

## 6 Pay for Performance (P4P)

### 6.1 Einleitung

*Pay for Performance* (P4P) bzw. qualitätsbezogene Vergütung ist neben *Public Reporting* eines der wichtigsten Instrumente zur Qualitätsverbesserung, das auf Systemebene einsetzbar ist. Zwar ist *Pay for Performance* (P4P) etwas in die Jahre gekommen, die Anfangseuphorie ist einer differenzierteren und realistischen Einschätzung gewichen, andererseits setzen aber gerade jetzt große und unterschiedlich strukturierte Gesundheitssysteme wie das der USA, Großbritanniens und Deutschlands auf die Bindung der Vergütung an die erbrachte Qualität, um die als einseitig empfundene Mengenkopplung der Vergütung zu relativieren. So beabsichtigt der Koalitionsvertrag der Großen Koalition vom 27.11.2013, Qualität über den Mehrerlösausgleich bei der Vergütung zu berücksichtigen (s.u. zu „Koalitionsvertrag 27.11.2013“). In den USA hat Medicare ein umfassendes *Value-Based Purchasing* Programm für den stationären Sektor aufgelegt, das auf den in den letzten 10 Jahren gesammelten Erfahrungen aufbaut (CMS 2011A, Belmont et al. 2011, Ryan und Blustein 2012, Chien und Rosenthal 2013) und im Jahr 2015 mit dem *Physician Value-Based Payment Modifier* (PVBPM) sogar auf den ambulanten Bereich ausgeweitet wird (Ryan et al. 2014A). In Großbritannien setzt der *National Health Service* das 2004 in Kraft getretene *Quality and Outcome Framework* (QOF) mit jährlich neu justierten Indikatoren fort (NHS 2013B).



### **Koalitionsvertrag 27.11.2013, S. 78:**

*„In einer Qualitätsoffensive werden wir die Qualität der stationären Versorgung verbessern. Qualität wird als weiteres Kriterium für Entscheidungen der Krankenhausplanung gesetzlich eingeführt (§ 1 KHG). (...) Gute Qualität muss sich für die Krankenhäuser auch finanziell lohnen. Die Menge soll künftig nur da berücksichtigt werden, wo sie entsteht. Das heute bestehende System der Mehrleistungsabschläge wollen wir dabei differenzieren: Leistungen mit nachgewiesener hoher Qualität können von Mehrleistungsabschlägen ausgenommen werden, für besonders gute Qualität sind Zuschläge möglich. Umgekehrt sollen bei unterdurchschnittlicher Qualität für einzelne Leistungen auch höhere Abschläge möglich sein. Die Qualität soll dabei risikoadjustiert und anhand wesentlicher Indikatoren gemessen werden.“*

Diese aktuelle Hinwendung zu P4P resultiert nicht aus den kurzfristigen (z.B. Campbell et al. 2007, Lindenauer et al. 2007) und langfristigen Ergebnissen der Evaluationsstudien (z.B. Campbell et al. 2009, Ryan et al. 2012B), denn diese haben die anfänglich hochfliegenden Erwartungen, zumindest auf den ersten Blick, nicht erfüllt. Der Grund für die gestiegene Bedeutung des Themas P4P im Gesundheitswesen besteht vielmehr darin, dass sich der Blick auf P4P geändert hat: P4P wird nicht mehr als *magic bullet* angepriesen, das alle Probleme um die Qualität der Gesundheitsversorgung schlagartig behebt (Sorbero et al. 2006), sondern wird als ein Vergütungsinstrument unter vielen betrachtet, das gut abgestimmt mit anderen Vergütungselementen und eingebettet in eine langjährig entwickelte qualitätsorientierte Strategie zur Weiterentwicklung des Gesundheitswesens seine Wirksamkeit erzielen kann. Diese Sichtweise erfordert für die Implementierung von P4P jedoch mehr als ein paar technische Details und eine Verständigung über Indikatorspezifikationen, so wie es in Deutschland gerne gesehen wird. Für eine wirksame Implementierung von P4P ist vielmehr ein konzeptioneller Rahmen (*conceptual framework*) notwendig, der zum Beispiel über die Wechselwirkung mit anderen Vergütungselementen oder über die Integration in eine generelle Qualitätsstrategie Auskunft gibt. Ein solches Rahmenkonzept stellt auch das Ziel in den ersten drei Teilen dieses Gutachtens dar, um daraus dann Empfehlungen ableiten zu können.

### **Zusammenfassung:**

*Pay for Performance* spielt international eine große Rolle und ist neben *Public Reporting* eines der wichtigsten Instrumente zur Qualitätsverbesserung, das auf Systemebene einsetzbar ist. Wenngleich die langfristigen Evaluationsstudien nicht so klare Ergebnisse erbracht haben wie zunächst erhofft, setzen

gerade jetzt große und unterschiedlich strukturierte Gesundheitssysteme wie das der USA, Großbritanniens und Deutschlands auf die Bindung der Vergütung an die erbrachte Qualität, um die als einseitig empfundene Mengenkopplung der Vergütung zu relativieren (Koalitionsvertrag der Großen Koalition vom 27.11.2013, *Value-Based Purchasing* Programm für den Krankenhausbereich und das *Physician Value-Based Payment Modifier* (PVBPM) für den ambulanten Bereich in den USA, *Quality and Outcome Framework* (QOF) mit jährlich neu justierten Indikatoren in Großbritannien).

## 6.2 P4P: Begriffsbestimmung und Definition

Unumgänglich für die aktuelle Diskussion ist eine Aktualisierung der Begriffsbestimmung von P4P. Die Definition des Sachverständigenrates von 2008 (SVR 2008, Nr. 732) als „Vergütungssystem, das die Qualität der Leistungserbringer in den Mittelpunkt stellt“ (vgl. unten zu „Erste Annäherung an den Begriff – Was ist Pay for Performance (P4P)?“) kann als Ausgangspunkt dienen, man muss aber in mindestens zweierlei Richtungen über Weiterungen nachdenken und in einem dritte Sinne eine Abgrenzung vornehmen:

- Vor allem in den USA hat man von Beginn an Definitionen verwendet, die sich nicht nur auf die gesonderte Vergütung von Qualität beziehen, sondern die Effizienz (das Verhältnis von Qualität und den angefallenen Kosten, *value*) in den Mittelpunkt stellen (Dudley u. Rosenthal 2006; Tompkins et al. 2009, aber auch schon Wenzel 1992). Diese Sichtweise hat auch in die Bezeichnung für das umfassende P4P-Programm von Medicare „*Hospital Inpatient Value-Based Purchasing*“ Eingang gefunden (CMS 2011A). In Analogie wäre hier eigentlich der Begriff „*Pay for Value*“ (P4V) angezeigt, ist aber nicht gebräuchlich.
- Die zweite Weiterung des Begriffes stammt aus der ökonomischen Theorie (*Principal-Agent-Theorie*, Agenturtheorie) und betrachtet das Thema unter dem Managementblickwinkel. Der Auftraggeber (*Principal*) gibt dem *Agent*, der einen Informationsvorsprung hat (z.B. einem Arzt, einem Krankenhaus), den Auftrag, in seinem Sinne eine Tätigkeit (eine Behandlung) auszuführen, die er selbst (der *Principal*) nicht ausführen kann, und die er auch nicht vollständig kontrollieren kann. P4P ist in diesem Verständnis ein Mittel des *Principal*, sich der Dienste des *Agent* in seinem Sinne zu sichern, oder zumindest eine größere Chance dazu zu haben, dass der *Agent* in seinem Sinne handelt (vgl. Prendergast 1999; Wodchis et al. 2007, weitere Implikationen s. Kap.8.5.4).
- In Hinblick auf die Diskussion in Deutschland scheint die Abgrenzung von Vergütungsbestandteilen des DRG-Systems von Bedeutung. In einem Gutachten des BQS-Institutes (Veit et al. 2012) werden z.B. die Regelungen zur Fallzusammenführungen bei Wiederaufnahme nach Krankenhausbehandlung als qualitätsorientierte Vergütung bezeichnet. Sicherlich besteht hier ein Zusammenhang zu Qualitätsproblemen (z.B. Wie-

deraufnahme nach Komplikationen), andererseits besteht der Sinn dieser Regelung in der Verhinderung einer Mengenausweitung und nicht primär in einer Qualitätsverbesserung. In der vorliegenden Arbeit wird daher eine **spezifische Definition** verwendet, die von der Qualität und der Sicherheit der Versorgung ausgeht und

- die explizite Messung von Qualität mittels definierter Indikatoren verbunden mit
- einer nachvollziehbaren Kopplung dieser Indikatoren an monetäre Vergütungsbestandteile in den Mittelpunkt stellt.



### ***Erste Annäherung an den Begriff – Was ist Pay for Performance (P4P)?***

***P4P oder qualitätsbezogene Vergütung „ist ein Vergütungssystem, das die Qualität der Leistungserbringer in den Mittelpunkt stellt“ (SVR 2008, Nr. 732). In den USA wird zunehmend „Value“ vergütet, also nicht nur die Qualität, sondern das Verhältnis von Qualität und den eingesetzten Kosten (strenggenommen Pay for Value, so im umfassenden „Hospital Inpatient Value-Based Purchasing“ (VBP)-Programm von Medicare [CMS 2011A]). Ökonomisch stammt der Begriff aus der Principal-Agent-Theory (vgl. Prendergast 1999). Die entscheidende Abgrenzung ist gegenüber mengenorientierten Vergütungssystemen vorzunehmen (z.B. DRG). Die hier verwendete Definition basiert auf der expliziten Qualitätsmessung durch Indikatoren und der nachvollziehbaren Kopplung an monetäre Vergütungsbestandteile (Definition s. Text).***

P4P besteht also aus zwei Systemen, zum einen aus einem System der Qualitätsmessung, zum anderen aus einem System der finanziellen Bewertung. Vergütungssysteme, die auf der Effizienz der Versorgung basieren, also auf dem Verhältnis von Qualität und den aufgewendeten Kosten, sind ebenfalls als P4P zu verstehen, soweit sie Qualität (und die Kosten) explizit messen und eine nachvollziehbare Kopplung an die Vergütung aufweisen. P4P betrifft fast immer nur einen Teil der Gesamtvergütung und ist somit in andere Vergütungssysteme eingebettet. In Erweiterung der SVR-Definition wird in der vorliegenden Arbeit daher folgende Definition verwendet:

Das Interesse in der Gesundheitspolitik an P4P fällt in eine Zeit, das darf nicht übersehen werden, in der grundsätzliche Vorbehalte gegen jegliche nachvollziehbare und valide Messung der Qualität im Gesundheitswesen (wieder) zunehmen, in der grundlegende Zweifel an den wissenschaftlichen Zahlen zur Häufigkeit von Problemen der Patientensicherheit wieder hoffähig werden (Anonymous 2014; Swinglehurst et al. 2014), alles einhergehend mit einer allgemeinen Kritik an der „Bürokratisierung“ der Gesundheitsversorgung und der vorgeblichen Standardisierung bzw. „Ökonomisierung“ durch die evidenzbasierte Medizin (arte 17.9.2013). Qualität ließe sich gar nicht oder nur sehr schwer messen, die methodischen Probleme seien zu zahlreich und im Grun-

de nicht lösbar. Qualität stehe einer patientenorientierten Medizin entgegen und bilde einen Gegensatz zu Empathie und Einfühlung (Swinglehurst et al. 2014). Es wird auch wieder von der „Kunst der Qualitätsmessung“ (und nicht deren exakten Methodik) gesprochen (Anonymous 2011). Wenn „Kunst“ im Spiel ist, das kennen wir aus der Patientensicherheitsdiskussion um den Begriff des „Kunstfehlers“, dann bezieht man sich auf ein implizites Qualitätsverständnis, das dem Patienten und der Öffentlichkeit nicht zugänglich ist.



**Pay for Performance (P4P) im Gesundheitswesen basiert auf einer Qualitätsmessung durch definierte Indikatoren und koppelt die Qualität der Versorgung an Vergütungsbestandteile mit dem Ziel, die Qualität der Gesundheitsversorgung zu verbessern. P4P kann sich auch auf Qualität im Verhältnis zu den Kosten (Effizienz, value) beziehen.**

**Zusammenfassung:**

Eine aktualisierte Definition von P4P umfasst zwei Systeme, sie geht von der Qualität und der Sicherheit der Versorgung aus und stellt die explizite Messung von Qualität mittels definierter Indikatoren verbunden mit einer nachvollziehbaren Kopplung dieser Indikatoren an monetäre Vergütungsbestandteile in den Mittelpunkt. In Erweiterung der Definition des Sachverständigenrates 2008 wird hier die folgende Definition gewählt: „Pay for Performance (P4P) im Gesundheitswesen basiert auf einer Qualitätsmessung durch definierte Indikatoren und koppelt die Qualität der Versorgung an Vergütungsbestandteile mit dem Ziel, die Qualität der Gesundheitsversorgung zu verbessern. P4P kann sich auch auf Qualität im Verhältnis zu den Kosten (Effizienz, value) beziehen.“

### 6.3 Wirkung von P4P: erste Phase der Evaluation

Wir kennen das Phänomen von anderen Methoden im Gesundheitswesen: anfangs ist alles Gold. Es folgt ein Auf und Ab von Ernüchterung, dann wieder positiver Einschätzung, nochmals Skepsis usw., bis man dann nach einer gewissen Zeit zu einer adäquaten Beurteilung der Wirksamkeit der Methode kommt, die der jeweiligen Anwendung und ihrem Kontext angemessen ist (s. Abb. 32). So verhielt es sich auch bei P4P (Rosenthal et al. 2007); die früheren Studien und Reviews kamen zu besseren Ergebnissen als die späteren Evaluationen, und aus letzteren ergaben sich wiederum Hinweise, die in den USA und in Großbritannien zur Entwicklung weiterführender Programme wie dem Value-Based Purchasing-Programm von Medicare (CMS 2011B) und dem Quality and Outcome Framework des NHS führten [NHS 2013A]). Auch Einzelstudien neueren Datums kamen wieder zu ermutigenderen Ergebnissen (z.B. Bardach et al. 2013; Calikoglu et al. 2013; Casale et al. 2007; Petersen et al. 2013), allerdings sind in alle diese neueren Projekte die Erfahrungen aus vorangegangenen „Gehversuchen“ eingeflossen.

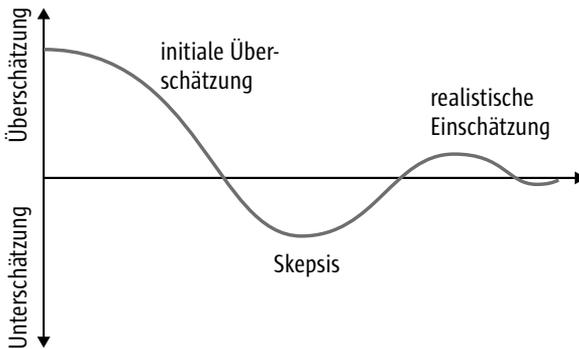


Abb. 32 Innovationen im Gesundheitswesen. Wechsel von initialer Überschätzung einer Methode über eine skeptische Beurteilung bis zu einer realistischeren Einschätzung

In seinem **frühen Systematischen Review** (Studien bis 2006) zur kurzfristigen Wirksamkeit von P4P auf definierte Versorgungsziele konnte der SVR Gesundheit 28 kontrollierte Studien identifizieren, wovon 21 Studien einen positiven Effekt zeigten (SVR 2008, Nr. 738). Jedoch war bereits in diesem frühen Review auffällig, dass zwar 100% aller methodisch schwächeren Studien einen durchschlagenden Erfolg zeigten, dies jedoch nur bei 6/9 Studien mit methodisch stringenterem, randomisierten Design der Fall war. Diese Verschiebung der Effektmaße hin zu den methodisch schwächeren Studien könnte als ein Zeichen für die mangelnde interne Validität der dem Review zugrundeliegenden Studien gewertet werden: methodisch schwächere Studien überschätzen gern das Ergebnis (s. Abb. 33). Der SVR empfahl daher nur vorsichtig „die schrittweise Einführung von Elementen dieser Vergütungsform mit Pilotierung und intensiver Evaluation“ (SVR 2008, Kurzfassung, Nr. 132).

Die weitere Entwicklung gab dieser vorsichtigen Einschätzung zunächst recht. Im Systematischen Review von Van Herck et al. (2010) wurde auf der Basis von 128 Studien positive Ergebnisse (Wirksamkeit von P4P) überwiegend bei klinischen, eindimensionalen Endpunkten (z.B. Impfrate) gefunden, bei komplexen Endpunkten (Koordination, Kontinuität der Behandlung, Patientenzentriertheit, Kosteneffektivität) waren die Ergebnisse hingegen widersprüchlich. Auch diese Konstellation muss vorsichtig stimmen, denn eindimensionale Endpunkte sind gerade bei Interventionen wie P4P in der Praxis weit weniger adäquat als komplexe Endpunkte. Zwei kleinere, methodisch aber sehr strenge Systematische Reviews zeigten geringe positive Auswirkungen (Mehrotra et al. 2009, 8 Studien, ausschließlich HQIP-Krankenhäuser) an einem Teil der untersuchten Endpunkte (Scott et al. 2011, Cochrane-Review, 7 Studien). (Die älteren Systematischen Reviews sind an dieser Stelle nicht nochmals aufgeführt, vgl. SVR 2008; die Systematischen Reviews von Gillem et al. (2012) und Langdown et al. (2013) waren nur auf das britische QOF- Programm ausgerichtet). Ein Meta-Review, der 22 Systematische Reviews umfasste, kam zu keinen definitiven Ergebnissen (Eijkenaar et al. 2013).



## SVR-Review: Ergebnisse

28 Studien wurden identifiziert

- 21/28 Studien wiesen einen positiven Effekt aus
  - 3 Studien gemischtes Ergebnis
  - 4 Studien negatives Ergebnis
- positives Ergebnis bei 15/19 Studien mit einfachen und 6/9 mit komplexen Endpunkten
- positives Ergebnis bei 14/18 Studien, bei denen P4P die alleinige Intervention war, und 7/10 Studien mit komplexer Intervention

Einfluss des Studiendesigns:

- 9 Studien randomisiertes Design
  - 6/9 Studien positives Ergebnis (2-, 1+/-)
- 4 quasiexperimentelles Design
  - 2/4 Studien positives Ergebnis (1-, 1+/-)
- 3 Case Control Design
  - 1/3 Studien positives Ergebnis (2-)
- 12 historisch kontrollierte Studien
  - 12/12 Studien positives Ergebnis

Abb. 33 Systematischer Review des SVR (2008): Nur kontrollierte Studien berücksichtigt; mäßiger Einfluss des Studiendesigns

Im Jahr 2012 wurde ein Review des BQS Institut für Qualität und Patientensicherheit GmbH veröffentlicht, das vom Bundesministerium für Gesundheit beauftragt war (Veit et al. 2012). Dieser Review basierte auf einer weiter gefassten Operationalisierung des Begriffes P4P, indem er sich auf das „Leistungsniveau“ und nicht auf die Qualität als Zielparameter bezog, und nutzte außerdem Einschlusskriterien für die untersuchten Studien, die auch Fallserien und Fallberichte einschlossen. Die Wertung der Studienlage und die Empfehlungen bleiben zurückhaltend, die langfristigen Effekte werden jedoch zutreffend und kritisch dargestellt.

### Zusammenfassung:

Die frühen Erfahrungen mit P4P waren sehr ermutigend, die ersten Studien und Reviews kamen zu besseren Ergebnissen als die späteren Evaluationen. Der Sachverständigenrat sprach sich auf der Basis seines frühen Systematischen Reviews (Studien bis 2006) zur kurzfristigen Wirksamkeit von P4P auf definierte Versorgungsziele vorsichtig für einen Einsatz von P4P aus, da 21 von 28 kontrollierten Studien einen positiven Effekt zeigten.

## 6.4 Langfristige Effekte

### 6.4.1 HQIP und VBP in den USA

Vor allem aber waren es die langfristigen Ergebnisse der beiden großen Studien, die in den USA (im Krankenhausbereich) und in Großbritannien (im

ambulanten Bereich) durchgeführt wurden, die die Grenzen eines isolierten und voraussetzungslosen P4P-Einsatzes vor Augen führten, gleichzeitig aber wichtige Hinweise für die weitere Entwicklung und einen integrierten Einsatz gaben. Auf diesen Erfahrungen basiert vor allem das *Value-Based Purchasing* Programm (VBP) von Medicare, an dem alle Akutkrankenhäuser der USA teilnehmen (s. Kap. 4.5). Das größte P4P-Projekt im Krankenhausbereich war das *Premier Hospital Quality Incentive Demonstration Project* (HQIP) von Premier und den *Centers of Medicare and Medicaid Services* (CMS) (Grossbart 2006; Lindenauer et al. 2007):



*Das Premier Hospital Quality Incentive Demonstration Project (HQIP) wurde seit 2003 von den Centers of Medicare and Medicaid (CMS) und der Premier Healthcare Alliance, einer Vereinigung von 2900 Krankenhäusern zum Zweck der Qualitätsverbesserung (2013), betrieben (Lindenauer et al. 2007). Die Teilnahme war freiwillig. Alle teilnehmenden 266 Krankenhäuser (63% der in Frage kommenden) hatten bereits an einem Benchmarking Projekt von Premier teilgenommen (Public Reporting). Bis zu 2% des Budgets wurde nach den Ergebnissen von 34 Prozessindikatoren als Zu- (bestes Dezil) oder Abschlag (schlechtestes Dezil) umverteilt (relativer Anreiz). Ab dem vierten Jahr (2007) wurde auch das Erreichen eines Performance-Medians finanziell vergütet (absoluter Anreiz) (Werner et al. 2011). Die Indikatoren gehörten zu klassischen Sets der JCAHO, der AHRQ und des NQF, sie betrafen ambulant erworbene Pneumonie, Herzinfarkt, chronische Herzinsuffizienz, Bypass-OP und Knie/Hüft-TEP. Insgesamt wurden in den ersten 5 Jahren 48 Mill. \$ umverteilt. Das HQIP diente zur Vorbereitung des Medicare Value-Based Purchasing Program, das seit 2012 läuft.*

Im Zweijahreshorizont konnte hier bei 266 Krankenhäusern, die P4P kombiniert mit *Public Reporting* einsetzten, an Prozessindikatoren zu fünf Krankheitsgruppen und an einem *Composite*-Indikator eine recht deutliche Verbesserung gegenüber einer Kontrollgruppe, die lediglich *Public Reporting* allein einsetzte, beobachtet werden (Lindenauer et al. 2007). Allerdings gab es bereits zu diesem Zeitpunkt bezüglich der Indikatoren zum Herzinfarkt eine große unabhängige Studie, die eine positive Wirkung von P4P in Abrede stellte (Glickman et al. 2007). In der Folge wurde die initial positive Wertung von HQIP durch drei große Untersuchungen mit langfristiger Perspektive relativiert:

- **In einer ersten Untersuchung (Werner et al. 2011)**, die einen um zwei Jahre längeren Zeitraum als die Studie von Lindenauer et al. (2007) umfasste, war beim *Composite*-Indikator kein Vorteil der P4P-Häuser mehr festzustellen, allerdings konnten die P4P-Häuser die Verbesserung zwei Jahre schneller als die Kontrollen umsetzen (nach drei Jahren hatten 56% der P4P-Häuser einen Score von 90% erreicht gegenüber 32% der



Kontrollen). Gerade bei P4P-Krankenhäusern, die hohe Zahlungen erhalten hatten (da sie viele Medicare-Patienten behandelt haben), war der Unterschied zu Kontrollhäusern, die bei Teilnahme an HQIP gleich hohe Zahlungen erhalten hätten, besonders groß, ein indirekter Hinweis auf die Bedeutung der Höhe des Anreizes. Außerdem waren die Qualitätsverbesserungen in Märkten mit geringer ausgeprägter Konkurrenz (regional wenige Häuser vorhanden, oder: Monopolstellung durch Konzentration) deutlich höher als in Märkten mit ausgeprägter Konkurrenz, da in den wettbewerbsintensiven Märkten Qualitätsverbesserungen bereits durch *Public Reporting* stattgefunden hatten. Auch waren die Verbesserungen bei Krankenhäusern mit einer gesunden finanziellen Grundausstattung stärker als bei finanziell schwach aufgestellten Häusern (Werner et al. 2011).

- **Die erkrankungsbezogene Langfristauswertung des HQIP durch Ryan et al. (2012B)** zeigte dann über einen 6-Jahres-Zeitraum (2003–2009) keine Verbesserung der P4P-Häuser (P4P plus *Public Reporting*) gegenüber gemachten Kontrollhäusern (*Public Reporting* allein); alle beobachteten Krankenhäuser wurden besser in ihren Qualitätsdaten, ohne dass sich die beiden Gruppen jedoch unterschieden. Allerdings war bei den Interventionshäusern eine deutliche Abnahme der Varianz der Ergebnisse zu beobachten. Anders als vermutet, verbesserten sich die anfangs schlechteren Häuser nicht stärker als die Top-Häuser. Dies war auch zu beobachten, nachdem man 2006 begonnen hatte, neben der relativen Position der Häuser (an der Spitze Zuschläge, am „schlechten Ende“ Abschläge) zusätzlich auch das Überschreiten eines definierten Schwellenwertes (Median der Performance in den Vorjahren aller Häuser) zu belohnen, damit auch die im Mittelfeld stehenden Krankenhäuser einen Anreiz haben, sich zu verbessern (zur monetären Bewertung s. Kap. 8.5.2, Abb. 39).
- **In der dritten Studie** konnte ebenfalls keine Verbesserung festgestellt werden, und zwar im Vergleich zwischen den 252 HQIP-Krankenhäusern und 3363 Kontrollhäusern (*Public Reporting* allein) anhand der 30-Tage-Mortalität, also eines wichtigen Outcome-Parameters. Auch hier zeigte sich keine Veränderung bei den „*poor performers*“ (Jha et al. 2012).

Jenseits der Datenlage lässt sich die eigentliche Bedeutung des *Premier Hospital Quality Incentive Demonstration Project* (HQIP) jedoch erst in seinem langfristigen **gesundheitpolitischen Zusammenhang** abschätzen. Wie in Kapitel 4.5 ausgeführt, gehen die Wurzeln gehen bis in die 90er-Jahre des letzten Jahrhunderts, HQIP war letztlich das Pilotprojekt und die Vorbereitung für das *Value-Based Purchasing* (VBP)-Programm, eine der zentralen Säulen der umfassenden Neuorientierung, in dem sich das Gesundheitswesen der USA derzeit befindet (s. Kap. 4.5).

### ! Wichtigste Gesetzgebung mit gesundheitspolitischen Schwerpunkt in den USA aus den letzten Jahren (aus Kap. 4.5):

- **Balanced Budget Act (1997)**
- **Medicare Prescription Drug, Improvement and Modernization Act (MMA) (2003)**
- **Deficit Reduction Act (2005)**
- **Tax Relief and Health Care Act (2006)**
- **Patient Protection and Affordable Care Act (2010)**

Das **aktuelle VBP-Programm** ist sehr breit aufgestellt, neben dem Krankenhaussektor (*Hospital-VBP*) ist auch die Versorgung chronisch Nierenkranker mit einem eigenen Programm vertreten (*QIP: End-Stage Renal Disease Bundled-Payment and Quality Incentive Program*), und auch die ambulante Versorgung wird über den *Physician Value-Based Payment Modifier (PVBPM)* adressiert, der erste finanzielle Auswirkungen im Jahr 2015 haben wird (VanLare et al. 2012, Tompkins et al. 2009, CMS 2011A, CMS 2011B, Chien und Rosenthal 2013, Ryan et al. 2014A).

Inhaltlich bezieht sich das VBP-Programm auf die **sechs „Qualitätsdomänen“** der *National Quality Strategy* des *Department of Health and Human Services (HHS)* (VanLare et al. 2012), die einerseits Patientensicherheit und Klinische Versorgung betonen, andererseits aber mindestens gleichwertig die Patienten- und „Caregiver“-Erfahrungen (einschließlich der *Patient-Reported Outcomes*), die Koordination der Behandlung und die „*Population or Community Health*“ unter Einschluss des Zugangs zur Versorgung (*Access to Care*) sowie letztendlich die Effizienz und Kostenreduktion in den Vordergrund stellen.

Alle Akutkrankenhäuser, die Medicare-Patienten behandeln, sind zur **Teilnahme am Hospital-VBP verpflichtet**. Derzeit werden bis zu 1% der Erlöse in die Zu- und Abschläge einbezogen, dieser Wert soll bis 2017 auf 2% steigen, zu diesem Zeitpunkt werden dann jährlich knapp 2 Mrd. \$ nach Qualitätskriterien verteilt (Ryan und Blustein 2012C, zu weiteren Daten s. auch Brown et al. 2014). Es werden weitgehend die bereits aus dem Hospital-IQR Programm bekannten **Indikatoren** eingesetzt (s. unten zu „Indikatoren des Hospital Value-Based Purchasing Programms der CMS“), in erster Linie

- Prozessindikatoren und
- Patientenerfahrungen (im ersten Jahr nur diese beiden), weiterhin
- Patientensicherheitsindikatoren und
- *Hospital-Acquired Conditions*,
- nur in ausgewählten Fällen risikoadjustierte Ergebnisindikatoren.

Im ersten Jahr wurden allein prozess- und patientenseitige Indikatoren im Verhältnis von 70 zu 30% verwendet, ab dem 2. Jahr (2014) werden alle 5 Grup-

pen je 20% gewichtet (Ryan und Blustein 2012C). Bemerkenswert ist wiederum die große Bedeutung der Patientenerfahrungen sowie der große Anteil von insgesamt 60%, den die Patientensicherheitsindikatoren ausmachen, zu denen ja auch die *Hospital-Acquired Conditions* und *Mortality*-Indikatoren hinzurechnen sind. Ergebnisse können naturgemäß zum jetzigen Zeitpunkt noch nicht vorliegen, allerdings gibt es bereits jetzt Modellrechnungen über die finanziellen Auswirkungen. Es wird diskutiert, ob diese ausreichen, um wirklich Verbesserungen erreichen zu können (Werner et al. 2012).



### **Indikatoren des Hospital Value-Based Purchasing Programms der CMS (n. Ryan und Blustein 2012C)**

#### **1. Clinical process**

- **Acute myocardial infarction:** fibrinolytic therapy received within 30 minutes of hospital arrival; primary percutaneous coronary intervention received within 90 minutes of hospital arrival
- **Heart failure:** discharge instructions
- **Pneumonia:** blood cultures performed in the emergency department before initial antibiotic received in hospital; initial antibiotic selection for community-acquired pneumonia in immunocompetent patients
- **Infections associated with health care:** prophylactic antibiotic received within 1 hour before surgical incision; prophylactic antibiotic selection for surgical patients; prophylactic antibiotics discontinued within 24 hours after surgery end time; cardiac surgery patients with controlled 6 a.m. postoperative serum glucose levels
- **Surgery:** surgery patients taking a beta-blocker before arrival who received a beta-blocker during the perioperative period; recommended venous thromboembolism prophylaxis ordered; appropriate venous thromboembolism prophylaxis received no more than 24 hours before surgery and up to 24 hours after surgery

#### **2. Patient experience**

- communication with nurses
- communication with physicians
- responsiveness of hospital staff
- pain management; communication about medicines
- hospital cleanliness and quietness
- discharge instructions
- overall rating of hospital

### 3. Patient mortality

- mortality at 30 days, acute myocardial infarction
- mortality at 30 days, heart failure
- mortality at 30 days, pneumonia

### 4. Hospital-acquired conditions

- foreign object retained after surgery
- air embolism
- blood incompatibility
- pressure ulcer stages 3 and 4
- falls and trauma (includes fracture, dislocation, intra-cranial injury, crushing injury, burn, electric shock)
- infections associated with vascular catheters
- urinary tract infections associated with catheters
- manifestations of poor glycemic control

### 5. Patient safety

- composite measure of complication and patient-safety indicators (pressure ulcer, iatrogenic pneumothorax, bloodstream infections related to central venous catheters, postoperative hip fracture, postoperative pulmonary embolism or deep vein thrombosis, postoperative wound dehiscence, and accidental puncture or laceration)
- mortality for selected medical conditions (composite)

### Zusammenfassung:

Das weltweit größte P4P-Projekt im Krankenhausbereich war das *Premier Hospital Quality Incentive Demonstration Project* (HQIP) in den USA, 266 Krankenhäuser mit *Public Reporting* und P4P (*dual use*) wurden mit *Public Reporting* allein verglichen. Nach ermutigenden Zweijahresergebnissen zeigten drei Langzeitauswertungen für die *dual use*-Häuser keinen Vorteil mehr, insbesondere ergab sich keine Verbesserung für die *poor performers*, auch nachdem man neben der relativen Position zusätzlich auch das Überschreiten eines definierten Schwellenwertes belohnte. Allerdings war bei den Interventionshäusern eine deutliche Abnahme der Varianz der Ergebnisse zu beobachten, und die P4P-Häuser konnten die Verbesserungen zwei Jahre schneller als die Kontrollen umsetzen. Bei P4P-Krankenhäusern, die hohe Zahlungen erhalten hatten (viele *Medicare*-Patienten), war ein Unterschied gegenüber Kontrollhäusern, die bei Teilnahme an HQIP gleich hohe Zahlungen erhalten hätten, zu beobachten, ein indirekter Hinweis auf die Bedeutung der Höhe des Anreizes. Die Qualitätsverbesserungen in Märkten mit geringer ausgeprägter Konkurrenz waren höher als in Märkten mit ausgeprägter Konkurrenz (dort Qualitätsverbesserungen bereits durch *Public Reporting*). Finanziell gut aufgestellte Häuser profitierten stärker.

Das aus HQIP entwickelte VBP-Programm (*Value Based Purchasing*) umfasst neben dem Hospital-VBP auch die ambulante Versorgung (*Physician Value-Based Payment Modifier* [PVBPM]). Alle Akutkrankenhäuser, die Medicare-Patienten behandeln, sind zur Teilnahme am Hospital-VBP verpflichtet. Derzeit werden bis zu 1% der Erlöse in die Zu- und Abschläge einbezogen, dieser Wert soll bis 2017 auf 2% steigen, zu diesem Zeitpunkt werden dann jährlich knapp 2 Mrd. \$ nach Qualitätskriterien verteilt. Es werden Prozessindikatoren und Indikatoren aus Patientenerfahrungen eingesetzt, weiterhin Patientensicherheitsindikatoren und *Hospital-Acquired-Conditions*, nur in ausgewählten Fällen risikoadjustierte Ergebnisindikatoren.

### 6.4.2 QOF in Großbritannien

Das *Quality and Outcome Framework* (QOF) des *National Health Service* (NHS) in Großbritannien stellt das zweite, wissenschaftlich breit evaluierte P4P-Projekt dar. Es betrifft den ambulanten Primärarztsektor (*General Practitioner*) und bezieht sich ebenso wie das HQIP in erster Linie auf Prozessindikatoren. Ähnlich wie beim VBP-Programm des CMS hat auch der NHS dem QOL eine umfassende Strategie hinterlegt (Anonymous 1999; Campbell et al. 2007; NHS 2013B), die (ohne eine derart explizit systemtheoretische Fundierung wie in den USA) **fünf Domänen umfasst** und parallel zu entsprechenden *Frameworks* für Public Health und *Adult Social Care* implementiert wurde:

- Domain 1: Preventing people from dying prematurely
- Domain 2: Enhancing quality of life for people with longterm conditions
- Domain 3: Helping people to recover from episodes of ill health or following injury
- Domain 4: Ensuring that people have a positive experience of care
- Domain 5: Treating and caring for people in a safe environment; and protecting them from avoidable harm

Das QOL startete im Jahr 2004 nach umfänglichen Diskussionen zwischen NHS und der britischen Ärzteschaft, die nur nach der Bereitschaft der englischen Regierung, das QOL-Programm mit erheblichen zusätzlichen Mitteln auszustatten, zu einem erfolgreichen Ende geführt werden konnten (Roland 2004). Im Jahr 2013/2014 waren **insgesamt 123 Indikatoren** (NHS 2013A) in Kraft, die sich wie folgt zusammensetzen:

- 95 klinische Indikatoren (s.u.),
- Public Health Domain (18 Indikatoren): Primärprävention KHK (2), Bluthochdruck (1), Adipositas (1), Rauchen (5), Screening Cervix-Ca. (4), Kindergesundheit (1), Müttergesundheit (1), Kontrazeption (3),
- Quality and Productivity Domain (9 Indikatoren, z. B. Teilnahme an Peer Review-Verfahren, Leitlinien, Patientensicherheit),
- Patient Experience Domain (1 Indikator).



### **Klinische QOF-Indikatoren 2013/2014 (NHS 2013A):**

- Vorhofflimmern (4 Indikatoren)
- Sek. Prävention der KHK (6)
- chronische Herzinsuffizienz (6)
- Bluthochdruck (5)
- periphere arterielle Verschlusskrankheit (4)
- Schlaganfall/transit.-ischäm. Attacke (7)
- Diabetes mellitus (16)
- Schilddrüsenunterfunktion (2)
- Asthma (4)
- chronisch-obstr. Lungenerkrankung (6)
- Demenz (3)
- Depression (2)
- psychische Erkrankungen (10)
- Krebs (2)
- chronische Niereninsuffizienz (4)
- Epilepsie (3)
- Lernstörungen (2)
- Osteoporose: sek. Frakturprävention (3)
- rheumatoide Arthritis (4)
- Palliativversorgung (2)

Über diese Indikatoren werden pro Praxis bis zu 1.050 Punkte verteilt, die entsprechend vergütet werden. Es fällt sofort auf, dass das **Spektrum der klinischen Indikatoren** sehr viel breiter und sehr viel mehr auf chronische Erkrankungen ausgerichtet ist als beim HQIP-Programm, wo ja nur 5 klinische (v.a. akute) Krankheitsbilder einbezogen wurden; dieser Umstand ist für die allgemeine GP-Versorgung sicherlich adäquat. Es handelt sich fast ausschließlich um Prozessindikatoren. Für ein Teil der klinischen Indikatoren wird die Bereitstellung von Daten für Erkrankungsregister (z.B. chronische Herzinsuffizienz) gesondert vergütet. Weiterhin fällt die große Bedeutung der *Public Health*- sowie der organisatorischen Indikatoren auf (zu Beginn einschließlich IT-Ausstattung), außerdem ein hoch bewerteter *Composite*-Indikator zu *Patient-Reported Outcomes*. Ziel des Programms ist eine strukturierte, teambasierte und auf evidenzgestützte Qualitätsindikatoren ausgerichtete Versorgung; seit 2009 liegt die Indikatorenentwicklung in den Händen des *National Institute for Health and Clinical Excellence* (NICE) (Gillem et al. 2012).

Ein weiterer deutlicher Unterschied zu HQPI ist die Einführung von sog. *Exception Rules* (s.u.), die es unter bestimmten Bedingungen erlauben, Patienten aus dem P4P-Programm herauszunehmen, insbesondere wenn patientenseitige Gründe für ein Nicht-Erreichen des angestrebten Ziels (z.B. Compliance bzgl. Impfungen) vorliegen (NHS 2013A). Diese Regelung soll einer Risikose-



lektion und Schlechterstellung (bzw. niedrigerer Attraktivität) von Praxen in sozial unterprivilegierten Wohngebieten entgegenwirken.



***Exception Rules (2013/14) (NHS 2013A): Patients may be excepted if they meet the following criteria:***

- A. Patients who have been recorded as refusing to attend review (...)*
- B. Patients for whom it is not appropriate to review the chronic disease parameters due to particular circumstances (...)*
- C. Patients newly diagnosed or who have recently registered with the contractor (...)*
- D. Patients who are on maximum tolerated doses of medication whose levels remain suboptimal.*
- E. Patients for whom prescribing a medication is not clinically appropriate (...)*
- F. Where a patient has not tolerated medication.*
- G. Where a patient does not agree to investigation or treatment (informed dissent) (...)*
- H. Where the patient has a supervening condition which makes treatment of their condition inappropriate (...)*
- I. Where an investigative service or secondary care service is unavailable*

Die **zusätzlich ausgeschütteten Mittel sind sehr hoch**, bereits im ersten Jahr konnte eine Praxis bis zu 42.000 Engl. Pfd. (etwa 50.000 €) zusätzlich einnehmen, wenn alle 1.050 Punkte erreicht wurden. In der ersten Periode der Umsetzung wurde 1 Mrd. Engl. Pfd. qualitätsbezogen verteilt, das entspricht einer 20%-igen Erhöhung des Budget für die *Family Practitioners im United Kingdom* (Roland 2004). Im Median wurden 83,4% der klinischen Qualitätsziele und im Median insgesamt 1.003 Punkte erreicht, 230 der 8.105 auswertbaren Praxen erreichten die maximale Punktzahl von 1.050. Für im Median 6% der Patienten kam ein *exception reporting* zur Anwendung. In den Jahren 2009–2010 kamen die Praxen auf im Durchschnitt 937 Punkte bei einem „Punktwert“ von 130 Engl. Pfd. (Gillem et al. 2012).

Ähnlich wie beim HQIP-Programm kann auch hier zwischen den **frühen Evaluationen**, die eine gewisse Wirksamkeit zeigten, und den **langfristigen Untersuchungen** unterschieden werden, die den Optimismus etwas dämpften. In der Evaluation von Campbell et al. (2007) war in einer *interrupted time-series* Analyse die Qualitätsverbesserung insbesondere bei den Krankheitsbildern Asthma und Typ 2-Diabetes von dem langjährigen säkularen Trend einer langsamen auch ohne Intervention sichtbaren Verbesserung deutlich abzugrenzen, dieser Effekt war aber in den Jahren danach nicht mehr nachweisbar (Campbell et al. 2009). Dieser Effekt wurde als *ceiling*-Effekt interpretiert, eine

Anhebung der Grenzwerte empfohlen (Doran und Roland 2010). Vor allem aber waren die nicht direkt vom P4P-Ansatz betroffenen Diagnosen von dem Verbesserungseffekt ausgeschlossen. Während die patientenorientierte Outcomes z. B. hinsichtlich Zugang und Erreichbarkeit besser wurden, war für die Kontinuität der Behandlung (vor allem hinsichtlich „persönlicher Arzt“) eine Verschlechterung zu verzeichnen. Besondere Aufmerksamkeit erlangte eine Studie, in der Kaiser Permanente in Californien im Primärzktbereich zwei Indikatoren aus dem P4P-Programm herausnahm, die auch im NHS gleichzeitig aktiv waren (Screening des Augenhintergrunds bei Diabetes mellitus, Screening für Zervix-Carcinom), und zeigte, dass die erreichten Verbesserungen wieder auf den Ausgangsbereich zurückfielen (Lester et al. 2010).

Zwei **systematische Reviews, die sich ausschließlich auf Studien zum QOL-Programm** bezogen, erhärteten diese Ergebnisse (Gillem et al. 2012; Langdown et al. 2013). Die Verbesserungen werden ähnlich wie im HQIP zeitlich vorgezogen, Teamarbeit und Veränderung der Rollen gefördert – andererseits beschreiben Patienten eine Verschlechterung in der persönlichen Kontinuität der Betreuung als „Kosten“ für die bessere Erreichbarkeit, nicht angereizte Krankheitsbilder profitieren nicht, und es gibt Befürchtungen, dass soziale Unterschiede, trotz *exception rules*, verstärkt werden.

### **Zusammenfassung:**

Das *Quality and Outcome Framework* (QOF) des NHS im ambulanten Primärzktsektor in Großbritannien stellt das zweite langfristige evaluierte P4P-Projekt dar. Es bezieht sich ebenso wie das HQIP in erster Linie auf Prozessindikatoren, ist sehr auf die Betreuung chronischer konservativer Erkrankungen abgestellt, umfasst organisatorische und *Public Health*-Aspekte sowie Patientenerfahrungen. Seit 2009 liegt die Indikatorenentwicklung in den Händen des NICE. Die zusätzlich ausgeschütteten Mittel sind hoch, in der ersten Periode der Umsetzung wurde 1 Mrd. Engl. Pfd. qualitätsbezogen verteilt (20%-ige Erhöhung des Budgets für die *Family Practitioners*). Im Median wurden 83,4% der klinischen Qualitätsziele erreicht, 230 der 8.105 auswertbaren Praxen erreichten die maximale Punktzahl. Für im Median 6% der Patienten kam ein *exception reporting* zur Anwendung (Ausschluss). Auch hier zeigten kurzfristige Evaluationen eine gewisse Wirksamkeit, während die langfristigen Untersuchungen den Optimismus dämpften. Verbesserungen werden ähnlich wie im HQIP zeitlich vorgezogen, Teamarbeit und Veränderung der Rollen gefördert – andererseits beschreiben Patienten eine Verschlechterung in der persönlichen Kontinuität der Betreuung, nicht angereizte Krankheitsbilder profitieren nicht, und es gibt Befürchtungen, dass soziale Unterschiede verstärkt werden.

### **6.4.3 Zusammenfassung: langfristige Effekte**

Führend sind Prozessindikatoren, Patientensicherheitsindikatoren und Indikatoren auf der Basis von Patientenerfahrungen, im Mittelpunkt stehen zu-



nehmend Probleme der ambulanten Versorgung und Koordination. Ergebnisindikatoren werden risikoadjustiert eingesetzt. Enttäuschenderweise wurde kein *spill over*-Effekt zur Verbesserung von nicht angereizten Indikatoren beobachtet, dafür aber ein deutlicher *ceiling*-Effekt; die *poor performers* profitierten anders als vermutet nicht stärker als die Top-Einrichtungen. Nach Beendigung des Anreizes scheint die Qualität der Versorgung wieder auf das Ausgangsniveau zurückzufallen. Die zusätzlichen (oder einbehaltenen) Zahlungen sind teilweise recht hoch.

Die **Gründe für die geringe langfristige Wirkung von P4P** sind darin zu suchen, dass die Indikatoren bereits langfristig bekannt waren, es zu einem *ceiling*-Effekt kam, dass durch eine mangelhafte Kopplung von Vergütung und Qualitätssystematik nicht zu einer Verbesserung der *poor performers* gekommen war (s. Kap. 8.5.2), dass es einen gleichzeitigen Einsatz von *Public Reporting* und P4P gab (*dual use*), dass die Vergütung (immer noch) zu niedrig war (vor allem unter Berücksichtigung der Opportunitäts- und Grenzkosten) und dass man die Risikoaversion der Entscheidungsträger unterschätzte, ebenso wie die Dominanz des grundlegenden Vergütungssystems (DRG). Diese Faktoren werden in Kapitel 8 ausführlich hergeleitet und diskutiert.

## 6.5 Unerwünschte Nebeneffekte von P4P

Bei jedem Vergütungs- und Anreizsystem muss mit unerwünschten Nebeneffekten gerechnet werden (Brown et al. 2014). Auch bei P4P sind zahlreiche solcher Wirkungen bereits früh beschrieben worden (SVR 2008, Nr. 741ff, Roland 2004) und werden hier zunächst systematisch und synoptisch aufgeführt, um im späteren Verlauf in die Analyse und die Empfehlungen eingearbeitet werden zu können. Ganz entscheidend werden dabei die Frage der professionellen Motivation der Angehörigen der Gesundheitsberufe, die mögliche Verstärkung sozialer Ungleichheit, der Anreiz zur Risikoselektion und die Frage der Mengenausweitung sein, vor allem in Kombination mit anderen Vergütungssystemen. Zu beachten ist, dass sich für eine große Zahl von unerwünschten Effekten Studienergebnisse finden lassen, es aber auch einen gewissen Bias zugunsten ungünstiger Ergebnisse gibt, man muss die Ergebnisse daher wirklich in ihrem Kontext sehen. In Erweiterung der Darstellung des SVR (s. Abb. 33) kann man die Nebeneffekte folgendermaßen systematisieren:

- Zugang zur Versorgung einschl. Kontinuität und Risikoselektion
- professionelle Faktoren
- institutionelle Ebene
- Datenqualität
- Qualität der Versorgung im gleichen Sektor
- Qualität der Versorgung auf Systemebene
- Kosteneffektivität der Versorgung



### **Mögliche Einschränkungen des Zugangs zur Versorgung unter P4P:**

- **Benachteiligung von Bevölkerungsgruppen**
    - ethnische Gruppen
    - sozial benachteiligte Gruppen
    - Exception Reporting
    - Gender
  - **Kontinuität der Versorgung**
  - **Risikoselektion**
- Der **Zugang zur Versorgung** mit P4P-bezogenen Leistungen (s.o.) kann im Einzelfall besser werden (Wodchis et al. 2007), vor allem soweit es sich, wie beim QOF-Programm, um eine versteckte Einzelleistungsvergütung handelt (s. Kap. 8.5.4). Im Allgemeinen wird aber eine Verschlechterung der Versorgung für sozial, ethnisch oder aus anderen Gründen benachteiligten Gruppen befürchtet (Rosenthal und Dudley 2007). Eine Benachteiligung ethnischer Gruppen konnte in den USA bislang nicht belegt werden (Blustein et al. 2011, Casalino und Elster 2007, Ryan 2010), wenngleich immer wieder betont wird, dass Aufmerksamkeit angebracht ist. Bezüglich allgemeiner sozialer Unterschiede ist die Tendenz noch kritischer (Alshamsan et al. 2010; Casalino u. Elster 2007; Jha et al. 2010; Langdown u. Peckham 2013; McLean et al. 2006; Millett et al. 2007). Ein *exception reporting* ist häufiger in Regionen mit niedrigem Einkommen (QOL-Programm in Großbritannien (Doran et al. 2006), in einem Systematischen Review konnte jedoch eine leichte Verbesserung dieser Differenz nachgewiesen werden (Gillam et al. 2012). Besondere Bedeutung kommt der Analyse von Ryan (2013) über das erste Jahr des VBP-Programms von Medicare zu, in der ein inverser Zusammenhang zwischen sozialer Ungleichheit und Zahlungen durch das VBP-Programm nahegelegt wird. Eine Studie berichtet über Nachteile von weiblichen Patientinnen (Millett et al. 2007). Zur Zugangsproblematik gehört auch die **Kontinuität der Versorgung** (Roland 2004). Während sich allgemein die *patient-reported outcomes* unter P4P verbessern, wird die Kontinuität in einigen Studien und Befragungen schlechter (Campbell et al. 2009; Gillem et al. 2012; Maisey et al. 2008). Patienten kommen „nicht zu ihrem Recht“ (Chew-Graham et al. 2013), ob dies an arbeitsteiligeren Konzepten in der ärztlichen Versorgung, die Übernahme ärztlicher Tätigkeiten durch Pflegende oder anderen Ursachen liegt, ist noch nicht abschließend geklärt. Ein weiterer und im Kontext von P4P äußerst wichtiger Aspekt des Zugangs zur Versorgung ist eine mögliche **Risikoselektion**, weil auf diese Weise für bestimmte Patienten bzw. Patientengruppen die Versorgung erschwert sein kann (Brown et al. 2014; Casalino und Elster 2007; Shen 2003). In den USA wurde keine Risikoselektion beobachtet (Ryan u. Blustein 2012C), allerdings wird als Reaktion

auf den *readmission*-Indikator eine Ausdünnung der Notfallaufnahmen befürchtet (Ryan u. Mushlin 2014B). Eine Risikoselektion gehört prinzipiell zu den unmittelbaren Wirkungen eines Finanzierungssystems, das entweder die umfassende Behandlung von Erkrankungsgruppen oder Populationen (z.B. Disease Management oder Managed Care) in den Mittelpunkt stellt (s. Kap. 8.6) oder auf den Behandlungserfolg (Ergebnisindikatoren) setzt; bei Verwendung von Prozessindikatoren ist sie weniger stark ausgeprägt (Mehta et al. 2008). Die Tendenz zur Risikoselektion stellt daher eines der entscheidenden Argumente bei der Auswahl der verwendeten Indikatoren und der Integration von P4P in bestehende Vergütungssysteme dar (s. auch Kap. 8.6).

Zusammenfassend kann festgehalten werden, dass das Thema einer Verstärkung sozialer Benachteiligung durch P4P besonderer Beachtung bedarf, ebenso muss die Kontinuität der Versorgung und die Gefahr einer Risikoselektion mit daraus folgender Zugangsbeschränkung beachtet werden. Die Kontrolle aller drei Formen ist eine wichtige Aufgabe der politischen Instanzen bei Implementierung von P4P.

#### **Gliederung (s.o.): Unerwünschte Nebeneffekte von P4P**

- **Zugang zur Versorgung einschließlich Kontinuität und Risikoselektion**
  - *professionelle Faktoren*
  - *institutionelle Ebene*
  - *Datenqualität*
  - *Qualität der Versorgung im gleichen Sektor*
  - *Qualität der Versorgung auf Systemebene*
  - *Kosteneffektivität der Versorgung*
  
- **Professionelle Faktoren** (zur Gliederung s. nochmals oben) beziehen sich in erster Linie auf die professionelle Motivation der unterschiedlichen Gesundheitsberufe, die Berufsbilder und die Koordination der Arbeit auf Teamebene (s.u. zu „Unerwünschte Effekte von P4P. Professionelle und institutionelle Faktoren einschl. Datenqualität“). Grundsätzlich muss bei jedem Vergütungssystem, das im beruflichen Kontext angewandt wird, mit Auswirkungen auf Motivation, Rolle und Zusammenarbeit gerechnet werden (Prendergast 1999). Zur **Motivation** der Ärzte besteht die wichtigste Befürchtung darin, dass die interne Motivation (professionelle Einstellung), durch die externe Motivation i.S. einer qualitätsbezogene Vergütung gemindert wird (sog. „*crowding out*“; Berenson et al. 2013; Cassel u. Jain 2012). Empirisch ist dies jedoch nicht zu belegen, wie Untersuchungen z.B. im QOL-Projekt in Großbritannien zeigen (McDonald et al. 2007), und auch von theoretischer Seite gibt es

Gegenargumente (Prendergast 1999, Staehle 1999, S. 242). Dieses häufig geäußerte Argument ist also nicht so durchschlagend, wie man es auf den ersten Blick meint. Allerdings kommt es zu Verschiebungen in den beruflichen **Rollen**, insbesondere ändert sich im Rahmen einer P4P-Einführung die Arbeitsteilung zwischen Ärzten und Pflege. Durch die Übernahme vorher ärztlicher Tätigkeiten durch die **Pflege** kann es dort zu zusätzlicher Belastung und Unzufriedenheit kommen (Kurtzman et al. 2011, Maisey et al. 2008, McDonald et al. 2007). Eventuell wird die **Teamarbeit** erschwert (Maisey et al. 2008), wenngleich im Allgemeinen eine Verbesserung berichtet wird.

- Auf der **institutionellen Ebene** (s.u. zu „Unerwünschte Effekte von P4P. Professionelle und institutionelle Faktoren einschl. Datenqualität“) kommt es durch eine P4P-Einführung zu „organisatorischem Stress“. In der internen Steuerung wird jetzt nicht nur über Mengen und Kosten (bzw. interne Leistungsverrechnung) diskutiert, sondern es muss über die Qualität der Leistungserbringung gesprochen werden, da sie jetzt vergütungsrelevant ist (Corrigan u. McNeill 2009). Dies beginnt mit der Erhebung der Qualitätsdaten, eine Ergänzung oder gar Restrukturierung der IT-Ausstattung wird oft notwendig. Zentral ist dann die Frage, ob die Organisation oder die einzelnen für die Behandlung verantwortlichen Ärzte oder Abteilungen Adressaten der P4P-Vergütung sind (Rosenthal u. Dudley 2007). Die Vergütung muss dabei mit dem fachlichen Verantwortungsbereich kongruent sein. Während im Allgemeinen die Organisation die Zahlungen empfängt (ohne dass sie weitergeleitet werden), gibt es Hinweise darauf, dass (relativ kleine) Zahlungen direkt an die Behandler sehr effektiv sind (Chung et al. 2010; Torchiana et al. 2013). Das Konzept der Ärztlichen Leitung (*medical leadership*) wird ein bedeutender Faktor (Frolich et al. 2007), da hier diese Diskussionen entscheidend mit gestaltet werden. Durch die organisatorische Kleinteiligkeit wird die Situation im ambulanten Bereich noch komplizierter, weil die Verantwortung für die Qualität der Leistung auf verschiedene Akteure verteilt ist (Pham et al. 2007).
- Die **Datenqualität** betrifft in diesem Zusammenhang in erster Linie die Reliabilität (Zuverlässigkeit) der Daten (Begriffe und Abgrenzung zur Validität s.o.). Unter mangelnder Reliabilität versteht man die zwischen Personen und zu unterschiedlichen Zeiträumen inkonstante Erfassung der Daten, vor allem aber auch Verzerrungen durch Mängel der Dokumentation und Fälschung der Daten (sog. *gaming*, s.u. zu „Unerwünschte Effekte von P4P. Professionelle und institutionelle Faktoren einschl. Datenqualität“). Während die erste Form durch Verbesserung der Definition und der Spezifikationen der verwendeten Indikatoren anzugehen ist, ist dies im Fall des *gaming* schwieriger. Der Grund besteht darin, dass dem *gaming* eine skeptische oder ablehnende Grundhaltung gegenüber der Qualitätsmessung zugrundeliegt, und zwar sowohl bei den Gesundheitsberufen als auch auf der Ebene der Organisation. Es gibt Studien, in denen Ärzte in Befragungen ein *gaming* offen zugeben (Maisey et al. 2008). Da in

diesen Fällen Qualitätsmessung und -vergleich weder zum professionellen Selbstverständnis noch zur Qualitätskultur der Einrichtung gehört, werden Ausflüchte und Erklärungen gesucht (Mears u. Webley 2010). Man unterscheidet analog zur Steuerhinterziehung eine partiell und eine komplett ablehnende Haltung, bei ersterer wird bis zu einem gewissen Punkt Compliance geübt und erst darüber hinaus Datenmanipulation betrieben, bei der zweiten Haltung werden von vorneherein die Daten verfälscht. In den USA gibt es im Bereich der Datenerhebung zur Infektionsepidemiologie die Einrichtung eines „klinischen Vetos“ (*clinician veto*) gegen die Erhebung von nosokomialen Infektionen, in denen das nach infektionsepidemiologischen Kriterien erhobene Ereignis durch klinische Einschätzungen in Frage gestellt wird (und meist trotzdem antibiotisch behandelt wird); die Infektionsepidemiologen warnen dort vor einer Gefährdung der für infektiologische Zwecke optimierten Erhebungsmethodik (Talbot et al. 2013). In jedem Fall ist die Frage der Qualitäts- bzw. Sicherheitskultur, einer der Eckpfeiler der Organisationskultur und damit zentraler Baustein des organisatorischen Zusammenhalts, bei der Beurteilung des *gaming* von zentraler Bedeutung (s. organisatorisches Lernen).

### **Unerwünschte Effekte von P4P. Professionelle und institutionelle Faktoren einschl. Datenqualität**

#### **Professionelle Faktoren:**

- **Motivation**
- **Rollenverständnis der Gesundheitsberufe**
- **Teamarbeit**

#### **Institutionelle Faktoren:**

- **IT-Ausstattung**
- **Verantwortung für Qualität**
- **Adressaten der Zahlungen**
- **Konzept des medical leadership**

#### **Datenqualität:**

- **technische Erhebungsprobleme**
- **Gaming**

Zusammenfassend sind auf professioneller und institutioneller Ebene Veränderungen der beruflichen Rollen beschrieben, die Unsicherheit erzeugen. Verlässliche Daten zur Infragestellung der internen Motivation der Ärzte durch den externen finanziellen Anreiz gibt es kaum. vergütungsrelevante Qualitätsaspekte stellen große institutionelle Herausforderungen für die interne Steuerung dar. Die Reliabilität der Daten betrifft vor allem das *gaming*, begleitet von zentralen Defiziten im Hinblick auf die Organisations- und Sicherheitskultur.

- Die **Qualität auf der Ebene des gleichen Sektors** ist gefährdet, wenn die Anreizwirkung eines Instrumentes wie P4P trotz erhaltenem Zugang zur Versorgung und professioneller sowie institutioneller Akzeptanz nicht zur gewünschten Verbesserung der Versorgung führt, ohne dass dies auf methodische Ursachen wie z.B. inadäquate Kopplung von Qualität und Vergütung zurückzuführen wäre (s.u. zu „Unerwünschte Effekte von P4P: Sektor- und Systemebene“). Einer der hier relevanten Mechanismen wird als *teaching to the test* bezeichnet. Man versteht darunter die einseitige Ausrichtung der Aktivitäten auf den Indikator unter Vernachlässigung der anderen Aufgaben, um „gute Ergebnisse“ zu erzielen (Brown et al. 2014, Casalino und Elster 2007). Durch dieses Verhalten wird die „Stellvertreterfunktion“ des Indikators, der als gut messbarer Parameter Qualitätsprobleme in einem Versorgungsbereich vorhersagen soll, untergraben (s. Kap.2.4.1). Es handelt sich hierbei um eine dem Indikatorkonzept immanente Gefahr, denn Indikatoren machen Qualitätsprobleme explizit und lösen Vermeidungsreaktionen aus. Bei P4P und anderen Formen des „Qualitätswettbewerbs“ (zur Systematik s. Kap. 4.4) ist diese Gefahr noch größer, denn P4P tendiert dazu, finanzielle Ressourcen von nicht angereizten Leistungen zu angereizten Leistungen zu allozieren (Ryan 2010). Allerdings ist dieses Phänomen zumindest in den amerikanischen Programmen nicht in großem Umfang zu beobachten gewesen (Ryan und Blustein 2012). Ganz entscheidend ist die Frage, ob ein *spill over* zu anderen, nicht angereizten Leistungsbereichen nachzuweisen ist, und ob sich durch P4P im Sinne der lernenden Organisation eine bleibende Verbesserung der Versorgung erreichen lässt, auch wenn das P4P-Programm beendet wird. Im QOL-Programm in Großbritannien konnte gezeigt werden, dass nicht angereizte Leistungsbereiche sich nicht verbessern, obwohl dies eigentlich angenommen worden war (Gillem et al. 2012, Langdown und Peckahm 2013). Zumindest in einer Studie konnte nachgewiesen werden, dass nach Beendigung von P4P-Anreizen die Qualität der von P4P betroffenen Leistung wieder zurückgeht (Lester et al. 2010).
- Die **Qualität auf Systemebene** wird hier von der potenziellen Qualitätsverschlechterung auf der Ebene des gleichen Sektors abgegrenzt, um die Gefährdung allgemeiner Gesundheitsziele (adäquate Behandlung chronisch und mehrfach erkrankter Patienten, Mengenanreiz, Versorgungsengpässe) zu kennzeichnen (s.u. zu „Unerwünschte Effekte von P4P: Sektor- und Systemebene“). Man kann das *ceiling*, den ausbleibenden Verbesserungseffekt durch mangelndes Verbesserungspotenzial, auch als methodisches Problem ansehen, aber es hat in jedem Fall systemische Auswirkungen, weil das Instrument in Verruf gerät und Ressourcen für die Messung und gesonderte Vergütung von Indikatoren verwendet werden, bei denen gar keine Möglichkeit einer Verbesserung mehr besteht (Bhattacharyya et al. 2009; Doran und Roland 2010). *Ceiling* tritt besonders dann auf, wenn die für P4P verwendeten Indikatoren bereits lan-



ge im Gebrauch sind und auch schon für andere Anreizsysteme (in erster Linie *Public Reporting*) verwendet wurden (Ryan u. Blustein 2012). Insbesondere bei ungenügend risikoadjustierten P4P-Konzepten auf Basis von Ergebnisindikatoren kann es zu einer **Benachteiligung kleiner Einrichtungen** kommen (Davidson et al. 2007), so dass in ländlichen Regionen sogar die Gefahr einer Unterversorgung entsteht. Ist die Abgrenzung zur Einzelleistungsvergütung ungenügend, kann der **Mengenanreiz**, z.B. bereits bestehend durch das DRG-System, noch verstärkt werden. Wenn P4P nicht zielgerichtet und strategisch eingesetzt wird, kann die Versorgung von Patientengruppen, die von besonderem Interesse sind (z.B. die Versorgung von multimorbiden chronisch Kranken) unter P4P sogar leiden (Cannon 2006), insbesondere falls sich darunter die Tendenz zur **Risikoselektion** verbirgt (s.o.). In einer bemerkenswerten Stellungnahme von H. Luft wird darauf hingewiesen, dass die Datenmenge und -verfügbarkeit, die für das Funktionieren von P4P notwendig sind (insbesondere für die Risikoadjustierung), eine ernsthafte Bedrohung des **Datenschutz** darstellen kann (Luft 2012).



### ***Unerwünschte Effekte von P4P: Sektor- und Systemebene***

#### *gleicher Sektor*

- *teaching to the test*
- *spill over*
- *bleibende Verbesserung*

#### *System*

- *ceiling*
  - **Benachteiligung kleiner Einrichtungen**
  - **Mengenanreiz**
  - **Versorgung chronisch Kranker**
  - **Datenschutz**
- Die **Kosteneffektivität von P4P** ist letztendlich auch als Systemparameter anzusprechen, denn wenn die Kosten für die Durchführung von P4P derart hoch sind, dass sie an anderer Stelle sinnvoller zur Qualitätsverbesserung hätten eingesetzt werden können, ist das Ziel einer besseren und kostenbewussteren Gesundheitsversorgung nicht erreicht (Roland 2004). Die Kosteneffektivität ist aber sehr wenig untersucht (Cannon 2006), in Großbritannien wird eine gemäßigte Kosteneffektivität berichtet (Gillem et al. 2012).

Zusammenfassend kann auch auf Systemebene ein Auftreten von unerwünschten Wirkungen von P4P nicht ausgeschlossen werden. In Verbindung von Risikoselektion, Benachteiligung kleiner Einrichtungen durch statistische Effekte und resultierender Verschlechterung der wohnortnahen Versorgung kann gerade die Versorgung von chronisch, multimorbiden älteren Patienten leiden, die eigentlich durch Instrumente zur Qualitätsverbesserung besonders unterstützt werden sollten. Die Kosteneffektivität der P4P-Programme scheint gegeben zu sein.