

Qualitätsmanagement im Gesundheitswesen: der weite Weg der Evidenzbasierung in der Versorgungsforschung

Norbert Schmacke

Prof. Dr. med. Norbert Schmacke ist Hochschullehrer für Gesundheitswissenschaften an der Universität Bremen und leitet die Arbeits- und Koordinierungsstelle Gesundheitsversorgungsforschung und ist unparteiisches Mitglied im Gemeinsamen Bundesausschusses

Ein soeben erschienener Sammelband präsentiert eine Bilanz von Projekten zur Verbesserung der Versorgung aus Sicht eines großen Verbandes der gesetzlichen Krankenkassen. Der Autor des nachfolgenden Beitrags war von 1999 bis 2003 Leiter des Stabsbereichs Medizin beim AOK-Bundesverband und während dieser Zeit u.a. im Koordinierungsausschuss tätig, welcher die ersten DMP auf den Weg brachte. Er ist seit 2004 unparteiisches Mitglied des G-BA, welcher in seiner Verfahrensordnung den Prinzipien der evidenzbasierten Medizin verpflichtet ist und setzt sich in seinem Diskussionsbeitrag überwiegend kritisch mit den Inhalten dieses Sammelbandes auseinander.

Seit etwa drei Jahren findet eine lebhaft und auch kontroverse Debatte innerhalb der universitären Forschung zu Methoden der Versorgungsforschung statt, die bezüglich quantitativer Methoden ihren Niederschlag vor allem in einem Memorandum des Deutschen Netzwerks Versorgungsforschung e.V. (Neugebauer et al 2010) gefunden hat. Zu den qualitativen Methoden sind für 2012 weitere Publikationen geplant. Die im vorliegenden Memorandum vertretene Haltung zu den angemessenen Methoden wird vom Netzwerk evidenzbasierte Medizin kritisiert (http://www.ebm-netzwerk.de/netzwerkarbeit/images/Stellungnahme-Memorandum_20100826.pdf).

Der Streit rankt sich nicht zuletzt um die Debatte, ob die Versorgungsforschung eigene oder die Anpassung vorhandener Forschungsmethoden benötigt, weil sich vor allem der Alltag der Versorgung in den klassisch einschlägigen RCT nicht abbilde. Jüngst hat die Deutsche Forschungsgemeinschaft eine Stellungnahme zur Versorgungsforschung vorgelegt, welche auf die Einhaltung methodischer

Standards dringt (Raspe et al 2010). Auch die gesetzlichen Krankenkassen haben begonnen, stärker Einfluss auf die Entwicklung von Versorgungsforschung zu nehmen und sich unmittelbar an der Entwicklung von Forschungsfragen und Studiendesigns beteiligt. Ein in dieser Zeitschrift schon kurz besprochener Sammelband (Knieps in Heft 6/2010 zu Klusen et al 2011) ist im Kontext dieser Methodendebatte besonders interessant und geeignet, einige Grundfragen der Versorgungsforschung am Beispiel der hier präsentierten Projekte zur Verbesserung der Versorgung noch einmal zu diskutieren

Thomas Ballast macht in der Einleitung des Bandes klar, dass methodisch gut gemachte Evaluationen für Optimierungsprojekte in der GKV einfach deshalb erforderlich sind, weil Erfolge nicht an der Aussage gemessen werden können, man sei sehr zufrieden. Er spricht vom Niveau »oral history«. Damit steht sofort die Frage im Raum: welche Forschungsdesigns können und welche müssen erwartet werden, wenn – aufsetzend auf einem ja hohen Ni-

veau der Versorgung, wie Ballast betont – Innovationen geltend gemacht werden. Auf diese Frage geben die Einzelbeiträge des Bandes leider nur in Ansätzen Antworten, und warum das ein Problem ist, soll exemplarisch erläutert werden.

Mahltig und Voermans machen auf einen zentralen Punkt aufmerksam: in der betriebswirtschaftlichen Perspektive erscheint Qualität als ein Instrument zur Verbesserung der Wettbewerbssituation. Die Ziele werden häufig aus Plausibilitätsüberlegungen heraus entwickelt. Schnell ist man dann beim Qualitätswettbewerb und »vergisst« unterwegs den Zusammenhang zu patientenrelevanten Messpunkten. Aus der immer neuen Darlegung der Schnittstellenproblematik erwuchs so – gestützt auf das Sachverständigenratsgutachten von 1994 – das Konzept der integrierten Versorgung. Die Politik begann, die Rahmenbedingungen für die Sektorenüberwindung und selektives Kontrahieren zu schaffen: mit frischem Geld begann der IV-Motor zu laufen. Kölner Versorgungsforscher haben auf das Problem hingewiesen, dass die IV-Verträge nicht mit Evaluationsauflagen verbunden sind, wozu die außerordentlich ernüchternden Erfahrungen mit der Qualität der Evaluation von Modellvorhaben nach §§ 63ff. eigentlich Veranlassung gegeben hätten (Wiethege et al 2010). Mahltig und Voermans wagen sich in diesem Sammelband am weitesten vor: wissenschaftlichen Grundsätzen genügende Evaluationskonzepte fehlen weiterhin, auch wenn vom SVR später angemahnt. Stattdessen herrscht »impressionistische Evidenz«.

In dem Artikel von **Ernst Thiel** wird der Stellenwert der Prozessqualität herausgehoben. Interdisziplinäre Fallbesprechungen, fraglos ein wichtiges Instrument, sind nach Auffassung des Autors im Krankenhaus selbstverständlich, hoffentlich bald auch über das Vehikel MVZ im ambulanten Sektor. Nun spricht die internationale Qualitätsforschung ausgerechnet an diesem Punkt überhaupt nicht dafür, dass Interdisziplinarität zur Medizinkultur gehört. Hier besteht dringender Forschungsbedarf, insbesondere wenn auch noch die Erwartung geweckt wird, dass die verschiedenen Fachberufe besser als früher miteinander kom-

munizieren sollen, was bei den Stichworten Delegation und Substitution ja eigentlich nahe liegt.

An zwei im Sammelband abgearbeiteten prominenten Handlungsfeldern lässt sich zeigen, dass Versorgungsforschung im Kontext der Leistungsversprechungen der Sozialversicherung immer auch damit zu tun hat, dass Politik ‚Steine ins Wasser wirft‘, um dann abzuwarten, wie das System G-BA und Vertragspartner darauf reagiert. Dabei fußen Reformkonzepte eben nur selten auf solider Versorgungsforschung, sondern verweisen auf gewaltige Forschungsdefizite, die dann im Rahmen der Sozialversicherung nur ausgesprochen schwer zu kompensieren sind.

Erstes Beispiel: Wie weit der Weg zur sektorenübergreifenden Qualitätssicherung im Sinne des SGB V ist, wird aus dem Artikel von **Fahlbusch und Nobmann** deutlich. Die Regelungsdichte des Gesetzes steht – diplomatisch formuliert – in gewissem Kontrast zur Datenbasis, mit der über Erfolg oder Misserfolg der verschiedenen Ansätze der Überwindung der Sektorengrenzen diskutiert werden könnte. Immerhin steht dem Gemeinsamen Bundesausschuss jetzt mit dem AQUA-Institut ein wissenschaftlicher Berater zur Seite, der das Fundament für diese neue Stufe der QS legen soll.

Zweites Beispiel: Es schwingt in der Politik ein gewisser Stolz mit, dass durch die SGB-V-Reformen das monolithische System der GKV aufgebrochen worden ist. Man möchte so gern den Optimismus teilen, der sich durch die Erfahrungsberichte des Sammelbandes zieht. Dies betrifft auch die jüngste Entdeckung »Pay for Performance«, die **Klusen, Meusch und Piesker** unter der Überschrift »Weder Königs- noch Holzweg« ankündigen. Es ist ausdrücklich der Einschätzung zuzustimmen, dass P4P noch kein fertiges Konzept ist. Die umfangreiche Literatur zu diesem Ansatz der Qualitätssteigerung wird nur cursorisch abgehandelt, so dass die Gefahr besteht, dass nicht nur der Gesetzgeber sondern auch die Kassen die Probleme des P4P-Ansatzes grandios unterschätzen. Im Band berichten Klusen et al über Public-Repor-

ting, dem P4P verwandt und weisen auf eine in einen Klinikführer integrierte Patientenbefragung hin, die sich – so steht zu vermuten – am Vorbild des Picker-Instruments orientiert und in regelmäßigen Abständen wiederholt werden soll, wie in Erfahrung gebracht werden konnte. Das ist fraglos innovativ. Wenn freilich der Eindruck erweckt wird, dies sei bereits der Weg zur Qualitätsverbesserung, so wird die – leider außerordentlich ernüchternde – Studienlage zu Public Reporting nicht berücksichtigt. Fung et al (2008) haben 45 verwertbare Studien gefunden, die keine belastbaren Belege dafür ergaben, dass Public Reporting Patientenströme beeinflusst, Versorgungseffizienz verbessert oder Patientensicherheit vergrößert. Die positive Deutung der Studienlage lautet: möglicherweise ist Public Reporting ein Stimulus zum internen Qualitätsmanagement im stationären Sektor. Die Auto-

Dabei fußen Reformkonzepte eben nur selten auf solider Versorgungsforschung, sondern verweisen auf gewaltige Forschungsdefizite.

rengruppe fordert methodisch gut gemachte Studien, um säkulare Trends und Verzerrungsfaktoren so gut es geht auszuschließen.

Damit landet man bei einem Kernthema der Zukunft: besteht Bereitschaft, sich über die erforderliche Studienqualität in der Versorgungsforschung zu unterhalten und im Lichte vorliegender, oft außerordentlich mangelbehafteter Forschung die Weichen für Pionierstudien zu stellen, die mehr Licht ins Dunkel bringen? Wie wichtig dies ist, verdeutlicht die von **Klusen et al** zitierte Begleitforschung zu einem Schmerztherapie-Vertrag, der Aufgreifkriterien für Schmerzkrankte definiert, eine intensive stationäre Phase interdisziplinärer Diagnostik und Therapie vorsieht und in eine multimodale ambulante Versorgungsphase überleitet. Die außerordentlich aufwändige Begleitforschung hat den Autoren zufolge ein Traumergebnis erzielt: In der Interventionsgruppe wurden die Kosten im ambulanten Bereich um 31,5 Prozent, im sta-

tionären Sektor um 21,5 Prozent und im rehabilitativen Bereich um 50,6 Prozent gesenkt. Wenn man der Forschergruppe von Göbel et al (2009) folgt, dann ist zudem in der Interventionsgruppe der Kopfschmerzpatienten die Zahl der Arbeitsunfähigkeitstage nach Aufnahme der Behandlung stark zurückgegangen, die Einkommenssituation hat sich gegenüber der Kontrollgruppe deutlich verbessert und die mittels SF 36 gemessene Lebensqualität hat in allen Dimensionen deutlich zugenommen. Wenn man sich die Herstellung der Kontrollgruppe anschaut, dann imponiert zunächst, wie viele Dimensionen in die Zuordnung von »virtuellen Zwillingen« zu den Patienten der Interventionsgruppe eingeflossen sind. Das soweit gewählte Matched-Pair-Verfahren, das als second-best-design prinzipiell wählbar ist, wird allerdings noch einmal wie folgt relativiert: »Die virtuellen Patienten werden – zumeist über Durchschnittsbildung – aus den entsprechenden Twins für die jeweiligen Patienten der Schmerzklinik XY generiert« (Roth und Rüschemann 2002, 49) Es stellt sich

Klusen et al. schreiben selber, dass P4P noch kein fertiges Konzept sei und dass »exorbitantes Know-how« erforderlich sei, um dieses Vergütungsinstrument zu evaluieren.

die Frage, warum bei dieser hoch relevanten Studie die Chance vertan wurde, eine ganz klassische kontrollierte Studie mit individueller Randomisierung der Patienten durchzuführen, welche tatsächlich eine Aussage ermöglicht hätte, ob ein kausaler Zusammenhang zwischen dieser vorbildlich erscheinenden Versorgungsform und den gemessenen Endpunkten besteht. Mit Blick auf den Charakter »Schmerztherapiestudie« ist weiter anzumerken, dass hier zwei komplexe Interventionen in einem Modell auf ihren Nutzen hin getestet werden sollten: zum einen natürlich das neue Versorgungsmodell einer integrierten Schmerzbehandlung, zum anderen aber gleichzeitig der damit für die Professionellen im Interventionsarm verbundenen finanziellen Anreize. Damit ist diese

Studie »hyperkomplex«, und die extrem positiven Ergebnisse sind mit großer Skepsis zu betrachten. **Klusen et al** schreiben am Ende des auf P4P abzielenden Artikels selber, dass P4P noch kein fertiges Konzept sei und dass »exorbitantes Know-how« erforderlich sei, um dieses Vergütungsinstrument zu evaluieren. Dies ist in der Tat der publizierten Literatur zu P4P zu entnehmen (zuletzt van Herck et al 2010, davor z.B. Petersen et al 2006). Damit spricht alles dafür, die Tür zu P4P erst dann aufzumachen, wenn es in Deutschland methodisch gut gemachte Pilotprojekte mit angemessener Evaluation gibt. Und es spricht gerade wegen der Fallstricke, die mit diesem Anreizsystem verbunden sind (so Mitnahmeeffekte, so auch Versanden kurzfristiger Qualitätseffekte) alles dafür, die Forschung im Rahmen solcher Pilotmodelle auszuschreiben. Dabei ist auch zu klären, ob nicht für zentrale Fragestellungen kassenartenübergreifende Modelle unabdingbar sind. Die P4P-Debatte wirft aber auch die ganz prinzipielle Frage auf, wie hochwertige Grundlagenforschung und anwendungsorientierte Versorgungsforschung besser ineinander greifen können als bisher: sie müssen es, weil hochwertige Versorgungsforschung nur auf methodisch solider Grundlagenforschung aufsetzen kann. Sie kann anders gesagt nicht im Bypass methodische Grundprobleme lösen.

Auch bei auf den ersten Blick weniger ambitionierten Modellvorhaben zu P4P wie einem Praxinetz (Beitrag von **Maaß und Möws**), wird man gespannt sein auf die noch nicht vorliegenden Evaluationsergebnisse. Immerhin werden als Ziele angegeben: Vermeidung von Fehl- oder Nichtdiagnosen, Vermeidung unnötiger Facharztconsultationen, Vermeidung von Doppeluntersuchungen, Vermeidung oder Verkürzung stationärer Aufenthalte und Abbau von Über- oder Unterversorgung bei Diagnostik und Medikation. Es wird nicht mitgeteilt, wie die Überlegenheit dieses mittels selektiven Kontrahierens organisierten Modells empirisch ermittelt werden soll. In einem weiteren Artikel (**Aalst, Brose, Schöpker, Schulte-Florian**) werden grundsätzliche Kriterien für qualitätsso-

rientierte Verträge zur integrierten Versorgung genannt. Wörtlich heißt es: »Bei jedem Vertrag werden zunächst die Behandlungsziele definiert. In einem nächsten Schritt ist die zur Zielerreichung erforderliche Struktur-, Prozess- und/oder Ergebnisqualität zu bestimmen und anhand verschiedener messbarer Kriterien zu operationalisieren« (149). Beispielhaft werden dann Befragungen nach Krankenhausaufenthalt, Skalen wie die Hamilton-Depressionskala oder AU-Tage als Messinstrumente erwähnt. In jedem Fall stellt sich die Frage, wie die nötige Kontrollgruppe(n) definiert werden sollen. Für ein derartiges Projekt gibt die Homepage der Krankenkasse (<https://www.tk.de/tk/hamburg/kinder-jugendliche-und-familie/babyfocus/11610?selectedPage=1>)

Auskünfte über den erwarteten Nutzen für schwangere Frauen, wie z.B.

- Sie werden von erfahrenen Hebammen und Gynäkologen betreut. Diese arbeiten Hand-in-Hand und erfüllen hohe Qualitätsstandards.
- Das geburtshilfliche Team im Krankenhaus bietet eine umfassende und persönliche Beratung.
- Falls Beschwerden oder Komplikationen auftreten, können Sie mit einer Überweisung von Ihrer Hebamme oder dem Frauenarzt die Gynäkologen der Partnerkliniken in Anspruch nehmen.
- Sie können kostenlos Angebote,
 - wie Geburtsvorbereitungskurse mit Ihrem Partner
 - Stillberatung
 - und Rückbildungsgymnastik nutzen.
- Nach der Geburt wird unter anderem durch ein erweitertes Neugeborenen-Screening und einen Hörtest überprüft, ob der Säugling gesund ist.

Es ist nur schwer vorstellbar, wie in diesem Fall das innovative Element in der Intervention gegenüber der Regelversorgung definiert und wie die intendierte bessere Vernetzung der Akteure dabei organisiert werden soll: ohne Klärung derartiger grundsätzlicher Fragen ist ein Ausrollen solcher Versorgungsmodelle aber nicht vorstellbar – deshalb hier dieser kleinteilige Blick auf den Ansatz.

Der Beitrag von **Hennicke, Meyer und Suchant** über »integra« beleuchtet ein weiteres mit hohen Erwartungen besetz-

tes Feld: nämlich Verträge über stationersetzende Operationen. Hierbei ging es darum, »diejenigen stationären Behandlungsfälle zu ersetzen, für die bezogen auf die Indikation und Fallschwere keine regulären Krankenhausbehandlungen notwendig, ambulante Operationen aber nicht möglich sind« (182). Das Spektrum der Operationen umfasste u.a. Kniearthroscopien, Stammvarikosis und Septumoperationen (HNO). Die Kalkulation der Kosten für die integra-Patienten erbrachte im stationären Bereich 68 Prozent der DRG-Vergütung und korrespondierte insoweit mit dem im Projekt vereinbarten 30-prozentigen Abschlag auf die DRG-Fallpauschalen. Die Gesundheitsökonominnen errechneten auch unter Berücksichtigung der DRG-Abschläge bei kurzen Verweildauern eine Kostenersparnis von 83 Euro pro Fall, denen 90 Euro Mehrausgaben für ambulante Arztkosten entsprachen. Weiter errechneten sie eine Reduktion der Arzneimittelkosten von integra-Patienten von 720 Euro, eine um gut 12 Tage reduzierte Arbeitsunfähigkeitsdauer, aus der allein ein volkswirtschaftlicher Gewinn von 2,73 Millionen Euro errechnet wurde. Zu den Grenzen diesen Ergebnisse finden sich im Abschlussbericht zwar einige wichtige Informationen, die nach Ansicht der Evaluatoren aber das Ergebnis nicht infrage stellen: so eine hohe Rate nicht-erfasster Patienten (v.a. im ambulanten, aber auch im stationären Sektor), so die fehlende Erfassung von Komorbiditäten. Insgesamt fehlen im Abschlussbericht vor allem ausreichend transparente Einschluss- und Ausschlusskriterien für die Interventionsgruppe.

Durch die letzten SGB-V-Reformen ist ohne Frage Bewegung in die Versorgungslandschaft gekommen und mit einiger Verzögerung sind jetzt von großen Kassenverbänden, nicht nur den hier genannten, relevante Modellvorhaben auf den Weg gebracht worden. Es ist Zeit, so die Schlussfolgerung des Autors, noch einmal grundlegend über methodische Anforderungen an die Evaluation neuer Versorgungsformen nachzudenken, die Überlegenheit gegenüber der Regelversorgung reklamieren. Das beginnt damit, zu akzeptieren, dass derartige Modelle als komplexe Interventionen zu bezeichnen sind, die prinzipiell genau so gut wie »Schlichtinterventionen« vom Typ Arzneimittelstudie

mittels randomisierter, kontrollierter Studien untersucht werden können, wobei häufig die Cluster-Randomisierung einzusetzen ist und auch die Option präferenzbasierte RCT zum Tragen kommen kann (Campbell et al 2000, Campbell et al 2007, Eccles et al 2003). Überlegenheit von Versorgungsformen wird mit der geringsten Irrtumswahrscheinlichkeit durch RCT belegt, das ist in der internationalen Versorgungsforschung unstrittig (s. z.B. Steinwachs und Hughes 2008). Auch in Bereichen, die scheinbar besonders heikel für die Durchführung von RCT sind, wie die Versorgung psychisch Kranker, sind gut gemachte RCT allemal gegenüber noch so gut gemachten Vorher-Nachher-Studien und/oder dem damit verbundenen »Herausrechnen« von Verzerrungsfaktoren durch die Verwendung nicht-randomisierter Kontrollgruppen zu bevorzugen (Gilbody und Whitty 2002).

Wenn immer wieder Gründe angeführt werden, warum RCT nicht durchführbar sind, so bleibt doch wahr: second-best-designs leben immer mit dem Problem nicht-erkannter Störfaktoren. Laien meinen gelegentlich, mit statistischen Kontrollverfahren ließen sich derartige Störfaktoren heute elegant beseitigen: dies ist ein schwerer Irrtum (s. z.B. Christenfeld et al 2004, Wunsch et al 2006). Es ist auch eine Fehlannahme, dass RCTs prinzipiell teurer und aufwändiger seien als gut gemachte Zeitreihen (interrupted time-series) oder mathematische Modellierungen. Etwas ganz anderes ist die Tatsache, dass Versorgungsforschung im Kontext von Qualitätsmanagement sich natürlich auch nicht in RCT erschöpft. So wird die Rolle von qualitativen Studien vermutlich nach wie vor unterschätzt (s. Pope und Mays 2006). Und die Verwendung von Routinedaten hat ihren festen Platz zur Interpretation des Ordnungsgeschehens und auch in der Evaluation von QM-Maßnahmen.

Kritik an second-best-designs findet sich auch beim *Wissenschaftlichen Institut der TK für Nutzen und Effizienz im Gesundheitswesen*. So schreiben Linder et al in einer Arbeit zur Evaluation des DMP Typ-2-Diabetes mellitus, dass weder die bisherigen Qualitätsberichte zu den DMP noch die ELSID-Studie der Heidelberger Versorgungsforscher wegen eines mangelhaften Kontrollgruppendesigns den Anforderungen an eine

wissenschaftlich fundierte Evaluation genügen (Linder et al 2011, 155). Gleichzeitig wird in einer eigenen DMP-Evaluation der von Linder et al verwendete Ansatz des »propensity score interval matching« (von Rosenbaum und Rubin 1983 publiziert) ausdrücklich als »intelligentes Kontrollgruppendesign« bezeichnet, das für künftige Versorgungsforschung vorbildlich und dem in der ELSID-Studie gewählten Matching-Verfahren überlegen sein soll. Dies ist immanent unlogisch, da das von der Heidelberger Arbeitsgruppe gewählte individuelle Matching-Verfahren und die zum Vergleich herangezogenen klinischen Parameter keinesfalls minderer Güte sind. Die Kritik ist im Licht der Debatte um angemessene Methoden der Versorgungsforschung aber auch insofern unfair, als dem unkundigen Leser die Geschichte des Scheiterns einer DMP-RCT (s. Studiendesign bei Beyer et al 2006) unterschlagen wird. Der Autor könnte interessierten Leserinnen und Lesern nähere Auskunft über die damalige Aufstellung von Befürworter und Gegnern einer DMP-RCT geben.

Politische Parteien und Gesetzgeber fragen nicht vertieft nach wissenschaftlichen Belegen, wenn sie Reformen anstoßen oder einfordern, es entspricht nicht ihrer Logik (Schmacke 2010); das kann hilfreich oder schädlich sein – ein weites Feld für die Politikfolgenforschung. Hier geht es aber darum, noch einmal darüber nachzudenken, wie es um die Kultur im Bereich der Versorgungsforschung bestellt ist. Dazu gibt die hier besprochene Publikation vielfältige Anregungen. Aus wissenschaftlicher Sicht wäre es reizvoll, Einschätzungen zusammen zu tragen, welche methodisch hochwertigen Studien heute für die Weiterentwicklung der GKV fehlen und danach zu überlegen, ob es tatsächlich so utopisch ist, in Deutschland eine entsprechende Forschungskultur mit ausreichend langem Atem ins Werk zu setzen. Größere Praxisnähe und Erhöhung der Effizienz sind kein Widerspruch zu methodisch hohen Anforderungen. Versorgungsforschung kann keine neuen Forschungsmethoden erfinden, wenn die einschlägigen Methoden nicht bekannt oder nicht eingesetzt worden sind.

Literatur:

- Beyer M, Gensichen J, Szucsényi J, Wensing M.** Wirksamkeit von Disease-management-Programmen in Deutschland – Probleme der medizinischen Evaluationsforschung anhand eines Studienprotokolls. *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen* 2006; 100: 355-63
- Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandecrook P, Spiegelhalter D, Tyrer P.** Framework for design and evaluation of complex interventions to improve health. *British Medical Journal* 2000; 321: 694-6
- Campbell NC et al.** Designing and evaluating complex interventions to improve health care. *British Medical Journal* 2007; 334: 455-9
- Christenfeld NJ, Sloan RP, Carroll D, Greenland S.** Risk Factors, Confounding, and the Illusion of Statistical Control. *Psychosomatic Medicine* 2004; 66, 868-875
- Eccles M, Grimshaw J, Campbell M et al.** Research designs for studies in evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care* 2003; 12: 47-52
- Felder F, Mayrhofer T.** Evaluation des integrierten Versorgungsprojekts »integra – Neue Wege zur Gesundheit«. FEISA, 2008 (von den Autoren freundlicherweise zur Verfügung gestellt).
- Gilbody S, Whitty P.** Improving the delivery and organisation of mental health services: beyond the conventional randomised controlled trial. *British Journal of Psychiatry* 2002; 180: 13-8
- Göbel H, Heinze A, Heinze-Kuhn K, Henkel K, Roth A, Rüschemann HH.** Entwicklung und Umsetzung der integrierten Versorgung in der Schmerztherapie. Das bundesweite Kopfschmerzbehandlungszentrum. *Der Schmerz* 2009; 6: 653-670.
- Neugebauer EAM et al.** Memorandum III: Methoden der Versorgungsforschung (Teil 2). *Gesundheitswesen* 2010; 72: 739-748
- Herck P van, De Smedt D, Annemans L, Remmen R, Rosenthal MB.** Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Services Research* 2010; 10: 247 (www.biomedcentral.com/1472-6963/10/247)
- Klusen N, Meusch A, Thiel E (Hrsg.).** Qualitätsmanagement im Gesundheitswesen. Nomos Verlag Baden Baden, 2011.
- Petersen LA, Woodard D, Urech T, Daw C, Sookanan S.** Does Pay-for-Performance Improve the Quality of Health Care? *Annals of Internal Medicine* 2006; 145: 265-272
- Raspe H, Pfaff H, Härter M et al.** Versorgungsforschung in Deutschland. Stand – Perspektiven – Förderung. DFG Stellungnahme, 2010.
- Pope C, Mays N.** Qualitative methods in health research. In: Pope C, Mays N. *Qualitative Research in Health Care*. BMJ Books 2006, 1-11
- http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2010/stellungnahme_versorgungsforschung.pdf
- Roth A, Rüschemann H-H.** Karrieren chronisch schmerzkranker Patienten in Schleswig-Holstein. Eigendruck 2002 (von den Autoren freundlicherweise zur Verfügung gestellt).
- Schmacke N.** Gesundheitsreform: welche Logik zählt? In: Gerlinger T, Kümpers S, Lenhardt U, Wright M. *Politik für Gesundheit. Fest- und Streitschriften zum 65. Geburtstag von Rolf Rosenbrock*. Huber-Verlag, 2010: 112-121
- Steinwachs DM, Hughes RG.** Health Services Research: Scope and Significance. In: Hughes RG, editor. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. Rockville, AHRQ, 2008, Chapter 8
- Wiethöge J, Ommen O, Ernstmann N, Pfaff H.** Transparenz als Voraussetzung für Innovation in der Evaluationsforschung: Defizite am Beispiel der Evaluation von Managed-Care-Modellvorhaben. *Gesundheitswesen* 2010; 72: 722-8
- Wunsch H, Linde-Zwirble WT, Angus DC.** Methods to adjust for bias and confounding in critical care health services research involving observational data. *Journal of Critical Care* 2006; 21: 1-7