

Krankenhaus-Report 2011

„Qualität durch Wettbewerb“

Jürgen Klauber / Max Geraedts /
Jörg Friedrich / Jürgen Wasem (Hrsg.)

Schattauer (Stuttgart) 2011

Auszug Seite 117-130



9	Pay for Performance im Krankenhaus: Erfahrungen in den USA.....	117
	<i>Ateev Mehrotra, Cheryl L. Damberg, Melony E.S. Sorbero, Stephanie S. Teleki und Sören Mattke</i>	
9.1	Einleitung	118
9.2	Vorgehensweise	119
9.2.1	Merkmale der drei P4P-Programme für stationäre Krankenhausbehandlung.....	119
9.2.2	Ergebnisse der veröffentlichten Artikel und methodische Fragen	121
9.3	Diskussion	126
9.4	Literatur	129

9 Pay for Performance im Krankenhaus: Erfahrungen in den USA^{1,2}

Ateev Mehrotra, Cheryl L. Damberg, Melony E.S. Sorbero, Stephanie S. Teleki und Sören Mattke

Abstract

In den USA gibt es inzwischen mehr als 40 Pay-for-Performance (P4P)-Programme im privaten Krankenhausesektor, und das amerikanische Parlament erwägt die Einführung eines P4P-Programms für Medicare-finanzierte Krankenhausbehandlungen. Angesichts des wachsenden Interesses an P4P im Krankenhaus wird in diesem Beitrag anhand eines systematischen Literaturreviews untersucht, welche Wirkung P4P hinsichtlich klinischer Prozessqualität, Behandlungsergebnisse, Patientenerfahrungen, Sicherheit und Ressourcenauslastung hat. Noch gibt es wenige formelle Bewertungen von P4P im Krankenhaus, und die meisten der acht bisher veröffentlichten Studien weisen methodische Mängel auf. Die stringentesten Studien konzentrieren sich auf die klinische Prozessqualität. Es zeigt sich, dass Krankenhäuser, die an der Premier Hospital Quality Incentive Demonstration, einem P4P-Programm der Centers for Medicare und Medicaid Services (CMS), teilgenommen hatten, eine Qualitätssteigerung aufwiesen, die um zwei bis vier Prozentpunkte höher lag als bei den Krankenhäusern der Kontrollgruppe. Um die Wirkungen von P4P im Krankenhaus zu bewerten und zu beurteilen, ob der Nutzen von Investitionen in P4P die damit verbundenen Kosten aufwiegt, bedarf es weiterer systematischer Evaluation.

More than 40 private sector hospital pay-for-performance (P4P) programs now exist, and Congress is considering initiating a Medicare hospital P4P program. Given the growing interest in hospital P4P, this systematic review of the literature examines the current state of knowledge about the effect of P4P on clinical process measures, patient outcomes and experience, safety, and resource utilization. Little formal evaluation of hospital P4P has occurred, and most of the eight published studies have methodological flaws. The most rigorous studies focus on clinical process measures and demonstrate that hospitals participating in the Centers for Medicare and Medicaid Services Premier Hospital Quality Incentive Demonstration, a P4P program, had a 2- to 4-percentage point greater improvement than the improvement observed in control hospitals. There is a need for more systematic evaluation of hospital P4P to understand its effect and whether the benefits of investing in P4P outweigh the associated costs.

1 Dieser Beitrag beruht auf einem im American Journal of Medical Quality erschienenen Artikel der Autoren.

2 Übersetzung aus dem Englischen: Susanne Sollmann

9.1 Einleitung

Es gilt als erwiesen, dass die Ausgestaltung des Vergütungssystems Einfluss auf die Art der Leistungserbringung hat. Vor diesem Hintergrund liegt der Gedanke einer qualitätsbezogenen Vergütung nahe, das heißt einer Leistungshonorierung, die gute Qualität mit einer entsprechend besseren Vergütung verbindet und auf diese Weise einen Anreiz zur Qualitätssicherung und -steigerung setzt. Angesichts dieser Überlegung überrascht wie wenig solche Vergütungsansätze bisher in die Tat umgesetzt werden. Dabei muss freilich in Betracht gezogen werden, dass die in der Theorie so einleuchtende Verbindung von Qualitätsstandards und Vergütungshöhe hinsichtlich ihrer praktischen Umsetzung eine Reihe von Herausforderungen zu meistern hat.

In anderen Ländern, insbesondere in den USA und in Großbritannien, wurde bereits eine erhebliche Zahl von P4P-Programmen implementiert. Hinsichtlich einer stärkeren Nutzung dieses Ansatzes lohnt es sich daher, die dortigen Erfahrungen einer eingehenden Analyse zu unterziehen. Der vorliegende Beitrag referiert die Ergebnisse von Studien zu P4P-Programmen in den USA, wo derartige Programme mittlerweile eine weite Verbreitung gefunden haben. Es existieren schätzungsweise vierzig Programme mit einem Fokus speziell auf Krankenhauspatienten.

Trotz der enthusiastischen Einschätzung der P4P-Krankenhausprogramme ist ihre Wirkung bisher nicht hinreichend bewiesen. Frühere Untersuchungen, die sich auf P4P-Programme im ambulanten Bereich konzentrierten, fanden – wenn überhaupt – nur geringe Qualitätszuwächse (Rosenthal et al. 2006b; Sorbero et al. 2006; Petersen et al. 2006). Teilweise wurden Befürchtungen geäußert, dass P4P-Programme Datenmanipulationen fördern und andere unerwünschte Konsequenzen haben könnten (Petersen et al. 2006; Casalino und Elster 2007). Soweit uns bekannt ist, existieren bisher keine Überblicksartikel über Studien zu P4P-Programme für die stationäre Behandlung. Um zur Diskussion über die Vorteile von P4P-Programmen für Medicare und andere private Kostenträger beizutragen, unterzogen wir die publizierte Literatur zu den Auswirkungen von P4P auf die stationäre Behandlung einem systematischen Review.

Dabei richteten wir besonderes Augenmerk auf das Umfeld, in dem die Programme initiiert wurden. Seit 2004 sammeln und veröffentlichen die Centers for Medicare and Medicaid Services (CMS) in Zusammenarbeit mit der Hospital Quality Alliance, einer Partnerschaft von Anbietern, Versicherungen und Forschern zur Verbesserung der Versorgungsqualität im Krankenhaus, und der Joint Commission, einer Organisation für die Akkreditierung von Krankenhäusern, Qualitätsinformationen auf Krankenhausebene für die Versorgung des akuten Myokardinfarkts, der ambulant erworbenen Pneumonie und der Herzinsuffizienz (Medicare Prescription Drug, Improvement and Modernization Act 2003; Jha et al. 2005). Mehr als 98% der US-amerikanischen Akutkliniken übermitteln Leistungsdaten an die CMS, und seit Beginn des Programms konnte eine wesentliche Verbesserung bei der Behandlung dieser Krankheiten beobachtet werden (Lindenauer et al. 2007). Daher beachteten wir bei unserer Untersuchung insbesondere darauf, ob die Analysen für den Einfluss dieses aktuellen Trends zu Qualitätssteigerung kontrollierten.

9.2 Vorgehensweise

Unser Überblick bezieht alle englischsprachigen Beiträge in Zeitschriften mit Peer Review ein, die die Auswirkungen von P4P-Programmen auf die stationäre Versorgung beschreiben. Dazu gehören alle randomisierten, kontrollierten Studien, quasi-experimentellen Untersuchungen sowie Prä- und Postinterventionsstudien. Wir begrenzten unsere Suche auf Artikel, die Ergebnisse über den Einfluss der Vergütung auf Verbesserungen der klinischen Behandlungsqualität, von Patientenerfahrungen, der Sicherheit oder des Ressourcenverbrauchs auswiesen. Ausgeschlossen wurden Artikel, die sich auf die Auswirkungen größerer Umstellungen des Krankenhausvergütungssystems bezogen, wie zum Beispiel den Wechsel hin zu einem prospektiven Vergütungssystem, sowie Artikel, die sich auf P4P im ambulanten Bereich (also auch eventuelle ambulante Leistungen der Krankenhäuser) oder auf Anreize zur vermehrten Organspende bei Krankenhauspatienten konzentrierten.

Im ersten Suchlauf wurden 902 Artikel (ohne Duplikate) identifiziert. Nach Durchsicht der Abstracts wurden zehn Artikel für weitere gezielte Untersuchungen ausgewählt, von denen acht unsere Einschlusskriterien erfüllten. Die meisten nicht ausgewählten Beiträge behandelten das Thema Qualitätssteigerung ohne Bezug zu Anreizprogrammen. Die Konsultierung von P4P-Experten und eine Überprüfung einschlägiger Reports führten nicht zur Identifizierung von weiteren Artikeln.

Die acht Artikel, die in unsere Analyse einbezogen wurden, befassen sich mit den Auswirkungen der folgenden drei P4P-Programme für die stationäre Versorgung: (1) das P4P-Programm „Hospital Quality Service and Recognition“ der Hawaii Medical Service Association, (2) das „Participating Hospital Agreement Incentive Program“ der gemeinnützigen Krankenversicherungsgesellschaft Blue Cross Blue Shield of Michigan und (3) die „Premier Hospital Quality Incentive Demonstration (PHQID)“ der CMS.

9.2.1 Merkmale der drei P4P-Programme für stationäre Krankenhausbehandlung

Tabelle 9–1 zeigt die wichtigsten Merkmale der drei P4P-Programme, die im Folgenden im Detail beschrieben werden. Alle drei Programme enthalten klinische prozessbezogene Indikatoren, die sich meist auf den akuten Myokardinfarkt, die ambulant erworbene Pneumonie und die Herzinsuffizienz beziehen – jene Diagnosen, für die die CMS Qualitätsinformationen veröffentlichen. Die Programme wenden auch andere Qualitätsindikatoren an (Donabedian 1988), wie Behandlungsergebnisse (z. B. Mortalität nach einer koronaren Bypass-Operation), Kosteneffizienz (z. B. Verweildauer nach einer Operation), Struktur (z. B. Teilnahme des Krankenhauses an einem Programm zur Förderung der Patientensicherheit), Patientensicherheit (z. B. postoperative Nachblutungen oder Hämatome) und Patientenerfahrungen (z. B. Patientenbefragungen). Einige Programme bewerten die Leistung eines Krankenhauses gemessen anhand einer absoluten Schwelle (z. B. 85 % der Patienten mit akutem Myokardinfarkt erhalten Betablocker), während andere die relative Performance der Krankenhäuser untersuchen (z. B. Vergleich mit dem besten Krankenhaus bezüglich der Abgabe von Betablockern).

Tabelle 9-1

Merkmale der bewerteten P4P-Programme für Krankenhäuser

P4P-Programm	Indikator						Ziel		Finanzieller Anreiz		
	Ergebnis	Prozess	Struktur	Patienten- erfahrungen	Patienten- sicherheit	Kosten- effizienz	absolut	relativ	Bonus	Einbe- haltung	Ab- schläge
Blue Cross Blue Shield of Michigan Participating Hospital Agreement Incentive Program		x	x		x		x	x	x		
Hawaii Medical Service Association Hospital Quality Service and Recognition P4P Program	x	x	x	x	x	x	x	x	x		
CMS* Premier Hospital Quality Incentive Demonstration	x	x			x			x	x		x

* CMS: Centers for Medicare and Medicaid Services

9.2.2 Ergebnisse der veröffentlichten Artikel und methodische Fragen

In diesem Abschnitt werden die drei Pay-for-Performance-Programme für Krankenhäuser beschrieben, die Ergebnisse zusammengefasst und die Grenzen des jeweils gewählten Studiendesigns diskutiert.

Das P4P-Programm „Hospital Quality Service and Recognition“ der Hawaii Medical Service Association

Zwei Artikel bewerteten den Effekt dieses P4P-Programms, das 2001 gestartet wurde und alle 17 Krankenhäuser auf Hawaii einbezog (Berthiaume et al. 2006; Berthiaume et al. 2004). Die 17 Krankenhäuser erhielten pro Jahr insgesamt etwa \$ 9 Millionen (im Mittel etwa \$ 500 000 pro Krankenhaus). Die Incentive-Zahlungen basierten auf den folgenden vier Faktoren:

(1) Teilnahme an einem Workshop „Get with the Guidelines – Coronary Heart Disease“ der American Heart Association. Die Krankenhäuser konnten Punkte sammeln, indem sie sich für den Workshop anmeldeten, als „leitlinienreues Krankenhaus“ anerkannt wurden, ein Patienten-Management-Tool für die Datenerhebung einsetzten und bei mindestens drei von fünf Prozessmaßnahmen zur Behandlung des akuten Myokardinfarkts eine 85%ige Leistung erreichten. (2) Die Komplikationsrate und durchschnittliche Verweildauer des Krankenhauses (adjustiert für die Krankheitsschwere) für Geburtshilfepatientinnen oder Patienten, die sich einer von 18 häufig vorgenommenen Operationen unterzogen. (3) Die Zufriedenheit der Patienten und der Ärzte mit der stationären und der Notfallbehandlung. (4) Der vom Krankenhaus selbst berichtete Erfolg bei der Umsetzung eines internen Qualitätssteigerungsprogramms.

Berthiaume et al. (2004) fanden, dass von 13 Krankenhäusern auf Hawaii, die mehr als 30 Aufnahmen aufgrund des akuten Koronarsyndroms aufwiesen, zehn (77%) an einem Leitlinien-Workshop teilgenommen hatten. Nach Ansicht der Autoren ist die hohe Beteiligung der in Frage kommenden Krankenhäusern bemerkenswert, weil dieser Prozentsatz der Teilnahme an dem Programm viel höher sei, als durch Diffusionsmodelle zur Verbreitung von Innovationen im Gesundheitswesen vorhergesagt worden war (Berthiaume et al. 2004, S. 503).

In einer weiteren Studie berichten Berthiaume et al. (2006) unter anderem über Verweildauerveränderungen, Komplikationsraten, Patientenzufriedenheit und die Nutzung von internen Initiativen zur Qualitätssteigerung. Die Studie verglich Veränderungen von 2001 bis 2004. Die Komplikationsrate für Patienten der Chirurgie und der Geburtshilfe sank um rund 2 Prozentpunkte und die Verweildauer von Chirurgie- und Geburtshilfepatienten um 1,2 bzw. 0,4 Tage. Die Zufriedenheit der Patienten mit der stationären Versorgung erhöhte sich von 78% auf 79% und die Zufriedenheit mit der Notfallbehandlung von 71% auf 75%. Die Bewertung für die Umsetzung der internen Qualitätsvorgaben verbesserte sich von 2003 auf 2004 von 4,25 auf 6,5 von 10 möglichen Punkten. Die Autoren geben nicht an, ob die zwischen den Zeiträumen beobachteten Unterschiede statistisch signifikant waren.

Die Studien (Berthiaume et al. 2006, 2004) über das „Hospital Quality Service and Recognition“-Programm der Hawaii Medical Service Association sind jedoch nur begrenzt aussagefähig, da es keine Vergleichsgruppe oder Trenddaten gab. Es

bleibt unklar, ob die beobachteten Veränderungen mit dem Anreizprogramm oder mit langfristigen Trends einer verbesserten Versorgung des akuten Myokardinfarkts und einer verkürzten Verweildauer innerhalb des betrachteten Zeitraums zu erklären sind (Lindenauer et al. 2007; Williams et al. 2005). Keine der beiden Studien berichtet über Veränderungen bei den klinischen Prozessmaßnahmen, die ein Teil des Programms waren.

Das „Participating Hospital Agreement Incentive Program“ von Blue Cross Blue Shield of Michigan

Drei veröffentlichte Artikel untersuchten die Wirkung dieses Programms in 86 Krankenhäusern, die einen Vertrag mit Blue Cross Blue Shield of Michigan geschlossen hatten (Nahra et al. 2006; Reiter et al. 2006; Sautter et al. 2007). Nahra et al. untersuchten die Wirtschaftlichkeit des Programms, Reiter et al. analysierten die Ergebnisse einer Umfrage unter den teilnehmenden Krankenhäusern, um die Gründe für beobachtete Veränderungen in der medizinischen Praxis zu finden, und Sautter et al. untersuchten die Ergebnisse von qualitativen Interviews in zehn Krankenhäusern. Die Kliniken erzielten einen Zuschlag von bis zu 4 % der fallpauschalenbasierten Vergütung; die Zuschläge im Jahr 2004 reichten von \$ 30 000 bis \$ 4 000 000 pro Krankenhaus. Ursprünglich basierten die Anreizzahlungen auf dem Abschneiden bei verschiedenen prozessbezogenen Qualitätsmaßen mit Bezug auf die Behandlung des akuten Myokardinfarkts und der Herzinsuffizienz. Im Laufe der Zeit wurde das Programm auf andere Krankheiten wie die ambulant erworbene Pneumonie ausgedehnt und belohnte auch die Beteiligung an Gesundheitsförderungsprojekten (z.B. Raucherentwöhnung) und Patientensicherheitsinitiativen (z.B. die Prophylaxe-Initiative im Rahmen des „Keystone Intensive Care Unit Project“ in Michigan, das vom Keystone Center der Health & Hospital Association in Lansing, Michigan finanziert wird). Ursprünglich basierten die Leistungszuschläge im Rahmen des Programms auf dem relativen Ranking eines Krankenhauses im Vergleich mit konkurrierenden Häusern; inzwischen hängen die Zahlungen vom Erreichen einer absolute Schwelle ab.

Die Studie von Nahra et al. (2006) schätzt die Wirtschaftlichkeit aus der Perspektive der Krankenversicherung ein. Zur Abschätzung der Kosten des Programms wurden die Zahlungen durch die Krankenversicherung und die Verwaltungskosten einbezogen. Der Nutzen des Programms wurde auf Basis der Anzahl der Herzinfarkt- und Herzinsuffizienz-Patienten, die eine verbesserte Behandlung erhielten, bewertet. Der Prozentsatz der Patienten, die eines der folgenden Arzneimittel erhielten, hat sich von 2000 auf 2003 wie folgt erhöht: Aspirin bei akutem Myokardinfarkt (von 87 % auf 95 %), Betablocker bei akutem Myokardinfarkt (von 81 % auf 93 %) und ACE-Hemmer bei Herzinsuffizienz (von 70 % auf 80 %). Basierend auf veröffentlichten klinischen Studien schätzten die Wissenschaftler, dass die beobachtete Verbesserung der Prozessqualität einem Gewinn von 733 bis 1 701 qualitätsadjustierten Lebensjahren (QALYs) entspricht. Insgesamt lagen die geschätzten Kosten je QALY zwischen \$ 12 967 und \$ 30 081, einem Bereich, der im Allgemeinen als kosteneffizient gilt (Ubel et al. 2003).

Reiter et al. (2006) befragten die teilnehmenden Krankenhäuser danach, ob das P4P-Programm strukturelle und prozessorientierte Veränderungen ausgelöst habe. Strukturelle Veränderungen umfassten die Einrichtung einer Stabsstelle für Quali-

tätsmanagement oder einen Personalwechsel beim Qualitätsbeauftragten. Prozessorientierte Veränderungen umfassten die Einführung eines computergestützten Systems für Arzneimittelverordnungen (CPOE-System – Computerized Physician Order Entry-System) oder die Schaffung von Case-Management-Teams. Von den 66 Befragten (Rücklaufquote 70%) gaben 48% an, dass in ihrem Krankenhaus eine strukturelle Veränderung vorgenommen wurde, 59% berichteten über eine Prozessveränderung. Insgesamt gaben 75% der antwortenden Krankenhäuser an, dass sie mindestens eine Veränderung vorgenommen hatten.

In einer qualitativen Studie befragten Sautter et al. (2007) Führungskräfte und Kardiologen an zehn Krankenhäusern in Michigan. Sieben Kliniken konnten ihre Leistung verbessern, zwei erbrachten bereits vorher sehr gute Leistungen und eine Klinik verschlechterte sich. Nur an zwei von zehn Krankenhäusern waren die Befragten der Ansicht, Anreize förderten eine Qualitätssteigerung. Die Autoren stellen jedoch in Frage, ob dies bedeutet, dass das P4P-Programm keine Qualitätssteigerung nach sich zog. Sie beobachteten, dass die Krankenhäuser die Zahlungen für Qualitätssteigerungsprojekte verwendeten und führen aus, dass „die Zuschläge einigen Krankenhäuser offensichtlich erst ermöglichten, in Qualität zu investieren“ (Sautter et al. 2007, S. 105). Die Autoren deuten dies in dem Sinne, dass „leistungsschwache Krankenhäuser, die über eine basale Infrastruktur für die Qualitätssteigerung verfügen, den größten Erfolg zu verzeichnen hatten, wenn sie Zahlungen erhielten“ (Sautter et al. 2007, S. 104).

Bei den oben zitierten Studien gibt es einige wichtige Einschränkungen. Bei den Studien von Nahra et al. 2006 und Reiter et al. 2006 fehlen eine Kontrollgruppe oder Trenddaten, um beurteilen zu können, ob die beobachteten Prozessverbesserungen oder die Anwendung von Qualitätssteigerungsmaßnahmen auf die Teilnahme am Programm zurückzuführen sind oder lediglich unabhängige Trends repräsentieren. Nahra et al. berücksichtigten außerdem nicht die Kosten der Krankenhäuser für die Erfassung von Qualitätsdaten und die Durchführung von Qualitätssteigerungsmaßnahmen. Obwohl diese Kosten nicht unmittelbar von der Krankenversicherung übernommen werden, fließen sie in der Regel in eine Wirtschaftlichkeitsanalyse aus der üblichen volkswirtschaftlichen Sichtweise ein. Wären diese Kosten berücksichtigt worden, würde das Programm möglicherweise nicht als kosteneffizient betrachtet. Und schließlich führen die von Reiter et al. und Sautter et al. beobachteten vermehrten Qualitätssteigerungsaktivitäten nicht unbedingt tatsächlich zu einer Verbesserung der Ergebnisqualität.

Die „Premier Hospital Quality Incentive Demonstration (PHQID)“ der CMS

Drei Artikel bewerten die Wirkung der Premier Hospital Quality Incentive Demonstration (PHQID) (Lindenauer et al. 2007; Glickman et al. 2007; Grossbart 2006). Dieses von den CMS finanzierte dreijährige P4P-Demonstrationsprojekt startete im Jahr 2003. Rund 450 Krankenhäuser, die das Datensystem der Firma Premier Inc (Charlotte, North Carolina) verwendeten, wurden zur Teilnahme aufgefordert, weitere 262 Krankenhäuser in 38 Staaten meldeten sich aus eigenem Antrieb. Die Krankenhäuser wurden nach ihrem zusammengefassten Leistungsindex für die Behandlung der folgenden fünf Diagnosen beurteilt: akuter Myokardinfarkt, ambulant erworbene Pneumonie, Herzinsuffizienz, koronare Bypass-Operation und Hüft- oder Kniegelenkersatz-Operation.

Die meisten Indikatoren bezogen sich auf klinische Prozesse, aber das Programm umfasste für einige Diagnosen auch Indikatoren, die Mortalität, Wiederaufnahme- und Komplikationsraten betrafen. Für jede Diagnose erhielten die Krankenhäuser des obersten Zentils einen zweiprozentigen Zuschlag auf ihre Medicare-Entgelte für Krankenhauspatienten, die aufgrund der entsprechenden Krankheiten aufgenommen worden waren. Krankenhäuser des zweiten Zentils erhielten einen einprozentigen Zuschlag. Die Zuschläge für Krankenhäuser betragen im Durchschnitt \$ 71 960 pro Jahr (Lindenauer et al. 2007). Krankenhäuser, die bis zum Ende des dritten Jahres die während des ersten Jahres erreichten Leistungen von Krankenhäusern der untersten zwei Zentile nicht überschritten, mussten einen Abschlag von 1 % bis 2 % auf die Medicare-Entgelte für die jeweilige Diagnose hinnehmen.

Grossbart (2006) untersuchte die Wirkung des PHQID-Programms auf Krankenhäuser der katholischen Healthcare Partners (Cincinnati, Ohio). Dabei stellte er vier Krankenhäuser, die an der PHQID teilgenommen hatten, sechs Kontrollkrankenhäusern gegenüber, die nicht teilgenommen hatten. Die teilnehmenden Krankenhäuser und die nicht teilnehmenden Kliniken ähnelten sich in einer Reihe von Merkmalen, unter anderem Durchführung von Operationen am offenen Herzen, Entlassungen pro Jahr und Leistungsspektrum. Die Analyse wurde auf eine Teilmenge von 17 von 34 Indikatoren der PHQID-Initiative begrenzt, die sowohl bei den beteiligten als auch den nicht beteiligten Krankenhäusern erhoben wurden. Sowohl die teilnehmenden als auch die nicht teilnehmenden Kliniken zeigten eine signifikante Verbesserung bei allen Indikatoren, wobei bei den teilnehmenden Krankenhäusern stärkere Verbesserungen zu verzeichnen waren (allgemeine Leistungssteigerung von 9,3 versus 6,7 Prozentpunkten; Differenz 2,6 Prozentpunkte, $P < .001$). Die Verbesserungen unter den teilnehmenden Krankenhäusern waren bei den folgenden beiden Indikatoren am größten: (1) Instruktionen für Patienten mit Herzinsuffizienz bei Entlassung aus dem Krankenhaus (40,1 versus 14,6 Prozentpunkte bei nicht teilnehmenden Krankenhäusern) und (2) Abgabe von Pneumokokken-Impfstoff an Patienten mit Lungenentzündung (31,6 versus 22,1 Prozentpunkte bei den nicht teilnehmenden Krankenhäusern).

Lindenauer et al. (2007) führten die gründlichste Evaluation der PHQID durch: Sie untersuchten Leistungsveränderungen über einen Zeitraum von zwei Jahren für 10 von 34 Indikatoren in 207 PHQID-Krankenhäusern und einer Kontrollgruppe von 406 Krankenhäusern. Die Kontrollgruppe wurde nach der Bettenzahl, dem Status als Lehrkrankenhaus (ja/nein), Region (Nordost, Mittlerer Westen, Süden oder Westen), Regionstyp (städtisch oder ländlich) und Eigentumsverhältnisse (gewinnorientiert oder nicht gewinnorientiert) ausgewählt. Bei einem Gesamtindikator, der sich aus zehn einzelnen Indikatoren zusammensetzte, verzeichneten die PHQID-Krankenhäuser eine um 4,3 Prozentpunkte stärkere Verbesserung als die Krankenhäuser der Kontrollgruppe. Dieser Unterschied war durchgehend für jede der drei Diagnosen (akuter Myokardinfarkt, ambulant erworbene Pneumonie, Herzinsuffizienz), für die meisten Einzelindikatoren und für einen Indikator für angemessene Versorgung, der den Prozentsatz der Patienten misst, die mithilfe aller für eine bestimmte Krankheit empfohlenen Maßnahmen behandelt werden, zu beobachten.

Die Autoren untersuchten, ob diese unterschiedlichen Verbesserungen durch eine Selbstselektion verursacht sein könnten, das heißt, ob diejenigen Krankenhäuser, die sich zur Teilnahme am PHQID-Programm bereiterklärt hatten, sind grund-

legend von den nicht teilnehmenden Krankenhäusern unterschieden. (Lindenauer 2007). Nach Berücksichtigung des Leistungsniveaus und der Fallzahlen zu Studienbeginn verringerte sich der Unterschied von 4,3 auf 2,9 Prozentpunkte. Bei einem weiteren Versuch zur Ermittlung möglicher Verzerrungen analysierten die Autoren alle Krankenhäuser, die für die Teilnahme an dem PHQID-Programm in Frage kamen, unabhängig davon, ob sie tatsächlich teilnahmen oder nicht. In Frage kommende Krankenhäuser hatten größere Qualitätssteigerungen zu verzeichnen als jene, die nicht teilnahmeberechtigt waren.

Schließlich untersuchten Glickman et al. (2007) die Wirkung der PHQID auf Krankenhäuser, die freiwillig an CRUSADE, einer landesweiten Initiative zur Qualitätssteigerung bei akutem Myokardinfarkt teilnahmen (Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes With Early Implementation of the American College of Cardiology/American Heart Association Guidelines – etwa: Kann eine schnelle Risikostratifizierung instabiler Angina-pectoris-Patienten unerwünschte Behandlungsergebnisse verhindern, indem die Behandlungsleitlinien des American College of Cardiology/der American Heart Association frühzeitig eingeführt werden?). Krankenhäuser, die an CRUSADE teilnehmen, erhalten ein Feedback über ihre Versorgung von Patienten mit Myokardinfarkt sowie verschiedene Schulungsmaßnahmen. Die Wissenschaftler verwendeten ein quasi-experimentelles Studiendesign, das die Tatsache ausnutzt, dass 54 der an CRUSADE beteiligten Krankenhäuser sich bereit erklärten, ebenfalls an der PHQID teilnehmen. Sie verglichen diese 54 Krankenhäuser, die sowohl an CRUSADE als auch an der PHQID teilnahmen, mit einer Kontrollgruppe von 446 Krankenhäusern, die zwar an CRUSADE teilnahmen, nicht jedoch an der PHQID. Zusätzlich zu einem Vergleich anhand von sechs auf den akuten Myokardinfarkt bezogenen Indikatoren der PHQID wurden die Krankenhäuser anhand von acht weiteren Myokardinfarkt-bezogenen Prozessindikatoren verglichen, die nicht Basis für Leistungszuschläge waren. Die Studie versucht zu ermitteln, ob die Teilnahme an der P4P-Intervention einen zusätzlichen Schub für Leistungssteigerungen gab, der über das hinausging, was die CRUSADE-Interventionen bewirkten. Außerdem wurde analysiert, ob die Teilnahme an dem P4P-Programm sich negativ auf die Leistungen bei Indikatoren auswirkte, die nicht als Grundlage für die Anreize dienten.

Die Leistungsunterschiede zwischen den PHQID-Krankenhäusern und den Krankenhäusern der Kontrollgruppe von Juli 2003 bis Juni 2006 bei jenen Indikatoren, die als Grundlage für Leistungszuschläge dienten, (Verbesserung um 7,2 bzw. 5,6 Prozentpunkte; Differenz von 1,6 Prozentpunkten) und bei den nicht zuschlagsrelevanten Indikatoren (Verbesserung um 13,6 bzw. 8,1 Prozentpunkte; Differenz 5,5 Prozentpunkte) waren statistisch nicht signifikant (Glickman et al. 2007). Die PHQID-Krankenhäuser hatten signifikant stärkere Verbesserungen bei drei Einzelindikatoren, darunter zwei, die zuschlagsrelevant waren (z.B. Aspirinverordnung bei Entlassung und Beratung zur Raucherentwöhnung für aktuelle Raucher oder Patienten, die vor Kurzem mit dem Rauchen aufgehört hatten) und einer, der nicht zuschlagsrelevant war (Verordnung von Lipidsenkern bei Entlassung). Es gab zwischen den zwei Gruppen bei den stationären Patienten keine statistisch signifikanten Mortalitätsunterschiede.

Die drei veröffentlichten Artikel, die die PHQID bewerten, sind die bisher umfassendsten Evaluationen von P4P für stationäre Krankenhausbehandlung (Lin-

denauer et al. 2007; Glickman et al. 2007; Grossbart 2006). Alle drei verwenden eine Kontrollgruppe, und in allen Fällen zeigten sich auch in dieser Kontrollgruppe signifikante Qualitätssteigerungen. Was zu den Qualitätssteigerungen in den Krankenhäusern der Kontrollgruppe geführt hat, ist nicht bekannt, es ist jedoch davon auszugehen, dass die öffentliche Berichterstattung der CMS eine Rolle gespielt hat. Unter den Krankenhäusern, die für das PHQID-Programm in Frage kamen, unterscheiden sich diejenigen, die teilgenommen haben, von denen, die nicht beteiligt waren. Bei Berücksichtigung der Verzerrungen durch die Krankenhausausswahl stellten Lindenauer et al. fest, dass die Wirkung des P4P-Programms erheblich von 4,3 auf 2,9 Prozentpunkte gesunken war.

Zu den Einschränkungen der Studien gehören auch die Untersuchungsgesamtheit der ausgewerteten Krankenhäuser. Grossbart und Glickman et al. untersuchten die Versorgung in ausgewählten Kliniken, einschließlich Häusern in katholischer Trägerschaft und in denen, die freiwillig an CRUSADE teilnahmen. Die Ergebnisse sind nur eingeschränkt verallgemeinerbar, da diese Krankenhäuser sich möglicherweise in ihrem Engagement und ihren Ressourcen für Qualitätssteigerungen von den anderen Krankenhäusern unterscheiden, die an PHQID teilgenommen haben. Die Studie von Grossbart ist zudem durch die geringe Stichprobengröße begrenzt sowie durch die Tatsache, dass der Ausgangswert für die Leistungen der Krankenhäuser nicht gemessen wurde. Glickman et al. prüften, ob P4P zu einer schlechteren Performance bei der Behandlung des akuten Myokardinfarkts bei den Indikatoren führte, die nicht als Grundlage für Leistungszulagen verwendet werden. Allerdings betreffen diese Indikatoren die gleiche Diagnose (akuter Myokardinfarkt), die im P4P-Programm verwendet wurde, sodass diese Studie keinen Aufschluss darüber gibt, ob es zu unbeabsichtigten positiven oder negativen Auswirkungen auf die Qualität bei der Behandlung anderer Krankheiten kam, die nicht zuschlagfähig waren. Zu guter Letzt sind die Krankenhäuser, die sich für eine Teilnahme an PHQID entschieden hatten, im Vergleich zu allen Krankenhäusern der USA größer und eher nicht profitorientiert (Lindenauer et al. 2007). Daher ist unklar, ob die Erfahrungen mit dem PHQID-Programm auf alle Krankenhäuser im Land übertragen werden können.

9.3 Diskussion

Obwohl der Einsatz von Krankenhaus-P4P-Programmen zur Förderung von Qualitätssteigerungen in den letzten zehn Jahren zugenommen hat, wurden nur wenige Programme systematisch untersucht und der Wissensstand über die Auswirkungen dieser Programme weist erhebliche Lücken auf. Mit Stand Juni 2007 waren nur acht Artikel mit Peer Review zu diesem Thema veröffentlicht worden (Tabelle 9–2). Bei den meisten fehlte eine Kontrollgruppe, was die Aussagekraft erheblich abschwächt, wenn man davon ausgeht, dass Krankenhäuser, die nicht an P4P-Programmen teilnahmen, wesentliche Qualitätssteigerungen bei den gleichen Indikatoren aufwiesen, die im Allgemeinen als Maßstab für Leistungszuschläge genutzt werden. Die drei Studien (Lindenauer et al. 2007; Glickman et al. 2007; Grossbart 2006), die eine Kontrollgruppe verwendeten, bewerten die Wir-

Tabelle 9–2

Studien zu P4P für Krankenhäuser

P4P-Programm	Quelle	Art der Studie	Bewertung von Leistungssteigerungen	Verwendung einer Kontrollgruppe
Blue Cross Blue Shield of Michigan Participating Hospital Agreement Incentive Program	Reiter et al. 2006	Erhebung von Veränderungen der medizinischen Praxis	entfällt ¹	nein
	Sautter et al. 2007	Semistrukturierte Interviews in zehn teilnehmenden Kliniken	entfällt ²	nein
	Nahra et al. 2006	Kosten-Nutzen-Analyse	ja	nein
Hawaii Medical Service Association Hospital Quality Service and Recognition P4P Program	Berthiaume et al. 2004	Beschreibung einer Programmkomponente und damit verbundene Ausgaben	nein	nein
	Berthiaume et al. 2004	Beschreibung von Trends anhand von Indikatoren	ja	nein
CMS* Premier Hospital Quality Incentive Demonstration	Grossbart 2006	Evaluation der Qualitätsverbesserungen der teilnehmenden Kliniken im Vergleich zur Kontrollgruppe	ja	ja
	Lindenauer et al. 2006	Evaluation der Qualitätsverbesserungen der teilnehmenden Kliniken im Vergleich zur Kontrollgruppe	ja	ja
	Glickman et al. 2007	Evaluation der Qualitätsverbesserungen der teilnehmenden Kliniken im Vergleich zur Kontrollgruppe	ja	ja

* CMS: Centers for Medicare and Medicaid Services

¹ Analyse von Maßnahmen zur Qualitätsverbesserung, nicht von Leistungssteigerungen.

² Qualitative Interviews mit Vertretern von Kliniken, die aufgrund von Leistungssteigerungen ausgewählt wurden.

kung des gleichen P4P-Programms (PHQID). Deshalb bezieht sich der aussagekräftigste Nachweis der Auswirkungen von Krankenhaus-P4P auf ein einziges Krankenhaus-P4P-Programm.

Die Auswertung des PHQID-Programms ergab eine Verbesserung, die um 2 bis 4 Prozentpunkte über der der Krankenhäuser der Kontrollgruppe lag (Lindenauer 2007; Glickman et al. 2007; Grossbart 2006). Diese geringe Steigerung deckt sich mit der Wirkung von Anreizprogrammen in anderen Ländern (Asch und Warner 1996). Für das PHQID-Programm konnte keine verringerte Mortalität bei stationären Herzinfarktpatienten nachgewiesen werden, obwohl es Meinungsverschiedenheiten über das Verhältnis zwischen Prozessqualität und Mortalität gibt (Werner und Bradlow 2006; Peterson et al. 2006b). Eine Studie (Petersen et al. 2006a) fand keinen Anhalt für die unerwünschte mögliche Folge, dass PHQID mit einer Verringerung der Bemühungen bei den nicht zuschlagfähigen Indikatoren einhergeht. Die Ergebnisse einer anderen Studie (Drake et al. 2007) (die nicht für dieses Review geeignet ist) legen nahe, dass die am PHQID-Programm

teilnehmenden Krankenhäuser im Bestreben, ihre Werte für den Einsatz von Antibiotika bei Behandlung der ambulant erworbenen Pneumonie zu verbessern, Antibiotika bei anderen Atemwegserkrankungen wie Asthma oder bei chronischer Herzinsuffizienz zu häufig verwendeten. Zudem müssen die Kosten für das Monitoring der Leistungen und die Missbrauchskontrolle in Betracht gezogen werden, wenn man über die Einführung von leistungsorientierten Vergütungssystemen nachdenkt (Asch und Warner 1996).

Viele Fragen zum Thema P4P für Krankenhäuser bleiben offen. Erstens evaluieren die vorliegenden Studien P4P im Rahmen der öffentlichen Berichterstattung über die Qualität von Krankenhäusern; die Wirksamkeit des Krankenhaus-P4P ohne öffentliche Berichterstattung ist deshalb nicht bekannt. Zweitens ist die Wirkung eines P4P-Programms in Form eines Bonusmodells (z. B. Zahlung von Zuschlägen, deren Höhe von den Verbesserungen eines Krankenhauses im Vergleich zum Vorjahr abhängt) verglichen mit einem Anreizsystem, das nicht nur die besten zwei Zentile belohnt, unklar, da keine der umfassenden Studien ein solches Modell evaluiert. Drittens gibt es kaum Hinweise darauf, ob ein P4P-Programms für die stationäre Krankenhausversorgung sich auf andere als prozessbezogene Indikatoren der Krankenhausqualität auswirkt. Teilweise wurde erwartet, dass P4P Kosteneinsparungen bewirkt, weil Qualitätssteigerungen zu weniger Komplikationen und weniger Ressourcenverbrauch führen. Vertreter der Firma Premier Inc berichteten auf einer nationalen Konferenz (Remus 2006) und im Rahmen einer Pressemitteilung, die PHQID habe zu verminderten Kosten und einer gesunkenen Mortalität geführt. Diese Ergebnisse sind noch nicht veröffentlicht worden und es scheint, dass keine Kontrollgruppe verwendet wurde, was nach unserer Ansicht als kritisch einzuschätzen ist. Viertens beschränken sich die als Grundlage für P4P verwendeten Diagnosen auf Herz-Kreislauf-Erkrankungen, die ambulant erworbene Pneumonie und Hüft- oder Kniegelenkersatz, deshalb ist nicht bekannt, ob P4P-Anreize bei anderen Krankheiten eine größere oder geringere Wirkung erzielen würden. Fünftens ist nicht bekannt, welche Effekte ein Krankenhaus-P4P-Programm hätte, das einen kleineren oder größeren Anreiz als den im PHQID-Programm gewährten (1 bis 2% der Medicare-Entgelte) nutzt.

Die Erkenntnisse aus diesem Literaturreview werfen für Krankenversicherungen, Krankenhäuser und die politischen Entscheidungsträger schwierige Fragen hinsichtlich eines P4P-Programms für Krankenhäuser auf. Sind die beobachteten Verbesserungen bei evidenzbasierten klinischen Prozessen durch Krankenhaus-P4P die Investitionen wert, wenn man sie mit anderen Interventionen vergleicht, die zu Qualitätssteigerungen beitragen könnten? Würde ein Krankenhaus-P4P-Programm, das andere Kennzahlen oder größere Anreize verwendet, zu größeren Verbesserungen führen? Wenn die meisten der beobachteten Qualitätssteigerungen durch eine gesetzlich verordnete öffentliche Berichterstattung induziert werden, ist es dann vorteilhaft, zusätzlich zur öffentlichen Berichterstattung ein weiteres P4P-Programm zu implementieren? Rechtfertigt eine Verbesserung dieser Größenordnung durch P4P für Krankenhäuser die Kosten der Aufrechterhaltung eines P4P-Programms und die – noch unbewiesenen – Bedenken, dass P4P unbeabsichtigte negative Konsequenzen haben kann? Wegen der begrenzten Datengrundlage sind die Antworten auf diese Fragen unklar. Dennoch gilt weiterhin als erwiesen, dass die Qualität der Gesundheitsversorgung unzureichend ist. Zukünftig wird es wich-

tig sein, P4P für Krankenhäuser auf breiterer Basis zu evaluieren, um auf die Gestaltung dieser Programme Einfluss zu nehmen und zu bestätigen, dass die Anreize tatsächlich die gewünschten Ziele erreichen.

9.4 Literatur

- Asch B, Warner J. Incentive systems: theory and evidence. In: Lewin D, Mitchell D, Zaidi M, eds. *Handbook of Human Resource Management*. Greenwich, CT: JAI Press 1996;175–215.
- Berthiaume JT, Tyler PA, NgOsorio J, LaBresh KA. Aligning financial incentives with “Get With the Guidelines” to improve cardiovascular care. *Am J Manag Care*. 2004;10 (7, pt 2): 501–4.
- Berthiaume JT, Chung RS, Ryskina KL, Walsh J, Legorreta AP. Aligning financial incentives with quality of care in the hospital setting. *J Healthc Qual*. 2006; 28 (2):36–44, 51.
- Casalino LP, Elster A. Will payforperformance and quality reporting affect health care disparities? *Health Aff (Millwood)*. 2007; 26 (3): w405–14.
- Damberg CL, Sorbero ME, Mehrotra A, Teleki SS, Lovejoy S, Bradley L. An environmental scan of pay for performance in the hospital setting: final report. November 2007. [http:// aspe.hhs.gov/ health/reports/08/payperform/PayPerform07 .html](http://aspe.hhs.gov/health/reports/08/payperform/PayPerform07.html). (16. Oktober 2008).
- DeFrances CJ, Hall MJ; Division of Health Care Statistics, Centers for Disease Control and Prevention. 2005 National Hospital Discharge Survey. July 12, 2007. <http://www.cdc.gov/nchs/data/ad/ad385.pdf> (16. Oktober 2008).
- Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988; 260(12):1743–8.
- Drake DE, Cohen A, Cohn J. National hospital antibiotic timing measures for pneumonia and antibiotic overuse. *Qual Manag Health Care*. 2007;16 (2):113–22.
- Dudley RA, Rosenthal M. *Pay for Performance: A Decision Guide for Purchasers*. Rockville, MD: Agency for Healthcare Research and Quality; 2006. AHRQ publication 06–0047.
- Glickman SW, Ou FS, DeLong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007; 297 (21): 2373–80.
- Grossbart SR. What’s the return? assessing the effect of “pay-for-performance” initiatives on the quality of care delivery. *Med Care Res Rev*. 2006; 63(1) (suppl):29S–48S.
- Hospital quality improving, cost, mortality rate trends declining for participants in Medicare pay-for-performance project; 2008 [press release]. Charlotte, NC: Premier Inc; 31. Januar 2008.
- Institute of Medicine. *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington, DC: National Academy Press 2006.
- Jha AK, Li Z, Orav EJ, Epstein AM. Care in U.S. hospitals: the Hospital Quality Alliance program. *N Engl J Med*. 2005; 353 (3): 265–74.
- Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med*. 2007; 356 (5): 486–496.
- Medicare Prescription Drug, Improvement, and Modernization Act of 2003. 501(b) (2003). [http:// www.ihs.gov/NonMedicalPrograms/chs/documents/PL108173%20%20Medicare%20Modernization%20Act%20Sec.pdf](http://www.ihs.gov/NonMedicalPrograms/chs/documents/PL108173%20%20Medicare%20Modernization%20Act%20Sec.pdf) (17. Oktober 2008).
- Medicare Payment Advisory Commission. *Report to the Congress: variation and innovation in Medicare*. Juni 2003. http://www.medpac.gov/documents/June03_Entire_Report.pdf 16. Oktober 2008.
- Mehrotra A, Epstein AM, Rosenthal MB. Do integrated medical groups provide higher-quality medical care than individual practice associations? *Ann Intern Med*. 2006;145(11): 826–833.
- Milgate K, Cheng SB. Pay-for-performance: the MedPAC perspective. *Health Aff (Millwood)*. 2006; 25 (2): 413–9.
- Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JR. Costeffectiveness of hospital payfor-performance incentives. *Med Care Res Rev*. 2006; 63 (1) (suppl): 49S–72S.
- Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does payforperformance improve the quality of health care? *Ann Intern Med*. 2006;145 (4): 265–72.

- Peterson ED, Roe MT, Mulgund J, et al. Association between hospital process performance and outcomes among patients with acute coronary syndromes. *JAMA*. 2006; 295 (16): 1912–20.
- Reiter KL, Nahra TA, Alexander JA, Wheeler JR. Hospital responses to payforperformance incentives. *Health Serv Manage Res*. 2006;19 (2):123–134.
- Remus D. Financial incentives work! results of the CMS Hospital Quality Incentive Demonstration Project Year One. Präsentiert beim Academy Health Annual Research Meeting; 25.–27. Juni 2006; Seattle, WA.
- 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.
- Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med*. 2006a; 355 (18):1895–02.
- Sautter KM, Bokhour BG, White B, et al. The early experience: of a hospital-based pay-for-performance program. *J Healthc Manag*. 2007; 52 (2):95–108.
- Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood)* 2005; 24 (1):138–50.
- Sorbero ME, Damberg CL, Teleki S, et al. Assessment of PayforPerformance Options for Medicare Physician Services: Final Report. RAND working paper prepared for the Assistant Secretary for Planning and Evaluation, US Dept of Health and Human Services. Santa Monica, CA: RAND 2006.
- Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn't it increase at the rate of inflation? *Arch Intern Med*. 2003;163 (14):1637–41.
- Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA* 2006; 296 (22): 2694–702.
- Williams SC, Schmaltz SP, Morton DJ, Koss RG, Loeb JM. Quality of care in U.S. hospitals as reflected by standardized measures, 20022004. *N Engl J Med*. 2005; 353 (3): 255–64.