatoren oder Risikoadjustierung uneingeschränkt Anwendung finden.

Spezifische Methoden und Instrumente der Patientensicherheitsforschung

Die Organisationen des Gesundheitswesens sind durch komplexe Aufgaben und damit durch eine Auseinandersetzung mit Risiken für Patienten geprägt. Die Patientensicherheitsforschung hat zum Ziel, wissenschaftliche Grundlagen für die Patientensicherheit zu schaffen und diese zu verbessern. Hierzu zählt die Anwendung geeigneter Methoden und Instrumente zur Identifikation von systembasierten Einflussfaktoren auf Fehler sowie die Entwicklung und Implementierung von Interventionen zur erfolgreichen Prävention und Bewältigung kritischer Ereignisse.

Die Patientensicherheitsforschung stellt zunehmend einen eigenständigen Schwerpunkt innerhalb der Versorgungsforschung dar. Verschiedene praxisorientierte und wissenschaftlich untersuchte Methoden und Instrumente stehen zur Identifizierung und Analyse von Qualitätsmängeln, Risiken, Zwischenfällen und/oder unerwünschten Ereignissen (UE) im Rahmen der Gesundheitsversorgung zur Verfügung [71]. Neben den etablierten Methoden und Instrumenten der Versorgungsforschung, werden Methoden und Instrumente der ‚Human Factors‘ Wissenschaften, die sich mit der Optimierung der Interaktion von Mensch und Technik beschäftigen, immer wichtiger.

Im Folgenden werden exemplarisch 6 spezifische Methoden und 3 Instrumente der Patientensicherheitsforschung skizziert.

Zu den zentralen Datenerhebungsmethoden zählen bspw.:

- (i) Befragungen zur Patientensicherheitskultur, die grundlegende Informationen über die wert- und einstellungsbezogenen sowie motivationalen Voraussetzungen für eine Gewährleistung und Optimierung der Patientensicherheit liefern;
- (ii) Beobachtungsstudien zum Verhalten des klinisch tätigen Personals in der Versorgungspraxis, um sicherheitsförderliche Verhaltensweisen zur Prävention und Bewältigung unerwünschter Ereignisse zu identifizieren und durch Prozessoptimierung und Trainingsmaßnahmen systematisch zu fördern;
- (iii) Die Fehlermöglichkeits- und Einflussanalyse (englisch: failure mode and effects analysis (FMEA)), um Risiken und Fehler zu identifizieren, bevor sie sich in einem konkreten Ereignis manifestieren;
- (iv) Die Fehler(ketten)analyse, um retrospektiv Systemursachen kritischer Behandlungsereignisse zu analysieren;
- (v) Audits, um den Grad der Erfüllung zuvor definierter Auditkriterien im Rahmen der internen Qualitätssicherung zu bestimmen;
- (vi) Peer Reviews zur interkollegialen Beratung von Ärzten, die speziell für medizinische Behandlungsprozesse entwickelt worden sind.

Methoden der Patientensicherheitsforschung

- (i) Befragungen zur Patientensicherheitskultur

Sicherheitskultur ist ein relativ stabiles, multidimensionales Konstrukt, das auf geteilten Werten und Normen basiert. Diese etablierten Werte und Normen wirken sich auf die Einstellungen und Wahrnehmungen und letztlich auf das Verhalten der Mitarbeiter aus. Entsprechend stellt Sicherheitskultur ein übergeordnetes Konzept hinter dem Sicherheitsklima dar, welches die Summe dieser Einstellungen und Wahrnehmungen von Mitarbeitern bezüglich sicherheitsrelevanter Aspekte in ihrer Arbeitsumgebung zu einem bestimmten Zeitpunkt umfasst [72–76]. Diese Wahrnehmungen liefern den Bezugsrahmen, an dem sich die Mitarbeiter bei der Erfüllung ihrer Aufgaben und im Umgang mit Sicherheitsfragen orientieren [77].

Bereits im Jahr 2000 forderte das Institute of Medicine (IOM) die Etablierung einer Sicherheitskultur zur Verbesserung der Patientensicherheit [14]. Seither wurde eine Vielzahl an Instrumenten zur Messung von Sicherheitskultur insbesondere für die stationäre Versorgung entwickelt. Zu den bekanntesten und international am weitesten verbreiteten Instrumenten gehören der Hospital Survey on Patient Safety Culture (HSOPS) [78] und der Safety Attitudes Questionnaire (SAQ) [79]. In Deutschland ist der HSOPS zwischenzeitlich für eine Befragungen von Führungskräften [80] sowie für Befragung von Mitarbeitern im Krankenhaus (z. B. WorkSafeMed-Studie) adaptiert worden. Der SAQ wurde von Hoffmann et al. (2009) [81] für die Nutzung im ambulanten Sektor angepasst.

Zwischenzeitlich gibt es eine Vielzahl von Studien, die den Zusammenhang zwischen Sicherheitskultur und Patientensicherheit untersucht haben [82–84]. Aufgrund unterschiedlicher Studiendesigns und verschiedener Erhebungsmethoden sind die Ergebnisse nach wie vor diskrepant [85–87]. Dennoch konnte in einigen Studien ein Zusammenhang zwischen Sicherheitskultur und Patientensicherheit nachgewiesen werden [88].

- (ii) Beobachtungen

Beobachtungen werden in der Patientensicherheitsforschung genutzt, um Beteiligte der Krankenversorgung wie Pflegekräfte, Ärzte

oder Patienten in ihrem Wirkungsbereich zu beforschen [89]. Das Spektrum reicht neben der Untersuchung des Umgangs mit administrativen Prozessen, über Teamsituationen bis hin zur Beobachtung von speziellen Anwendungen, wie bspw. im Bereich der Handhygiene. Sie finden vielfach in risikobehafteten Bereichen Anwendung, wie im Operationssaal, auf Intensivstationen oder während Übergaben. Ziel ist es, Verhaltens- und Kommunikationsmuster aufzudecken, die mit effektiver und sicherer Versorgung von Patienten im Zusammenhang stehen sowie Rahmenbedingungen zu ermitteln, die solches Verhalten begünstigen.

Grundsätzlich unterscheidet man zwischen teilnehmenden Beobachtungsstudien, bei denen der Forscher als Beobachter in der beobachteten Situation als Teilnehmer involviert ist und nicht-teilnehmenden Beobachtungen, bei denen weder eine aktive noch eine passive Intervention erfolgt. Im Gesundheitswesen werden meist offene Beobachtungen (im Gegensatz zur verdeckten Beobachtung) durchgeführt, bei der den beobachteten Personen die Präsenz des Forschers und das Studienziel bekannt sind. Zur Sicherung der internen Validität sind potenzielle Verzerrungseffekte wie die Überbewertung der Handlungen durch die Reihenfolge, der Halo-Effekt und damit die Überstrahlung einzelner positiver oder negativer Merkmale sowie die Tendenz zur Mitte, Kontrastfehler und Projektionsfehler angemessen zu berücksichtigen [90]. Je nach Beobachtungsform sind sowohl Ablenkungen als auch die stetige Fokussierung auf das zu untersuchende Phänomen zusätzliche Herausforderungen. Bei direkten Beobachtungsverfahren im Rahmen der Patientensicherheitsforschung ist verstärkt auf mögliche Verzerrungen und ethische Aspekte zu achten, da der Forscher als Beobachter Teil der Beobachtung ist und ggf. Zeuge risikobehafteter Vorgänge werden könnte [91, 92]. Zudem stellt die Beobachtung selber eine Kontextänderung und Intervention dar, deren Effekte es zu berücksichtigen gilt. Für die anschließende Analyse wird ein Beobachtungsplan bzw. -protokoll erstellt, der bzw. das je nach Art und Umfang der Studie standardisiert und strukturiert ist [93].

Im Rahmen der Patientensicherheitsforschung wurden für unterschiedliche Versorgungssituationen Beobachtungssysteme entwickelt, deren Ziel entweder die systematische Beschreibung von Handlungsabläufen und Interaktionsprozessen oder die Bewertung von Handlungsweisen und den ihnen zugrundeliegenden Fähigkeiten ist (vgl. Behavioural Marker Systeme zur Bewertung von nicht-technischen Fertigkeiten [90]). Bspw. wurden in einer prospektiven Studie Übergaben anhand vorab definierter Verhaltensweisen beobachtet und zu Qualitätsbewertungen seitens des klinischen Personals in Bezug gesetzt. Die Analyse der insgesamt 117 postoperativen Übergaben ergab, dass die Kommunikation der klinischen Einschätzung und zu erwartender Probleme ausschlaggebend sind für eine positive Bewertung der Übergabequalität. Diese Inhalte sind jedoch in den wenigsten standardisierten Übergabeprotokollen bislang enthalten und haben zudem Implikationen für Trainingskonzepte.

Direkte Beobachtungsverfahren werden auch eingesetzt, um UE oder Fehler durch trainierte Beobachter zu identifizieren [94]. Die Art der Fehler wird vor der Beobachtung exakt definiert. Bspw. betreffen Medikationsfehler die Verordnung, die Verabreichung oder die Medikationskontrolle [92, 94]. Diese Verfahren sind im Rahmen der Patientensicherheitsanalyse weit verbreitet und werden von Flynn et al. (2002) als Goldstandard im Vergleich zu wei-

teren Methoden der Datenerfassung z. B. Krankenaktenanalysen betrachtet [92, 95]. Barker et al. (2002) haben bspw. eine prospektive Kohortenstudie in 36 von der Joint Commission akkreditierten Krankenhäusern und Pflegeheimen zur Identifizierung von Medikationsfehlern durchgeführt. In dieser Beobachtungsstudie wurde ein 2-stündiges direktes Beobachtungsverfahren, einschließlich der Dokumentation der zu verabreichenden Medikamente durchgeführt. Die Dosierung der Medikamente wurde im Anschluss mit den Verordnungen des Arztes verglichen und Unterschiede beschrieben und kategorisiert. Zudem wurde danach geschaut, ob weitere Medikamente hätten verabreicht werden müssen. Im Anschluss an die Fehleridentifizierung wurden alle Daten an einen verblindeten unabhängigen Pharmakologen weitergeleitet, um Inkonsistenzen der Daten zu identifizieren [92]. Die Studie konnte zeigen, dass Medikationsfehler häufig waren: bei 19 % der Verabreichungen (605/3216) wurden Fehler entdeckt. Die häufigsten Fehler pro Kategorie waren die Verabreichung der Medikamente zum falschen Zeitpunkt (43 %), Weglassen (30 %), falsche Dosierung (17 %), und die Gabe von unautorisierten Medikamenten (4 %). 7 % der Fehler wurden als unerwünschte Arzneimittelereignisse eingestuft [92].

(iii) Fehlermöglichkeits- und Einflussanalyse

Die Fehlermöglichkeits- und Einflussanalyse (FMEA) ist eine Methode zur Identifikation von Risiken und Fehlern bevor sie sich in einem konkreten Ereignis manifestieren. Diese Methodik zur Risikobeurteilung hat ihren Hintergrund in der Raumfahrt sowie der Automobilindustrie in den 1960er Jahren [96]. Seither wurde das Konzept in viele Industrien, einschließlich des Gesundheitswesens, übertragen. Die FMEA ist ein Konzept, das prospektiv potenzielle Schwachstellen identifiziert, um so Fehlern und deren Auswirkungen vorzubeugen. Häufig findet die FMEA Anwendung bei komplexen Prozessen, indem der Gesamtprozess in seine Teilprozesse aufgegliedert wird, um diese vertieft auf potenzielle Risiken zu untersuchen und die Folgen für den Gesamtprozess zu analysieren und zu bewerten [97]. Weitestgehend folgt die FMEA einem schrittweisen Ablauf, bei dem potenzielle Fehler identifiziert werden, mögliche Auswirkungen auf Teilprozesse und den Gesamtprozess registriert werden, diese nach ihrer Schadenhöhe sowie Eintrittswahrscheinlichkeit bewertet werden, mögliche Ursachen ermittelt und Maßnahmen zur Vermeidung der identifizierten Fehler oder Risiken abgeleitet werden.

Im Unterschied zur Fehlerkettenanalyse liegt der Fokus der FMEA also vielmehr auf einer prospektiven Einschätzung der Auswirkungen der Risiken als auf den Ursachen. Sie nutzt insbesondere die Fachexpertise der am Prozess beteiligten Mitarbeiter und das Wissen der Organisation. Die identifizierten Risiken werden an Hand ihrer Bedeutung, ihrer Eintretenswahrscheinlichkeit sowie ihrer Wahrscheinlichkeit entdeckt zu werden, quantifiziert, Verantwortlichkeiten bestimmt sowie Maßnahmen definiert, implementiert und überwacht [97, 98].

Im Gesundheitswesen fand die Methodik der FMEA z. B. in einer Analyse der Informationsweitergabe in der Versorgung chirurgischer Patienten Anwendung, die das Potenzial der FMEA zur Aufdeckungen und proaktiven Bewältigung von Schwachstellen im System veranschaulicht [99]. Im Rahmen der multidisziplinären FMEA wurde der chirurgische Versorgungsprozess in die Teilprozesse präoperati-

ves Assessment, präoperatives Teamwork, postoperative Übergabe und stationäre Versorgung unterteilt. Der Großteil der Fehlermöglichkeiten wurde in der ersten Phase identifiziert. 41 von 132 Fehlermöglichkeiten wurden als kritisch eingestuft, wobei 26 bereits durch bestehende Protokolle abgedeckt wurden. Für die 15 weiteren Fehlermöglichkeiten wurden Empfehlungen erarbeitet.

(iv) Fehlerkettenanalyse

Zur Identifikation der Systemursachen von unerwünschten Ereignissen und Fehlern in der Patientenversorgung sowie der Planung und Optimierung von sicherheitskritischen Systemen, wird in der Regel das Konzept der Fehlerketten angewendet, das verdeutlicht, dass Fehler häufig das Ergebnis einer Verkettung mehrerer Rahmenbedingungen und Ereignisse sind [100].

Ziel der Fehlerkettenanalyse ist es, ein tiefergehendes Verständnis über Sicherheitslücken und -barrieren im gesamten Versorgungsprozess zu erhalten. Durch eine prozessorientierte Fehlerketten-Analyse können Sicherheitslücken und fehlende bzw. nicht wirksame Sicherheitsbarrieren in der Versorgungspraxis systematisch identifiziert werden. Damit wird über die bisher verbreitete isolierte Betrachtung einzelner Prozessschritte hinausgegangen, da die isolierte Betrachtung die Komplexität des Fehlergeschehens im gesamten Versorgungsprozess nicht angemessen widerspiegelt und die Interdependenzen zwischen den einzelnen Phasen bzw. Schnittstellen vernachlässigt.

Die Studie "Identifying Medication Error Chains From Critical Incident Reports: A New Analytic Approach" von Huckels-Baumgart & Manser (2014) [101] hat erstmalig einen prozessorientierten Ansatz zur Ereignisanalyse von Medikationsfehlern mit Schwerpunkt auf Fehlerketten angewendet. In der Studie wurden 1591 Medikationsfehler, die von 2009 bis 2012 in einem krankenhausweiten Fehlermeldesystem (CIRS) berichtet wurden, zunächst anhand des Medikationsfehler-Index NCC MERP (NCC MERP 2001) sowie der WHO Klassifikation für Patientensicherheit [102] kategorisiert. Zur Identifikation von Fehlerketten wurden die Medikationsfehler den jeweils involvierten Phasen des Medikationsprozesses zugeordnet (Verordnung, Übertragung, Richten/Zubereitung und Verabreichung). Es zeigte sich, dass trotz teils vorhandener Sicherheitsbarrieren nur wenige Fehler innerhalb eines Prozessschritts und damit vor Erreichen des Patienten abgefangen wurden. Der Großteil der berichteten Medikationsfehler (74%) war Teil einer Fehlerkette über 2 oder mehr Phasen. Dies verweist auf die notwendige Verbesserung von Sicherheitsbarrieren in allen Phasen des Medikationsprozesses zur gezielten Unterbrechung von Fehlerketten.

(v) Audits

Audits sind systematische, unabhängige, dokumentierte Prozesse, die dazu dienen, den Grad der Erfüllung zuvor definierter Auditkriterien zu bestimmen. Es wird zwischen internen Audits, die durch die Organisation selbst durchgeführt werden und externen Audits unterschieden, die durch Interessengruppen der Organisation oder unabhängige externe Organisationen im Rahmen einer Zertifizierung (z. B. ISO, KTQ®, ProCumCert, QEP®, EPA oder JCI) durchgeführt werden [7]. Bspw. beinhaltet die KTQ® Zertifizierung Kriterien der Sicherheit und des Risikomanagements, wie Risiken in den Bereichen Arbeitsschutz, Notfallmanagement, Patientensicherheit, Hygiene, Anwendung von Arzneimitteln sowie Blut und Blutpro-

dukten [103]. Die DIN EN 15224:2012 (s. o.) enthält Patientensicherheit explizit als eines der Auditkriterien für Gesundheitseinrichtungen [104]. Die Merkmale der Norm wurden teilweise bereits um 2000 als formulierte Qualitätsmerkmale anerkannt [105].

Ebenso stellt die aktuelle Fassung der DIN EN ISO 9001:2015 ein integriertes Managementsystem (Risikomanagement als Teil des Qualitätsmanagements) in den Fokus. Als neuer Begriff gegenüber den Vorgänger-Versionen wird das „risikobasierte Denken“ eingeführt, welches auf allen Ebenen und in allen Bereichen der Organisation in Strukturen und Leistungsprozessen zu realisieren ist. Damit ist dieser für die Gesundheitsversorgung seit jeher bedeutsame qualitätsimmanente Aspekt besonders hervorgehoben (proaktiv) und findet seine Entsprechung in Bezug auf das klinische Risikomanagement im Kapitel 8.7 der Norm „Steuerung nichtkonformer Ergebnisse“ (reaktiv).

(vi) Peer Reviews

Von den zuvor genannten Audits sind „Clinical“ oder „Medical“ Audits zu unterscheiden. Hinter den Begriffen verbirgt sich kein Auditverfahren, sondern ein sogenanntes Peer Review Konzept: eine kontinuierliche, systematische und kritische Reflexion durch mehrere Leistungserbringer über die eigene Leistungsfähigkeit und die der Kollegen – unter Verwendung eines strukturierten Prozesses und mit dem Ziel einer kontinuierlichen Verbesserung der Qualität der Patientenversorgung [7]. Dieses Instrument ist speziell für medizinische Behandlungsprozesse entwickelt worden und dient der interkollegialen Beratung von Ärzten und weiteren Gesundheitsberufen („peers“ = „Gleichrangige“) [7]. Dennoch wird der Begriff Clinical Audit in der angelsächsischen Literatur nicht einheitlich verwendet. Ebenso wird von umfassenden Clinical Audit Projekten gesprochen, die ein gesamtes Team der Gesundheitsversorgung umfassen können mit dem Ziel, definierte Qualitätskriterien zu überprüfen unter Verwendung von Methoden der Organisationsentwicklung, statistischen Methoden und Methoden des Informationsmanagements [106].

Über die Methoden eines Clinical Audits, wie es im obigen Konzept beschrieben wird, herrscht kein einheitliches Bild und wenig Evidenz [107, 108]. Shaw et al. (1989) stellten 7 Prinzipien vor und formulieren damit eine „guideline“ für diese Methode: (1) Definition von Verantwortlichkeiten, (2) Selbstorganisation der Leistungserbringer, (3) Priorisierung, (4) Güte des Audits mit dem Potenzial zur Qualitätsverbesserung, (5) Bereitstellung von Ressourcen, (6) Dokumentation und (7) Evaluation (Struktur, Prozess, Outcome) [109]. Als eine Methode des Clinical Audit Konzeptes wird Peer Review als „kritische Evaluation eines bestimmten Aspektes der Leistungserbringung von professionellen Kollegen“ verstanden [108]. Diese Evaluation kann ein kriterienbasiertes Audit-Projekt umfassen oder auch die Auswertung von Berichten zu bedeutenden Ereignissen, deren Anwendbarkeit von der Schulung der Leistungserbringer abhängen könnte [108]. Die Studie von MacKay et al. (2005) zeigte, dass Allgemeinmediziner Schwierigkeiten hatten, die Methoden anzuwenden, womit Auswirkungen auf die Wirksamkeit der Methode verbunden sein könnten [108].

Aktuell hat diese Methode in Deutschland im Bereich der Erwachsenen-Intensivmedizin (DIVI), in der Initiative Qualitätsmedizin (IQM) und weiteren medizinischen Fächern (Geburtshilfe, Neonatologie, Kardiologie, Pädiatrische Intensivmedizin) an Bedeutung

zugenommen [110, 111]. Eine Evaluation hinsichtlich der Wirksamkeit steht jedoch noch aus. Zumindest im Rahmen von Modellprojekten hat das Peer Review auch Einzug in den ambulanten Bereich gehalten (s. z. B. Zentrum Qualität ÄK Niedersachsen). Weniger stark formalisierte und strukturierte Peer Besuche in Form von Hospitationen existieren seit langem insbesondere im Bereich der Primärversorgung (Niederländisches Visitation [112] oder UK-NHS Primary Care Practice Visits Programme, DEGAM Label Hospitatio-praxis).

Instrumente der Patientensicherheitsforschung

Zu den zentralen Instrumenten der Patientensicherheit zählen bspw.

- (i) Patientensicherheitsindikatoren, anhand derer durch zuvor festgelegte Referenzwerte Probleme in der Versorgung, z. B. Unerwünschte Ereignisse identifiziert, analysiert und vorhergesagt werden können;
- (ii) Schadenserfassungsinstrumente, bspw. Trigger Tools, die der Erfassung von Patientenschäden und (seltenen und schwerwiegenden, sog. Sentinel Events) UE dienen;
- (iii) Berichts- und Lernsysteme, die dazu dienen, dass Fehler, Risiken, kritische Ereignisse und Beinahe-Schäden durch die Leistungserbringer gemeldet und analysiert werden können. Patientenschäden werden über diese Systeme nicht gemeldet.

(i) Patientensicherheitsindikatoren

Patientensicherheitsindikatoren (PSI) werden im Rahmen der internen und/oder externen Qualitätssicherung eingesetzt und sind Kennzahlen, anhand derer durch zuvor festgelegte Referenzwerte Probleme in der Versorgung, z. B. UE identifiziert, analysiert und vorhergesagt werden können [4, 113, 114]. Diese Kennzahlen können bspw. anhand von Krankenaktenüberprüfungen, die als Goldstandard angesehen werden, oder auf Basis von Routinedaten erhoben werden [115–117]. Der Erhebungsaufwand bei Krankenaktenüberprüfungen ist gegenüber routinedatenbasierten Erhebungen vergleichsweise hoch. Routinedaten hingegen sind schnell verfügbare administrative Patienten- oder Strukturdaten, bspw. Diagnosen und Prozeduren oder auch Verweildauerdaten bzw. Wiederaufnahmeraten, die zu Abrechnungszwecken an die Krankenkassen weitergeleitet werden und zur Identifizierung von UE computerbasiert erhoben werden können [114].

Bei PSI spielt eine hohe Sensitivität, d. h. eine große Reichweite der Messung, mit einer hohen Rate richtig positiver Fälle und dem Ziel der Vollständigkeit, eine wichtige Rolle bei der Datenerhebung [4]. Mehrere Studien zur Bewertung der Güte von routinedatenbasierten Erhebungen, die auf Vergleichen mit Krankenakten beruhen, wurden durchgeführt [z. B. 115, 116, 118]. Bei der Auswahl einer Datenerhebungsmethode muss einerseits der Aufwand der Aktenkontrolle gegen die Verfügbarkeit von Routinedaten abgewogen werden. Andererseits ist zu beachten, dass die Datenqualität, wie Vollständigkeit oder Verfügbarkeit, aber auch die Qualität der Indikatordefinition, die Aussagekraft der Ergebnisse stark beeinflussen kann [118]. Quan et al. (2013) haben bspw. in ihrer Studie die PSI der „Agency for Healthcare Research and Quality“ (AHRQ) mithilfe von Krankenaktenüberprüfungen und Routinedatenerhebungen an 490 zufällig ausgewählten Patienten untersucht,

bei denen in Routinedaten sicherheitsrelevante Ereignisse identifiziert wurden. Sie kamen zu dem Schluss, dass die ausgewählten AHRQ-PSI zur Messung von UE zwar geeignet sind, jedoch im Hinblick auf die Datenerhebungsmethode weitere Forschung notwendig ist [115]. In der Literatur wird darum eine Kombination von Routinedatenerhebung und Krankenaktenüberprüfungen zur Identifizierung und Analyse von UE empfohlen [114, 118].

(ii) Schadenserfassungsinstrumente, bspw. Trigger Tools

Trigger Tools wurden ursprünglich zur Identifizierung und Analyse von unerwünschten Arzneimittelereignissen genutzt. Im Jahre 2006 wurden sie zu „Global Trigger Tools“ (GTT) weiterentwickelt und dienen nun der Erfassung von Patientenschäden und (seltenen und schwerwiegenden, sog. Sentinel Events) UE [119–121]. Die GTT werden im Rahmen der internen Qualitätssicherung mittels Krankenaktenüberprüfungen an geschlossenen Akten (retrospektiv) durchgeführt. Von einer Verwendung des GTT zum Vergleich von Krankenhäusern wird auf der Basis einer Studie von Schildmeijer et al. (2012) aufgrund der schlechten Interrater-Reliabilität der Reviewerteams abgeraten [120]. Im Rahmen der Anwendung des GTT werden die Akten unabhängig voneinander von mehreren Reviewern zufallsbasiert und anhand eines zuvor definierten Sets von Kriterien (Trigger) ausgewählt. Jedes gefundene UE dient als ein Warnsignal und führt zu einer tiefergehenden Analyse. Die Ergebnisse der Reviewer werden verglichen, die Schäden abschließend kategorisiert und die Anzahl der Schäden in Raten dargestellt [121]. Die Güte dieses Instrumentes (Validität und Reliabilität) wurde in verschiedenen Studien untersucht und eine solide Reliabilität, hohe Spezifität und stabile Sensitivität festgestellt [122, 123]. Bspw. wurden in der retrospektiven Studie von Sharek et al. (2011) an 10 % von 2400 Akten die Untersuchung der Interrater-Reliabilität und an 5 % der Akten die Intrarater-Reliabilität geplant. Eine Validierung des GTT wurde zudem anhand von 10 % der Akten durch erfahrene interne und externe Teams geplant. Letztlich sind 88–98 % der Akten untersucht wurden. Die Reliabilitätswerte variierten zwischen $K = 0,40$ und $0,60$. Die Spezifitätswerte umfassten für das interne Team 94 % und für das externe Team 93 %. Die Sensitivität betrug für das interne Team 49 % und für das externe Team 34 %. Auf Grundlage der Ergebnisse wurde darauf aufmerksam gemacht, dass dieses Verfahren zwar reliabel und praktikabel, jedoch als Goldstandard kritisch zu hinterfragen ist. Insgesamt wird das GTT, trotz seines Aufwandes, zur Erfassung von Patientenschäden empfohlen [123–125]. In einer Studie empfehlen Wong et al. (2015) zur Verbesserung der Güte die Entwicklung eines Rahmenkonzepts zur besseren Klassifizierung der Ereignisse [126].

(iii) Berichts- und Lernsysteme

Bei den Berichts- und Lernsystemen wird zwischen verpflichtenden und freiwilligen Systemen unterschieden. Im Rahmen verpflichtender Meldesysteme werden in der Regel unerwünschte Ereignisse, im Rahmen freiwilliger Meldesysteme werden bspw. Fehler, Risiken, und/oder Beinahe-Schäden gemeldet. Die Meldungen können seitens der Leistungserbringer und/oder seitens der Organisation erfolgen [127, 128]. Sie sind zwar im Falle selektiver freiwilliger Meldungen nicht für klassische Benchmarkings geeignet, da die „Gesamtpopulation“ letztlich unbekannt bleibt, dennoch kann ein einrichtungsübergreifendes Lernen durch sie unterstützt werden

[129], z. B. in Form von sog. „Safety Alerts“. Erwähnenswert ist in diesem Zusammenhang die Studie von Tamuz et al. (2004), die im Rahmen von semi-strukturierten Interviews (N = 86) untersucht hat, inwieweit die Definitionen und Klassifikationen von sicherheitsrelevanten Ereignissen bspw. die organisationale Routine der Datensammlung und der Datenanalyse beeinflussen. Sie fanden heraus, dass in Abhängigkeit des Verständnisses von Definitionen und der Wahl der Kategorien die Datenerfassung und Datenanalyse, somit die Möglichkeiten aus Fehlern zu lernen, beeinflusst werden [130]. Grundsätzlich ist das Risiko im Hinblick auf Verzerrungen der Daten (Bias) bei freiwilligen Meldesystemen groß, da die Daten unvollständig oder selektiv sein können [71]. Dieses Verfahren ist auch nicht dazu geeignet, Daten im zeitlichen Verlauf zu sammeln, da auch das Meldeverhalten stark biasanfällig ist: eine höhere Aufmerksamkeit für eine bestimmte Thematik kann zu einem stärkeren Meldeverhalten führen. Aufgrund mangelnder Ressourcen in Einrichtungen bleibt eine tiefergehende Analyse und die Entwicklung von Interventionen häufig aus [129]. Der wissenschaftliche Wert dieser Daten ist jedoch unbestritten, da durch sie insbesondere unerwartete Fehler aufgedeckt werden können und die Daten zudem dazu dienen können, Forschungsbereiche zu priorisieren [71]. Sowohl in Deutschland als auch international werden neuerdings fachspezifische Patientenregister auch zur Analyse von sicherheitsrelevanten Ereignissen genutzt. Publierte Erfahrungen liegen bspw. in der Rheumatologie [131] und in der Dermatologie vor (Rabbit, DRFZ Berlin, www.biologika-register.de; PsoBest, Hamburg, www.psobest.de; www.psonet.eu).

Vergleich der Methoden und Instrumente zur Erfassung sicherheitsrelevanter Ereignisse

Verschiedene Studien haben eine oder mehrere der bisher genannten Methoden und Instrumente im Hinblick auf die Erfassung von UE untersucht [114, 132, 133]. Naessens et al. (2009) verglichen PSI, GGT und Meldungen im Hinblick auf die Häufigkeit der Erfassung von UE. Sie kamen zu dem Schluss, dass die verschiedenen Datenerhebungsmethoden im großen Maße verschiedene UE erfassen und empfehlen die Ansätze im Rahmen der internen Qualitätssicherung miteinander zu kombinieren [132]. Auch Zhan et al. 2003 empfahlen eine Triangulation von Krankenaktenüberprüfungen und Datenerhebungen mit Routinedaten mit weiteren Verfahren, z. B. Fehlermeldungen [114]. Beckmann et al. (2003) verglichen „geförderte“ Meldungen als besondere Form der Meldeverfahren mit Krankenaktenüberprüfungen und empfahlen ebenfalls eine Kombination [133]. Flynn et al. (2002) verglichen direkte Beobachtungen mit Krankenaktenanalysen und Meldeverfahren. Sie fanden im Rahmen ihrer Studie zur Erfassung von Medikationsfehlern heraus, dass die direkte Beobachtung den beiden anderen Verfahren überlegen war [95]. Dennoch werden direkte Beobachtungen bspw. mit Krankenaktenüberprüfungen kombiniert [134]. Eine Kombination verschiedener Datenquellen wird empfohlen [94].

Fazit – Spezifische Methoden und Instrumente der Patientensicherheitsforschung

Zusammenfassend belegen die hier dargestellten spezifischen Methoden und Instrumente der Patientensicherheitsforschung deren Nutzen für systematische Analysen von Risiken und Fehlern im Versorgungsprozess, die Identifikation von Bewältigungsstrategien

mittels Prozessoptimierung und Trainingsmaßnahmen sowie die Entwicklung einer sicherheitsförderlichen Kultur durch deren Messung und deren Einbezug bei der Planung von Maßnahmen zur Optimierung der Patientensicherheit.

Letztlich erscheint vor dem Hintergrund wissenschaftlicher Literatur eine Kombination ausgewählter Verfahren im Hinblick auf die Erfassung und Analyse von UE im Rahmen der internen Qualitätssicherung als sinnvoll, um ein umfassendes Bild über UE zu erzielen. Im Rahmen einer externen Qualitätssicherung ist unter Berücksichtigung der jeweiligen Stärken und Schwächen sowie aus Machbarkeitsgründen die Verwendung einzelner Verfahren in Erwägung zu ziehen.

Methoden der Evaluation von Interventionen

„Doing an improvement project is fundamentally different from ‘studying’ it“ [135].

Der Sachverständigenrat Gesundheit forderte bereits im Jahr 2001, Maßnahmen der Qualitätsverbesserung konsequent zu evaluieren („Qualitätssicherung der Qualitätssicherung“) (SVR 2001, Bd. II Nr. 274), insbesondere da entsprechende Maßnahmen, durchaus einsichtig und von „gutem Willen“ getragen, in einigen Fällen nicht den erwünschten, sondern sogar paradoxe Effekte zeigen und einen inadäquat hohen Aufwand aufweisen können [136]. Die entsprechende Methodik einer solchen Evaluation unterscheidet sich nicht grundsätzlich von anderen Problemstellungen, denen sich die Versorgungsforschung gegenüber sieht; in erster Linie muss der „doppelten Komplexität“ [3, 137] von Kontext (z. B. Intensivabteilung) und Intervention (z. B. Einführung einer Checkliste) Rechnung getragen werden, indem die quantitativen Verfahren um qualitative Herangehensweisen ergänzt werden. In den vergangenen Jahren sind mehrere Empfehlungen zur Methodik von Implementierung und Evaluation in der Versorgungsforschung veröffentlicht worden [138–141], die hier nicht ausführlich dargestellt werden können.

Im Bereich Qualität und Patientensicherheit sind allerdings Besonderheiten zu beachten, von denen einige in der Lage sind, den kombinierten Einsatz quantitativer und qualitativer Methoden zu behindern, während andere Faktoren diesen durchaus fördern können.

Hinsichtlich der hemmenden Faktoren kann genannt werden [142, 143]:

- QPSF-Interventionen weisen – wie oben bereits angemerkt – sehr häufig eine hohe Augenscheinvalidität auf, wodurch die Notwendigkeit einer Evaluation von den Beteiligten infrage gestellt wird – trotzdem ist sie wichtig, wie zahlreiche Beispiele zeigen (paradoxe Effekte bei CPOE-Einführung, s. u.);
- Interventionen zeigen für sich genommen oft nur kleine Effekte, daher greift man zu Mehrfachinterventionen, die durchaus wirksam erscheinen, jedoch größere Anforderungen an die Evaluation stellen (z. B. bundled interventions wie zur Verminderung nosokomialer Infektionen [144]);
- Interventionen unterliegen in hohem Maße dem Phänomen der sozialen Erwünschtheit, außerdem ist ein ausgeprägter Hawthorne-Effekt (unspezifische Wirkung einer Intervention) zu beobachten, der jedoch nur eine limitierte zeitliche Konstanz hat;

- einige Phänomene der Patientensicherheit (z. B. Seitenverwechslung) sind sehr selten und können daher statistisch nur schwer direkt gefasst werden.

Auf der anderen Seite sind Aspekte zu nennen, die der doppelten Komplexität gegenüber adäquat erscheinen und den kombinierten Einsatz qualitativer und quantitativer Methoden bei der Evaluation komplexer Interventionen fördern können:

- Untersucher sind zumindest im institutionellen Bereich vertraut mit
 - Veränderungsbedingungen in komplexen Zusammenhängen sowie
 - komplexen Mehrfach-Interventionen (CMCI) wie sie z. B. zur Verminderung von Katheterinfektionen eingesetzt werden [144], zum Begriff s. unten und Guise et al. 2014 [145], und sind daher auf methodisch-technische Schwierigkeiten in der Durchführung von Evaluationen, auf paradoxe Effekte und eine geringe Ausprägung der Effektmaße vorbereitet;
- im Bereich Qualität und Patientensicherheit besteht eine besondere Nähe zu explorativen Methoden und damit eine Akzeptanz von qualitativen Verfahren, die die quantitative Evaluation ergänzen und zusätzliche Aspekte aufzeigen können, die sich quantitativen, klinisch-epidemiologischen Verfahren verschließen, da sie zunächst nicht bekannt sind (sog. Unknown Unknowns):
 - die regelmäßig angewandten Methoden wie Brainstorming, Prozessdarstellung, Prozessanalyse, Fehlerkettenanalyse und FMEA (Methodik s. o.) im Qualitätsmanagement können blinde Flecken identifizieren und einer Analyse zuführen;
 - besonders im Risikomanagement werden solche „generierenden Ansätze“ praktiziert (CIRS, Fehler-Ursachen-Analyse, Beschwerdemanagement, Analyse von Freitexten in Patienten- und Mitarbeiterbefragungen etc.) (zum Begriff des „generierenden Verfahrens“ s. Schrappe 2014, S 80);
- im Qualitäts- und Risikomanagement sind Untersucher (insbesondere auf institutioneller Ebene) oft selbst in den zu evaluierenden qualitätsverbessernden Prozess involviert und daher in der Lage, über ihre Rolle Auskunft zu geben, so wie es im Rahmen qualitativer Methoden gefordert wird.

Die vorgenannten hemmenden und fördernden Bedingungen spezifizieren das evaluative Herangehen in der Qualitäts- und Patientensicherheitsforschung. Grundsätzlich überwiegen jedoch die Überschneidungen mit anderen Forschungsgebieten in der Versorgungsforschung, insbesondere was die große Bedeutung komplexer Umgebungsfaktoren (Kontext) und die Komplexität der Intervention selbst angeht. Ein komplexes System besteht aus zahlreichen, in ihrer Zahl schwankenden, nicht-linear per multiplem Feedback miteinander verbundenen Teilen, die zu in Zeit und Stärke unvorhersehbaren Ereignissen führen, einzelne, nicht-explicite und veränderbare interne Regeln kennen und zur Selbstorganisation, Adaptation an die Umwelt sowie zu Lernprozessen in der Lage sind [146, 147]. Entsprechend der systemtheoretischen Provenienz ist das System größer als die Summe der Einzelteile (sog. Emergenz). Komplexe Systeme müssen von komplizierten Systemen abgegrenzt werden, deren Regeln im Prinzip bekannt und nachvollziehbar sind.

Die Diskussion, inwieweit die Erforschung komplexer Sachverhalte im Sinne einer „Linearisierung“ überhaupt möglich ist [148], kann hier nicht ausgeführt werden. Parallelen zu anderen Wissenschaftsbereichen geben jedoch zu einer konstruktiven Haltung Anlass, als Beispiel sei hier die Human Factor and Ergonomics-Forschung (HFE) genannt [149]. In jedem Fall ist es allerdings notwendig, Interventionen und den damit verbundenen Kontext genau zu spezifizieren, um in der Qualitäts- und Patientensicherheitsforschung evaluative Untersuchungen durchführen und v. a. diese bewerten zu können [150]. Folgende 7 Aspekte sollen hier angesprochen werden:

Intervention und Kontext sind gleichberechtigte Einflussfaktoren

Lineare Input-Output-Modelle sind nicht hinreichend und können – obwohl die Evaluation positive Ergebnisse erbringt – sogar paradoxen, negativen Auswirkungen Vorschub leisten (z. B. Einführung von Computer-Assisted Order Entry Systemen (CPOE) – Anstieg der Mortalität durch Veränderung der Dokumentation [151]). Neben der eigentlichen Intervention ist daher der Kontext als Einflussfaktor mit in den Blick zu nehmen, es handelt sich um sog. „CMO-Interventionen“ (Context + Mechanism = Outcome, in Abgrenzung zu OXO Observe – Perturbation – Observation [1, 152, 162]). Gleichzeitig ist die Intervention an den Kontext anzupassen. Auch der Prozess der Implementierung spielt eine Rolle und muss bei der Beurteilung der Wirksamkeit einer Qualitätsverbesserungs-Maßnahme berücksichtigt werden.

Zunächst wurde die Rolle des Kontexts unter dem Rubrum „externe Validität“ gefasst; diese Sichtweise reicht jedoch nicht aus, weil sie impliziert, Kontext sei eine passive Umgebungskonstante; stattdessen muss davon ausgegangen werden, dass der Kontext eine aktive und sehr veränderliche Größe darstellt [142]. Zusammenfassend ergeben sich folglich 5 Untersuchungsdomänen [141]:

- Intervention mit ihren core elements (die „peripheren Eigenschaften“ der Intervention können evtl. angepasst werden),
- äußerer Kontext (z. B. soziales oder politisches Umfeld),
- interner Kontext (z. B. organisatorische Struktur),
- beteiligte Individuen, insbesondere ihre Einstellungen und die geübte Praxis, und letztlich
- der Prozess der Implementierung (z. B. Schulungen).

Bei der Planung der Evaluation müssen also Kontext, individuelle Einstellungen und Implementierung mit einfließen, was durchaus erhebliche Auswirkungen auf Ressourcen und Zeitbedarf haben kann. Ganz entscheidend ist natürlich die Frage, auf welcher Ebene die Evaluation durchgeführt wird (Kontextebene), in erster Linie unterscheidet man

- Ebene des Gesundheitssystems (Qualitätsverbesserung auf Systemebene)
- institutionelle Ebene (Qualitätsmanagement)
- individuelle Ebene (Verhaltensänderung).

Typische Beispiele für komplexe Qualitätsverbesserungs-Initiativen auf Systemebene sind die Einführung von eHealth-Instrumenten [153] oder die qualitätsorientierte Vergütung (P4P).

Methodik und Untersuchungsgegenstand sind interdependent

Im klassischen Ansatz der biomedizinischen Forschung betrachtet man Untersuchungsgegenstand und Untersuchungsmethodik als

Z voneinander unabhängige Größen; der Methodik (z. B. Randomisation zum Ausschluss von Confoundern) wird kein eigener Einfluss auf den Untersuchungsgegenstand zugeschrieben. In der Versorgungsforschung kann diese Trennung nicht vollständig aufrechterhalten werden; stattdessen geht man davon aus, dass die gewählte Untersuchungsmethode zusätzlich zu Intervention und Kontext auf das Ergebnis einwirkt. Die Wahl der adäquaten Methodik hängt also nicht nur von der untersuchten Intervention ab, sondern die Methodik ist selbst ein aktiver Bestandteil der Intervention, die Auswahl der untersuchten Intervention hängt sogar erheblich von der eingesetzten Methodik ab. So könnte z. B. ein ausschließlicher Einsatz der randomisierten Kontrollgruppenbildung (Methodik) dazu führen, dass im Endeffekt nur Wissen zu isolierten, biomedizinisch dominierten Interventionen zur Verfügung steht, weil komplexe Interventionen schlechter im randomisierten Design zu untersuchen sind bzw. dort geringere Effektmaße ergeben: „that is where the evidence is“ [154].

Die seit über 10 Jahren geführte Auseinandersetzung über die methodischen Standards in der Qualitäts- und Patientensicherheitsforschung [154–156] kann also dahingehend aufgelöst werden, dass in Abhängigkeit vom Untersuchungsgegenstand sowohl quantitative als auch qualitative Methodenarsenale Verwendung finden müssen, um ein breites Feld von Interventionen untersuchen zu können und um keine Intervention von der Evaluation auszuschließen [157].

Evaluation: Adäquate Theoriebildung ist unbedingte Voraussetzung

In den Empfehlungen zur Methodik in der Versorgungsforschung wird grundsätzlich auf die große Bedeutung einer a priori stattfindenden Theoriebildung verwiesen, die den Untersuchungsgegenstand strukturiert, Interventionen und Kontext-Bedingungen darstellt sowie Aussagen über sinnvolle Endpunkte macht [17]. Dieses Vorgehen erscheint vor dem Hintergrund der genannten Wechselwirkungen von Kontext, Intervention und Methodik verständlich, allerdings fließt natürlich ein Vorwissen des Untersuchers in die Evaluation mit ein. An einigen Beispielen lässt sich jedoch zeigen, dass eine fehlende bzw. inadäquate Theoriebildung keine Alternative darstellt:

- Bei der Evaluation von Strategien zur Qualitätsverbesserung auf Systemebene wie Public Reporting („Qualitätsbericht“) spielt es eine große Rolle, ob man das Rational Choice Modell (individuelle Nutzenmaximierung) oder weiter gefasste Modelle (z. B. verhaltensökonomischer Herkunft) zugrunde legt, denn im ersten Fall wird man annehmen, dass Patienten den Qualitätsbericht konsequent zur Nutzenmaximierung verwenden, im zweiten Fall wird man auch Verhaltensalternativen (z. B. soziale Rollen) in Betracht ziehen.
- Bei der Betrachtung von Pay for Performance kann man davon ausgehen, dass ein finanzieller Anreiz jeglicher Größe (auch kleine Beträge) einen Effekt ergibt, oder man kann entsprechend der ökonomischen Theorie (z. B. Opportunitätskosten) die entgangenen Erlöse aus alternativen Verhaltensweisen (z. B. Generierung von Erlösen durch Fallzahlsteigerung) in den Mittelpunkt stellen und evtl. sogar Erkenntnisse aus der Principal Agent Theorie (Informationsasymmetrie) hinzuziehen – in jedem Fall wird man völlig unterschiedliche Settings

untersuchen (kleine Zahlungen vs. größere Zahlungen vs. Einsatz in Bereichen unvollständigen Wissens).

- In der Untersuchung von organisatorischen Veränderungsprozessen spielt es eine große Rolle, welches Verständnis man vom organisatorischen Wandel hat. Hängt man z. B. der sog. Professionalismus-Theorie [158] an, wird man andere, nämlich arzt-zentrierte Endpunkte untersuchen, als wenn man bspw. systemtheoretische Modelle verwendet und in der Folge z. B. Teamfaktoren, Wissensvorrat und Normen in der Organisation in den Mittelpunkt stellt [159].
- Bei der Evaluation von institutionellen QM-Interventionen kann man den Fehler begehen, Daten aus generierenden Verfahren (z. B. CIRS) als Messmethode zu verwenden, auch wenn bekannt ist, dass diese wegen ihrer mangelnden Sensitivität hierzu nicht geeignet sind. Selbst Indikatoren mit hoher Sensitivität und mittlerer Spezifität sind mit Vorsicht einzusetzen (s. u.). Aus diesen Gründen sollten wissenschaftliche Forschungsprojekte/Studien in diesem Themenfeld unbedingt auch Angaben zu den theoretischen Annahmen der Wirksamkeit enthalten.

Interventionen sind im Regelfall komplex

Die in der Qualitäts- und Patientensicherheitsforschung verwendeten Interventionen sind Kontext-sensibel und stehen in wechselseitigem Verhältnis zur eingesetzten Methodik. Beide, Kontext und Intervention, sind komplex (sog. „doppelte Komplexität“), es hat sich der Begriff der complex multicomponent interventions (CMCI) eingebürgert [1]. Der Begriff multicomponent macht hierbei einen wichtigen Aspekt der Komplexität verständlich, nämlich den gleichzeitigen Einsatz mehrerer Interventionen mit dem Effekt zahlreicher kleinster (inkrementeller) Veränderungen, so wie er immer wieder am Beispiel der im Laufe der Jahre verbesserten Sicherheit in der Anästhesie diskutiert wird. So sind auch die sog. bundled-interventions wie z. B. in der Klinischen Infektiologie (z. B. Vermeidung von Katheter-assoziierten Infektionen durch ein „Bündel“ von Interventionen wie Einhaltung von mehreren Hygiene-Maßnahmen, Veränderungen der safety culture usw.) als typische CMCI zu bezeichnen [144].

Methodisch von großem Interesse ist die Frage, ob „sich das Bündel aufschnüren“ lässt. In der Diskussion um die Testung und Isolierung von MRSA-Infektionen, einem der wichtigsten Probleme der Krankenhaushygiene, lässt sich dieser Konflikt gut nachzeichnen. Einerseits gibt es die bereits genannte große Multicenter-Studie [144], die durch den Einsatz des sog. „MRSA-bundle“ beeindruckende Ergebnisse im Sinne einer Senkung der MRSA-Infektion um 2/3 erbracht hat. In einer ex post-Analyse dieser Studie zeigt die Einzelmaßnahme von Testung/Isolierung jedoch nur eine marginale statistische Signifikanz [160], weswegen in Deutschland die Jain-Studie zur Begründung für dieses Vorgehen infrage gestellt wird [161]. Wegen der komplexen Zusammenhänge, die eine CMCI „zusammenhält“, muss es jedoch als äußerst problematisch gelten, eine solche komplexe Mehrfach-Intervention zu „dekonstruieren“.

CMCI stellen den Hauptteil der in der Qualitäts- und Patientensicherheitsforschung evaluierten Interventionen dar (z. B. Einführung von Checklisten, Evaluation von technologischen Verfahren wie Patientenarmband). Allerdings ist damit der Begriff der komplexen Intervention nicht vollständig beschrieben. Von großer

Wichtigkeit sind solche Interventionen, bei denen weniger die Multiplizität, sondern die Komplexität der Wirkung im Vordergrund steht. Dies ist z. B. der Fall bei edukativen Maßnahmen wie Kommunikations- oder Teamtraining und dem sog. human factors-Training mit Vermittlung von Führungs- und Teamkompetenz. Diese fortgeschritteneren Interventionen weisen eine tiefgehende Integration in den jeweiligen Kontext auf und involvieren die Untersucher maßgeblich. Letztendlich können solche Interventionen auch als soziale Konstrukte aufgefasst werden, die von den Vorstellungen der Beteiligten abhängen; umso wichtiger ist die Art der theoretischen Grundannahmen, die solchen Interventionen zugrunde liegen [162].

Datenquellen und Endpunkte

Bei aller Komplexität muss das Ergebnis einer Intervention sichtbar gemacht werden, der Rückgriff auf Daten bzw. Endpunkte der Evaluation kann nicht ausbleiben. Zum einen ist in der Qualitäts- und Patientensicherheitsforschung die Frage der Surrogat-Endpunkte von Bedeutung, die ja auch in der klinisch-evaluativen Forschung bzw. klinischen Epidemiologie kritisch diskutiert wird, in dem hier diskutierten Zusammenhang aber eine besondere Relevanz erhält, die 2 Aspekte betrifft:

- einerseits ist die Relevanz darin begründet, dass die Aussagekraft des Surrogates im komplexen Kontext noch weitaus kritischer zu beurteilen ist als im linearen klinischen Versuch. Eines der am meisten diskutierten Beispiele ist der Endpunkt Müdigkeit und Fehlerfreiheit von Ärzten in Abhängigkeit von Arbeitsbelastung und Arbeitszeit, denn die auf den ersten Blick einleuchtenden Zusammenhänge können durch negative Effekte an anderer Stelle (erhöhte Übergabefrequenz, Informationsverlust, Diskontinuität der Behandlung) wieder aufgehoben werden (Diskussion s. [163]). Wenn man mit Surrogat-Parametern arbeitet, muss auf jeden Fall sichergestellt werden, dass der ganze infrage stehende Prozess bzw. – im Falle Sicherheits-relevanter Fragen – die Fehlerkette mit einbezogen wird.
- andererseits gibt es gerade in der Patientensicherheitsforschung Fragestellungen, bei denen der beobachtete Endpunkt derart selten ist, dass er für statistische Zwecke nicht sinnvoll einzusetzen ist (z. B. Patientenverwechslungen im OP). Es ist daher unvermeidbar, hier Surrogate zu verwenden (z. B. unvollständige Patientenidentifikation), die den eigentlichen Endpunkt vorhersagen.

Ein zweiter Punkt, der die Verwendung von Endpunkten in der Qualitäts- und Patientensicherheitsforschung betrifft, reflektiert das Erkenntnisinteresse und somit die 3 typischen Konstellationen, in denen Qualitäts- und Sicherheitsdaten vorliegen ([4], S. 78ff):

- am schlechtesten für Evaluationszwecke geeignet, aber charakteristisch für Qualitäts- und Patientensicherheitskontexte sind Daten, die aus sog. generierenden Verfahren entstehen, mit denen das „unbekannt Unbekannte“ beschrieben wird, damit sie anschließend Gegenstand präventiver und analytischer Verfahren werden (z. B. CIRS). Diese Datenquellen weisen eine sehr geringe Sensitivität auf (nicht über 1 %), können also nicht zur Quantifizierung bzw. zur Evaluation verwendet werden. Allerdings sind sie – wie eingangs dargestellt – sehr gut geeignet, in einer qualitativen Vorfeldanalyse auf Parameter aufmerksam zu machen, die in der

Evaluation eine Rolle spielen könnten, von deren Existenz man aber nichts wusste.

- gut für Evaluationszwecke geeignet sind quantitative Verfahren, die eine gleichermaßen (möglichst) hohe Sensitivität und Spezifität aufweisen, also zum „Zählen geeignet“ sind. Hierunter sind z. B. klinisch-epidemiologische Falldefinitionen zu verstehen, wie sie u. a. die Infektionsepidemiologie seit 40 Jahren verwendet (z. B. Zahl der intravenösen Katheterinfektionen pro 1000 Patiententage mit Exposition).
- fraglich sind für Evaluationszwecke solche Indikatoren im engeren Sinne zu verwenden, deren eigentliches Einsatzgebiet das Monitoring von Prozessen und Ergebnissen darstellt. Indikatoren sind in erster Linie hoch-sensitiv eingestellt, auf Kosten ihrer Spezifität. Diese Eigenschaft resultiert aus ihrer Aufgabe, unerwünschte Verläufe i. S. einer Ampelfunktion vorherzusagen, um Gegenmaßnahmen einleiten zu können: es soll kein Warnhinweis übersehen werden [16], falsch-positive Ergebnisse werden jedoch toleriert. Um bei dem Beispiel der Katheterinfektion zu bleiben: als quantitativer Parameter „zählt“ sie die Infektionen, als Indikator sagt sie die Qualität der Intensivstation voraus. Allerdings kann eine nicht optimal eingestellte Spezifität bei der Evaluation von Interventionen sehr störend sein, denn hier sollten nicht Ereignisse fälschlicherweise als problematisch erkannt werden (falsch-positiv), weil das Evaluationsergebnis dadurch verfälscht wird.

In der Konsequenz beruht die Evaluation in der Regel auf quantitativen Verfahren, die eine gleichermaßen hohe Sensitivität und Spezifität aufweisen. Generierende Verfahren wie das CIRS sind geeignet, in einer qualitativen Analyse auf die unknown Unknowns aufmerksam zu machen und Indikatoren sind in der Lage, nach Abschluss der Evaluation z. B. in der Implementierungs- und Erhaltungsphase der Qualitätsverbesserung zu einem Monitoring zu kommen.

Ein dritter Aspekt betrifft die Art der Endpunkte. So ermöglicht der Einsatz qualitativer Methoden die Beobachtung von Endpunkten, deren Fokus über die biomedizinisch definierten und klinisch-epidemiologischen Endpunkte hinausgeht. V. a. aber macht es der breiter gefasste Fokus, also die Einbeziehung des Kontextes und die Komplexität der Interventionen, notwendig, zusätzlich zu klassischen quantitativen Endpunkten auch Endpunkte, die das Verhalten (sog. behavioral markers) und die Einstellungsebene beschreiben, einzubeziehen (Systematischer Review hierzu s. [164]). Ganz im Vordergrund stehen in diesem Zusammenhang die patient reported outcomes (PRO), die langfristig in der Lage sein könnten, den Fokus der Qualitäts- und Sicherheitsforschung von der Anbieter- hin zur Patientenorientierung zu verschieben [165].

Ein vierter Aspekt wird in seiner Wichtigkeit in Zukunft deutlich an Bedeutung zunehmen: die Verfügbarkeit sowie Analysierbarkeit großer Datenmengen durch Internet und *big data* [166]. Hier wird ein entscheidendes Forschungsfeld liegen, wobei erhofft werden kann, dass man zu neuen Ansätzen kommen wird, die „doppelte Komplexität“ von Interventionen im Gesundheitswesen zu evaluieren.

Methodische Instrumente zur Evaluation von Interventionen im Einzelnen

Qualitative Methoden

Die methodischen Instrumente, die zur Evaluation von Interventionen im Bereich der Qualitäts- und Patientensicherheitsforschung

zur Verfügung stehen, verwenden das im Allgemeinen genutzte Repertoire der qualitativen und quantitativen Methoden der QPSF. Die qualitativen Methoden umfassen u. a. Interviewtechniken, Gruppentechniken und teilnehmende Beobachtung i.S. ethnografischer Untersuchungen, Dokumentenanalysen und die Grounded Theory Methodik zur Theoriebildung [167]. Die AG Qualitative Methoden des DNVF fasst in ihrem ersten Diskussionspapier die Herangehensweise zusammen: Subjektorientierung, Interpretation von Bedeutungen, induktiver, re-iterativer Ansatz, mögliche Modifikation der Forschungsfrage und -methoden im Verlauf, Integration von Einstellungen und Werthaltungen der Forscher, Widerstandsorientierung und offenes Datenformat [168].

Wie eingangs dargestellt, stehen im Bereich Qualität und Patientensicherheit zudem einige als qualitativ zu bezeichnende Methoden zur Verfügung, die insbesondere aus der Notwendigkeit heraus entwickelt wurden, im organisatorischen Kontext Zugang zu Daten zu erhalten, die als „unbekannte Unbekannte“ in der Analyse von Prozessen, Schadensfällen oder in multivariate klinisch-epidemiologische Modelle integriert werden konnten. Es handelt sich hier um (Aufzählung nicht vollständig, s. o.)

- Prozessanalyse mit vorgeschaltetem brainstorming,
- Fehlerkettenanalyse z. B. nach dem London-Protokoll [169]
- Ursachenanalyse nach Ischikawa oder der Failure Mode and Efficiency-Analysis (FMEA),
- Critical Incident Reporting Systems (CIRS) als freiwilliges Fehlermeldesystem im Risikomanagement,
- andere generierende Verfahren wie Analyse von Freitexteinträgen von Patienten- oder Mitarbeiterbefragungen,
- Daten aus Obduktionen, Peer Review-Verfahren oder Morbidity Mortality Konferenzen.

Quantitative Methoden

In der quantitativen Erfassung werden zunächst unkontrollierte von kontrollierten Verfahren unterschieden (Zusammenfassung s. [143, 170]). Zur ersten Gruppe gehören viele Beobachtungsstudien (Prävalenz oder Inzidenz, retro- oder prospektiv) mit Modellierung und Analyse der Risikofaktoren (z. B. Unerwünschte Arzneimittelereignisse auf Intensivstation [171]).

Kontrollierte Studien können in erster Linie historisch oder parallel kontrolliert sein, wobei die Aussagekraft historisch kontrollierter Studien eher als schwach angesehen wird, da beobachtete Unterschiede trotz einer Kontrolle gegenüber bekannten Störgrößen (Confounder) auf nicht bekannte Störgrößen zurückzuführen sein können. Allerdings gibt es Situationen, in denen dieses Studiendesign unvermeidbar ist, wenn z. B. nationale Kampagnen [172, 173] oder Veränderungen in einzelnen Institutionen evaluiert werden müssen. Dieser Studientyp hat eine höhere Aussagekraft [143], wenn

- mehrere Messungen vor und nach der Intervention durchgeführt wurden (time series Design),
- die Größe des Effekts überzeugend ist,
- die Intervention eine hohe Augenscheinplausibilität aufweist, oder
- die Ergebnisse mit anderen zeitgleichen Studienergebnissen vereinbar sind (sog. Triangulation der Evidenz).

Es gibt zahlreiche Studien, die im historischen Vergleich unter Verwendung eines time series Designs maßgebliche Verbesserun-

gen durch Mehrfachinterventionen zeigen konnten [174, 175]. Als für den Bereich Qualitäts- und Risikomanagement typischer Sonderfall des time series-Designs kann die sog. statistische Prozesskontrolle (report oder control cards) gelten, in der ebenfalls zahlreiche Messungen hintereinander durchgeführt werden, wodurch die Regression to the mean weniger wahrscheinlich ist als bei Einzelmessungen [143].

Bei kontrollierten Parallelgruppenstudien kommt als weiteres Unterscheidungsmerkmal die Methode der Zuteilung zur Interventions- und Kontrollgruppe hinzu. Bei nicht randomisierten Studien kann diese Zuteilung z. B. durch die natürlichen Umstände erfolgen, z. B. Patienten der Station A vs. Station B (sog. quasi-experimentelles Design). Diese Studien können als Querschnittsstudien (eine gleichzeitig durchgeführte Einzelbeobachtung in der Interventions- und Kontrollgruppe) oder als parallel durchgeführte longitudinale Untersuchung konzipiert sein. In der letztgenannten Form, die gerade bei Systeminterventionen sehr oft verwendet wird, werden sowohl in der Interventions- als auch Kontrollgruppe eine oder besser mehrere Beobachtungen jeweils vor und nach der Intervention vorgenommen (z. B. [176] zur Wirksamkeit von P4P-Programmen).

Randomisierte Studien können wiederum auf Patientenebene und auf Leistungserbringerebene unter Verwendung von sog. Clustern stattfinden. Der Vorteil der (Cluster-) Randomisierung ist, dass für bekannte und unbekannte Confounder kontrolliert wird und daher die Risikoadjustierung weniger relevant wird – vorausgesetzt die Fallzahl ist hoch genug.

Klassische randomisierte Studien mit individueller Randomisierung auf Patientenebene sind für komplexe Interventionen jedoch nur im Ausnahmefall anwendbar und werden meist für biomedizinische Einzelinterventionen genutzt (z. B. Anwendung von Mupirocin-Nasensalbe bei der MRSA-Sanierung [177]). Das Problem besteht in der sog. Kontamination durch die nur eingeschränkt mögliche Zuordnung und Verblindung gegenüber komplexen (System-) Interventionen: auf der Patientenebene ist es z. B. kaum möglich, die Einführung einer Checkliste zufällig zuzuordnen und verdeckt zu halten.

Übersteigt die Kontaminationsrate einen gewissen Prozentsatz (z. B. 30 % aller Zuordnungen [143]), dann ist es unvermeidlich, eine Cluster-Randomisierung einzusetzen. Cluster-Randomisierungen können auf der Ebene des Behandlungsortes bzw. der Behandlungsinstitution vorgenommen werden; auch können behandelnde Ärzte als Cluster gelten. Cluster-Studien können, je nach Ausprägung der sog. Intraklassen-Korrelation (ICC), eine geringere Power als Studien mit Randomisierung auf individueller Ebene haben.

2 Formen des Cluster-Designs sind zu unterscheiden: erstens das Querschnittsdesign (in beiden Gruppen werden zu bestimmten Zeitpunkten Querschnittsmessungen gemacht, die Patienten sind aber nicht die gleichen) – im engeren Sinne könnte man bei einem solchen Design auch von einer Querschnittsstudie mit random sampling sprechen. Zweitens das Kohortendesign (die gleichen Patienten in beiden Gruppen werden longitudinal untersucht). Die Cluster-Randomisierung ist in der Qualitätsforschung oft mit Erfolg angewandt worden (z. B. critical pathways für die ambulant erworbene Pneumonie [178], Ernährung auf Intensivstation [179], Einführung eines Notfallteams [180]). Ist die Intervention so komplex, dass sie nicht zu einem Zeitpunkt eingeführt wer-

den kann, wird das sog. stepped wedge Design mit schrittweiser Einführung der Intervention empfohlen [143].

Unabhängig von der Patientenzuordnung spielt ebenso wie bei den Therapiestudien die Verblindung eine wichtige Rolle. Eine Verblindung ist insbesondere bei der Auswertung und Analyse notwendig, weil sonst die Beobachter die Ergebnisse beeinflussen können, wenn sie z. B. wissen, ob es sich um einen Patienten mit unerwünschtem Ereignis oder nicht handelt (sog. hindsight bias). Zudem muss der Hawthorne-Effekt beachtet werden, der die unspezifische Beeinflussung des Ergebnisses durch die Studiendurchführung selbst beschreibt. Wenn eine Verblindung nicht möglich ist (z. B. Untersuchung des Einflusses der Arbeitszeit auf die Patientensicherheit), dann müssen zumindest die Beobachtungsendpunkte verblindet sein.

Zusammenwirken qualitativer und quantitativer Methoden
Der kombinierte Einsatz von qualitativen und quantitativen Methoden, oft als mixed methods design [181] oder „Triangulation“ [182] bezeichnet, wirft in der konkreten Umsetzung zahlreiche Fragen auf. Oft wird ein sequenzieller Einsatz impliziert: Problemstellung klären, Kontext analysieren, Evaluation entwickeln und beschreiben, Modell etablieren – zusammengefasst unter dem Begriff der Präimplementierungsforschung [1, 142] – und danach folgt die quantitative Evaluation. Es wird aber auch ein kombiniert parallelsequenzieller Einsatz von qualitativen und quantitativen Ansätzen diskutiert ([183], s. Box).

WAYS IN WHICH QUALITATIVE METHODS CAN BE USED ALONGSIDE RANDOMISED CONTROLLED TRIALS

Before a trial

- To explore issues related to the healthcare question of interest or context of the research
- To generate hypotheses for examination in the randomised controlled trial
- To develop and refine the intervention
- To develop or select appropriate outcome measures

During a trial

To examine whether the intervention was delivered as intended, including describing the intervention as delivered

- To “unpack” processes of implementation and change
- To explore deliverers’ and recipients’ responses to the intervention

After a trial

- To explore reasons for the findings of the trial
- To explain variations in effectiveness within the sample
- To examine the appropriateness of the underlying theory
- To generate further questions or hypotheses

Umfassende Ansätze zu einer Qualitäts- und Patientensicherheitsforschung, die auch niedrigere Grenzwerte für Evidenz, Einbeziehung statt Ausschluss von Bias-Formen (z. B. Nutzung als local wisdom) und Berücksichtigung von Stimmung und Atmosphäre einbeziehen, stammen z. B. von Berwick [1]. Eine hervorragende Zusammenfassung der möglichen Studienansätze einer „improvement science“ auf der Basis eines systematischen Reviews wird

durch Portela et al. (2015) [182] gegeben. Ein gutes Beispiel für ein kombiniert qualitativ-quantitatives Vorgehen stellt die Studie von Marwick et al. (2014) [184] zur Prävention der nosokomialen Sepsis dar.

Veröffentlichung und Wissenssynthese

In zusammenfassenden Darstellungen von Qualitätsforschungsergebnissen, insbesondere auch von mixed-methods-Forschung, werden schon seit längerer Zeit Qualitätsmerkmale dieser Studien diskutiert, die Ansatzpunkte für ein rating bieten können [183]. Das Instrument SQUIRE 2.0 gibt in Anlehnung an das CONSORT-Instrument Empfehlungen für die Veröffentlichung von Qualitätsforschung [135]. Aus Deutschland stammt das Instrument „STROSA“, das Empfehlungen zur Veröffentlichung von Sekundärdaten-basierten Studien gibt [185]. Diese Diskussion steht jedoch erst am Anfang; gleiches gilt für die Methodik zur Anfertigung systematischer Reviews dieser Studien [186]. Eventuell kommt dem handsearching und dem snowballing größere Bedeutung zu als bei klassisch klinischen Fragestellungen [187].

Forschungsbedarf in der Qualitäts- und Patientensicherheitsforschung

Im Vergleich zur angloamerikanischen Qualitäts- und Patientensicherheitsforschung (QPSF) lässt sich für den deutschsprachigen Raum erheblicher Nachholbedarf bei der Beforschung von Qualität und Patientensicherheit konstatieren, dem zahlreiche Ursachen zugrunde liegen, von denen die wichtigsten benannt werden sollen. Gleichzeitig soll aber darauf hingewiesen werden, dass für die kommenden Jahre aufgrund der gesundheitspolitischen Zielsetzungen mit „Qualität“ als Leitkriterium der Gesundheitsversorgung und den Ansätzen zur qualitätsorientierten Krankenhausplanung und Versorgungssteuerung die klare Erwartung besteht, dass die QPSF in Deutschland einen beträchtlichen Aufschwung erleben wird:

Erstens kommt erst ab 2016 mit dem Innovationsfonds der politische Wille zum Ausdruck, die (Gesundheits-)Versorgungsforschung, der die Qualitäts- und Patientensicherheitsforschung zuzuordnen ist, umfassend zu fördern. Etablierte deutsche Förderinstitutionen wie die DFG bevorzugten bislang Forschungsanträge zur Grundlagenforschung [188], öffnen sich aber zunehmend der Gesundheitsversorgungsforschung. Es besteht daher die Hoffnung, dass in Zukunft die oftmals komplexen Fragestellungen und Interventionen der QPSF trotz komplexer Umgebungen qualitativ hochwertig, mit Primär- und Sekundärdaten beforscht werden können.

Zweitens sind Routinedaten, die man mit überschaubaren finanziellen Mitteln beforschen könnte, zwar aus den verschiedenen Sektoren vorhanden, allerdings lückenhaft bzw. aus datenschutzrechtlichen Aspekten z. T. nur schwer zugänglich. Inwieweit die zunehmend sich etablierenden Patientenregisterdaten beforscht werden können und welchen Erkenntniswert diese gerade von chirurgischen Fächern propagierten Register gegenüber RCTs für die QPSF haben können, bleibt genauso abzuwarten wie die Evaluation des Potenzials der „Big Data“-Ansätze in der Gesundheitsversorgungsforschung.

Drittens existiert auf der Ebene des deutschen Gesundheitssystems – im Gegensatz zu anderen Ländern, die nationale Gesund-

heitsqualitätsberichte publizieren (z. B. USA, Canada, Niederlande) – weder ein konsistentes Konzept zur systematischen Beurteilung der verschiedenen Qualitätsdimensionen nach Donabedian oder zur Berücksichtigung der Patientenperspektive, noch eine über die oftmals kritisierte so genannte externe Qualitätssicherung der Krankenhäuser hinausgehende, methodisch fundierte Herangehensweise, die geeignet wäre, die Ergebnisse verschiedener Institutionen sektorübergreifend und risikoadjustiert sowie die Ergebnisse verschiedener Regionen miteinander zu vergleichen.

Viertens ist in Deutschland bislang die u. a. von Aiken et al. (2014) [189] geforderte, prozessorientierte Herangehensweise der Patientenbehandlung kaum verbreitet, die v. a. die Interdisziplinarität der beteiligten Berufsgruppen sowie deren explizite Qualifikationen berücksichtigt.

Fünftens muss in Anbetracht der Vielzahl an aktuellen Aktivitäten zur Qualitäts- und Patientensicherheitsförderung aus Sicht der Versorgungsforschung kritisch angemerkt werden, dass die meisten Aktivitäten nicht hinreichend wissenschaftlich evaluiert wurden. Hierzu gehören u. a. die gesetzlichen, dokumentationsintensiven Qualitätssicherungsmaßnahmen oder auch die Zertifizierung von Organzentren, die hinsichtlich Aufwand und Nutzen bzw. Evidenz für Wirksamkeit nicht systematisch untersucht wurden. Der nunmehr politisch eingeschlagene Weg, Qualitätsaspekte zur Grundlage von Vergütung und Versorgungsplanung heranzuziehen, hat in Deutschland bislang keine wissenschaftliche Grundlage. Internationale Studien legen aufgrund kurzfristiger und/oder nicht einheitlich gerichteter Effekte nahe, sehr differenziert an diese Thematik heranzugehen [4].

Sechstens streifen die in Deutschland favorisierten, noch sehr traditionell angelegten Qualitätssicherungsansätze zwar implizit Fragen zur Über-/Unter-/Fehlversorgung und zur Priorisierung/Rationierung, aber Empfehlungen zur Fokussierung auf populationsorientierte, auf vernetzte Regionen bezogene und sektorenübergreifende Qualitätsindikatoren sowie zur stärkeren Einbindung von Patienten-berichteten Ergebnissen (PRO) sind noch zu wenig berücksichtigt [4].

Siebtens wäre eine systematische Berücksichtigung der Perspektive aller am Versorgungsprozess Beteiligten aus allen Versorgungsbereichen wünschenswert. Potenzielle Zusammenhänge zwischen Faktoren, die das behandelnde Personal einerseits und Behandlungsergebnisse andererseits betreffen, sollten verstärkt untersucht werden, insbesondere in Zeiten der sich verändernden personellen Anforderungen. Forschungssettings sollten neben der akutstationären Versorgung auch rehabilitative oder pflegerische Institutionen sein.

Achtens fehlt einer Vielzahl an Initiativen zur Qualitätsverbesserung ein umfassendes Rahmenkonzept. Dieses ließe Einzelmaßnahmen und ihre Auswirkungen besser einordnen und beurteilen.

Aus dieser exemplarischen Aufzählung der Defizite lassen sich Forschungsthemen ableiten, die die Qualitäts- und Patientensicherheitsforschung (QPSF) in Deutschland vordringlich in den Fokus nehmen sollte:

- a) Forschung zu Methoden, die dabei helfen, die (erreichbaren) Ziele für Qualität und Sicherheit auf Systemebene und für einzelne Leistungserbringer unter Einbeziehung aller Akteure zu definieren,

- b) Forschung zur optimalen Auswahl von Daten für die Beschreibung der Ist-Situation – Analyse von Datenquellen, die Datenqualität, Stichprobenrepräsentativität und Aussagekraft bei gleichzeitiger Ressourcensparsamkeit garantieren,
- c) Forschung zur Entwicklung risikoadjustierter Qualitätsindikatoren (QI) für die unterschiedlichen Einsatzzwecke: planungsrelevante QI, QI für Zu- und Abschläge der Vergütung, QI für regionenbezogene Qualitätsmessungen („Area-QI“), usw.
- d) Forschung zur Entwicklung von Scores bzw. Indices aus einzelnen QI (Gewichtung usw.) und zur Bewertung der Güte und Aussagekraft von Qualitätsindikatorensets für spezifische medizinisch-pflegerische Themen und Versorgungsbereiche
- e) Forschung zur Zuschreibbarkeit von Qualitätsergebnissen, insbesondere bei sektorübergreifenden Qualitätsmessungen,
- f) Forschung zur Effektivität und Effizienz qualitäts- und sicherheitsrelevanter Interventionen in den Einrichtungen der Gesundheitsversorgung, insbesondere auch der eingeführten, flächendeckenden Verfahren im ambulanten und stationären Sektor,
- g) Forschung zur Kombination von Patientenerfahrungen und klinischen Qualitätsmessungen,
- h) Forschung zu den Voraussetzungen und zur Förderung einer Qualitäts- und Sicherheitskultur im Gesundheitswesen,
- i) Forschung zur Beurteilung von Qualität und Sicherheit aus Patienten- und Mitarbeitersicht,
- j) Forschung zur laienverständlichen Aufbereitung von Qualitäts- und Sicherheitsdaten für Informationsportale
- k) Forschung zur Eignung von „Big-Data“-Ansätzen zur Beantwortung von Fragestellungen der QPSF

Diese Liste potenzieller Forschungsthemen der QPSF ist sicher nicht erschöpfend, sondern skizziert nur Ausgangspunkte zukünftiger QPSF-Forschung. Dabei soll das vorliegende Memorandum dazu beitragen, dass die bereits etablierten Methoden und Ergebnisse ausreichend berücksichtigt werden.

Entwicklung, Kommentierung und Abstimmung des Memorandums

Alle Mitglieder des Deutschen Netzwerks Versorgungsforschung wurden zur Mitwirkung in der AG Qualitäts- und Patientensicherheitsforschung eingeladen, die das Memorandum federführend koordiniert hat. Innerhalb der AG wurde die Grundstruktur des Memorandums abgestimmt und es wurden Autorengruppen gebildet, deren erster Entwurf allen Mitgliedern mit einer Frist von drei Monaten zur Kommentierung zugesandt wurde. Nach Sichtung und Auswertung der eingegangenen Änderungs- und Ergänzungsvorschläge seitens der Autorengruppen wurde eine finale zur Publikation erstellte Fassung erstellt und an die ordentlichen und korrespondierenden institutionellen Mitglieder des DNVF mit der Einladung zur „Mitzeichnung“ versandt.

Folgende Fachgesellschaften haben sich im Rahmen des Kommentierungsprozesses am Methodenmemorandums zur Qualitäts- und Patientensicherheitsforschung beteiligt:

- Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin
- Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V.

- Deutsche Gesellschaft für Hebammenwissenschaft
- Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e.V.
- Deutsche Gesellschaft für Medizinische Soziologie e.V.
- Deutsche Gesellschaft für Pflegewissenschaft e.V.
- Deutsche Gesellschaft für Prävention und Rehabilitation von Herz- und Kreislauferkrankungen e. V.
- Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde
- Deutsche Ophthalmologische Gesellschaft e.V.
- Deutsche Sepsis-Gesellschaft
- Deutsche Gesellschaft für Unfallchirurgie e.V.

Folgende Wissenschaftliche Institute und Forschungsverbände haben sich im Rahmen des Kommentierungsprozesses am Methodenmemorandum zur Qualitäts- und Patientensicherheitsforschung beteiligt:

- AQUA - Institut für Angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH
- Kompetenzzentrum Versorgungsforschung in der Dermatologie
- LVR-Institut für Versorgungsforschung
- Zentrum für Bevölkerungsmedizin und Versorgungsforschung der Universität zu Lübeck
- Gesundheit Österreich GmbH (GÖG)
- Zentrum für Versorgungsforschung Köln
- Institut der Deutschen Zahnärzte (IDZ)

Zudem haben verschiedene Einzelmitglieder des DNVF, die hier nicht alle aufgeführt werden können, hilfreiche Kommentare geliefert.

Verantwortlichkeit der Autoren für die einzelnen Kapitel

Alle Autoren zeichnen für den gesamten Text verantwortlich. Die Federführung bei den einzelnen Kapiteln war wie folgt verteilt: „Einleitung“: MG, SD, JS und MS; „Zur Aktualität des Forschungsgebietes“: BS; „Methoden der Messgrößenentwicklung und –güteprüfung“: JS, KD und MG; „Methoden der Risikoadjustierung“: GH; „Spezifische Methoden und Instrumente der Patientensicherheitsforschung“: SK und TM; „Methoden der Evaluation von Interventionen“: MS und ME-G; und „Forschungsbedarf in der Qualitäts- und Patientensicherheitsforschung“: MG und SD.

Danksagung

Wir danken Herrn Prof. Dr. J. Schmitt für hilfreiche Kommentare zum Manuskript und Frau S. Grobosch, Frau M. Stolte und Herrn M. Darbinjan sowie Herrn J. Pollmanns für redaktionelle Überarbeitungen des Manuskripts. Insbesondere gilt unser Dank allen Personen und/oder Gesellschaften, Organisationen und Institutionen die durch ihre kritischen Beiträge im gesamten Entwicklungsprozess und der Kommentierungsphase zu dieser Publikation beigetragen haben.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Das Memorandum III, Teil 3 wird von folgenden ordentlichen und korrespondierenden institutionellen Mitgliedern des Deutschen Netzwerks Versorgungsforschung e. V. getragen*

Von folgenden Mitgliedern der Sektion „Fachgesellschaften“:

- Deutsche Diabetes Gesellschaft e.V. (DDG)
- Deutsche Gesellschaft für Allergologie und klinische Immunologie e.V. (DGAKI)
- Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin e.V. (DEGAM)
- Deutsche Gesellschaft für Ambulante Allgemeine Pädiatrie e.V. (DGAAP)
- Deutsche Gesellschaft für Angiologie – Gesellschaft für Gefäßmedizin e.V. (DGA)
- Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin e.V. (DGAI)
- Deutsche Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V. (DGAUM)
- Deutsche Gesellschaft für Chirurgie e.V. (DGCH)
- Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e.V. (DGGG)
- Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e.V. (DGHO)
- Deutsche Gesellschaft für Innere Medizin e.V. (DGIM)
- Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e.V. (DGK)
- Deutsche Gesellschaft für Kinderchirurgie e.V. (DGKCH)
- Deutsche Gesellschaft für Kinder- und Jugendmedizin e.V. (DGKJ)
- Deutsche Gesellschaft für Medizinische Psychologie e.V. (DGMP)
- Deutsche Gesellschaft für Medizinische Soziologie e.V. (DGMS)
- Deutsche Gesellschaft für Mund-, Kiefer und Gesichtschirurgie e.V. (DGMKG)
- Deutsche Gesellschaft für Neurochirurgie e.V. (DGNC)
- Deutsche Gesellschaft für Orthopädie und Orthopädische Chirurgie e.V. (DGOOC)
- Deutsche Gesellschaft für Pflegewissenschaft e.V. (DGP)
- Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e.V. (DGP)
- Deutsche Gesellschaft für Prävention und Rehabilitation von Herz-Kreislauferkrankungen e.V. (DGPR)
- Deutsche Gesellschaft für Psychiatrie, Psychotherapie, Psychosomatik und Nervenheilkunde e.V. (DGPPN)
- Deutsche Gesellschaft für Psychoanalyse, Psychotherapie, Psychosomatik und Tiefenpsychologie e.V. (DGPT)

* Alle ordentlichen und korrespondierenden institutionellen Mitglieder des DNVF e.V. wurden zur Mitzeichnung des Memorandums eingeladen, aufgeführt werden die Mitglieder, deren positive Rückmeldung bis zum 04.09.17 in der DNVF-Geschäftsstelle einging.

- Deutsche Gesellschaft für Psychosomatische Medizin und Ärztliche Psychotherapie e.V. (DGPM)
- Deutsche Gesellschaft für Schlafforschung und Schlafmedizin e.V. (DGSM)
- Deutsche Gesellschaft für Senologie e.V. (DGS)
- Deutsche Gesellschaft für Sozialmedizin und Prävention e.V. (DGSMP)
- Deutsche Gesellschaft für Thorax-, Herz- und Gefäßchirurgie e.V. (DGTHG)
- Deutsche Gesellschaft für Unfallchirurgie e.V. (DGU)
- Deutsche Gesellschaft für Zahn-, Mund- und Kieferheilkunde e.V. (DGZMK)
- Deutsche Ophthalmologische Gesellschaft e.V. (DOG)
- Deutsche Krebsgesellschaft e.V. (DKG)
- Deutsche Schlaganfall-Gesellschaft (DSG)
- Deutsche Schmerzgesellschaft e.V. (DGSS)
- Deutscher Verband für Gesundheitswissenschaften und Public Health e.V. (DVGPH)
- Deutsches Kollegium für Psychosomatische Medizin (DKPM)
- Gesellschaft für Arzneimittelanwendungsforschung und Arzneimittel-epidemiologie e.V. (GAA)
- Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e.V. (GQMG)
- Schleswig-Holsteinische Krebsgesellschaft e.V.

Von folgenden Mitgliedern der Sektion „Wissenschaftliche Institute und Forschungsverbände“:

- Abteilung Allgemeinmedizin und Versorgungsforschung, Universitätsklinikum Heidelberg
- BQS Institut für Qualität und Patientensicherheit GmbH
- Center for Health Care Research, Universitätsklinikum Hamburg-Eppendorf & Universität Hamburg
- Centre for Health and Society, Universitätsklinikum Düsseldorf
- Fachgebiet Management im Gesundheitswesen, Technische Universität Berlin
- Institut der Deutschen Zahnärzte (IDZ)
- Institut für Forschung in der Operativen Medizin (IFOM) der Universität Witten/Herdecke gGmbH
- Institut für Hausarztmedizin, Universitätsklinikum Bonn
- Institut für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaft (IMVR) der Universität zu Köln
- Institut für Patientensicherheit (IfPS) der Universität Bonn
- Institut für Pflegewissenschaft und -praxis der Paracelsus Medizinischen Privatuniversität Österreich
- Institut für Versorgungsforschung in der Onkologie GbR (InVO)
- Landesinstitut für Gesundheit des Bayerischen Landesamts für Gesundheit und Lebensmittelsicherheit (LGL)
- LVR-Institut für Versorgungsforschung (LVR-IVF)
- Medizinische Hochschule Brandenburg (MHB), Theodor Fontane
- Picker Institut Deutschland gGmbH
- Sektion Versorgungsforschung und Rehabilitationsforschung, Universitätsklinikum Freiburg (SEVERA)
- Tumorzentrum Regensburg, Institut für Qualitätssicherung und Versorgungsforschung der Universität Regensburg
- Universitäres Zentrum für Versorgungsforschung (UZVF) der Universität Leipzig

- Universität Potsdam – Professur für Rehabilitationswissenschaften
- Wissenschaftliches Institut der Niedergelassenen Hämatologen und Onkologen (WINHO)
- Zentrum für Bevölkerungsmedizin und Versorgungsforschung (ZBV) der Universität zu Lübeck
- Zentrum für Evidenzbasierte Gesundheitsversorgung (ZEGV) der Medizinischen Fakultät Carl Gustav Carus der Technischen Universität Dresden und des Universitätsklinikums Carl Gustav Carus Dresden
- Zentrum für Medizinische Versorgungsforschung (ZMV) Universitätsklinikum Erlangen, Psychiatrische und Psychotherapeutische Klinik
- Zentrum für Versorgungsforschung Köln (ZVFK)

Von folgenden Mitgliedern der Sektion „Juristische Personen und Personenvereinigungen“:

- 4QD – Qualitätskliniken.de GmbH
- AXA Konzern AG
- BARMER
- Berlin-Brandenburger Herzinfarktregister e.V. (B₂HIR)
- Berufsverband der Kinder- und Jugendärzte e.V. (bvkj)
- Berufsverband für Orthopädie und Unfallchirurgie e.V. (BVOU)
- BKK Dachverband e.V.
- Bundesverband Deutscher Privatkliniken e.V. (BDPK)
- Deutsche Rheuma-Liga Bundesverband e.V. (DRL)
- Gesundheit Österreich GmbH
- gevko GmbH
- INSIGHT Health GmbH & Co. KG
- Interessenverband zur Qualitätssicherung der Arbeit niedergelassener Uro-Onkologen in Deutschland e.V. (IQUO)
- Kassenärztliche Bundesvereinigung (KBV)
- Landeszahnärztekammer Baden-Württemberg
- OptiMedis AG
- Techniker Krankenkasse (TK)

Von folgenden „korrespondierenden“ Mitgliedern:

- Aktionsbündnis Patientensicherheit e.V. (APS)

Literatur

- [1] Berwick DM. The science of improvement. JAMA 2008; 299: 1182–1184
- [2] Marshall MN, Pronovost P, Dixon-Woods M. Promotion of improvement as a science. Lancet 2013; 381: 419–421
- [3] Schrappe M, Pfaff H. Versorgungsforschung vor neuen Herausforderungen: Konsequenzen für Definition und Konzept. Gesundheitswesen 2016; 78: 689–694
- [4] Schrappe M. Qualität 2030 – die umfassende Strategie für das Gesundheitswesen. Mit einem Geleitwort von Ulf Fink und Frans Dormann. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2015
- [5] McGinnis JM, Diaz A, Halfon N. Systems strategies for health throughout the life course. JAMA 2016; 316: 1639–1640
- [6] Institute of Medicine. Vital signs: Core metrics for health and health care progress. Washington, DC: The National Academies Press; 2015

- [7] Sens B, Fischer B, Bastek A et al. Begriffe und Konzepte des Qualitätsmanagements – 3. Auflage. GMS Med Inform Biom Epidemiol 2007; 3: Doc05
- [8] Gesundheitsqualitätsgesetz Österreich. Im Internet: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=22000388&ShowPrintPreview=True> Stand: 12.07.2017
- [9] Manser T, Frings J, Heuser G et al. The German clinical risk management survey for hospitals: Implementation levels and areas for improvement in 2015. Z Evid Fortbild Qual Gesundheitswes 2016; 114: 28–38
- [10] Sens B, Bamberg C, Kurscheid-Reich D et al. Von der Strategie zur kennzahlengestützten Umsetzung – in zehn Schritten zum Erfolg. Heidelberg: Medhochzwei Verlag; 2015
- [11] Manky T, Völzke T, Nimptsch U. Improving outcomes using German inpatient quality indicators in conjunction with peer review procedures. Z Evid Fortbild Qual Gesundheitswes 2015; 109: 662–670
- [12] Sens B, von Lewinski G, Albrecht UV. „TEP-App“ für die Endoprothetik. Dtsch Arztebl 2015; 112: A830
- [13] Langer M, Schuster S, Dormann H et al. eMediPlan – ein sektorenübergreifender Ansatz zum elektronischen Medikationsplan. In: Duesberg F, (Hrsg.). E-Health. 2015: Solingen: Medical future Verlag; 2014: 111–115
- [14] Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001
- [15] Carinci F, Van Gool K, Mainz J et al. Towards actionable international comparisons of health system performance: expert revision of the OECD framework and quality indicators. Int J Qual Health Care. 2015; 27: 137–146
- [16] Joint Commission on Accreditation of Healthcare Organizations (JCAHO). Primer on Indicator Development and Application. Oakbrook Terrace, Illinois: JCAHO; 1990
- [17] Medical Research Council. Developing and evaluating complex interventions: New guidance (2008). Im Internet: www.mrc.ac.uk/complexinterventionsguidance Stand: 12.07.2017
- [18] Altenhofen L, Brech W, Brenner G et al. Beurteilung klinischer Messgrößen des Qualitätsmanagements – Qualitätskriterien und -Indikatoren in der Gesundheitsversorgung. Konsenspapier der Bundesärztekammer, der Kassenärztlichen Bundesvereinigung und der AWMF. Z Arztl Fortbild Qualitätssich 2002; 96: 2–15
- [19] Schmitt J, Petzold T, Eberlein-Gonska M et al. Anforderungsprofil an Qualitätsindikatoren. Relevanz aktueller Entwicklungen der Outcomes Forschung für das Qualitätsmanagement. Z Evid Fortbild Qual Gesundheitswes 2013; 107: 516–522
- [20] Schmidutz F, Beirer M, Braunstein V et al. The Munich Shoulder Questionnaire (MSQ): development and validation of an effective patient-reported tool for outcome measurement and patient safety in shoulder surgery. Patient Saf Surg 2012; 6: 9
- [21] National Quality Forum. Composite performance measure evaluation guidance (08.04.2013). Im Internet: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx Stand: 12.07.2017
- [22] Shwartz M, Restuccia JD, Rosen AK. Composite measures of health care provider performance: A description of approaches. Milbank Q 2015; 93: 788–825
- [23] De Vet HC, Terwee CB, Mokkink LB et al. Measurement in medicine - a practical guide. New York: Cambridge University Press; 2011
- [24] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) - Ad hoc Kommission „Gemeinsam Klug Entscheiden“. Manual Entwicklung von Empfehlungen im Rahmen der Initiative Gemeinsam Klug Entscheiden. Version 1.1, 2016. Verfügbar: <http://www.awmf.org/medizin-versorgung/gemeinsam-klug-entscheiden.html> Stand 12.07.2017
- [25] Lutzner J, Schmitt J, Lange T et al. Wann ist der Ersatz angebracht? Vorstellung von Evidenz- und konsensbasierten Indikationskriterien zur Knie totalendoprothese. Dtsch Arztebl 2016; 113: A1983–A1985
- [26] Schmitt J, Petzold T, Deckert S et al. Recommendations for quality indicators in German S3 guidelines: A critical appraisal. Gesundheitswesen 2014; 76: 819–826
- [27] National Health Service, Institute for Innovation and Improvement. The Good Indicators Guide: Understanding how to use and choose indicators (29.01.2008). Im Internet: <http://webarchive.nationalarchives.gov.uk/20170106081009/http://www.apho.org.uk/resource/view.aspx?RID=44584> Stand: 12.07.2017
- [28] Bundesministerium für Gesundheit. Eckpunkte der Bund-Länder-AG zur Krankenhausreform 2015 (05.12.2014). Im Internet: http://www.bmg.bund.de/fileadmin/dateien/Downloads/B/Bund_Laender_Krankenhaus/Eckpunkte_Bund_Laender_Krankenhaus.pdf Stand: 12.07.2017
- [29] Reiter A, Fischer B, Kötting J et al. QUALIFY: Ein Instrument zur Bewertung von Qualitätsindikatoren. Im Internet: https://www.researchgate.net/publication/229317447_QUALIFY_Ein_Instrument_zur_Bewertung_von_Qualitätsindikatoren Stand: 12.07.2017
- [30] Ärztliches Zentrum für Qualität in der Medizin. Programm für Nationale Versorgungs-Leitlinien von BÄK, KBV und AWMF – Qualitätsindikatoren – Manual für Autoren (äzq Schriftenreihe Band 36). Berlin: Ärztliches Zentrum für Qualität in der Medizin; 2009
- [31] Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen (AQUA). Allgemeine Methoden im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach §137a SGB V. Version 4.0 (17.02.2015). Im Internet: <https://sqq.de/sqq/upload/CONTENT/Hintergrund/Methodenpapier/AQUA-Methodenpapier-4.0.pdf> Stand: 12.07.2017
- [32] Reiter A, Geraedts M, Jäckel W et al. Selection of hospital quality indicators for public disclosure in Germany. Z Evid Fortbild Qual Gesundheitswes 2011; 105: 44–48
- [33] Agency for Healthcare Research and Quality. Appendix A. Measure evaluation framework (10.2014). Im Internet: <http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/perfmeasappa.html> Stand: 12.07.2017
- [34] Brook RH, Chassin MR, Fink A et al. A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care 1986; 2: 53–63
- [35] Fitch K, Bernstein SJ, Aguilar MD et al. The RAND/UCLA. Appropriateness Method User's Manual. Santa Monica, CA: RAND; 2001
- [36] Asch SM, Kerr EA, Hamilton EG et al. Quality of care for oncologic conditions and HIV: A review of the literature and quality indicators. Santa Monica, CA: RAND; 2000
- [37] McGlynn EA, Kerr EA, Damberg CL et al. Quality of care for women: A review of selected clinical conditions and quality indicators. Santa Monica, CA: RAND; 2000
- [38] Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen (AQUA). Bericht zur Prüfung und Bewertung der Indikatoren der externen stationären Qualitätssicherung hinsichtlich ihrer Eignung für die einrichtungsbezogene öffentliche Berichterstattung (17.05.2015). Im Internet: https://www.g-ba.de/downloads/39-261-2274/2015-06-18_Freigabe-AQUA-Bericht_Ver%C3%B6ffentlichung%20QS-Indikatoren.pdf Stand: 12.07.2017
- [39] De Koning J. Development and validation of a measurement instrument for appraising indicator quality: Appraisal of Indicators through Research and Evaluation (AIRE) instrument (06.09.2007). Im Internet: www.egms.de/de/meetings/gmds2007/07gmds798.shtml Stand: 12.07.2017
- [40] De Koning J, Smulders A, Klazinga N. Appraisal of Indicators through Research and Evaluation (AIRE), version 2.0 (01.2007). Im Internet: <http://zorginzicht.garansys.nl/kennisbank/PublishingImages/Paginas/AIRE-instrument/AIRE%20Instrument%202.0.pdf> Stand: 08.01.2017

- [41] National Quality Forum. Measure evaluation criteria. Im Internet: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx Stand: 12.07.2015
- [42] Jacobs R, Goddard M, Smith PC. Public services: Are composite measures a robust reflection of performance in the public sector (16.06.2006). Im Internet: <http://www.york.ac.uk/che/pdf/rp16.pdf> Stand: 12.07.2017
- [43] Becker A, Stausberg J, Fischer B et al. Risikoadjustierung von Qualitätsindikatoren – Eine Positionsbestimmung der DKG-Experten-Gruppe „Qualitätsmessung und Risikoadjustierung“. Das Krankenhaus 2016; 11: 954–963
- [44] Hedderich J, Sachs L. Angewandte Statistik. Methodensammlung mit R. 15. Aufl. Berlin, Heidelberg: Springer Verlag; 2015
- [45] Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: The problem with small sample size. JAMA 2004; 292: 847–851
- [46] Heller G. Zur Messung und Darstellung von medizinischer Ergebnisqualität mit administrativen Routinedaten in Deutschland. Bundesgesundheitsbl 2008; 51: 1173–1182
- [47] Heller G. Qualitätssicherung mit Routinedaten – Aktueller Stand und Weiterentwicklung. In: Klauber J, Geraedts M, Friedrich J, (Hrsg.). Krankenhaus-Report 2010. Stuttgart, New York: Schattauer Verlag; 2011: 239–253
- [48] Heller G, Rochon J, Tasche K et al. Risikoadjustierung. In: Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen., (Hrsg.). Qualitätsreport 2011. Göttingen: Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen; 2012: 208–210
- [49] Heller G, Thomas T, Konheiser S et al. Qualitätsindizes. In: Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen., (Hrsg.). Qualitätsreport 2011. Göttingen: Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen; 2012: 212–213
- [50] Geschäftsstelle Qualitätssicherung Hessen (GQH). Externe Qualitätssicherung in der stationären Versorgung Geburtshilfe (Modul 16/1). Jahresauswertung 2010. Qualitätsindikatoren (18.03.2011). Im Internet: http://www.gqhnet.de/auswertungen-strukturierter-dialog/auswertungen/jahresauswertungen-2010/2010_16_1_qi Stand: 12.07.2017
- [51] Heller G, Schnell R. Hospital mortality risk adjustment using claims data. JAMA 2007; 297: 1983–1984
- [52] Nicholl J. Case-mix adjustment in non-randomised observational evaluations: The constant risk fallacy. J Epidemiol Community Health 2007; 61: 1010–1013
- [53] Kessner DM, Kalk CE, Singer J. Assessing health quality-the case for tracers. N Engl J Med 1973; 288: 189–194
- [54] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22: 719–748
- [55] Lux G, Biermann J, Dahl H et al. Risikoadjustierung und Komorbiditäten. In: Swart E, Ihle P, Gothe H, Matusiewicz D, (Hrsg.). Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. 2., vollst. überarb. Aufl. Bern: Verlag Hans Huber; 2014: 411–423
- [56] Iezzoni LI. Conceptual and practical Issues in Developing Risk Adjustment Methods. In: Iezzoni LI. Ed. Risk Adjustment for Measuring Healthcare Outcomes. 4th Edition Chicago: Health Administration Press; 2013: 195–223
- [57] Wissenschaftliches Institut der AOK. QSR – Verfahren: Indikatorenhandbuch für Leistungsbereiche ohne Berichterstattung im AOK – Krankenhausnavigator. Verfahrensjahr 2015 (10.2015). Im Internet: http://qualitaetssicherung-mit-routinedaten.de/imperia/md/qsr/methoden/indikatorenhandbuch_ohne_berichterstattung_2015.pdf Stand: 12.07.2017
- [58] Heller G, Misselwitz B. Risikoadjustierte Analyse mit Routinedaten (06.10.2008). Im Internet: <http://www.egms.de/de/meetings/dkvf2008/08dkvf058.shtml> Stand: 12.07.2017
- [59] Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd Edition New York: Wiley Series in Probability and Statistics. Wiley & Sons; 2013
- [60] Mansky T, Nimptsch U. Notwendigkeit eines ungehinderten Zugangs zu sozial- und krankheitsbezogenen Versichertendaten für die Bundesärztekammer und andere ärztliche Körperschaften sowie wissenschaftliche Fachgesellschaften zur Optimierung der ärztlichen Versorgung. Expertise im Rahmen der Förderinitiative zur Versorgungsforschung der Bundesärztekammer (2008). Im Internet: http://www.bundesaeztekammer.de/fileadmin/user_upload/downloads/Datenzugang-2.pdf Stand: 12.07.2017
- [61] Swart E, Ihle P, Gothe H et al., (Hrsg.). Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. 2., vollst. überarb. Aufl; Bern: Verlag Hans Huber; 2014
- [62] Heller G. Sekundärnutzung der Routinedaten gemäß § 21 KHEntG (und § 301 SGB V) für DEK. Bund – Land – Treffen Dekubitusprophylaxe 2012 (08.03.2012). Im Internet: <https://www.sgg.de/downloads/Termine/2012/Dekubitus/Heller.pdf> Stand: 12.07.2017
- [63] Renner D. Entwicklung eines risikoadjustierten Qualitätsindikators zur Dekubitusprophylaxe im Rahmen der externen stationären Qualitätssicherung (äzq Schriftenreihe Band 39). Berlin: Ärztliches Zentrum für Qualität in der Medizin; 2012
- [64] Heller G, Broge B, Szecsenyi J. Nutzung von Sekundärdaten in der gesetzlichen Qualitätssicherung. In: Swart E, Ihle P, Gothe H, Matusiewicz D, (Hrsg.). Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. 2., vollst. überarb. Aufl. Bern: Verlag Hans Huber; 2014: 460–473
- [65] Heller G. Die Risikoadjustierung im Leistungsbereich Dekubitusprophylaxe. Laienverständlich erklärt. Bund-Land-Konferenz Dekubitusprophylaxe 2016 (26.01.2016). Im Internet: http://www.iqtig.org/downloads/bula_dek_2016/06_Risikoadjustierung%20laienverst%C3%A4ndlich_Heller.pdf Stand: 12.07.2017
- [66] Dimick JB, Ghaferi AA, Osborne NH et al. Reliability adjustment for reporting hospital outcomes with surgery. Ann Surg 2012; 255: 703–707
- [67] Ash A, Shwartz M, Peköz EA et al. Comparing outcomes across providers. In: Iezzoni LI. Ed. Risk Adjustment for Measuring Healthcare Outcomes. 4th Edition Chicago: Health Administration Press; 2013: 335–378
- [68] Charlson ME, Pompei P, Ales KL et al. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 1987; 40: 373–383
- [69] Elixhauser A, Steiner C, Harris DR et al. Comorbidity measures for use with administrative data. Med Care 1988; 36: 8–27
- [70] Schneeweis S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol 2000; 29: 891–898
- [71] Brown C, Hofer T, Johal A et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. Qual Saf Health Care 2008; 17: 170–177
- [72] Hammer A. Zur Messung von Sicherheitskultur in deutschen Krankenhäusern. Köln: Universität zu Köln; 2012
- [73] Gershon RR, Stone PW, Zeltser M et al. Organizational climate and nurse health outcomes in the United States: a systematic re-view. Ind Health 2007; 45: 622–636
- [74] Seo DC, Torabi MR, Blair EH et al. A cross-validation of safety climate scale using confirmatory factor analytic approach. J Safety Res 2004; 35: 427–445
- [75] Zohar D. Safety climate in industrial organizations: Theoretical and applied implications. J Appl Psychol 1980; 65: 96–102

- [76] Zhang H, Wiegmann DA, Thaden TL et al. Safety culture. A concept in chaos? Proc. 46th Annu. Santa Monica: Human Factors and Ergonomics Society; 2002
- [77] Schneider B. Organizational climates: An essay. *Personnel Psychology* 1975; 28: 447–479
- [78] Sorra J, Nieva V. Hospital survey on patient safety culture. Rockville, MD: Agency for Healthcare Research and Quality; 2004. Im Internet <http://www.ahrq.gov/qual/patientsafetyculture/usergd.htm> Stand 11.01.2017
- [79] Sexton J, Helmreich R, Neilands T et al. The Safety Attitudes Questionnaire: Psychometric properties, benchmarking data, and emerging research. *BMC. Health Serv Res* 2006; 6: 44
- [80] Hammer A, Ernstmann N, Ommen O et al. Psychometric properties of the Hospital Survey of Patient Safety Culture for hospital management (HSOPS_M). *BMC. Health Serv Res* 2011; 11: 165
- [81] Hoffmann B, Domanska O, Müller V. Entwicklung des Fragebogens zum Sicherheitsklima in Hausarztpraxen (FraSiK): Transkulturelle Adaptation - ein Methodenbericht. *Z Evid Fortbild Qual Gesundheitswes* 2009; 103: 521–529
- [82] Colla JB, Bracken AC, Kinney LM et al. Measuring patient safety climate: A review of surveys. *Qual Saf Health Care* 2005; 14: 364–366
- [83] Fliin R, Burns C, Mearns K et al. Measuring safety climate in health care. *Qual Saf Health Care* 2006; 15: 109–115
- [84] Singla AK, Kitch BT, Weissman JS et al. Assessing patient safety culture: A review and synthesis of the measurement tools. *J Patient Saf* 2006; 2: 105–115
- [85] Groves PS. The relationship between safety culture and patient outcomes: Results from pilot meta-analyses. *West J Nurs Res* 2014; 36: 66–83
- [86] Hammer A, Manser T. The use of the Hospital Survey on Patient Safety Culture in Europe. In: Waterson P, (Hrsg.). *Patient Safety Culture: Theory, Methods and Application*. Farnham: Ashgate; 2014: 229–262
- [87] Morello R, Lowthian J, Barker A et al. Strategies for improving patient safety culture in hospitals: A systematic review. *Qual Saf Health Care* 2013; 22: 11–18
- [88] The Health Foundation. Evidence scan. Does improving safety culture affect patient outcomes? London: The Health Foundation; 2011
- [89] Hofinger G. Kommunikation. In: Badke-Schaub P, Hofinger G, Lauche K, (Hrsg.). *Human Factors. Psychologie sicheren Handelns in Risikobranchen*. Heidelberg: Springer Medizin Verlag; 2008: 131–150
- [90] Manser T, Perry J, Schmutz J. Verhalten ist messbar. Behavioural Marker Systeme und Kompetenzentwicklung. In: Pierre MS, Breuer G, (Hrsg.). *Simulation in der Medizin*. Heidelberg: Springer Medizin Verlag; 2013: 169–179
- [91] Lingard L, Espin S, Whyte S et al. Communication failures in the operating room: An observational classification of recurrent types and effects. *Qual Saf Health Care* 2004; 13: 330–334
- [92] Barker KN, Flynn EA, Pepper GA et al. Medication errors observed in 36 health care facilities. *Arch Intern Med* 2002; 162: 1897–1903
- [93] Bortz J, Döring N. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer Medizin Verlag; 2007
- [94] Murff HJ, Patel VL, Hripcsak G et al. Detecting adverse events for patient safety research: A review of current methodologies. *J Biomed Inform* 2003; 36: 131–143
- [95] Flynn EA, Barker KN, Pepper GA et al. Validity of methods for detecting medication errors in 36 hospitals and skilled-nursing facilities. *Am J Health Syst Pharm* 2002; 59: 436–446
- [96] McDermott RE, Beauregard MR, Mikulak R. *Basics of FMEA*. 2nd Edition New York: Productivity Press; 2009
- [97] Von Eiff W., (Hrsg.). *Risikomanagement: Kosten-Nutzen-basierte Entscheidungen im Krankenhaus (Schriftenreihe Gesundheitswirtschaft Band 2)*. Wegscheid: WIKOM-Verlag; 2007
- [98] Paparella S. Failure mode and effects analysis: A useful tool for risk identification and injury prevention. *J Emerg Nurs* 2007; 33: 367–371
- [99] Nagpal K, Vats A, Ahmed K et al. A systematic quantitative assessment of risks associated with poor communication in surgical care. *Arch Surg* 2010; 145: 582–588
- [100] Reason J. Human error: Models and management. *BMJ* 2000; 320: 768–770
- [101] Huckels-Baumgart S, Manser T. Identifying medication error chains from critical incident reports - a new analytic approach. *J Clin Pharmacol* 2014; 54: 1188–1197
- [102] World Health Organization. Conceptual framework for the International Classification for Patient Safety. Version 1.1 (01.2009). Im Internet: http://www.who.int/patientsafety/taxonomy/icps_full_report.pdf Stand: 12.07.2017
- [103] Seyfarth-Metzger I, Vogel S, Krabbe-Berndt A. Neue Herausforderungen an das Qualitätsmanagement: Wirtschaftlichkeit und Patientensicherheit. *Das Krankenhaus* 2005; 9: 757–764
- [104] Sens B. DIN EN 15224:2012 – eine neue Zertifizierungsnorm für Qualitätsmanagementsysteme in Gesundheitseinrichtungen. *Interdisziplinäre Beiträge zum Krankenhaus-Management* (27.05.2013). Im Internet: <http://www.clinotel-journal.de/article-id-011.html> Stand 12.07.2017
- [105] Kohn LT, Corrigan J, Donaldson MS. *To err is human: Building a safer health system*. Washington, DC: National Academies Press; 2000
- [106] National Institute for Clinical Excellence. *Principles for Best Practice in Clinical Audit*. Abingdon: Radcliffe Publishing 2002;
- [107] Johnston G, Crombie I, Alder E et al. Reviewing audit: Barriers and facilitating factors for effective clinical audit. *Qual Health Care* 2000; 9: 23–36
- [108] McKay J, Bowie P, Lough M. Variations in the ability of general medical practitioners to apply two methods of clinical audit: A five-year study of assessment by peer review. *J Eval Clini Pract* 2006; 12: 622–629
- [109] Shaw CD, Costain DW. Guidelines for medical audit: Seven principles. *BMJ* 1989; 299: 498–499
- [110] Eberlein-Gonska M, Budic A, Winklmaier C et al. 5 Jahre Peer Review Verfahren bei der Initiative Qualitätsmedizin (IQM). *Z Evid Fortbild Qual Gesundheitswes* 2015; 109: 650–659
- [111] Chop I, Eberlein-Gonska M. Übersichtsartikel zum Peer Review Verfahren und seine Einordnung in der Medizin. *Z Evid Fortbild Qual Gesundheitswes* 2012; 106: 547–552
- [112] Van den Hombergh P, Grol R, van den Hoogen H et al. Practice visits as a tool in quality improvement: mutual visits and feedback by peers compared with visits and feedback by non-physician observers. *Qual Health Care* 1999; 8: 161–166
- [113] Campbell S, Braspenning J, Hutchinson A et al. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002; 11: 358–364
- [114] Zhan C, Miller M. Administrative data based patient safety research: A critical review. *Qual Saf Health Care*. 2003; 12: ii58–ii63
- [115] Quan H, Eastwood C, Cunningham CT et al. investigators I. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ open* 2013; 3: e003716
- [116] Romano PS, Mull HJ, Rivard PE et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health services research* 2009; 44: 182–204
- [117] Humphries KH, Rankin JM, Carere RG et al. Co-morbidity data in outcomes research Are clinical data derived from administrative databases a reliable alternative to chart review? *J Clin Epidemiol* 2000; 53: 343–349
- [118] Maass C, Kuske S, Lessing C et al. Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. *Int J Qual Health Care* 2015; 27: 305–313

- [119] Institute for Healthcare Improvement (IHI). Aktives Messinstrument der Patientensicherheit – das IHI Global Trigger Tool. Interreg 4A Projekt: Stärkung der Patientensicherheit zwischen Dänemark und Deutschland. Cambridge, MA: :: IHI; 2009
- [120] Schildmeijer K, Nilsson L, Årestedt K et al. Assessment of adverse events in medical care: Lack of consistency between experienced teams using the Global Trigger Tool. *BMJ Qual Saf* 2012; 21: 307–314
- [121] Von Plessen C, Kodal AM, Anhøj J. Experiences with global trigger tool reviews in five Danish hospitals: An implementation study. *BMJ Open* 2012; 2: pii:e001324
- [122] Sharek PJ, Parry G, Goldmann D et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res* 2011; 46: 654–678
- [123] Naessens JM, O'Byrne TJ, Johnson MG et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care* 2010; 22: 266–274
- [124] Resar R, Rozich J, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003; 12: ii39–ii45
- [125] De Wet C, Bowie P. The preliminary development and testing of a global trigger tool to detect error and patient harm in primary-care records. *Postgrad Med J* 2009; 85: 176–180
- [126] Wong BM, Dyal S, Etchells EE et al. Application of a trigger tool in near real time to inform quality improvement activities: A prospective study in a general medicine ward. *BMJ Qual Saf* 2015; 24: 272–281
- [127] Stavropoulou C, Doherty C, Tosey P. (2015) How Effective Are Incident-Reporting Systems for Improving Patient Safety? A Systematic Literature Review. *Milbank Quarterly* 2015; 93: 826–866
- [128] Bommersholdt MB. Key findings and recommendations on Reporting and learning systems for patient safety incidents across Europe (2014). http://ec.europa.eu/health/sites/health/files/patient_safety/docs/guidelines_psqcwg_reporting_learningsystems_en.pdf Stand 12.07.2017
- [129] Pham JC, Girard T, Pronovost PJ. What to do with healthcare incident reporting systems. *J Public Health Res* 2013, doi:10.4081/jphr.2013.e27
- [130] Tamuz M, Thomas E, Franchois K. Defining and classifying medical error: Lessons for patient safety reporting systems. *Qual Saf Health Care* 2004; 13: 13–20
- [131] Meissner Y, Richter A, Manger B et al. Serious adverse events and the risk of stroke in patients with rheumatoid arthritis: results from the German RABBIT cohort. *Ann Rheum Dis*. 2017, doi:10.1136/annrheumdis-2017-211209
- [132] Naessens JM, Campbell CR, Huddleston JM et al. A Validity of of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009; 21: 301–307
- [133] Beckmann U, Bohringer C, Carless R et al. Evaluation of two methods for quality improvement in intensive care: Facilitated incident monitoring and retrospective medical chart review. *Crit Care Med* 2003; 31: 1006–1011
- [134] Lisby M, Nielsen LP, Mainz J. Errors in the medication process: Frequency, type, and potential clinical consequences. *Int J Qual Health Care* 2005; 17: 15–22
- [135] Ogrinc G, Davies L, Goodman D et al. SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): Revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* 2015; 24: 466–473
- [136] Eberlein-Gonska M, Petzold T, Helaß G et al. Häufigkeiten und Determinanten von Dekubitalulzera in der stationären Versorgung. *Dtsch Arztebl* 2013; 110: 550–556
- [137] Shojania KG. Conventional evaluations of improvement interventions: More trials or just more tribulations? *BMJ Qual Saf* 2013; 22: 881–884
- [138] Campbell M, Fitzpatrick R, Haines A et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000; 321: 694–696
- [139] Greenhalgh T, Robert G, Macfarlane F et al. Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Q* 2004; 82: 581–629
- [140] Medical Research Council. 2008; Developing and evaluating complex interventions: new guidance 2008. www.mrc.ac.uk/complexinterventionsguidance Stand: 12.07.2017
- [141] Damschroder L, Aron D, Keith R et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* 2009; 4: 50–65
- [142] Brown C, Hofer T, Johal A et al. An epistemology of patient safety research: A framework for study design and interpretation. Part 1. Conceptualising and developing interventions. *Qual Saf Health Care* 2008; 17: 158–162
- [143] Brown C, Hofer T, Johal A et al. An epistemology of patient safety research: A framework für study design and interpretation. Part 2. Study Design. *Qual. Saf. Health Care* 2008; 17: 163–169
- [144] Jain R, Kralovic SM, Evans ME et al. Veterans affairs initiative to prevent methicillin-resistant staphylococcus aureus infections. *N Engl J Med* 2011; 364: 1419–1430
- [145] Guise JM, Chang C, Viswanathan M et al. Systematic Reviews of Complex Multicomponent Health Care Interventions. Rockville (MD): Agency for Healthcare Research and Quality (US); (2014); Im Internet: <https://www.ncbi.nlm.nih.gov/books/NBK194846/> Stand 09.07.2017
- [146] Plsek P. Redesigning Health Care with Insights from the Science of Complex Adaptive Sys-tems. In: Institute of Medicine., (Hrsg.). Crossing the Quality Chasm: A New Health System for 21th Century. Washington, DC: National Academies Press; 2001: 309–322
- [147] Richardson KA. Managing complex organizations: Complexity thinking and the science and art of management. *E:CO* 2008; 10: 13–26
- [148] Cohn S, Clinch M, Bunn C et al. Entangled complexity: Why complex interventions are just not complicated enough. *J Health Serv Res Policy* 2013; 18: 40–43
- [149] Hignett S, Jones EL, Miller D et al. Human factors and ergonomics and quality improvement science: Integrating approaches for safety in healthcare. *BMJ Qual Saf* 2015; 24: 250–254
- [150] Leviton L. Reconciling complexity and classification in quality improvement research. *BMJ Qual Saf* 2011; 20: i28–i29
- [151] Han YY, Carcillo JA, Venkataraman ST et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005; 116: 1506–1512
- [152] Pawson R, Tilley N. Realist Evaluation (2004). Im Internet: http://www.communitymatters.com.au/RE_chapter.pdf Stand: 12.07.2017
- [153] Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? Alternative set of guiding principles. *PLoS Med* 2010; 7: e1000360
- [154] Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. *JAMA* 2002; 288: 501–507
- [155] Wachter RM. Patient safety at ten: Unmistakable progress, troubling gaps. *Health Aff* 2010; 29: 165–173
- [156] Shojania KG, Duncan BW, McDonald KM et al. Safe but sound. Patient safety meets evidence-based medicine. *JAMA* 2002; 288: 508–513
- [157] Alexander JA, Hearld LR. Review: What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev* 2009; 66: 235–271
- [158] Freidson E. Professionalism. The Third Logic. Chicago: The University of Chicago Press; 2001

- [159] Corrigan J, McNeill D. Building organizational capacity: A cornerstone of health system reform. *Health Aff* 2009; 28: w205–w215
- [160] Gurieva T, Bootsma MCJ, Bonten MJM. Successful veterans affairs initiative to prevent methicillin-resistant *Staphylococcus aureus* infections revisited. *Clin Infect Dis* 2012; 54: 1618–1620
- [161] Fätkenheuer G. Screening auf MRSA – unnützlich, teuer und potenziell gefährlich. *Dtsch Med Wochenschr* 2015; 140: 1449–1450
- [162] Pawson R, Tilley N. Realist Evaluation (2004). Im Internet: http://www.communitymatters.com.au/RE_chapter.pdf Stand: 12.07.2017
- [163] Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. *N Engl J Med* 2007; 357: 608–613
- [164] Dietz AS, Pronovost PJ, Benson KN et al. A systematic review of behavioral marker systems in healthcare: What do we know about their attributes, validity and application. *BMJ Qual Saf* 2014; 23: 1031–1039
- [165] Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013; 346: f167
- [166] Hwang TJ, Bourgeois FT, Seeger JD. Drug safety in the digital age. *N Engl J Med* 2014; 370: 2560–2562
- [167] Karbach U, Stamer M, Holmberg C et al. Qualitative Studien in der Versorgungsforschung – Diskussionspapier, Teil 2: Stand qualitativer Versorgungsforschung in Deutschland – ein exemplarischer Überblick. *Gesundheitswesen* 2012; 74: 516–525
- [168] Meyer T, Karbach U, Holmberg C et al. Qualitative Studien in der Versorgungsforschung – Diskussionspapier, Teil 1: Gegenstandsbestimmung. *Gesundheitswesen* 2012; 74: 510–515
- [169] Taylor-Adams S, Vincent C. Systems analysis of clinical incidents. The London Protocol (2004). Im Internet: https://www1.imperial.ac.uk/resources/C85B6574-7E28-4BE6-BE61-E94C3F6243CE/londonprotocol_e.pdf Stand: 12.07.2016
- [170] Schrappe M. Patientensicherheitsforschung. In: Pfaff H, Glaeske G, Neugebauer E, Schrappe M, (Hrsg.). *Lehrbuch Versorgungsforschung*. Stuttgart: Schattauer Verlag; 2011: 246–252
- [171] Valentin A, Capuzzo M, Guidet B et al. Errors in administration of parenteral drugs in intensive care units: Multinational prospective study. *BMJ* 2009; 308: b814
- [172] Haynes AB, Weiser TG, Berry WR et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009; 360: 491–499
- [173] Rhodes P, Giles SJ, Cook GA et al. Assessment of the implementation of national patient safety alert to reduce wrong site surgery. *Qual Saf Health Care* 2008; 17: 409–415
- [174] Krinsky WS, Mroz IB, McIlwaine JK et al. A model for increasing patient safety in the intensive care unit: Increasing the implementation rates of proven safety measures. *Qual Saf Health Care* 2009; 18: 74–80
- [175] Pronovost P, Needham D, Berenholtz S et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006; 355: 2725–2732
- [176] Lindenauer PK, Remus D, Roman S et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007; 356: 486–496
- [177] Bode LGM, Kluytmans JAJW, Wertheim HFL et al. Preventing surgical-site infections in nasal carriers of *Staphylococcus aureus*. *N Engl J Med* 2010; 362: 9–17
- [178] Marrie TJ, Lau CY, Wheeler SL et al. A controlled trial of critical pathway for treatment of community-acquired pneumonia. *JAMA* 2000; 283: 749–755
- [179] Doig GS, Simpson F, Finfer S et al. Effect of evidence-based feeding guidelines on mortality of critically ill adults. *JAMA* 2008; 300: 2731–2741
- [180] Hillman K, Chen J, Cretikos M et al. Introduction of the Medical Emergency Team (MET) system: A cluster-randomised controlled trial. *Lancet* 2005; 365: 2091–2097
- [181] Curry LA, Krumholz HM, O’Cathain A et al. Mixed methods in biomedical and health services research. *Circ Cardiovasc Qual Outcomes* 2013; 6: 119–123
- [182] Portela MC, Pronovost PJ, Woodcock T et al. How to study improvement interventions: A brief overview of study types. *BMJ Qual Saf* 2015; 24: 326–336
- [183] Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: Methodological study. *BMJ* 2009; 339: b3496
- [184] Marwick CA, Guthrie B, Pringle JEC et al. A multifaceted intervention to improve sepsis management in general hospital wards with evaluation using segmented regression of interrupted time series. *BMJ Qual Saf* 2014; 23: e2
- [185] Swart E, Bitzer EM, Gothe H et al. A consensus German reporting standard for secondary data analyses, Version 2 (STROSA-Standardisierte BerichtsROUTine für Sekundärdatenanalysen). *Gesundheitswesen* 2016; 78: e145–e160
- [186] Guise JM, Chang C, Viswanathan M et al. Agency for Healthcare Research and Quality Evidence-based Practice Center methods for systematically reviewing complex multicomponent health care interventions. *J Clin Epidemiol* 2014; 67: 1181–1191
- [187] Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ* 2005; 331: 1064–1065
- [188] Gerhardus A, Becher H, Groenewegen P et al. Applying for reviewing and funding public health research in Germany and beyond. *Health Res Policy Syst* 2016; 14: 43
- [189] Aiken LH, Sloane DM, Bruyneel L et al. Nurse staffing and education and hospital mortality in nine European countries: A retrospective observational study. *Lancet* 2014; 383: 1824–1830