

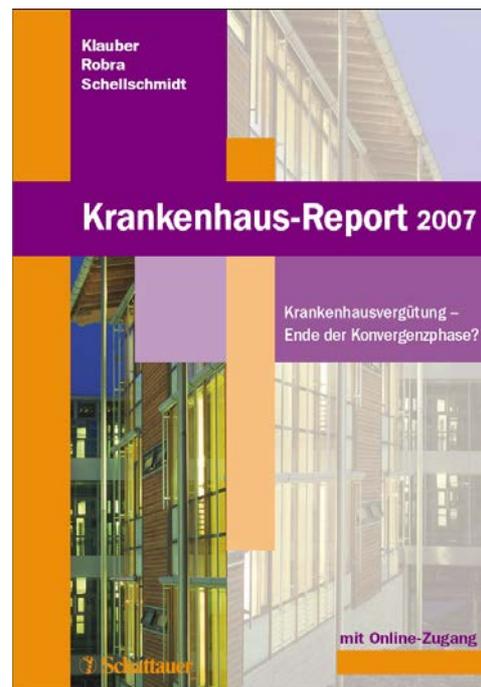
Krankenhaus-Report 2007

„Krankenhausvergütung – Ende der Konvergenzphase“

Jürgen Klauber / Bernt-Peter Robra /
Henner Schellschmidt (Hrsg.)

Schattauer (Stuttgart) 2008

Auszug Seite 157-169



9	Pay-for-Performance: Neue Impulse für den Wettbewerb zwischen Krankenhäusern?	157
	<i>Markus Lünen, Andreas Gerber und Karl W. Lauterbach</i>	
9.1	Pay-for-Performance als Neudefinition des Wettbewerbs	157
9.1.1	Definition und Entwicklung des Begriffs	157
9.1.2	Formale Wirkungen von Pfp	159
9.1.3	Indirekte Formen von Pfp	160
9.2	Anfänge und Erfahrungen im Ausland	160
9.2.1	Reviews über Pfp in der Literatur.....	161
9.2.2	Centers for Medicare & Medicaid Services (CMS)	162
9.2.3	Großbritannien	163
9.3	Probleme des Pfp und mögliche Übertragung auf das deutsche Gesundheitswesen	164
9.3.1	Kritik an Pfp	164
9.3.2	Relevante Fragen bei einer Übertragung nach Deutschland.....	165
9.4	Fazit	167
9.5	Literatur.....	168

9 Pay-for-Performance: Neue Impulse für den Wettbewerb zwischen Krankenhäusern?

Markus Lungen, Andreas Gerber und Karl W. Lauterbach

Abstract

Pay-for-Performance (PfP) bezeichnet im allgemeinen die Verknüpfung der Qualität einer medizinischen Leistung mit der Höhe der Vergütung. Im Ausland existieren bereits vielfältige PfP-Programme. Jedoch bleibt die Evidenz für eine erreichbare Effektivität und Kosteneffektivität bisher gering. Offen bleibt bisher, ob in einem ersten Schritt bereits die krankenhausspezifische Publikation von aussagekräftigen vorhandenen Qualitätsdaten ausreichen würde, um einen Qualitätswettbewerb zu entfachen. Eine Einführung von PfP in Deutschland könnte auf die Daten der externen Qualitätssicherung nach § 137 SGB V aufbauen.

Pay-for-Performance (PfP) is often defined as connecting quality of care with the level of reimbursement. Other countries have already introduced PfP-programs in recent years, but evidence for effectiveness and cost-effectiveness is still rare or missing. However, it is yet unknown whether public reporting of significant quality data suffices to ignite a competition for quality between providers. The use of PfP programs in Germany could be based on using data of the statutory quality assurance programs (§ 137 Code of Social Law).

9.1 Pay-for-Performance als Neudefinition des Wettbewerbs

9.1.1 Definition und Entwicklung des Begriffs

Für den Begriff des Pay-for-Performance (PfP; teilweise in der Literatur auch P4P; übersetzt etwa mit Ergebnis- oder qualitätsorientierte Vergütung) gibt es keine allgemein anerkannte Definition.¹ Meist wird darauf hingewiesen, dass PfP dem Leistungserbringer zusätzliche finanzielle Anreize bietet, um zuvor definierte (Qualitäts-) Ziele zu erreichen. Neben der Steigerung der Versorgungsqualität können

¹ Zuvor waren auch andere Begrifflichkeiten gebräuchlich, etwa fee-for-quality, paying-for-quality oder auch payment-by-results, die jedoch teilweise andere Inhalte vertreten.

die Ziele auch Effizienz und/oder ein verbesserter Zugang für Versicherte sein (Henley 2005, Baumann/Dellert 2006).² So ist denkbar, dass PfP-Programme der Krankenversicherungen (in den USA) die Inanspruchnahme der Versicherten zu vorher selektierten Leistungsanbietern lenken sollen, etwa über verringerte Zuzahlungen. In Deutschland würde dies in der Struktur einem Bonus-Programm entsprechen.

Hier soll unter PfP im engeren Sinne die direkte Ausrichtung der Vergütungshöhe an der Qualität der erbrachten Leistung eines Krankenhauses oder Arztes verstanden werden. Der Begriff wurde, wie viele andere Innovationen in der Leistungsvergütung der letzten 30 Jahre, geprägt von der MedPAC (Medicare Payment Advisory Commission) beziehungsweise den CMS (Centers for Medicaid & Medicare Services) (Milgate/Cheng 2006). Die CMS starteten zunächst ein Pilotprojekt für zehn große Gruppenpraxen im ambulanten Bereich. Die Bonuszahlungen sollten für die Einhaltung von Disease-Management und den Einsatz von EDV entrichtet werden. Es war vorgesehen, sie vorwiegend aus vorherigen Einsparungen zu generieren. Dadurch gab es von Beginn an den Vorbehalt, dass Kosteneinsparungen das eigentliche Ziel von PfP seien.³

Für den Krankenhausbereich kündigten die CMS im Frühjahr 2003 ein Projekt mit 278 Krankenhäusern an. Das Projekt wird weiter unten im Detail beschrieben. In der Folge traten vermehrt Managementgesellschaften auf, welche die Verbindung zwischen Leistungserbringer, Krankenversicherung und Qualitätsmessung herstellten. Sie übernahmen die Auswertung, Lenkung und Evaluation von PfP-Programmen.⁴ Oftmals bauten die Programme auf bestehenden Qualitätssicherungsinitiativen auf, etwa HEDIS (Health Plan Employer Data and Information Set) des NCQA (National Committee for Quality Assurance). Übertragen auf Deutschland würde dies in etwa bedeuten, dass die Messungen der BQS (Bundesgeschäftsstelle Qualitätssicherung) mit einer finanziellen Ausstattung versehen würden. Im Jahr 2005 schlug die MedPAC in einem Report vor, dass Medicare die Leistungsanbieter generell nach ihrer Leistung („performance“) bezahlen sollte. Zunächst sollten einfache Strukturvorgaben (Einsatz elektronischer Patientenakten etc.) belohnt werden. Für spätere Jahre wurden ergebnisorientierte Messwerte vorgeschlagen (Milgate/Cheng 2006).

2 Siehe die Definition bei den MeSH-Terms unter „Financial Incentives“. <http://www.ncbi.nlm.nih.gov/> (Zugriff Mai 2007).

3 Siehe die Stellungnahme der Vertragsärzte AAFP unter <http://www.aafp.org/online/en/home/policy/policies/p/payforperformance.html> (Zugriff Mai 2007) Dort werden Kriterien für PfP aus Sicht der Ärzte aufgestellt.

Siehe die Stellungnahme der American Medical Association (AMA) unter <http://www.ama-assn.org/ama1/pub/upload/mm/368/principles4pay62705.pdf> und Anforderungen an Programme unter <http://www.ama-assn.org/ama1/pub/upload/mm/368/guidelines4pay62705.pdf> (Zugriff Mai 2007).

4 Siehe etwa Bridges to Excellence (BTE) in Zusammenarbeit mit der NCQA unter <http://www.bridgestoexcellence.org/> (Zugriff Mai 2007).

9.1.2 Formale Wirkungen von PfP

Formal würde PfP eine Weiterentwicklung der Funktion der Vergütungshöhe für akutstationäre Versorgung in Deutschland bedeuten. Im System tagesgleicher Pflegesätze, gültig bis 2003/2004, war die Vergütungshöhe eine Funktion der Verweildauer und der Höhe der Abteilungs- und Basispflegesätze. Somit waren die Krankenhauskosten noch indirekt Parameter der Vergütungsfunktion (resource based payment system). DRGs modifizierten die Funktion, indem sie Diagnosen und Prozeduren als wesentliche Bestandteile der Vergütungsfunktion einführten und somit die Morbidität einzelner Patientenfälle. Daneben behielt die Verweildauer bei Kurz- oder Langliegern eine untergeordnete Funktion.

PfP würden die Vergütungsfunktion um Qualitätsindikatoren erweitern. Diese können sowohl die Struktur (und somit eher krankenhauses- oder abteilungsbezogene Ausstattungsmerkmale) als auch Prozesse oder Ergebnisse abbilden (Lüngen/Lauterbach 2002a). Die durch PfP gesetzten Anreize sollen generell bewirken, dass die Qualität der Versorgung eine größere Rolle bekommt. Bereits DRGs führten dazu, dass einige Formen der Fehlversorgung in der Vergütungsfunktion des Krankenhaus negative Auswirkungen hinterließen. So führten Komplikation nicht zwangsläufig zu höheren Entgelten, nicht indizierte Verfahren (etwa Labor, Bildgebung) resultierten in einer Minderung der Deckungsbeiträge. Jedoch konnte bei der Nutzung von DRG-Systemen auch Unterversorgung, etwa frühzeitige Entlassung oder Unterlassung angemessener Leistungen, auftreten (Lüngen/Lauterbach 2002b). Inwieweit PfP generell die mit DRGs begonnene Betonung von Effizienz und Qualität bei der Leistungserbringung fortführt oder aber eine notwendige Ergänzung der möglichen negativen Qualitätsauswirkungen von DRGs sind, hängt von der konkreten Ausgestaltung ab.

In der Literatur wird im Hinblick auf die formale Ableitung kritisch gesehen, dass ein abgeschlossenes theoretisches Modell, welches die Anreize von PfP erklären kann, nicht existiert (Frolich et al. 2007). Die Auswirkungen finanzieller Anreize wurden für den medizinischen Bereich zwar bereits vielfältig empirisch untersucht (Conrad et al. 2002, Armour et al. 2001). Unklar blieben jedoch die Auswirkungen der Rahmenbedingungen der Anreizsysteme sowie der Einfluss psychologischer Faktoren. So können Prozessparameter vom Arzt eher beeinflusst werden als Ergebnisparameter und stoßen daher auf größeres Verständnis. Ebenso reicht die in der Ökonomie verbreitete Grenzbetrachtung nicht aus, da auch die absolute Höhe der Vergütung und die absolute Anzahl betroffener Fälle, und nicht nur der Zuwachs, für einen Leistungsanbieter von Bedeutung sein können. Zudem kann bei einem Einzelvergütungssystem die zusätzlich erforderliche Leistung zur Bereitstellung hoher Qualität auch zusätzlich abgerechnet werden. Die Anreize sind daher abweichend von Systemen, in denen Pauschalierungen gelten und die Opportunitätskosten höher liegen. Hinzu kommen die in der Medizin ausgeprägten Prinzipal-Agent-Situationen mit Informationsasymmetrien, welche es erst erforderlich machen, dass Qualität statt einer Ex-ante-Definition einer Ex-post-Anreizbetrachtung unterliegt. (Frolich et al. 2007; Rosenthal & Frank 2006)

9.1.3 Indirekte Formen von PfP

In der ökonomischen Theorie gilt, dass der rational entscheidende Patient beziehungsweise Versicherte nach Studium der Qualitätsdaten eigenständig entscheiden kann, welche Leistung er bei welchem Anbieter in Anspruch nimmt. Veröffentlichung der Daten zu Qualität (Public Reporting, PR) würde ausreichen, eine darüber hinaus gehende Verknüpfung von Qualität mit der Erlöshöhe wäre nicht mehr erforderlich. Bereits die Patientenabwanderung beziehungsweise Mehrkosten, die durch geringe Qualität einer Leistung entstehen, wären ökonomische Anreizfaktoren. Das Pfp-Modell geht somit indirekt davon aus, dass diese indirekten Anreize (auch aus normativer Sicht) nicht ausreichend stark sind oder nicht wirken. Die Ursachen für eingeschränkte Wirksamkeit von PR sind in der Literatur teilweise extensiv beschrieben worden (Lungen/Lauterbach 2002a, Rosenthal et al. 2004). Unter anderem sind dies normativ-ethische Faktoren (einige Leistungen sollten nicht einer Qualitäts-/Preis-Abwägung unterliegen; generell sollte Qualität höher bewertet werden als Effizienz), Faktoren in der Person der Versicherten (Unfähigkeit zur Suche und Interpretation von Information; Immobilität beziehungsweise Notfalleigenschaften der Behandlung; Abhängigkeit von Prinzipal-Agent-Situation mit asymmetrischer Information und Anbieter-induzierter Nachfrage) und Faktoren im Gesundheitswesen (mangelnde Erkundung der Zahlungsbereitschaft; Moral Hazard vor dem Hintergrund des Versicherungsgedankens).

Aus diesen Gründen kann Pfp Qualität direkt fördern, um die unvollkommene Nachfragesituation auf Gesundheitsmärkten zu korrigieren oder abzumildern. Der Wettbewerb wird stärker auf Qualität gelenkt, beziehungsweise Veränderungen der erzielbaren Deckungsbeiträge werden für den Leistungserbringer im Hinblick/Verhältnis zu den Aufwendungen zur Veränderungen seiner Qualität exakter kalkulierbar.

9.2 Anfänge und Erfahrungen im Ausland

Pfp hat in den vergangenen Jahren erhebliche Aufmerksamkeit erhalten und scheint den Zenit der Umsetzung noch nicht überschritten zu haben.⁵ Während in PubMed, einer der wichtigsten Datenbanken gesundheitswissenschaftlicher Zeitschriftenartikel, 2004 erst 64 Beiträge zum Thema zu finden waren, stieg die Zahl auf 128 im Jahr 2005, 190 im Jahr 2006 und bereits 97 Beiträge bis Mai 2007. Dudley (2005) berichtet, dass in den USA 2004 bereits über 100 Initiativen zur Einführung von Pfp vorlagen, in denen 39 Mio. Versicherte einbezogen waren.

Nachfolgend sollen zunächst wissenschaftliche Überblicksarbeiten zu Pfp, insbesondere deren Effektivität, vorgestellt werden. Anschließend gehen wir auf die Programme der CMS und des englischen National Health Services (NHS) detaillierter ein.

5 Einen (tabellarischen) Überblick über verschiedene Länder bietet die Bertelsmann-Stiftung (2006) unter http://www.bertelsmann-stiftung.de/bst/de/media/Chartbook_Anreize.pdf (Zugriff Mai 2007).

9.2.1 Reviews über PfP in der Literatur

Die Beratungsfirma Medvantage hatte vor Jahren in einer damals viel beachteten Studie einen Überblick über Aktivitäten im Bereich PfP erstellt (MedVantage 2005). Die Mehrzahl der Programme richtete sich an Hausärzte, sah jährliche Bonuszahlungen vor und basierte auf HEDIS-Sets. Als Trends wurden die Ausweitung im Zuge des EDV-Einsatzes, die Einbeziehung von Fachärzten und die überragende Bedeutung der CMS-Initiative für Impulse der Umsetzung gesehen. Das Hauptproblem war der Umgang mit kleinen Fallzahlen innerhalb einer Arztpraxis.

Im Gegensatz zur Vielzahl der gelisteten Programme stand die geringe Zahl durchgeführter Kosten-Effektivitäts-Analysen. In einem neueren Review für die Jahre 1980 bis 2005 fanden Frohlich et al. (2007) lediglich 18 Studien (davon neun randomisierte kontrollierte Studien), die PfP untersuchten. Es ergab sich nur ein unsicherer Zusammenhang zwischen Höhe der Bonuszahlung, Art des Bonus und der erreichten Verbesserung der Qualität. Einige Programme zeigten keinerlei Auswirkungen.

Petersen et al. (2006) fanden 17 Studien zu PfP jeglicher Art, ebenfalls für den Zeitraum 1980 bis 2005. Sie kamen zu dem Ergebnis, dass die Programme mit Anreizen für individuelle Ärzte (fünf von sechs Studien in dem Bereich) und für Anbietergruppen (sieben von neun Studien) überwiegend auch die gewünschte Wirksamkeit entfalteten. Kritisch wurde gesehen, dass die Studien meist nur präventive Maßnahmen zum Inhalt hatten, und die Zahl der Studien bisher nicht ausreicht, um eine Evidenz über effektive (oder gar kosteneffektive) Ausgestaltungen von PfP-Programmen herzustellen.

Ebenfalls kritisch wird die Effektivität der PfP-Programme von Rosenthal und Frank (2006) gesehen. Sie weisen auf möglicherweise höhere notwendige Bonuszahlungen hin, um eine Effektivität zu erreichen. Zudem verweisen sie auf die Probleme, welche sich aus einer Zersplitterung der Krankenkassenlandschaft beziehungsweise kleiner betroffener Patientengruppen ergeben. Die Autoren fanden lediglich sieben Studien.

Rosenthal et al. (2004) fanden für einen früheren Review in den wissenschaftlichen Datenbanken lediglich einen Beitrag und weiteten ihre Suche auf die Tagespresse aus, um auch Berichte von Krankenversicherungen zu erfassen, die nicht wissenschaftlich ausgewertet wurden. Insgesamt wurden so 37 Ansätze ausfindig gemacht. Die Autoren stellten fest, dass mehrheitlich bekannte Indikatorensets eingesetzt wurden (insbesondere HEDIS) und Struktur- und Prozessmessungen überwogen. 56 % der Programme nutzen relative Schwellenwerte und waren damit stärker wettbewerblich ausgerichtet als die restlichen Programme, welche absolute Schwellenwerte für Bonuszahlungen einsetzten. Die für Anbieter erzielbaren Vergütungszuschläge lagen zwischen 1 % und 10 % der Budgetsumme, je nach Komplexität der Messung und Bedeutung der Indikatoren.

Weitere Schilderungen oder Auswertungen von einzelnen Programmen finden sich in der Literatur. Siehe etwa die Hinweise bei Rosenthal et al. 2004, Rosenthal et al. 2005 sowie Nalli et al. 2007.⁶

⁶ Siehe auch die vollständige Übersicht der 17 Studien von Petersen et al. (2006) unter <http://www.annals.org/cgi/data/145/4/265/DC/1/1> (Zugriff Juni 2007).

9.2.2 Centers for Medicare & Medicaid Services (CMS)

Wie bereits geschildert, war die Initiative der CMS ein Wegbereiter für eine Vielzahl von Initiativen und die Weiterentwicklung der theoretischen Grundlagen von PfP. Zunächst definierte die MedPAC Anforderungen an PfP (Milgate/Cheng 2006):

- Die Messindikatoren sollten evidenzbasiert sein.
- Datenerhebung und Analyse sollten mit angemessenem Aufwand erfolgen.
- Die angemessene Adjustierung des Qualitätsrisikos sollte möglich sein.
- Die Kriterien sollten durch den Erbringer in ihrer Ausprägung überhaupt beeinflussbar sein.
- Qualitätsverbesserungen sollten dort erfolgen, wo sie am notwendigsten sind.

Das Ziel der Vorgaben war es, PfP praktikabel umzusetzen. Das darauf aufbauende CMS-Projekt wurde mit Premier Inc, einer USA-weiten Organisation von Not-for-profit-Krankenhäusern, ab dem Frühjahr 2003 durchgeführt (Rosenthal et al. 2004). In dem Projekt wurden 278 freiwillig teilnehmende Krankenhäuser 33 Indikatoren für fünf Indikationen zusätzlich vergütet (Herzinfarkt, Bypassoperation, Herzinsuffizienz, Hüft- und Knieersatz sowie Lungenentzündung). Ein Minimum von 30 Fällen pro Indikation war Voraussetzung. Die Ergebnisse der Messung wurden publiziert.

Pro Jahr wurden 7 Mio. US\$ über insgesamt drei Jahre ausgeschüttet. Die 10 % besten Krankenhäuser pro Indikation sollten 2 % ihrer Budgetsumme zusätzlich erhalten. Krankenhäuser im nächsten Dezantil (also von 90 bis 80 % der Besten) erhielten 1 % der Budgetsumme zusätzlich. Im dritten Jahr des Pilotprojektes sollten die 10 % schlechtesten Krankenhäuser zudem einen Abschlag hinnehmen müssen (Rosenthal/Frank 2006).

Die CMS beschleunigten die Pfp-Diskussion nochmals durch ein Programm, welches ab dem 1.7.2007 niedergelassenen Ärzten auf freiwilliger Basis die Teilnahme ermöglicht.⁷ Ärzte können im Rahmen der Physician Quality Reporting Initiative (PQRI) dadurch eine um 1,5 % erhöhte Summe der Abrechnungen erzielen. Betroffen sind alle Abrechnungen, nicht nur solche, für die Qualitätsmessungen vorliegen oder ausgefüllt wurden.⁸ Betroffen sind auch Leistungen, die nicht unmittelbar von Ärzten, sondern von anderem Personal in Arztpraxen oder Heilmittelerbringern erbracht werden. Beispiel für eine Messung ist der Anteil der Patienten zwischen 18 und 75 Jahren mit Diabetes Typ I oder II, welche einen Hämoglobin A1c-Wert von über 9 % hatten.

Insgesamt sind 74 Messungen der Qualität vorgesehen.⁹ Die Indikatoren wurden auf der Basis von Evidenz und/oder Konsensus erstellt und messen insbesondere Prozesse, die einen engen Zusammenhang zum Ergebnis der Behandlung erwarten lassen. Für jeden Indikator wurde eine umfangreiche Dokumentation bereitgestellt,

7 Siehe den Gesetzestext unter http://www.cms.hhs.gov/PQRI/Downloads/PQRI_TaxReliefHealth_CareAct.pdf (Zugriff Mai 2007).

8 Die Inhalte der PQRI finden sich unter <http://www.cms.hhs.gov/PQRI/> (Zugriff Mai 2007)

9 Siehe für die vollständige Liste http://www.cms.hhs.gov/PQRI/Downloads/PQRI_MeasuresList.pdf (Zugriff Mai 2007).

die nicht nur die exakte Messung (Zeiträume, Einschlusskriterien für Patienten und Ärzte), sondern auch Auskunft über die Evidenz beziehungsweise Literatur gibt.¹⁰ Die Ergebnisse der Messungen werden nicht auf individueller Ebene der Ärzte veröffentlicht, vielmehr erhält der Arzt eine Rückmeldung über sein Abschneiden. Die Teilnahme der Ärzte erfolgt freiwillig, eine Anmeldung oder Registrierung ist nicht erforderlich. Jedoch muss die Dokumentation für mindestens 80% der in Frage kommenden Patienten erfolgen und für alle Indikatoren, welche in der Praxis des Arztes in Frage kommen.

Erklärtes Ziel der CMS ist es, alle Bereiche der Leistungserbringung in Initiativen des PfP einzubeziehen. Daher ist es konsequent, neben dem stationären Bereich auch ambulante Leistungen zu bewerten. Erwähnenswert ist zudem, dass neben PfP die bereits eingeführten Maßnahmen zur Qualitätssicherung weiter fortbestehen.¹¹ Diese sehen für Krankenhäuser beispielsweise bereits eine Publizierung der Qualitätsdaten vor (im sogenannten Hospital Quality Alliance Programm der CMS).

Lindenauer et al (2007) untersuchten die Effektivität des CMS-Programms erstmalig anhand von 207 teilnehmenden Krankenhäusern. Als Kontrollgruppe dienten 406 Krankenhäuser, welche lediglich ihre Qualität mit den identischen Kriterien veröffentlichten, jedoch keine finanzielle Bonuszahlung daran geknüpft hatten (Lindenauer et al. 2007). Die Abschätzung basierte auf 14 Qualitätsindikatoren über einen Zeitraum von zwei Jahren. Die Studie ergab, dass Teilnehmer des PfP größere Verbesserungen bei den Indikatoren erzielten, und zwar negativ korreliert mit dem Ausgangsniveau. Die Studie zeigt jedoch auch, dass die Kontrollgruppe ebenfalls erhebliche Verbesserungen der Qualität erzielte und bei vielen Indikatoren bereits über 90% der Patientenfälle erreicht wurden.

9.2.3 Großbritannien

Im Jahr 2004 führte der NHS (National Health Service) als zentrale Gesundheitsagentur ein PfP-Programm für den hausärztlichen Bereich ein. Im Vorfeld hatte es intensive Diskussionen um Möglichkeiten der wettbewerblichen Ausrichtung eines zentralistischen Gesundheitssystems gegeben (Stevens 2004). Das PfP-Programm sah vor, dass eine Messung von 146 Indikatoren für zehn chronische Erkrankungen erfolgte. In der Regel handelte es sich um Vom-Hundert-Schwellenwerte für Prozesse (beziehungsweise deren Ergebnisse), die der Arzt erbringen sollte. Ein Beispiel ist der Anteil der Diabetes-Patienten mit einem Blutdruck von unter 145/85 mm Hg. Je nach Anteilen wurden Punkte vergeben, deren Summe mit einem absoluten Geldbetrag bewertet wurden. Eine optimal abschneidende Praxis konnte ihr Einkommen um 25% steigern. Dies übertraf die Erwartungen bei weitem.

Da einige Patienten nur schlechte Therapietreue (s. unten zu Adherence) zeigten, konnten sie vom Arzt ausgeschlossen werden. Zu den Gründen zählten beispiels-

¹⁰ Das ausführliche Dokument befindet sich unter http://www.cms.hhs.gov/PQRI/Downloads/Specifications_2007-02-04.pdf (Zugriff Mai 2007).

¹¹ Siehe etwa für CMS: <http://www.hospitalcompare.hhs.gov/> (Zugriff Mai 2007).

weise das Nicht-Ansprechen auf Medikamente oder Nicht-Erscheinen nach dreimaliger Aufforderung. Eine Auswertung der Daten von 8105 Praxen ergab, dass nach einem Jahr die Rate der Patienten, die ausgeschlossen werden mussten, entscheidend dafür war, wie gut die Praxis abschneiden konnte (Doran et al. 2006). Da Qualitätsdaten für den Beginn der Programms nicht vorlagen, konnte nicht beurteilt werden, ob das Programm tatsächlich eine Qualitätsverbesserung gebracht hatte.

9.3 Probleme des PfP und mögliche Übertragung auf das deutsche Gesundheitswesen

Trotz oder wegen der dynamischen Entwicklung der Pfp-Programme insbesondere in den angelsächsischen Gesundheitssystemen wurde erhebliche Kritik geäußert, welche teilweise aus der noch mangelhaften Evaluation der Programme herrührt, teilweise jedoch auch grundsätzlicher Natur ist.

9

9.3.1 Kritik an Pfp

Eine grundsätzliche Kritik besteht darin, dass die Evidenz für die Wirksamkeit von Pfp bisher zu gering sei, um aufwändige Projekte zu starten (Epstein 2007). Insbesondere bestehen kaum randomisierte kontrollierte Studien oder belastbare Studiendesigns, sofern diese überhaupt durchgeführt wurden (Dudley 2005). Die Anzahl der Teilnehmer (also der Leistungserbringer) ist meist zu gering, die Übertragbarkeit auf andere Regionen, Indikatoren-Sets oder Gesundheitssysteme ist nicht gegeben. Diese Kritik kann zunächst bei allen neuen Versorgungsformen geäußert werden.

Eine Vielzahl von Kritiken bezieht sich auf ableitbare Anreize und Verteilungswirkungen in Folge der konkreten Ausgestaltung des Tarifs. So ist unklar, ob **zusätzliche Finanzmittel** bereit gestellt werden oder zuvor von einer Gesamtvergütung ein Pool abgezogen wird (analog der integrierten Versorgung in Deutschland). In Großbritannien wurden beispielsweise erhebliche Mittel zusätzlich bereitgestellt und Pfp war eine Möglichkeit, um die (politisch gewollte) Ausdehnung des Budgets des NHS zielgerichtet zu steuern. Ebenso wird bei nicht effektiven Programmen teilweise vermutet, dass die Höhe der zusätzlich über Pfp ausgezahlten Erlöse nicht angemessen war (Roski et al. 2003).

Unklar ist ebenfalls, ob **Qualitätsverbesserungen** oder aber **Qualitätsziele** (Schwellenwerte) vorgegeben werden sollen. Verbesserungen könnten auch in sozial schwachen Gebieten erzielt werden, würden auch bisher unterdurchschnittlich abschneidende Krankenhäuser ermuntern und umgekehrt gute Krankenhäuser nicht für eine unveränderte Versorgung belohnen (Rosenthal et al. 2004). Allerdings würden Krankenhäuser, die bisher nur geringe Anstrengungen in der Qualitätssicherung unternommen haben, eher belohnt (Rosenthal et al. 2005). Letztendlich handelt es sich auch um eine normative Entscheidung der Politik, ob ein egalitäres oder wettbewerbliches Gesundheitssystem zugrunde gelegt wird. So werden in den USA eher Veränderungen belohnt, in Großbritannien hingegen das Erreichen von Schwellenwerten.

Im Zusammenhang mit erreichbaren Qualitätsverbesserungen muss ebenfalls kritisch gesehen werden, ob sich lediglich die isoliert gemessenen Indikatoren (Strukturen oder Prozesse) oder aber auch das Versorgungsergebnis verbessert. Ebenso wird kritisch diskutiert, ob lediglich durch (legale oder illegale) Änderungen der Kodierungen die gewünschten Ergebnisse für PfP erzielt werden oder auch eine Selektion von Patientengruppen mit erhöhter Wahrscheinlichkeit für Bonuszahlungen stattfindet (Shen 2003, Norton 1992).¹² Diese Probleme sind aus dem Bereich der DRG-Nutzung bereits bekannt (**Upcoding, Gaming**) (Rosenthal & Frank 2006 unmittelbar zu PfP; zu DRG auch Lungen/Lauterbach 2000, Ellis 1998). Diskutiert wird, die Tarife auf einzelne Fälle zu beziehen, nicht auf Vom-Hundert-Anteile in Bezug auf Patientengruppen. Demnach würden Boni gezahlt für jeden optimal versorgten Patienten. Dies motiviert sowohl schwache Performer als auch Hochperformer. Festgelegt werden muss allerdings, was optimale Versorgung ist (Rosenthal/Dudley 2007).

Kritisch gesehen wird ebenso, dass **Alternativen zu PfP** eine bessere Kosteneffektivität aufweisen können, insbesondere PR (Lindenauer et al. 2007). Eine über PR hinaus gehende Vergütung der Qualität bewirkte in Studien nur noch geringe Verbesserungen. Die zusätzlichen Probleme des PfP sind jedoch erheblich, etwa die Generierung von neuen Finanzmitteln (beziehungsweise Kürzung bestehender Vergütungen) sowie die Entwicklung, Weiterentwicklung und Administrierung der Programme.

Sofern PfP als **kassenindividuelle Programme** angelegt werden, wird vor Trittbrettfahrerverhalten gewarnt. Der behandelnde Arzt könnte sein Verhalten entweder gar nicht ändern oder aber für alle Patienten (Pham et al. 2007). Dies kann entweder zu Trittbrettfahrerverhalten der Krankenkassen führen (und somit der zu geringen Einführung von sinnvollen Programmen) oder aber zu geringer Kosteneffektivität der Programme, wenn weniger Ärzte als möglich die Qualität verbessern. Insgesamt wird die Möglichkeit der Evaluation bei kassenindividuellen Programmen nahezu verhindert, ein Problem, welches bereits bei der Evaluation der Auswirkungen von Managed Care auftrat (Lungen/Stock 2006).

9.3.2 Relevante Fragen bei einer Übertragung nach Deutschland

Die Ausgestaltung der internationalen Programme und die in der Literatur geäußerte Kritik kann zu einem Fragenkatalog zusammengeführt werden, welcher vor einer Einführung von PfP in Deutschland beantwortet werden muss (Rosenthal/Dudley 2007, Rowe 2006, Dudley 2005, Lungen/Lauterbach 2002a).

1. **Individual- oder Gruppenauszahlung:** Die Form der Auszahlung wirkt sich auf die Motivation der einzelnen Leistungserbringer aus, ihre Qualität zu verbessern. Eine Gruppenauszahlung kann Vorteile haben im Hinblick auf statistische Verfahren und Vereinfachung der Administration, da Risiken gepoolt werden. In Deutschland scheint derzeit für den vertragsärztlichen Bereich eine

¹² Siehe Norton (1992) als eine frühe Studie im Bereich von Pflegeheimen in den USA, welche auch Anreize Risikoselektion bei Umstellung auf erfolgsbasierte und risikobasierte Vergütung thematisierte.

Gruppenauszahlung eher möglich. Aufgrund der Fallzahllhöhe und Verhandlungssituation von Krankenhäusern scheint dort eine Einzelauszahlung angemessener.

2. **Auswahl der Indikationen und Indikatoren:** Die Abwägung zwischen den Kosten der Datenerfassung, der Güte der Messung und damit des Zusammenhangs zwischen Daten, Qualität und Bonuszahlung ist in der Praxis der PfP-Programme bisher eher pragmatisch entschieden worden. Der Trend zur Nutzung von Routinedaten aus der Patientenabrechnung ist ebenso erkennbar wie die Vereinheitlichung der Indikatorensets (AOK-Bundesverband et al. 2007). Neuentwicklungen von Indikatoren sind aufwändig, schützen nicht vor neuerlichen kritischen Diskussionen und werden daher selten in Angriff genommen. Unklar ist auch, welche Evidenz für Indikatoren überhaupt hergestellt werden kann (Werner/Bradlow 2006). Betont wird bei allen Entwicklern die Notwendigkeit der Einbindung von betroffenen Ärzten und Leistungserbringern. Strittig ist, inwieweit auch Effizienzindikatoren mit den Qualitätsdaten verknüpft werden sollen. Mit den bisherigen Daten der externen Qualitätssicherung in Deutschland scheint zumindest eine Basis für eine Umsetzung gegeben (BQS fortlaufende Jahre). Der Sachverständigenrat stellt bereits vor zehn Jahren in seinem Sondergutachten 1997 recht optimistisch fest, dass die im Bereich der Ärztekammer Nordrhein damals schon seit Jahren laufenden Qualitätsmessungen der Schenkelhalsfraktur in die Vergütung einbezogen werden könnten (Sachverständigenrat 1998).
3. **Bonustarif:** Relative Schwellenwerte erhöhen den Wettbewerb, jedoch auch die Unsicherheit über die finanzielle Auszahlung für das Krankenhaus. Die Verweigerung des Bonus kann somit auch als indirekter Malus aufgefasst werden. Bei absoluten Schwellenwerten erhalten eine Anzahl Krankenhäuser zudem höhere Vergütungen, ohne dass sie ihr Versorgungsverhalten geändert haben. Auf der anderen Seite können schwache Krankenhäuser jede Anstrengung unterlassen, die Hürde zu erreichen, wenn diese zu hoch liegt. In der Praxis wird daher eine Abstufung der Schwellenwerte vorgeschlagen. Zudem kann die Entscheidung für einen Bonus oder Malus sich auch nach der Indikation beziehungsweise A-DRG richten. Ist eine Therapie bereits etabliert, können eher Maluszahlungen vorgesehen werden. Ist ein Verfahren noch in der Entwicklung, können eher Bonuszahlungen für eine besonders engagierte Weiterentwicklung der Qualität ausgesprochen werden. Notwendig dazu ist allerdings eine verstärkte Orientierung an den Ergebnissen der Behandlung (Lungen/Lauterbach 2002a).
4. **Höhe des Bonus:** Empirisch belegt ist, dass der Zusammenhang von Zahlbetrag und Effektivität des Programms nur schwach – wenn überhaupt – vorhanden ist. Ursache kann sein, dass die ökonomische Wirkung von Opportunitätskosten stärker beachtet werden muss. Entstehen beim Anbieter dafür, dass das Qualitätsziel erreicht werden kann, klar definierbare Kosten (höherer Personal- oder Sachaufwand, Folgekosten), sollten diese mindestens abgedeckt werden. Entstehen die Kosten einmalig (Investition in EDV, Räume etc.) und erfolgt die Bonuszahlung kontinuierlich pro Patient, können sich adverse Effekte ergeben. Ebenso kann es hinderlich sein, die Höhe der Prämie an den erwartbaren Einsparungen der Krankenversicherung festzumachen. Für die Entscheidung der Leistungsanbieter können diese ohne Bedeutung sein.

5. **Zielgruppe:** Weitgehend ungeklärt ist, wie die regional abweichende Adherence (Therapietreue) der Patienten sich auf die Qualität der Versorgung und die Bonus-Zahlungen niederschlägt. Verweigern Patienten die Kooperation, kann sich die Qualitätsausprägung verschlechtern. Dies hat negative Folgen für die Anbieter. Einerseits könnten relative Verbesserungen pro Anbieter zugrunde gelegt werden (Verbesserung der Anteile zwischen zwei Zeitpunkten), was jedoch Anbieter mit zuvor geringer Qualität bzw. wenig Anstrengungen eher belohnt. Andererseits kann die mangelnde Kooperation der Patienten (zumindest teilweise) statistisch standardisiert werden. Vorgeschlagen wird auch, die Therapietreue generell in eigenständigen Programmen zu fördern, das heißt außerhalb von PfP-Programmen (Rosenthal/Dudley 2007). Inwieweit dies ausreichend ist, bleibt offen. Casalino/Alster 2007 warnen davor, dass PfP die Ungleichheiten in der gesundheitlichen Versorgung sogar noch vergrößern können. Basieren PfP-Ergebnisse wesentlich auf Faktoren, die durch Patienteneigenschaften gesteuert werden, können Ärzte beziehungsweise Anbieter in sozial schwachen Regionen benachteiligt werden, beziehungsweise entsprechende Patientengruppen indirekt einer negativen Selektion unterliegen.

9.4 Fazit

PfP weist eine hohe Sichtvalidität auf und ist als Modell mit Anreizwirkung unmittelbar überzeugend. Der Gedanke, dass nicht mehr das Erbringen einer Leistung, sondern deren Erfolg beziehungsweise die Qualität vergütet wird, ist eingängig. Wettbewerb um Qualität hat auch für Patienten eine höhere Attraktivität als Wettbewerb ausschließlich um Effizienz. Es ist vorhersehbar, dass PfP noch stärkere Verbreitung finden wird.

Einige Aspekte lassen sich daher bereits für eine Ausgestaltung in Deutschland umreißen. Eine Verknüpfung der **Vergütungssystematik von DRG** und PfP ist generell möglich. Die starke Fixierung der externen Qualitätssicherung auf DRG kann hierbei genutzt werden. Wird in der Perspektive die Kontrahierung von Krankenhäusern selektiv durch Krankenversicherungen erfolgen, können Verknüpfungen von Qualität und Tarif leichter umgesetzt werden. Nachteilig ist jedoch, dass der Marktanteil der Krankenkasse entscheidend dafür sein wird, welche Wirkung das Programm entfalten kann.

Ein weiterer Nachteil besteht darin, dass die Begrenzung von PfP auf die akutstationäre Periode zu Ausweichreaktionen führen kann. Die geplante Neuausrichtung der externen Qualitätssicherung durch den Gemeinsamen Bundesausschuss im Rahmen des GKV-WSG kann genutzt werden, um Weichen zu stellen (Neufassung von § 137 SGB V). Statt einer relativ begrenzten akutstationären Periode könnte durch die Zusammenführung von Qualitätssicherung im ambulanten und stationären Bereich eine Zeitepisode Grundlage der Bonusvergütung sein (Davis 2007). Aus Patientensicht hätte dies Vorteile, da die sektorenüberreifende Versorgung gestärkt würde. Notwendig ist jedoch die entsprechende Anpassung der Finanzierung, beispielsweise im Rahmen der integrierten Versorgung (§ 140 a bis d SGB V) mit regionalem Bezug oder von Disease-Management-Programmen (§ 137f SGB V).

Letztendlich wird auch die **Krankenhausplanung** nicht umhin können, die Versorgung der Bevölkerung mit Krankenhäusern zu gewährleisten, welche in der Qualitätssicherung gut abschneiden. Ein faktischer Druck zur Berücksichtigung von Qualitätsergebnissen in der Krankenhausplanung ist aus unsrer Sicht einleuchtend begründbar.

Ungeachtet dieser Möglichkeiten überwiegen in der **wissenschaftlichen Literatur** bisher die Stimmen, welche sich kritisch bis abwartend zu PfP äußern. Einerseits liegt diese Kritik im Mangel an Erfahrungen begründet. Es wird beispielsweise darauf verwiesen, dass chirurgische Verfahren auf der Grundlage der geringen Evidenz, wie sie für PfP vorliegt, nicht eingeführt würden (Epstein 2007). Andererseits bestehen bei allen neuen Versorgungsformen zunächst viele offene Fragen zur Ausgestaltung. Die bedeutsamste Kritik lautet, dass ähnliche Qualitätsverbesserungen wie durch PfP bereits durch die **öffentliche Verbreitung von Qualitätsdaten** (PR) erzielt werden können. Inwieweit dies für alle Patientengruppen gilt oder aber eine gewisse Fähigkeit zur Informationsverarbeitung bei den Versicherten voraussetzt (literacy) oder regionale Ungleichheiten der Versorgung hervorruft, bleibt zumindest für Deutschland noch offen.

Die rasche Verbreitung von PfP kann jedoch nicht darüber hinwegtäuschen, dass in naher Zukunft weiterhin **pragmatische Entscheidungen** über die Ausgestaltung der Programme getroffen werden müssen. Die Festlegung des Tarifs und die Beseitigung von Nachteilen durch die Sozialstruktur der Patienten sind dabei die größten Herausforderungen. Eine hochwertige Evaluation und die nachfolgende Anpassung sind unabdingbar, um PfP langfristig von einer Sichtvalidität zu einer Akzeptanz auch bei den Leistungserbringern zu verhelfen.

9.5 Literatur

- American Medical Association (AMA). Guidelines for Pay-for-Performance Programs. Stand 21. Juni 2005.
- AOK-Bundesverband, FEISA, HELIOS, WIdO (Hrsg.). Qualitätssicherung der stationären Versorgung mit Routinedaten (QSR). Abschlussbericht. Bonn 2007.
- Armour BS, Pitts MM, Maclean R, Cangialose C, Kishel M, Imai H, Etchason J. The effect of explicit financial incentives on physician behavior. *Arch Intern Med.* 2001; 161 (10): 1261–6.
- Baker G, Carter B. Provider Pay-for-Performance Incentive Programs: 2004 National Study Results. San Francisco: Medvantage Inc. 2005.
- Baumann MH, Dellert E. Performance measures and pay for performance. *Chest.* 2006 Jan; 129 (1): 188–91.
- Bertelsmann-Stiftung. Anreize zur Verhaltenssteuerung im Gesundheitswesen. Effekte bei Versicherten und Leistungsanbietern. Gütersloh 2006.
- Bundesgeschäftsstelle Qualitätssicherung (BQS). BQS-Qualitätsreport. Düsseldorf: fortlaufende Jahre.
- Conrad DA, Sales A, Liang SY, Chaudhuri A, Maynard C, Pieper L, Weinstein L, Gans D, Piland N. The impact of financial incentives on physician productivity in medical groups. *Health Serv Res.* 2002; 37 (4): 885–906.
- Davis K. Paying for care episodes and care coordination. *N Engl J Med* 2007; 356 (11): 1166–8.
- Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006; 355 (4): 375–84.

- Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006; 355 (4): 375–84.
- Dudley RA. Pay-for-performance research: how to learn what clinicians and policy makers need to know. *JAMA* 2005; 294 (14): 1821–3.
- Ellis RP. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *J Health Econ* 1998; 17 (5): 537–55.
- Epstein AM. Pay for performance at the tipping point. *N Engl J Med*. 2007; 356 (5): 515–7.
- Frolich A, Talavera JA, Broadhead P, Dudley RA. A behavioral model of clinician responses to incentives to improve quality. *Health Policy* 2007; 80 (1): 179–93.
- Henley E. Pay-for-performance: what can you expect? *J Fam Pract* 2005; 54 (7): 609–12.
- Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007; 356 (5): 486–96.
- Lüngen M, Lauterbach K. Upcoding –eine Gefahr für den Einsatz von DRGs (Diagnosis Related Groups)? *Dtsch Med Wschr* 2000; 125: 852–6.
- Lüngen M, Lauterbach K. Ergebnisorientierte Vergütung bei DRG. Qualitätssicherung bei pauschalierender Vergütung stationärer Krankenhausleistungen. Heidelberg: Springer 2002.
- Lüngen M, Lauterbach K. Qualitätssicherung auf der Basis der DRG-Finanzierung. *Zeitschrift für Sozialreform* 2002; 48 (2): 133–63.
- Lüngen M, Stock S. USA. In: Lauterbach K, Brunner H, Stock S. Gesundheitsökonomie für Mediziner. Bern: Huber 2006; 259–68.
- MedVantage Inc. Provider Pay-for-Performance Incentive Programs. 2004 National Study Results. First Quarter 2005.
- Milgate K, Cheng SB. Pay-for-performance: the MedPAC perspective. *Health Aff (Millwood)* 2006; 25 (2): 413–9.
- Nalli GA, Scanlon DP, Libby D. Developing a performance-based incentive program for hospitals: a case study from Maine. *Health Aff (Millwood)* 2007; 26 (3): 817–24.
- Norton EC. Incentive regulation of nursing homes. *J of Health Economics* 1992; 11: 105–28.
- Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med* 2006; 145 (4): 265–72.
- Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. *N Engl J Med* 2007; 356 (11): 1130–9.
- Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA*. 2007; 297 (7): 740–4.
- Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood)* 2004; 23 (2): 127–41.
- Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA* 2005; 294 (14): 1788–93.
- Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev* 2006; 63 (2): 135–57.
- Rowe JW. Pay-for-Performance and Accountability: Related Themes in Improving Health Care. *Ann Int Med* 2006; (145) 9: 695–9.
- Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen. Gesundheitswesen in Deutschland, Kostenfaktor und Zukunftsbranche, Sondergutachten 1997. Baden-Baden: Nomos 1998: 380.
- Shen Y. Selection Incentives in a Performance-Based Contracting System. *Health Services Research* 2003; 38 (2): 535–52.
- Stevens S. Reform strategies for the English NHS. *Health Aff (Millwood)*. 2004; 23 (3): 37–44.
- Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA* 2006; 296 (22): 2694–702.

Wir danken einem anonymen Reviewer für wertvolle Hinweise.