

IPP-SCHRIFTEN 15

Norbert Schmacke

Pay for Performance (P4P)

**Bilanz einer gesundheitsökonomischen
Ideologie.**

Ein narrativer Review

**Schriftenreihe des Instituts für Public Health und Pflegeforschung (IPP)
der Universität Bremen**

**Herausgeber*innen: Prof. Dr. Gabriele Bolte, Prof. Dr. Karin Wolf-Oster-
mann, Prof. Dr. Ansgar Gerhardus**

Februar 2019

Impressum

IPP-Schriften, Ausgabe 15 / 2019

Titel: Pay for Performance (P4P)
Bilanz einer gesundheitsökonomischen Ideologie. Ein narrativer Review

Autor: Prof. Dr. Norbert Schmacke

Kontakt Autor: Prof. Dr. Norbert Schmacke, Institut für Public Health und Pflegeforschung (IPP), Universität Bremen.
E-Mail: schmacke@uni-bremen.de

Herausgeber*innen IPP-Schriften: Direktorium des Instituts für Public Health und Pflegeforschung (IPP), Fachbereich 11 Human- und Gesundheitswissenschaften, Universität Bremen: Prof. Dr. Gabriele Bolte (geschäftsführend), Prof. Dr. Karin Wolf-Ostermann, Prof. Dr. Ansgar Gerhardus.

Kontakt Herausgeber*innen: Prof. Dr. Gabriele Bolte (Geschäftsführende Direktorin), c/o Geschäftsstelle des Instituts für Public Health und Pflegeforschung (IPP), Fachbereich 11, Universität Bremen, Grazer Straße 4, 28359 Bremen.
Fon: (+)49 0421 – 218 68980/ -68880, E-Mail: hmertesa@uni-bremen.de, www.ipp.uni-bremen.de.

Bremen, Februar 2019
Umschlaggestaltung: IPP
Download der Online-Ausgabe unter: www.ipp.uni-bremen.de

ISSN 1867-1942 (Internetausg.)

V.i.S.d.P.: Verantwortlich für die publizierte Fassung ist der Autor der vorliegenden Ausgabe.

Datenschutzerklärung: Es gilt die Datenschutzerklärung der Universität Bremen (Online im Internet unter <https://www.uni-bremen.de/datenschutz/>)

Haftungsausschluss: Es gilt der nachfolgende Haftungsausschluss der Universität Bremen (Online im Internet unter <https://www.uni-bremen.de/datenschutz/haftungsausschluss/>):

1. Inhalt des Onlineangebotes

Die Universität Bremen übernimmt keinerlei Gewähr für die Aktualität, Korrektheit und Vollständigkeit der bereitgestellten Informationen. Haftungsansprüche gegen die Universität Bremen, die sich auf Schäden materieller oder immaterieller Art beziehen, die durch die Nutzung oder Nichtnutzung der dargebotenen Informationen bzw. durch die Nutzung fehlerhafter und unvollständiger Informationen verursacht wurden, sind grundsätzlich ausgeschlossen, sofern seitens der Universität Bremen kein nachweislich vorsätzliches oder grob fahrlässiges Verschulden vorliegt. Die Universität Bremen behält es sich ausdrücklich vor, Teile der Seiten oder das gesamte Angebot ohne gesonderte Ankündigung zu verändern, zu ergänzen, zu löschen oder die Veröffentlichung zeitweise oder endgültig einzustellen.

2. Verweise und Links

Die Universität Bremen hat keinerlei Einfluss auf die aktuelle und zukünftige Gestaltung und auf die Inhalte fremder Internetseiten, auf die direkt oder indirekt verwiesen wird (Links). Deshalb distanziert sie sich hiermit ausdrücklich von allen Inhalten aller gelinkten/verknüpften Seiten. Dies gilt auch für alle Fremdeinträge in von der Universität Bremen eingerichteten Gästebüchern, Diskussionsforen und Mailing-Listen.

3. Urheber- und Kennzeichenrecht

Die Universität Bremen hat in allen Publikationen die Urheberrechte der Autoren der verwendeten Grafiken, Tondokumente, Videosequenzen und Texte zu beachten. Sie ist bestrebt, möglichst von ihr selbst erstellte Grafiken, Tondokumente, Videosequenzen und Texte zu nutzen oder auf lizenzfreie Grafiken, Tondokumente, Videosequenzen und Texte zurückzugreifen. Eine Vervielfältigung oder Verwendung der innerhalb des Internetangebotes der Universität Bremen verwendeten Grafiken, Tondokumente, Videosequenzen und Texte in anderen elektronischen oder gedruckten Publikationen ist ohne ausdrückliche Zustimmung des Autors nicht gestattet, sofern diese nicht ausdrücklich als lizenzfrei erklärt sind.

4. Datenschutz

Sofern innerhalb des Internetangebotes die Möglichkeit zur Eingabe persönlicher oder geschäftlicher Daten (E-Mail-Adressen, Namen, Anschriften) besteht, so erfolgt die Preisgabe dieser Daten seitens des Nutzers auf ausdrücklich freiwilliger Basis.

5. Urheberrecht (Copyright - Universität Bremen)

Alle Rechte vorbehalten, all rights reserved, Universität Bremen, D-28359 Bremen.

Die Online-Dokumente und Webseiten einschließlich ihrer Teile sind urheberrechtlich geschützt. Sie dürfen nur zum privaten, wissenschaftlichen und nicht-gewerblichen Gebrauch zum Zweck der Information kopiert und ausgedruckt werden, wenn sie den Copyright-Hinweis enthalten. Die Universität Bremen behält sich das Recht vor, diese Erlaubnis jederzeit zu widerrufen. Ohne vorherige schriftliche Genehmigung der Universität Bremen dürfen diese Dokumente/Webseiten nicht vervielfältigt, archiviert, auf einem anderen Server gespeichert, in News-Gruppen einbezogen, in Online-Diensten benutzt oder auf einer CD-ROM gespeichert werden. Sie dürfen jedoch in einen Cache oder einen Proxy-Server zur Optimierung der Zugriffsgeschwindigkeit kopiert werden.

Wir erlauben ausdrücklich und begrüßen das Zitieren unserer Dokumente und Webseiten sowie das Setzen von Links auf unsere Website.

Norbert Schmacke

Pay for Performance (P4P)

Bilanz einer gesundheitsökonomischen Ideologie. Ein narrativer Review¹

Februar 2019

¹ Erweiterte Version eines Reviews für den Gemeinsamen Bundesausschuss vom Oktober 2018: "Pay for Performance (P4P). Aufarbeitung von internationalen Ansätzen, Möglichkeiten und Grenzen im Rahmen von Pay for Performance und Ergebnisqualitätsmessung im Rahmen der gesundheitlichen Versorgung."

Inhaltsverzeichnis

Zusammenfassung

Einleitung

Die deutsche Debatte um P4P bis zum BQS-Gutachten

Neuere Studien und Reviews zu P4P und VBP

1. Hämo- und Peritonealdialyse im internationalen Vergleich
2. Anreize von Prävention
3. Das englische Großprojekt QOF
4. Stationärer Sektor: Der Indikator Wiedereinweisungsraten
5. VBP und Mortalität

Konsequenzen aus der publizierten Literatur zu VBP-Programmen

Versorgungsqualität: ein Aufruf zur Verbesserung der Studienqualität

Bilanz

Nachbetrachtung

Über den Autor

Zusammenfassung

Die systematische Erhebung von Qualität(smängeln) in der Versorgung ist seit gut 20 Jahren Teil der internationalen gesundheitspolitischen Debatte. Große Erwartungen wurden und werden von Gesundheitsökonomen und Politikern in die Einführung ökonomischer Anreize zur Erbringung dokumentierter, transparenter Qualität gesetzt. Die publizierte Literatur spricht dafür, dass dieser hoch plausible Ansatz der Verbindung von Vergütung und Qualitätsverbesserung im Alltag der Versorgung die erwünschten Effekte nicht stabil erzeugen kann. Es ist die Frage, ob die darauf fußenden, methodisch nicht ausreichend begründeten Vergütungsmodelle durch weitere Forschung optimiert werden sollen oder ob andere Wege zur Verbesserung patientenrelevanter Outcomes nicht deutlich Erfolg versprechender sind. In jedem Fall stellt sich die Frage nach der Beziehung zwischen Wissenschaft und Politik: werden die skeptischen Einwände der Wissenschaft gehört oder setzen Gesundheitssysteme weiter auf plausible Vorannahmen, die vermeintlich nicht noch einmal hinterfragt werden müssen. Die Gesundheitspolitik hat mit P4P eine Lösung gefunden, welche die nachhaltige Kritik an Versorgungsmängeln und die stets unbeliebten Bemühungen um Preisregulierung in der Forderung nach qualitätsorientierter Vergütung populär zusammenführt.

Einleitung

Unter Pay for Performance (P4P) werden Vergütungsformen im stationären und ambulanten Versorgungsbereich verstanden, die über Zu- und/oder Abschläge die Höhe der Vergütung von der Erreichung einer vorab definierten Qualität der Leistungen abhängig machen. Es handelt sich basierend auf diesem Generalprinzip um eine Vielzahl unterschiedlicher Vertragsformen; eine allgemeingültige Definition von P4P fehlt insofern. Eine der vielen Definitionen lautet:

“In the healthcare industry, pay for performance (P4P), also known as "value-based purchasing (VBP)", is a payment model that offers financial incentives to physicians, hospitals, medical groups, and other healthcare providers for meeting certain performance measures. Clinical outcomes, such as longer survival, are difficult to measure, so pay for performance systems usually evaluate process quality and efficiency, such as measuring blood pressure, lowering blood pressure, or counseling patients to stop smoking. This model also penalizes health care providers for poor outcomes, medical errors, or increased costs”².

Ausgehend von den USA hat die Kernidee von P4P in den letzten zwei Jahrzehnten auch in Europa in unterschiedlichen Gesundheitssystemen eine gewisse Verbreitung gefunden. Ein starker Impuls für die Entwicklung dieses gesundheitsökonomischen Konzeptes war die Debatte um vermeidbare Fehler in der Versorgung, die datengestützt vom Institute of Medicine in den USA geführt wurde:

„In 1996 the Institute of Medicine launched the Quality Chasm Series, a series of reports focused on assessing and improving the nation’s quality of health care. Preventing Medication Errors is the newest volume in the series. Responding to the key messages in earlier volumes of the series—To Err Is Human (2000), Crossing the Quality Chasm (2001), and Patient Safety (2004)—this book sets forth an agenda for improving the safety of medication use. It begins by providing an overview of the system for drug development, regulation, distribution, and use. Preventing Medication Errors also examines the peer-reviewed literature on the incidence and the cost of medication errors and the effectiveness of error prevention strategies. Presenting data that will foster the reduction of medication errors, the book provides action agendas detailing the measures needed to improve the safety of medication use in both the short- and long-term. Patients, primary health care providers, health care organizations, purchasers of

² [https://en.wikipedia.org/wiki/Pay_for_performance_\(healthcare\)](https://en.wikipedia.org/wiki/Pay_for_performance_(healthcare))

*group health care, legislators, and those affiliated with providing medications and medication-related products and services will benefit from this guide to reducing medication errors”.*³

Die Rationale aller auf wirtschaftliche Anreiz setzenden Vergütungssysteme lautet:

1. Die intrinsische Motivation der Gesundheitsfachberufe, voran der Ärzteschaft, sei kein ausreichender Garant für die Erbringung guter Qualität und müsse durch die Aussicht auf höhere Einkommen (bzw. Renditen bei privaten Unternehmen) unterstützt werden.
2. In Verbindung mit der Publikation der ermittelbaren Qualitätsunterschiede zwischen den Leistungserbringern entstehe ein starker Impuls zur Optimierung der eigenen Ergebnisse.
3. Schließlich lasse sich zeigen, dass derartige Reformen zu höherer Effizienz im Gesundheitssystem führen.

Die deutsche Debatte um P4P bis zum BQS-Gutachten

Bereits im Jahr 1997 plädierte der Sachverständigenrat (SVR 1997, Nr. 137) für “ergebnisbezogene Bonuszahlungen im Rahmen mehrschichtiger Vergütungssysteme mit einer Verknüpfung der Vergütung an zu entwickelnde Leitlinien”. Im SVR-Gutachten 2007 folgte dann ein systematischer Review, der unter dem Begriff der qualitätsorientierten Vergütung 28 kontrollierte Studien einschloss und eine Wirkung auf Prozess- und Ergebnisparameter zeigen zu können meinte.

Im Jahr 2012 wurde ein Gutachten des BQS Institut für Qualität und Patientensicherheit GmbH veröffentlicht, das vom Bundesministerium für Gesundheit beauftragt war.⁴ Dieser Review basierte auf einer weiter gefassten Definition des Begriffes P4P, indem er sich auf das “Leistungsniveau” und nicht allein auf die Qualität als Zielparameter bezog. Das Kernergebnis des Gutachtens lautete:

³ <https://www.nap.edu/catalog/11623/preventing-medication-errors>

⁴ <https://www.bqs.de/images/downloads/P4P-Gutachten-BQS.compressed.pdf>

„Eine kritische Sichtung der Literatur mit den Schwerpunkten hausärztlicher Bereich und Krankenhausbereich sowie Kosteneffektivität von P4P kommt - auch unter Berücksichtigung der neueren Literatur von 2010 und 2011 - zum gleichen Ergebnis wie andere Reviews zuvor: Bislang ist es in Studien noch nicht gelungen, zweifelsfrei die Wirksamkeit von P4P-Projekten nachzuweisen. Zwar gibt es Belege für die Effektivität kombinierter Verfahren unter Einschluss nicht-finanzieller Anreize. Der Anteil der finanziellen Anreize am Erfolg ließ sich bislang jedoch nicht eindeutig nachweisen. Dies widerspricht der allgemeinen Erfahrung, dass Geld durchaus ein sehr starker Anreiz ist und Änderungen der Vergütungssysteme stets intensive Reaktionen nach sich ziehen. Die Komplexität medizinischer Versorgung sowie die systemischen Optimierungs- und Umgehungsreaktionen beeinflussen die Ergebnisse einiger Projekte durch diverse Selektionseffekte sehr. Viel an Wirksamkeit wird dabei verdeckt oder nicht realisiert, was an einzelnen Beispielen gezeigt werden kann. Zahlreiche Organisatoren bestehender Projekte sind jedoch von ihrem Tun und ihren guten praktischen Erfahrungen überzeugt und führen die Projekte fort oder bauen sie weiter aus. Trotz der mangelnden überzeugenden Evidenz für die Wirksamkeit von P4P sind die meisten der Befragten in Deutschland der Ansicht, dass P4P in der Zukunft eine größere Rolle spielen wird. Unklar bleibt, in welchem Umfang dies derzeit bereits politisch umsetzbar ist.“

Weiter heißt es:

„Die meisten Probleme von P4P ergeben sich aus Problemen der Qualitätsmessung. Es ist wichtig, dass bei allen Projekten das Verhältnis von Machbarkeit, Sinnhaftigkeit und Angemessenheit gewahrt bleibt und dass die Qualitätsaussagen auf das beschränkt werden, was die Messinstrumente leisten können. Dazu zählt auch die Berücksichtigung des damit verbundenen Aufwands.“

Vor Darlegung neuerer Studien und Reviews seit Publikation des BQS-Gutachtens wird noch einmal kurz auf den Cochrane-Bericht von 2011⁵ eingegangen, der im BQS-Gutachten bereits Berücksichtigung gefunden hat.

Hier ging es um den Bereich Primary Care, aus dem sieben Studien unterschiedlicher Anreizsysteme ausgewertet wurden. Darunter auch die vielfach zitierten Daten aus

⁵ Scott A1, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, Furler J, Young D. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database Syst Rev. 2011 Sep 7;(9):CD008451

dem ‚Quality and Outcomes Framework (QOF) for General Practitioners (GPs)‘ in England. Die Kernaussage lautete:

“Despite the popularity of these schemes, there is currently little rigorous evidence of their success in improving the quality of primary health care, or of whether such an approach is cost-effective relative to other ways to improve the quality of care.”

Inhaltlich ging es um Raucherentwöhnung, Chlamydien-, Cervix-CA- und Mammographie-Screening, HbA1c und ein weiteres umfassenderes auf Diabetes abzielendes Programm, Kinderimpfungen, und Asthmabehandlung. Die Berechnung der teils – wenn auch geringfügig – positiven Outcomes litt unter deutlichen methodischen Problemen. Am bedeutsamsten ist vielleicht der Hinweis der Autorengruppe, dass keine der Studien das Selektionsbias problematisierte, da die GPs sich in das Anreizsystem ein- und auswählen konnten.

Ein weiterer Cochrane-Bericht von 2011⁶ - ebenfalls im BQS-Bericht gewürdigt – wertete P4P-Programme auf den Zusammenhang von Verhaltensbeeinflussung und Vergütung aus, 32 Studien wurden eingeschlossen. Das Kernergebnis lautete:

“Financial incentives may be effective in changing healthcare professional practice. The evidence has serious methodological limitations and is also very limited in its completeness and generalisability. We found no evidence from reviews that examined the effect of financial incentives on patient outcomes”.

Der jüngste einschlägige Cochrane-Bericht von 2017⁷ führt zu keinen neuen Erkenntnissen, wobei interessant ist, dass gezielt nach den Effekten aller klassischen Vergütungstypen gesucht wurde: Budgetierung, Capitation, Kopfpauschalen, Einzelleistungsvergütungen, P4P. Von dauerhafter Überlegenheit eines dieser Systeme könne nicht die Rede sein, allen noch so begründeten Vorannahmen zum Trotz. Die Kernaussage lautete:

⁶ Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer R. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. Cochrane Database Syst Rev .2011 ; (7): CD009255

⁷ Yuan B, He L, Meng Q, Jia L. Payment methods for outpatient care facilities. Cochrane Database of Systematic Reviews 2017, Issue 3. Art. No.: CD011153.

“Pay-for-performance systems probably have only small benefits or make little or no difference to healthcare provider behaviour or patients' use of healthcare services. We are uncertain whether they cause harm. We are uncertain about the benefits and harms of other payments systems because the research is lacking or of very low certainty.”

Interessant erscheint die Empfehlung der Autorengruppe an die Politik:

“When considering using P4P to improve the performance of health facilities, policymakers should carefully consider each component of their P4P design, including the choice of performance measures, the performance target, payment frequency, if there will be additional funding, whether the payment level is sufficient to change the behaviours of health providers, and whether the payment to facilities will be allocated to individual professionals. Unfortunately, the studies included in this review did not help to inform those considerations”.

Als weitere Übersichtsarbeit soll Mendelson u.a. (2017)⁸ zitiert werden. Immerhin 69 Studien wurden ausgewertet, darunter 58 in ambulanten Settings. 38 dieser Studien berichteten patientenrelevante Outcomes. Am Ende kommen die Autoren zu dem Schluss, dass es nur völlig unzureichende Hinweise für relevante klinische Verbesserungen gibt:

“Pay-for-performance programs may be associated with improved processes of care in ambulatory settings, but consistently positive associations with improved health outcomes have not been demonstrated in any setting”.

Es gibt anders gesagt keine relevanten neuen Informationen im Vergleich zu dem BQS-Bericht von Veit u.a. von 2012. Im Detail heißt es bei Mendelson et al:

“Overall, in the ambulatory setting, we found low-strength evidence that P4P programs may improve process-of-care outcomes over the short term (2 to 3 years). Evidence on the longer-term effects of P4P programs was limited. Many of the studies reporting positive findings were conducted in the United Kingdom, where incentives were much larger than any P4P programs

⁸ Mendelson A, Kondo K, Damberg C, Low A, Motúapuaka M, Freeman M, O'Neil M, Relevo R, Kansagara D. The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review. *Ann Intern Med.* 2017; 166:341-353

in the United States. The largest improvements were seen in areas where baseline performance was poor. We found low-strength evidence that P4P had little to no effect on intermediate health outcomes (changes in laboratory measures), though there were inconsistencies among study results. The evidence examining patient health outcomes was insufficient because few methodologically rigorous studies reported these outcomes. In the hospital setting, low-strength evidence showed that P4P had a neutral effect on patient health outcomes and a positive effect on reducing hospital readmissions”

Als Zwischenfazit ist festzuhalten: Es gibt nach über zwanzig Jahren Bemühungen um den Nachweis der Überlegenheit von P4P-Modellen gegenüber traditionellen Vergütungsformen keine wirklich bestätigenden Resultate, weder für den umfangreicher untersuchten ambulanten noch für den stationären Sektor. Mit Blick auf einzelne positiv wirkende Resultate wiederholen sich in allen Reviews die Forderungen nach methodisch besseren Studien mit längerer Laufzeit. Angesichts der hohen politischen wie dahinter auch gesundheitsökonomischen Wertschätzung von P4P ist dies ein außerordentlich wichtiger Befund.

Nachfolgend werden einige Einzelstudien beschrieben, welche diese globalen Aussagen noch einmal exemplarisch beleuchten sollen. Einen Mehrwert zu den zitierten systematischen Bewertungen der P4P-Ansätze können sie naturgemäß nicht liefern; es soll vielmehr dargestellt wird, welche Schwerpunkte aktuellere Modellvorhaben und deren Evaluation besetzen.

Neuere Studien und Reviews zu P4P und VBP

1. Hämo- und Peritonealdialyse im internationalen Vergleich

Die 2012 veröffentlichte Studie⁹ von Vanderholder et al. analysierte das Vergütungssystem für chronische Dialyse in Ontario/ Canada, Belgien, Frankreich, Deutschland, Niederlande und UK. Alle Länder seien hoch motiviert, die Kosten für die Dialyse (die zwischen den untersuchten Ländern um den Faktor 3,3 variieren) zu kontrollieren und verwendeten dabei unterschiedlichste Vergütungssysteme inklusive P4P-Ansätzen mit vorgegebenen Qualitätszielen.

Die Kernaussage lautete:

“Unfortunately, it is not yet clear that such differences (gemeint: die unterschiedlichen Dialyseverfahren, Erläuterung durch den Autor) save money or improve quality of care. Future research should focus on directly testing both outcomes.”

In diesem Behandlungsfeld gibt es, das macht das Beispiel so bedeutend, enorme Auffassungsunterschiede zu den Standards der Therapien, technischen Verfahren und Behandlungsziele. Ob ökonomische Anreize hierbei eine positive Rolle spielen können, ist momentan vollkommen unklar. Dass umgekehrt das Anzielen eines der häufig vereinbarten Qualitätsziele, nämlich eines besonders hohen Hämoglobinwertes, bei fragwürdiger klinischer Bedeutung mit deutlichen Kostensteigerungen wegen eines hohen Einsatzes von Erythropoietin verbunden sein kann, ist demgegenüber gesichert¹⁰.

⁹ Vanholder R, Davenport A, Hannedouche T, Kooman J, Kribben A, Lameire N, Lonnemann G, Magner P, Mendelssohn D, Saggi SJ, Shaffer RN, Moe SM, Van Biesen W, van der Sande F, Mehrotra R; Dialysis Advisory Group of American Society of Nephrology. Reimbursement of dialysis: a comparison of seven countries. *J Am Soc Nephrol.* 2012; 8:1291-8

¹⁰ Collister D, Komenda P, Hiebert B, Gunasekara R, Xu Y, Eng F, Lerner B, Macdonald K, Rigatto C, Tangri N. The Effect of Erythropoietin-Stimulating Agents on Health-Related Quality of Life in Anemia of Chronic Kidney Disease: A Systematic Review and Meta-analysis. *Ann Intern Med.* 2016 Apr 5;164(7):472-8. S. auch zuvor https://www.g-ba.de/downloads/39-261-1341/2011-06-23_AM-RL-IV_Erythropoese_BAnz.pdf

2. Anreizen von Prävention

In Houston/ Texas wurde von 2001 bis 2007 versucht, in kommunalen Gesundheitszentren in einem quasi-experimentellen Design den Effekt einer Verdopplung der ärztlichen Vergütung für Pap-Abstriche, Mammographien und Kinderimpfungen zu ermitteln. Die Wunsch-Zielwerte und die aktuell erreichten Werte wurden kontinuierlich zurückgemeldet. Einen Unterschied zwischen den Interventions- und den Kontrollzentren war nicht festzustellen¹¹.

3. Das englische Großprojekt QOF

2004 wurde im NHS das bis dahin größte P4P-Programm für Primary Care aufgelegt: das Quality and Outcomes Framework Projekt. Die GPs erhielten bis zu 25 % mehr Einkommen, wenn sie die Anforderungen eines aufwändig erarbeiteten Systems von klinisch relevanten Qualitätsindikatoren erfüllten. Es handelte sich um ein Programm des gesamten NHS. Die erhoffte breite Steigerung der Qualität blieb aus, es gab offenkundig erhebliche Mitnahmeeffekte, so dass die Kosteneffektivität auf jeden Fall nicht gegeben zu sein scheint, und der NHS revidierte den Ansatz nach einigen Jahren fundamental in Richtung Optimierung der Versorgung chronisch Kranker¹². Die Auseinandersetzung mit diesem Großvorhaben liefert für die Frage P4P im Bereich primärärztlicher Versorgung fraglos die relevantesten Daten, mit denen jedes Nachfolgeprojekt sich auseinandersetzen muss. Diese kritische Gesamtbewertung soll hier nicht im Mittelpunkt stehen; es sei vielmehr auf die hier genannte Arbeit von Roland und Guthrie (2016) hingewiesen. Hier werden folgende Anforderungen an P4P-Projekte im Gefolge von QOF gestellt:

¹¹ Gavagan TF, MD, Du H, Saver BG, Adams GJ, Graham DM, BA, McCray R, Goodrick GK. Effect of Financial Incentives on Improvement in Medical Quality Indicators for Primary Care. *J Am Board Fam Med* 2010; 23: 622-631

¹² Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? *BMJ*. 2016 Aug 4;354:i4060, weiter: Pandya A, Doran T, Zhu J, Walker s, Arntson M, Ryan AM. Modelling the cost-effectiveness of pay-for-performance in primary care in the UK. *BMC Med*. 2018; 16: 135. S. auch: Forbes LJ1, Marchand C1, Doran T2, Peckham S1. The role of the Quality and Outcomes Framework in the care of long-term conditions: a systematic review. *Br J Gen Pract*. 2017;67: e775-e784.

Martin Roland and Frede Olesen Can pay for performance improve the quality of primary care? *BMJ* 2016;354:i4058

- *Pay for performance should be seen as part of wider quality improvement efforts.*
- *Public reporting of information on quality of care may be an effective driver of change without pay for performance*
- *Alternative strategies should be used to improve quality for aspects of care not easily measured to avoid them being neglected*
- *Single disease indicators may not be appropriate for important patient groups such as complex patients with multimorbidity*
- *Clinicians should be strongly represented among those selecting indicators and designing the programme*
- *Technical expertise in developing and implementing indicators is needed so that they measure what they are intended to measure and reduce unintended consequences*
- *Indicators should represent aspects of quality that command wide professional support and, where available, be based on strong scientific evidence*
- *Payments should be large enough to change behaviour but not so large as to divert excessive effort onto incentivised aspects of care*
- *The payment formula should encourage improvement at all levels up to an agreed maximum, ensuring that low performing providers have an incentive to improve*
- *Unexpected consequences should be anticipated and continuously monitored*
- *The effect on inequalities in delivery of care should be monitored*

Die Schlussfolgerungen der Autoren aus ihrer Bewertung von QOF lauten (Fettmarkierungen durch den Autor):

- *Pay for performance can be a useful part of wider programmes to improve quality of care, though **quality gains** from financial incentives are generally **modest***
- ***Clinicians must be closely involved** in the development of quality indicators and pay-for-performance programmes*
- *All incentive schemes (financial and non-financial) have the **potential to produce unexpected or perverse outcomes**, and these should be anticipated and continuously monitored*

4. Stationärer Sektor: Der Indikator Wiedereinweisungsraten

Die Senkung (unerwünschter, bei optimaler Versorgung nicht erforderlicher: Ergänzung durch den Autor) Wiedereinweisungsraten im stationären Sektor ist einer der am besten untersuchten Ansätze für P4P im Krankenhausbereich, der häufig als Beleg für

den Nutzen von P4P zitiert wird. Fischer et al¹³ präsentierten 2014 eine empirische Untersuchung zur Belastbarkeit dieses Indikators (Datenbasis aus dem Zeitraum 2001 bis 2013). Sie kommen zu einer kritischen Bilanz, die – ohne damit dieses Qualitätsziel in toto desavouiert zu haben – folgende Schlüsselprobleme anspricht:

- *The assessment of the indicator shows difficulties, as the indicator definition is often not explicit enough to identify readmissions related to quality of care (avoidable readmissions).*
- *The data used to calculate the indicator is mainly administrative data, which generally includes incomplete and inaccurate data elements and lacks clinical information*
- *Furthermore, in many countries readmissions to other institutions cannot be followed. Readmission rates are influenced also by other factors than quality of hospital care, which include length of stay, (in-hospital) mortality and patient characteristics.*
- *The magnitude of influence is partly not know as data is missing to investigate the association (e.g. no post discharge mortality, no clinical characteristics).*
- *Further, the scientific evidence of the indicator is limited, as existing research shows conflicting results with regard to the influence of quality of hospital care on the readmission rate*

Aus Sicht der System- und Versorgungsforschung weist dieser Review in besonders eindrucksvoller Form auf ein grundlegendes Problem für grundlegende Umsteuerungen von Vergütungssystemen hin: Noch so plausible gesundheitspolitische Versorgungsziele sollten immer nur mit tatsächlich belastbaren Indikatoren erprobt und nachhaltig evaluiert werden, um nicht nach 10 oder mehr Jahren festzustellen, dass es eine Fülle von methodischen Limitationen gibt, welche eine ‚echte‘ Aussage über Nutzen und Kosten-Nutzen nicht erlauben. Der Indikator ‚Wiedereinweisungsraten‘ ist in der Literatur zur Qualitätsverbesserung der Versorgung entgegen vielfacher Annahmen mit anderen Worten nicht adäquat belegt.

Es geht im Bereich Wiedereinweisungen in jedem Fall um eine Betrachtung ganz konkreter Indikationsfelder: das macht noch einmal die Schwierigkeit der Implementation

¹³ Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, et al. (2014) Is the Readmission Rate a Valid Quality Indicator? A Review of the Evidence. PLoS ONE 9(11): e112282. doi:10.1371/journal.pone.0112282

von P4P-Programmen aus: die gesundheitspolitische Erwartung ist in der Regel umfassend, die Ergebnisse im Versorgungsalltag aber sind indikationsbezogen hochgradig unterschiedlich. Dazu exemplarisch noch einmal folgende Erfahrung aus der Chirurgie:

Das Hospital Readmission Reduction Program (HRPP) im Rahmen des Affordable Care Act belegte Krankenhäusern ab 2012 mit per definitionem unerwartet hohen Wiedereinweisungsraten mit Abzügen. Die Erwartung war, dass diese auf Hüft- und Kniegelenkersatz begrenzte Vergütungsregelung auch auf andere Eingriffe wie Darm- und Lungenoperationen oder koronare Bypass-Operationen positive Einflüsse haben würde¹⁴. Das Kernergebnis lautete:

“The rate of reduction was most prominent after announcement of the program between 2010 and 2012 for both targeted and nontargeted conditions. During the same time period, mean hospital length of stay decreased; nontargeted conditions (10.4-8.4 days) and targeted conditions (3.6-2.8 days). There was no correlation between hospital reduction in readmissions and use of observation-only admissions (Pearson correlation coefficient = 0.01) or discharge to a skilled nursing facility (Pearson correlation coefficient = 0.05)”

Man kann diese Resultate als Beleg für¹⁵ und gegen die P4P-Strategie interpretieren: entweder hat der ‚Bestrafungsimpuls‘ gewirkt, und das für ausgewählte und nicht ausgewählte Indikationen; oder es sind andere Gründe dafür verantwortlich, dass in dieser

¹⁴ Ibrahim AM1, Nathan H, Thumma JR, Dimick JB. Impact of the Hospital Readmission Reduction Program on Surgical Readmissions Among Medicare Beneficiaries. *Ann Surg.* 2017 Oct;266(4):617-624. Die Grenzen der Aussagekraft von Studien mit second best designs (häufig interrupted time series) werden auch von Autorengruppen benannt, die positive Einflüsse von P4P-Ansätzen auf Wiedereinweisungsraten ermittelt haben: s. jüngst Andrew M. Ryan, PhD¹; Sam Krinsky, MA¹; Julia Adler-Milstein, PhD^{1,2}; et al. Association Between Hospitals’ Engagement in Value-Based Reforms and Readmission Reduction in the Hospital Readmission Reduction Program. *JAMA Intern Med.* 2017;177:862-868. Ähnliche Schlussfolgerungen finden sich auch bei Jason H. Wasfy, MD, MPhil, Corwin Matthew Zigler, PhD, Christine Choirat, PhD, Yun Wang, PhD, Francesca Dominici, PhD, and Robert W. Yeh, MD, MS. Readmission Rates After Passage of the Hospital Readmissions Reduction Program. *Ann Intern Med.* 2017 March 07; 166(5): 324–331

¹⁵ Eine weitere Studie, die das US-amerikanische Programm der Senkung von Wiedereinweisungen (Hospital Readmission Reduction Program (HRRP) tendenziell positiv bewertet: Salerno AM, Horwitz LI, Kwon JY, Herrin Jm, Grady JN, Lin Z, Ross JS, Bernheim SM. Trends in readmission rates for safety net hospitals and non-safety net hospitals in the era of the US Hospital Readmission Reduction Program: a retrospective time series analysis using Medicare administrative claims data from 2008 to 2015. *BMJ Open* 2017;7:e016149. Die hier verwendeten Daten sind für die Debatte in Deutschland kaum nutzbar zu machen, da in den USA zwischen ‘Safety Net Hospitals’ und ‘Non-Safety Net Hospitals’ unterschieden wird; der erste Typus ist verpflichtet, alle Patienten unbeschadet ihres sozialen Status aufzunehmen. Die Bewertung der Autoren endet mit folgender grundsätzlicher Bemerkung: “Readmission rate, however, is only one outcome and further studies will be needed to assess

Studie die Wiedereinweisungsraten generell gesunken sind. Diese Studie verweist wiederum auf die Notwendigkeit, belastbare Studiendesigns gerade im Bereich der Anreize zu höherer Qualität zu fordern. Belastbarkeit bedeutet dabei auch, Zeiträume zu markieren, für die Ergebnisse aussagekräftig sein können, eine weithin unterschätzte Problematik von Studien im Bereich der Versorgungsforschung.

Dies belegt auch eine weitere Studie¹⁶, welche die Daten des HRRP-Programms auswertete, mit optimistischem Unterton bezüglich der Erreichung des intendierten Ziels:

“First, because this was an analysis of claims data, our results are subject to discrepancies due to inaccuracies of billing codes.

Second, we cannot be certain that decreases in readmission rates were due to genuine improvements in quality, particularly in other domains.

Third, we do not know whether these improved quality metrics will continue. These questions will be essential to examine in the future.

Fourth, as a study of inpatient readmissions, this analysis could not detect the extent to which hospitals may have decreased inpatient readmissions by increasing observation readmissions

Fifth, we did not model comorbidities as having potential time dependence in their effect on readmissions, which may have affected our results.”

Finally, although we demonstrated that lower-performing hospitals showed the most accelerated improvement in readmission rates, our analysis cannot distinguish between improvement caused by the magnitude of the penalty and improvement caused by changes in health status in different patient populations.”

Damit sind die Grenzen derartiger Vorher-Nachher-Studien ehrlich angesprochen: auch wenn das im Abstract in dieser Deutlichkeit nicht adressiert wird.

Es gibt auch deutlich skeptischer Einschätzungen des HRRP, welche z.B. eine höhere Mortalität für Patienten mit Herzinsuffizienz nach Entlassung ermittelt haben. Fonarow

HRRP’s effect on further patient-centric outcomes of mortality, patient satisfaction, emergency department visits and observation stays as well as its effect on potential hospital closures that would limit care”.

¹⁶ Jason H. Wasfy, MD, MPhil, Corwin Matthew Zigler, PhD, Christine Choirat, PhD, Yun Wang, PhD, Francesca Dominici, PhD, and Robert W. Yeh, MD, MSc. Readmission Rates After Passage of the Hospital Readmissions Reduction Program: A Pre–Post Analysis. *Ann Intern Med* . 2017 March 07; 166(5): 324–331

bemängelt vor allem, dass das Programm gestartet wurde, ohne klare und gut begründete Empfehlungen zur sicheren Umsetzung auszusprechen und ohne eine angemessene prospektive Datenerhebung einzufordern¹⁷.

Ryan et al¹⁸ publizierten 2017 Vierjahres-Ergebnisse des 2013 gestarteten 'Hospital Value-Based Purchasing (HVBP) Program', das bis zu ein Prozent der Medicare-Vergütungen in Akutkrankenhäusern von der Erreichung definierter Qualitätsindikatoren abhängig macht.

Qualitätsverbesserung wurde operationalisiert durch einen kombinierten Indikator klinischer Verbesserung, durch Patientenberichte sowie die 30-Tage-Mortalität von Patienten mit akutem Herzinfarkt, Herzinsuffizienz und Pneumonie. Verglichen wurden Matched Pairs aus Interventions- und Kontrollkrankenhäusern. Das Kernergebnis lautete:

„In our study, HVBP was not associated with improvements in measures of clinical process or patient experience and was not associated with significant reductions in two of three mortality measures“ (Ermittelt wurde eine ignifikante Verbesserung für Pneumonie-Patienten, die von den Autoren auf einen unerklärten Anstieg der Pneumonie-Mortalität in den Kontrollkliniken zurückgeführt wird. Ergänzung durch den Autor).

Es sei wieder auf die Schlussfolgerungen dieser Autorengruppe hingewiesen: Es könne sinnvoll sein, dass Medicare und Medicaid andere Value-Based-Payment Modelle erprobe, einschließlich HRPP oder auch gebündelte Programme. Die Suche nach ‚tatsächlich‘ wirksamen P4P-Modellen geht m.a.W. für diese Autoren weiter.

Saver et al¹⁹ formulierten wenige Jahre vorher in ähnlicher Form ihre Schlussfolgerungen aller bisherigen Bemühungen um fiskalische Anreize zu höherer Versorgungsqualität:

¹⁷ Fonarow GC. Unintended Harm Associated With the Hospital Readmission Reduction Program. JAMA 2018; 320: 2539-2540.

¹⁸ Ryan AM, Krinsky S, Maurer KA, Dimick JB. Changes in Hospital Quality Associated with Hospital Value-Based Purchasing. N Engl J Med. 2017 Jun 15;376::2358-2366.

¹⁹ Saver BG, Martin SA, Adler RN, Candib LM, Deligiannidis KE, Golding J, et al. (2015) Care that Matters: Quality Measurement and Health Care. PLoS Med 12(11):e1001902

- *There is limited evidence that many “quality” measures—including those tied to incentives and those promoted by health insurers and governments—lead to improved health outcomes.*
- *Despite the lack of evidence, these measures and comparative “quality ratings” are used increasingly.*
- *These measures are often based on easily measured, intermediate endpoints such as risk-factor control or care processes, not on meaningful, patient-centered outcomes; their use interferes with individualized approaches to clinical complexity and may lead to gaming, overtesting, and overtreatment.*
- *Measures used for financial incentives and public reporting should meet higher standards.*

Der Artikel endet mit dem Aufruf, bei der Qualitätsorientierung viel stärker auf Patientenzentrierung, Evidenzbasierung der Leistungen und Transparenz des Geschehens zu setzen:

“We must not continue to mismeasure quality by prioritizing time-and cost-effectiveness over principles of patient-centeredness, evidence-based interventions, and transparency.”

Es gibt aber in der Forschung auch immer wieder Stimmen, welche die P4P-Großprojekte in den USA als vorbildlich bezeichnen. Hierzu zählt Nikolas Matthes, der von „durchschlagendem Erfolg“ spricht²⁰ und hierfür u.a. den landesweiten Rückgang von Wiedereinweisungen reklamiert oder auch den Rückgang von Krankenhausinfektionen. Geht man nun in die zitierte Literatur, so weisen bezüglich der Veränderungen der Wiedereinweisungsraten die Autoren korrekt auf zahlreiche Limitationen ihrer Aussage hin, teils bezogen auf die verwendete Methode der retrospektiven Kohortenstudie, teils bezogen auf mögliche andere mit P4P nicht assoziierbare Einflussfaktoren²¹. Auch bezüglich der eindrucksvollen Senkung bestimmter Hospitalinfektionen verweist

²⁰ Matthes N. Pay for Performance. Klinikqualität made in USA. <https://www.gg-digital.de/2018/07/klinikqualitaet-made-in-usa/index.html>

²¹ Desai NR, Ross JS, Kwon JY, Herrin J, Dharmarajan K, Bernheim SM, Krumholz HM, Horwitz LI. Association Between Hospital Penalty Status Under the Hospital Readmission Reduction Program and Readmission Rates for Target and Nontarget Conditions. JAMA. 2016 Dec 27;316(24):2647-2656. doi: 10.1001/jama.2016.18533.

eine Autorengruppe²² wieder auf die Limitationen, die dem Studiendesign geschuldet sind:

“Finally, because we were evaluating the effect of a nationwide policy change, our study design was, by necessity, a pre-post comparison of outcomes. It is not possible to attribute the changes we observed to the policy alone, without consideration of other programs or events that encouraged hospitals and providers to enhance prevention efforts related to the targeted HACs”.

Es ist m.a.W. nicht möglich, aus Studien dieses Typs Aussagen zu kausalen Zusammenhängen zu machen: das ist aus wissenschaftlicher Sicht eigentlich trivial, wird aber in öffentlichen Verlautbarungen von Wissenschaftlern leider nicht immer in der notwendigen Klarheit kommuniziert.

5. VBP und Mortalität

Figuroa et al.²³ haben 2016 die US-amerikanische Literatur zum Outcome-Indikator ‚30-Tage-Mortalität‘ kritisch bewertet:

“Evidence shows that pay for performance programs are largely ineffective in improving patient outcomes. Hospital Value-Based Purchasing (HVBP) is a pay for performance program introduced in the United States in 2011 that incentivizes hospitals based on performance on clinical processes; patient outcomes, including 30 day mortality; patient experience; and cost efficiency. Early results in the first year of the program indicated no effect on clinical process measures or patient experience, but the impact of HVBP on patient mortality is unknown. The introduction of the HVBP program was not associated with an improvement in 30 day mortality of Medicare beneficiaries admitted to US hospitals. Nations considering similar pay for performance programs may want to consider alternative models to achieve improved patient outcomes.”

²²Waters TM, Daniels MJ, Bazzoli GJ, Perencevich E, Dunton N, Staggs VS, Potter C, Fareed N, Liu M, Shorr RI. Effect of Medicare's nonpayment for Hospital-Acquired Conditions: lessons for future policy. JAMA Intern Med. 2015; 175: 347-54

²³ Figuroa, Yusuke Tsugawa Y, Zheng J, Orav EJ, Ashish KJ. Association between the Value-Based Purchasing pay for performance program and patient mortality in US hospitals: observational study. BMJ. 2016; 353: i2214.

Hier findet sich ein erstes Mal in der einschlägigen Literatur die klare Empfehlung, angesichts der negativen Studienergebnisse über Alternativen zu P4P nachzudenken, wenn Qualitätsmängel in der stationären Versorgung systematisch adressiert werden sollen.

Konsequenzen aus der publizierten Literatur zu VBP-Programmen

Die Interdisziplinäre Forschergruppe von Chi et al.²⁴ hat 2016 im Lichte der USA-Erfahrungen mittels eines systematischen Reviews begründet, vor welchen Aufgaben alle – von ihnen potentiell aller publizierten Evidenz zum Trotz als zukunftsfähig eingestufte - Value-Based Purchasing Programme stehen, wenn sie die Erwartungen an einen Paradigmenwechsel in den Vergütungssystemen einlösen wollen. Die Kernaussagen lauten (Paraphrasierung durch den Autor):

- *Die Effektivität von VBP ist bislang von untergeordneter Bedeutung.*
- *Möglicherweise waren die wirtschaftlichen Anreize zu gering.*
- *Die eingesetzten Qualitätsindikatoren könnten für die Praxis zu komplex und zu intransparent gewesen sein.*
- *Der Zeitabstand zwischen Messung der Performance und Anreizen könnte unattraktiv gewesen sein.*
- *Medicare beziffert die Incentives als Prozent der Standardvergütung und nicht als Zusatzvergütung für höhere Leistungen.*
- *Die vielfältigen Programme unterschiedlicher Kostenträger sorgen für eine Verwirrung stiftende Gesamtlage.*

Des weiteren werden die mit dem Qualitätsbegriff verbundenen Probleme wie folgt beschrieben:

- *Es ist bislang strittig, in wieweit die verwendeten Sets von Qualitätsindikatoren von klinischer Relevanz sind.*

²⁴ Chee TT, Ryan AM, Wasfy JH, Borden WB. Current State of Value-Based Purchasing Programs. Circulation. 2016; 133: 2197-205.

- *Bezüglich der prinzipiell anzustrebenden Outcomeindikatoren stellt sich stets die Frage, ob die obligatorische Risikoadjustierung die tatsächlichen medizinischen Risiken auch abbilden.*
- *Noch unklarer ist die Frage einer fairen Berücksichtigung der sozioökonomischen Faktoren.*
- *VBP erfasst immer nur einen Teil der medizinischen Angebote einer Praxis oder Klinik, es bleibt insofern offen, ob sie für die Gesamtqualität stehen.*
- *Bisherige VBP-Programme kalkulieren nicht ausreichend die eigene Kosten-Nutzen Bilanz.*
- *Insgesamt ist die wissenschaftliche Fundierung der Qualitätsmessung zu verbessern.*

Es geht weiter um das Design der Anreizstrukturen:

- *Höhe, Empfänger (Individuen oder Organisationen), Zielerreichung des Standards oder Verbesserung, Häufigkeit von Datenrückmeldungen (Qualität und Anreize), Einschluss nichtökonomischer Wertschätzung stehen jeweils in Diskussion.*
- *Die Weiterentwicklung müsste Aufgabe wissenschaftlicher Studien sein, während bis dato die Anreizstrukturen administrativ festgelegt werden.*

Als nächstes werden Auswirkungen auf Gesundheitschancen diskutiert:

- *VBP können Anbieter bestrafen, die besonders benachteiligte Bevölkerungsgruppen versorgen. Die bestehenden Adjustierungen können dies nicht zuverlässig ausschließen.*
- *Der Sozialstatus beeinflusst die Ergebnisse der Erhebung von Patientenzufriedenheit, ohne dass die Programme dies bisher adäquat berücksichtigen können: dies kann Krankenhäuser mit striktem Versorgungsauftrag benachteiligen.*

Weiter: die Auswirkungen der Qualitäts-Fokussierung können problematisch sein:

- *Anbieter können verführt werden, nicht fokussierte Versorgungsbereiche zu vernachlässigen.*

- *Die Neuentwicklung von ‚broad outcomes measures‘ ist erforderlich, um derartige unerwünschte Effekte abzuschwächen.*

Weiter wird die große Varianz der Vergütungen problematisiert:

- *Selbst in einer umschriebenen Region driften die Vergütungen für die Anbieter weit auseinander; dies ist insbesondere zwischen Medicare- und Privatversicherungen spürbar. Im nationalen Rahmen ist dies noch auffälliger.*
- *Es stellt sich die Frage, ob die Höhe von Incentives dies berücksichtigen muss.*

Schließlich, so die Autoren, ist erst noch zu erproben, in welcher Balance die unterschiedlichen Elemente von VBP und traditionelle Fee-For-Service Vergütungen eingesetzt werden sollten. In ihren Schlussfolgerungen betonen die Autoren erneut, dass die bisherigen Ansätze von VBP nur „mixed and modest“ Ergebnisse hervorbringen konnten. Aus wissenschaftlicher Sicht sei zudem erheblich, dass einerseits viele Evaluationsstudien nur kurze Zeiträume abgegriffen haben, dass andererseits bei langer Laufzeit das Risiko neuer, initial unbekannter Confounder vergrößere.

Versorgungsqualität: ein Aufruf zur Verbesserung der Studienqualität

Grady et al²⁵ haben 2017 im JAMA mit Blick auf die bisher publizierten Daten zu P4P und VBP dazu aufgerufen, relevantere und methodisch belastbarere Forschungsprogramme zu entwickeln und als Standard zu definieren:

- *“Studies should document that the intervention actually resulted in better health outcomes, rather than focus only on changes in health care processes, use, or costs. Studies should examine potential adverse outcomes as well as benefits. In addition, we encourage the evaluation of costs, including the costs of the intervention, to enable the determination of value. At a minimum, quality improvement studies should have a trends, which can have a substantial association with the quality of health care. Better*

²⁵ Grady D, Redberg RF, O’Malley PG: Quality Improvement for Quality Improvement Studies. JAMA intern Med 2018; 178 (2): 187

yet, quality improvement studies should use randomization and blinding. Using the electronic health record to accomplish randomization and to collect study outcomes makes randomization and blinding more feasible than in the past.

- *Studies should be registered at [Clinicatrials.gov](https://www.clinicaltrials.gov) before the study starts, prespecifying the population to be enrolled and outcomes to be ascertained. We encourage higher-quality reporting guided by SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines*
- *For example, clustered randomization and randomized stepped wedge designs are often feasible when randomization of individual patients is not. And blinding of outcome assessment is almost always feasible, even when blinding of the intervention itself is not. In summary, the editors of JAMA Internal Medicine wish to initiate a quality improvement effort for quality improvement studies. We aim to ensure that the evidence supporting the effectiveness and safety of quality improvement interventions is rigorous and generalizable. We look forward to receiving reports of studies whose design, conduct, and analyses are most likely to generate compelling valid evidence, and help clinicians improve the health of their patients.”*

Diese Forderungen sind aus Sicht der Gesundheitswissenschaften selbsterklärend. Dass sie nicht längst in die Tat umgesetzt sind, ist Ausdruck eines gewissermaßen intuitiven Vertrauens in plausible Vorannahmen zum Nutzen ökonomischer Anreize. Insofern handelt es sich hierbei um ‚sympathische‘ Zeitgeistströmungen, nicht um klassisch wissenschaftlich begründete Handlungskonzepte.

Eine weitere aktuelle Bilanzierung der auf ärztliche Leistungen gerichteten Bemühungen um Anhebung der Qualität durch ökonomische Anreize findet sich bei Khullar u.a.²⁶. Sie bezweifeln die ökonomisch begründete Vorannahme, dass Qualität prinzipiell finanziell angereizt werden müsse, und zitieren den für seine dem Mainstream der Gesundheitsökonomien widersprechende Grundhaltung bekannten Ökonomen Uwe Reinhardt:

²⁶ Khullar D, Wolfson D, Casalino LP. Professionalism, Performance, and the Future of Physician Incentives. JAMA 2018; 320: 2419-2420

„The idea that everyone’s professionalism and everyone’s good will has to be bought with tips is bizarre“²⁷.

Die Autoren sehen fünf grundlegende Probleme im dominierenden P4P-Konzept:

1. Finanzielle Anreize können die intrinsische Motivation aushöhlen, vor allem bezüglich des sozialen Verhaltens und der komplexen kognitiven Aufgaben der Ärzteschaft. Es komme darauf an, die Ärztinnen und Ärzte in die Erkennung von Verbesserungsschwächen und deren Beseitigung einzubeziehen und dabei bestehende Hürden zu benennen. Als vorbildlich betrachtet diese Autorengruppe das Konzept der international tätigen Cleveland Clinic, in deren Mittelpunkt eine medizinische Fakultät besteht, die mit einem systematischen Erziehungsauftrag in Forschung und Lehre betraut ist. Als dessen wichtiger Kern wird – neben professionsübergreifenden Ansätzen der Fortbildung – das systematische und fortlaufende Training von Führungskräften betrachtet, die eine entsprechende qualitätsfördernde Kultur erst ermöglichen können²⁸.
2. Die klassische ökonomische Vorannahme lautet, dass Ärztinnen und Ärzte durch Anreize zu konsequenterem Arbeiten motiviert werden müssen. Damit werde außer Acht gelassen, dass Qualität stark von Prozessen auf der Systemebene geprägt werde. Mit Blick auf die immer noch stark vertretenen Kleinpraxen in der ambulanten Versorgung in den USA ermittelten Khullar u.a.²⁹ Unterstützungsbedarf der Praktiker v.a. im Management des Praxisablaufs, in der Datenverarbeitung, im Qualitätsbericht und der elektronischen Krankenakte: genau diese Defizite lassen sich durch unterschiedlich strukturierte und finanzierte Dienstleister beheben.
3. An die Stelle der Zusatzvergütung für Einzelpersonen könnten Anreize für die Teams treten, welche die Qualität verantworten. Als Referenz verweisen sie auf Geisinger Health System, ein großes Non-Profit-Unternehmen mit fast 1600 überwiegend angestellten Ärztinnen und Ärzten, deren Einkommen zu etwa 20

²⁷ Zitat nach Hartocolis A. Pay for Performance extends to health care in New York State experiment. New York Times, March 2, 2015

²⁸ Stoller J. The Cleveland Clinic: a distinctive model of American medicine. *Annals of Translational Medicine* 2014; 2: 33. Doi: 10.3978/j.issn.2305-5839.2013.12.02; Hess CA, Barss C, Stoller JK. Developing a leadership pipeline: the Cleveland Clinic experience. *Perspect Med Educ* 2014;3:383-90

²⁹ Khullara D, Burke GC, Cassalino LP. Can Small Physician Practices Survive? Sharing Services as a Path to Viability. *JAMA* 2018; 319: 1321-1322

Prozent vom Erreichungsgrad definierter Behandlungsziele ihrer Patientenkollektive abhing³⁰. Interessant ist nun, dass Geisinger diese Strategie 2016 zugunsten von oberhalb des nationalen Durchschnitts vereinbarten Festgehältern verlassen hat, gekoppelt an die Förderung einer Unternehmenskultur, die mit dem Begriff „social compact“ bezeichnet wird³¹. Geisinger zog die Konsequenzen aus der Analyse der weltweiten P4P-Erfahrungen. Interne Qualitätsziele bleiben gleichwohl Bestandteil der Unternehmenskultur, so mittels Anwendung der HEDIS-Qualitätsindikatoren³² und der Erhebung der Patientenzufriedenheit.

4. Als grundlegende Alternative wird der Abbau von Überversorgung bezeichnet. Das Konzept lehnt sich an Wennbergs seit Jahrzehnten erprobte Methode der Analyse unerklärter Leistungsvariation in benachbarten Versorgungsgebieten³³ an. Im Blue Cross Shields of Massachusetts werden Ärzte darüber informiert, wo sie auf der erhobenen Skala liegen. Und man setzt auf den Dialog der Meinungsführer der jeweiligen Fachgebiete mit den Versorgern. Es gibt Anhaltspunkte dafür, dass sich hiermit nicht medizinisch erklärbare hohe Leistungsdichten im Bereich Pharmakotherapie und endoskopischer Diagnostik reduzieren lassen³⁴. Die ‚Choosing Wisely‘-Initiative findet hier ihren empirischen Ursprung³⁵.
5. Inwieweit Impulse zur Verbesserung der Qualität von den Beschäftigten in der Praxis aufgegriffen werden, hängt – und dies führt zu einem weiteren grundlegenden Alternativvorschlag – entscheidend von dem Ineinandergreifen der organisatorischen Rahmenbedingungen mit der Förderung einer patientenzentrierten Kultur der jeweiligen Institution ab. In einem systematischen Review

³⁰ Lee TH, Bothe A, Steele GD. How Geisinger Structures Its Physician's Compensation To Support Improvement in Quality, Efficiency, And Volume. Health Affairs 2012; 31: 2068-2073

³¹ Butcher L. Geisinger Scraps Physicians Pay-for-Performance. HFMA. https://www.hfma.org/Leadership/Archives/2017/June/Geisinger_Scraps_Physician_Pay-for-Performance/

³² <https://www.ncqa.org/hedis/>

³³ Wennberg J, Gittelsohn A. Small Area Variation in Health Care Delivery. Science 1973; 182: 1102-1108

³⁴ Breene RA, Beckman HB, Mahoney T. Beyond the efficiency index. Finding a better way to reduce overuse and increase efficiency in physician care. Health Affairs 2008; 27: w250-259

³⁵ Colla CH, Morden NE, Sequeist TD, Schpero WL, Rosenthal MB. Choosing Wisely: Prevalence and Correlates of Low-Value Health Care Services in the United States. J Gen Intern Med 2014; 30: 221-228

haben Braithwaite u.a.³⁶ die verwertbaren Studien zum Zusammenhang von Organisationskultur und patientenrelevanten Outcomes (von Mortalitätsraten über Patientenzufriedenheit bis zu Hospitalinfektionen) ausgewertet und den vermuteten positiven Zusammenhang gesehen, wobei es bezeichnend ist, dass es lediglich 4 Interventionsstudien unter den 62 einbezogenen Studien gab und dass die Mehrzahl der Studien aus dem sozialwissenschaftlichen und nicht dem medizinisch-klinischen Milieu stammen. Khullar et al bewerten die Befunde wie folgt: „Organizations in which physicians are encouraged to identify problems, engage in care improvement processes, and discuss challenges with leadership likely to have more robust and lasting quality improvements than those relying on narrow financial incentives³⁷“.

Khullar et al betonen ausdrücklich, dass alle alternativen Modelle zur Verbesserung der Versorgungsqualität in experimentellen Designs untersucht werden müssen, vorzugsweise in clusterrandomisierten Studien. Und sie raten dazu, Ansätze zur Förderung patientenzentrierter Kultur nicht nur quantitativ zu evaluieren sondern auch qualitative Methoden einzusetzen.

Bilanz

Dieser narrative Review hat keinen Anhaltspunkt dafür gefunden, dass sich die Bewertung von P4P-Modellen seit dem Gutachten der BQS entscheidend geändert hat. Es kann selbstverständlich durch ein systematisches Review untersucht werden, ob hier wesentliche Studien übersehen worden sind: das ist relativ unwahrscheinlich, weil auch Studien bzw. Einschätzungen mit positiver Bewertung eingeschlossen wurden und echte Sprunginnovationen im gesundheitsökonomischen und gesundheitswissenschaftlichen Bereich rasch kommuniziert werden. Einigkeit zwischen Befürwortern und Skeptikern besteht darin, dass P4P kein homogenes Modell darstellt sondern eine Vielzahl unterschiedlicher Bonus- und Malusregelungen umschreibt. Ein Passepartout zur Anwendung von P4P in ambulanten oder stationären Settings unterschiedlicher

³⁶ Braithwaite J, Herkes J, Ludlow K, Testa L, Lamprell G. Association between organizational and workplace cultures, and patient outcomes: systematic review. *BMJ Open* 2017; 7: e017708. Doi: 10.1136/bmjopen-2017-017708

³⁷ Khullar et al, 2018, S. 2420

Systeme existiert nicht in Ansätzen. Ob diese alles in allem ernüchternde Bilanz nun trotzdem immer noch den Anfang einer neuen Vergütungswelt darstellt, die Qualitätsbewertungen direkt beim Abschluss von Verträgen ins Kalkül zieht, oder ob es sich am Ende doch bloß um eine Modeerscheinung handelt, mit der die irrige Erwartung verbunden ist, bisherige Vergütungsmodelle könnten damit eines Tages in erheblichem Maße ersetzt werden, ist mangels entsprechender prognostischer Fähigkeiten im wissenschaftlichen Sektor nicht endgültig bestimmbar.

Es zeichnen sich idealtypisch für die deutsche gesetzliche Krankenversicherung zwei Wege ab, soweit P4P nicht aus gutem Grund ad acta gelegt wird:

1. Politisch legitimierte Administration führt P4P-Ansätze ohne klare Evidenznachweise auf Grund des Plausibilitätsprinzips ein und versucht, durch nachgehende Evaluationen (etwa durch unterbrochene Zeitreihen und/oder Matchingverfahren) die Überlegenheit von P4P zu belegen. Von diesem Weg ist aus wissenschaftlicher Sicht grundsätzlich abzuraten.

2. Es werden einige ausgewählte prospektive, randomisierte Studien für P4P entwickelt incl. ausreichender Laufzeiten. Hierbei sind die Möglichkeiten der Orientierung auf klinisch relevante Outcomes auszuschöpfen. Transparenz ist auch über die Investitions- und Laufkosten derartiger P4P-Programme herzustellen. Ob ein solcher Weg gerechtfertigt ist, sollte angesichts der dürftigen bisherigen Resultate gründlich überlegt werden.

Ein weiterer Aspekt: es erscheint – nicht zuletzt auf Grund des Kassenwettbewerbs und der dadurch bedingten Hürden für die Definition repräsentativer Interventions- und Kontrollgruppen - wenig wahrscheinlich, dass in Deutschland methodisch hochwertige Studien zur Implementierung von P4P durchführbar sind.

Aus Sicht des Autors ist vor allem sinnvoll, die Frage der Entwicklung neuer Vergütungssysteme unter Qualitätsgesichtspunkten systematisch einzubinden in eine Gesamtbetrachtung der Faktoren, die nachweislich oder mutmaßlich Einfluss auf die klinisch relevante Qualität ambulanter wie stationärer Versorgung haben. Es ist zu erwarten, dass Optimierungsversuche durch VBP auch bei weiterer Entwicklung ihrer

Instrumente im günstigsten Fall einen außerordentlich geringen Grenznutzen erzeugen, wenn man dem Konzept von VBP andere qualitätssichernde Faktoren gegenüberstellt. Solche sind, ohne jeden Versuch von Vollständigkeit:

- Die möglichst umfassendere Bewertung aller für die GKV geltend gemachter Innovationen (produkt- wie systembezogen)
- Die Herausnahme nicht nutzenstiftender Methoden aus der Versorgung
- Anstrengungen zur Hebung und Sicherung des Bildungs- und Kompetenzniveaus der Fachberufe des Gesundheitswesens (von der Primärausbildung über Fortbildung und Supervision bis zur Rezertifizierung)
- Bessere Stratifizierung der Angebote zwischen Primary Health Care und zentralisierter stationärer Hochleistungsmedizin
- Förderung der Versorgungskultur in Klinik und Praxis über die bestehenden QM-Ansätze hinaus.

Es ist mit anderen Worten angeraten, weit über die P4P-Debatte hinaus das weitere Vorgehen zur Frage der systematischen Verbesserung der Qualität im Versorgungsalltag im Lichte der bisherigen Erfahrungen und der publizierten methodisch akzeptablen Literatur grundlegend neu zu diskutieren.

Nachbetrachtung

Man muss sich fragen, warum ein Konzept wie P4P Jahrzehnte überdauern kann, obwohl es zu keinem Zeitpunkt stabile Nutzenbelege gegeben hat. Es hat – wenn auch vereinzelt- schon frühzeitig Stimmen gegeben, welche vor der Weiterverfolgung des P4P Ansatzes gewarnt haben; hierfür sei hier Donald Berwick's Artikel „The Toxicity of Pay for Performance“ genannt³⁸. Eine weitere gewichtige Kritikerin des Konzeptes, Meredith Rosenthal, hat neben dem Hinweis auf den fehlenden Nutzen von P4P bei der Reform im Gesundheitswesen auch vergeblich darauf hingewiesen, dass die Umsetzung der Effizienzsteigerung durch ökonomische Anreize auch in anderen gesellschaftlichen Sektoren immer nur durchwachsene Ergebnisse gezeitigt habe³⁹.

Die vielleicht beste Analyse der P4P-Geschichte hat die Politikwissenschaftlerin Sandra Tanenbaum vorgelegt⁴⁰. Sie deutet die Kernidee von P4P als Geschenk an die Politik, die Themen Kostenausweitung und Versorgungsmängel mit der Losung „P4P“ gewissermaßen unangreifbar miteinander zu verbinden: „P4P offers to join cost and quality under the rubric of value“. Politik kann – anders formuliert – die Kritik an den allzu zögerlichen Bemühungen um höherer Qualität dergestalt mit der immer hoch umkämpften Vergütungsfrage kombinieren, dass nunmehr das Einkommen der Ärzte von der Qualität abhängig gemacht werden soll: Widerspruch zu P4P wird in einer tief verwurzelten Kultur des ‚most bang for the buck‘ nahezu unmöglich. Hauptprofiteur von P4P ist für Tanenbaum somit die Politik, die eine moralisch gerechtfertigte Alternative zu den immer hoch umstrittenen Versuchen der ordnungspolitischen Preisregulierung gefunden zu haben glaubt⁴¹.

Ein abschließender Hinweis sei darauf verwendet, dass fast wie selbstverständlich bei P4P immer an die Ärzteschaft gedacht wird. Soweit Qualitätsverbesserungen in der

³⁸ Berwick DM. The Toxicity of Pay for Performance. *Quality Management in Health Care* 1995; 4: 27-33

³⁹ Rosenthal MR. What is the Empirical Basis for Paying for Quality in Health Care? *Medical Care Research and Review* 2006; 63: 135-157

⁴⁰ Tanenbaum SJ. Pay for Performance in Medicare: Evidentiary Irony and the Politics of Value. *J Health Polit Policy Law* 2009; 34: 717-746

⁴¹ Tanenbaum SJ. What is the Value of “Value-Based Purchasing”? *J Health Polit Policy Law*. 2016 Oct;41:1033-45

Versorgung tatsächlich angestrebt werden, ist dies ohne Berücksichtigung der Pflege aber nicht vorstellbar. Beiträge aus der Pflege selbst, welche die spezifischen Belange der Pflege thematisieren, gehen in der Generaldebatte um P4P nahezu vollständig unter⁴². Nelson und Folbre haben auf einen geradezu makabren Teilaspekt dieser Debatte hingewiesen⁴³. Der Ökonom Anthony Heyes argumentierte 2005 in dem durchaus angesehenen ‚Journal of Health Economics‘, dass eine bessere Bezahlung von Beschäftigten der Pflegeberufe den Anteil von Pflegerinnen und Pfleger senke, die eine Begabung und Berufung für ihre Tätigkeit haben⁴⁴. Eine ähnliche Argumentation bezogen auf die Ärzteschaft oder gar Top-Verdiener der Gesellschaft ist in der P4P-Debatte noch nicht aufgetaucht. Demgegenüber liegt der Gedanke nahe, dass eine bessere Bezahlung von Pflegerinnen und Pflegern gerade der Wahrung und Verbesserung der Struktur-, Prozess- und Ergebnisqualität in der Versorgung dienlich sein dürfte. Dieser erweiterte Blick erscheint angesichts der Einführung der P4P-Debatte am Ende mehr als legitim.

⁴² Hellier S. Maybe Someone Should Tell the Nurses. *The Health Care Manager* 2010; 29: 172-178; Bodrock JA, Mion LC. Pay for Performance in Hospitals: Implications for Nurses and Nursing Care. *Q Manage Health Care* 2008; 17: 102-111

⁴³ Nelson JA, Folbre N. Why a Well-Paid Nurse is a Better Nurse. *Nurs Econ* 2006; 24: 127-130

⁴⁴ Heyes A. The economics of vocation or „why is a badly paid nurse a good nurse?“. *Journal of Health Economics* 2005; 24: 561-569

Über den Autor

Prof. Dr. med. Norbert Schmacke. Geb. 1948. Arzt für innere Medizin, öffentliches Gesundheitswesen und Sozialmedizin. Habilitation in Gesundheitswissenschaften/ Schwerpunkt Sozialmedizin an der Universität Bremen. Hochschullehrer am Institut für Public Health und Pflegeforschung der Universität Bremen (www.ipp.uni-bremen.de). Mitglied in der Deutschen Gesellschaft für Allgemeinmedizin und im Deutschen Netzwerk evidenzbasierte Medizin e.V. Bis Juni 2018 stellvertretendes unparteiisches Mitglied im Gemeinsamen Bundesausschuss.

Kontakt Daten: Marsse 48, 28719 Bremen. schmacke@uni-bremen.de



Institut für Public Health und Pflegeforschung (IPP)

Fachbereich 11 Human- und
Gesundheitswissenschaften
Grazer Straße 4
28359 Bremen

IPP Geschäftsstelle

Grazer Straße 4
28359 Bremen
Phone (+49) 0421 - 218 68980/-68880
E-Mail hmertesa@uni-bremen.de
www.ipp.uni-bremen.de

IPP-Schriften, Ausgabe 15

Norbert Schmacke: Pay for Performance (P4P)
Bilanz einer gesundheitsökonomischen Ideologie.
Ein narrativer Review

ISSN 1867-1942 (Internetausg.)

Februar 2019