

Kurt Hess

P4P in der ambulanten Medizin: ein Ansatz mit Potenzial

Zu Pay for Performance (P4P) gibt es bisher nur wenige quantitative Untersuchungen, für schlüssige Aussagen zur Wirkung auf die Qualität ist es noch zu früh. Dieser Beitrag gibt einen Überblick über den Stand des Wissens, die laufende Diskussion und die Argumentarien der Befürworter und Skeptiker.

Wer heute über die Prinzipien des Pay for Performance (P4P) spricht, ohne sie gleich in Bausch und Bogen zu verwerfen oder – am andern Ende der Skala – als Allweltslösung zu preisen, spricht bereits auf relativ hohem Level über die Geheimnisse der medizinischen Qualitätssicherung. Während diese in den vergangenen Jahren auch beim Gros der Leistungserbringenden eine gewisse Salonfähigkeit erreicht hat, nicht zuletzt aufgrund der immer vernehmlicheren Forderungen von politischer Seite, wird es beim Thema der Qualitätsbemessung schon deutlich enger. Und wenn dann gar die finanzielle Abgeltung der Leistungserbringung auf qualitativen Bemessungsergebnissen basieren soll, kommt es vielerorts zu offenem Widerstand, der an die frühesten Anfänge der Qualitätsbewegung erinnert.

P4P ist ein klarer Kontrapunkt zur Einzelleistungsvergütung.

Grundsätzlich verbindet der Begriff «Performance» (Leistung) Effizienz und Qualität der Leistungserbringung. Das Kürzel P4P verrät, woher dieses Prinzip stammt. Die ursprünglichste Form von P4P ist in industriellen Betrieben etabliert worden, um effizientes Arbeiten zu belohnen, etwa einen hohen Produktionsausstoss. Bei äusserst qualitätssensiblen Produkten – zum Beispiel in der Raumfahrt oder Elektronik – ist die Performance bald durch qualitative Dimensionen ergänzt oder gar abgelöst worden.

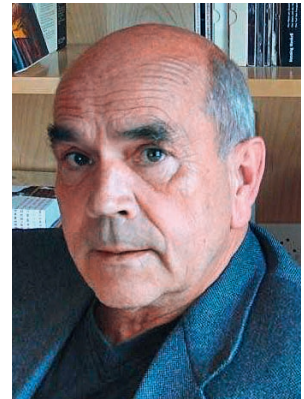
So auch in der medizinischen Leistungserbringung, wo eine ungenügende Leistungsmenge infolge der geltenden Anreize kaum ein Thema ist. Hier wird unter P4P in erster Linie die Honorierung für das Erreichen von Qualitätszielen verstanden. Dies in der Annahme, damit gute Qualität zu fördern. Diese Anreizsetzung ist ein klarer Kontrapunkt zur Einzelleistungsvergütung («Fee for Service»).

Die Ausgestaltung solcher Anreize kann – einzeln oder in Kombination – variable Formen beinhalten.

- Meist handelt es sich um die Belohnung von einzelnen Leistungserbringenden oder Ärztgruppen nach dem Grad der Zielerreichung vorgegebener Qualitätsstandards.
- Einige Modelle beschränken sich auf eine Bonusrichtung beim Erreichen globaler Qualitätsziele, oft im Zusammenhang mit der Patientenzufriedenheit.
- Zunehmend an Bedeutung gewinnen finanzielle Sanktionen beim Verfehlen von Mindestanforderungen. Bekannt geworden ist insbesondere die Ankündigung des US-Staates Massachusetts, durch seine Sozialversicherung «Medicaid» («HealthyMass») bei einstweilen 28 ernsthaften vermeidbaren Fehlern keine Honorare mehr zu bezahlen [1].
- Schliesslich wird als Anreiz die Veröffentlichung der Qualitätsbewertungen zunehmend forciert.

P4P ist in den vergangenen Jahren zu einer der prominentesten gesundheitspolitischen Initiativen im Qualitätsmanagement geworden und wurde von Beginn an – je nach Provenienz – optimistisch, skeptisch bis pessimistisch beurteilt. Umfangreiche Listen von Argumenten für und wider das Konzept wurden publiziert und diskutiert, noch bevor überhaupt Studien verfügbar waren [2].

Die Befürworter sehen in diesem Prinzip die einzige echte Alternative zum Fee for Service, wo die Anreize auf die Mengenausweitung ausgerichtet sind. Ausserdem werden eine deutliche Qualitätsorientierung, verbesserte Dokumentation in der Praxis sowie längerfristig auch tiefere Kosten erwartet. Die Gegner führen in erster Linie die



Kurt Hess

Eine nüchterne Betrachtung von P4P tut Not.

bislang unbefriedigende wissenschaftliche Evidenz ins Feld, ferner das Fehlen ausreichend realistischer Qualitätsindikatoren und Standards, wobei sich die vorhandenen zudem auf einige ausgewählte Diagnosen oder Indikationen beschränken. Sie befürchten, dass sich die

Ärztenschaft dann vorwiegend auf jene Fälle konzentriert, die «Qualitätspunkte» generieren, dass sie unter P4P schwerkranke und schwierige Patienten meiden und dass der einzelne Arzt nur begrenzten Einfluss auf die Outcomequalität hat, da die Behandlungskette – namentlich in komplexeren Fällen – diverse andere Leistungserbringende einschliesst. Eine nüchterne Betrachtung des fast schon modischen Begriffs tut also not – insbesondere was Ausgestaltung, Wirksamkeit und Praktikabilität betrifft.

P4P in den USA und in Grossbritannien

In den USA arbeiteten im Jahr 2006 gemäss einer nationalen Untersuchung bereits 52% der kommerziellen HMO nach P4P [3]. Die gleiche Studie beziffert den qualitätsabhängigen Anteil des ärztlichen Einkommens auf lediglich 5%, basierend auf nur fünf bis zehn Leitindikatoren. Ein typisches Indikatorenset, nämlich dasjenige der kalifornischen «Integrated Healthcare Association» (IHA), zeigt Tabelle 1. Mittlerweile bildet P4P in den USA einen üblichen Vertragsbestandteil zwischen privaten Versicherern und der Ärzteschaft. Auch die sozialen Programme der Regierung und der Bundesstaaten («Medicare and Medicaid») ziehen auf entsprechende Aufforderung durch den Kongress mit.¹

In Europa ist insbesondere das ambitionierte, umfassende und komplexe Qualitätsförderungsprogramm des briti-

In Europa ist insbesondere das komplexe Qualitätsförderprogramm des britischen NHS bekannt.

sehen «National Health Service» (NHS) bekannt, das seit 1997 mit einer Reihe von Reformen gestartet und 2004 mit dem sogenannten «New Contract» etabliert wurde [4]. Im Gegensatz zur eher bescheidenen Basis in den USA basiert die Honorierung der Grundversorgenden im Jahr 2005 auf 146 Indikatoren aus zehn chronischen Diagnosen bzw. Indikationen, die laufend ergänzt werden, zudem auf Indikatoren der Praxisorganisation und der subjektiven Bewertung durch die Patientinnen und Patienten. 2008 werden mit dem System bereits 19 Diagnosen bzw. klinische Indikationen erfasst, die

Tabelle 1.

Indikatorenset der Integrated Healthcare Association (IHA) 2004 (Klammer: Gewichtung in der Gesamtbewertung).

Klinische Indikatoren (40%):

Screening für Mammographie, Cervix-Karzinom, HbA1c bei Diabetikern, LDL-Cholesterol, Durchimpftrate bei Kindern.

Patientenbewertung (40%):

Mittelwert diverser Resultate aus der Patientenbefragung.

IT-Nutzung (20%):

Nutzung von elektronischen Daten zur Dokumentation, Point-of-Care-Testing (apparative und Laboruntersuchungen unmittelbar in der Praxis).

über 70% der Konsultationen abdecken. Von der Maximalpunktzahl, die das Honorar bestimmt, entfallen zwei Drittel auf klinische Indikatoren, der Rest auf solche der Organisationsqualität, auf die Patientenerfahrung und auf Zusatzleistungen (zervikale Abstriche, Kontrazeption usw.). Die Gewichtung dieser Domänen und die Punktwerte werden periodisch neu evaluiert, letztmals im Juni 2008.²

Die Datenerhebung der klinischen Indikatoren erfolgt durch direkten Onlinezugang der Gesundheitsbehörde in die elektronische Patientenkartei (QMAS: «Quality Management Analysis System»). Die Qualität der zu dokumentierenden Daten werden in einem jährlichen QOF-Besuch («Quality and Outcomes Framework») überprüft, bei dem ein Allgemeinarzt, eine nicht medizinische Person (Patientenvertretung) und ein Datenspezialist anwesend sind. Erreichte Qualitätsstandards generieren Punkte, die mit einem gewichteten Faktor und der Patientenanzahl multipliziert werden, um das Honorar zu berechnen.

Der qualitätsabhängige Anteil des ärztlichen Verdienstes im NHS ist ebenfalls massiv grösser als in den USA: Verdiente ein Familienarzt vor Einführung des New Contract noch jährlich rund 70 000 bis 750 00 £, so erhöhte sich das Mittel nach Einführung der Qualitätskomponente um 23 000 £. Die Tendenz ist durch Verfeinerung des Systems und durch Aufwertung der Qualitätspunkte steigend. Andererseits generierte dieses Programm massive Kosten, die nur bedingt auf die Qualitätskomponente, sondern stark auf die Investitionen in die Logistik (allem voran die unverzichtbare elektronische Dokumentation) und auf die für Abwicklung, Monitoring und Weiterentwicklung zuständigen Institutionen zurückzuführen sind.

¹ Centers of Medicare & Medicaid Services. www.cms.hhs.gov.

² Quality and Outcomes Framework Guidance for General medical Services, Contract 2008/2009. British Medical Association. 2008.

Die Qualitätseffektivität

Die Kernfrage in diesem Zusammenhang ist selbstverständlich jene nach der Qualitätseffektivität eines solchen Paradigmawechsels. Elliot S. Fisher sagte 2006 dazu: «The evidence based linking such programs to a better quality of care is thin» [5]. Dieses Fazit bedarf jedoch einer etwas genaueren Betrachtung.

Es gibt nur wenige relevante Studien, die sich mit gemessenen Wirkungen von P4P auf die Versorgungsqualität auseinandersetzen. 2006 hat Petersen in einer amerikanischen Review 17 dieser Arbeiten ausgewertet, wobei sich die meisten auf Indikatoren der Präventionsqualität beschränken [6]:

- Bei sechs Studien zu direkten finanziellen Anreizen für Einzelärzte wurde für fünf ein partieller bis positiver Effekt auf die Qualität gefunden.
- Von den neun Studien mit Anreizen auf HMO-Ebene zeigten sieben einen partiellen bis positiven Einfluss auf die Qualität.
- Es fanden sich zwei Studien mit Anreizen auf Versichererebene, wovon die erste einen besseren Zugang zur Versorgung nachwies.
- Vier Studien wiesen auf unerwartete Effekte des P4P hin, wovon eine Studie gewisse Hinweise für einen schlechteren Versorgungszugang für Schwerkranke zeigte.

Weitere häufig zitierte US-Studien auf Ebene von Health Plans weisen ähnliche Resultate auf. So hat die «Integrated Healthcare Association» (IHA) im Oktober 2006

Tendenziell gibt es Hinweise für eine Qualitätsverbesserung unter P4P.

eine erste Untersuchung auf Basis der Indikatoren gemäss der gezeigten Tabelle publiziert, die sich auf seit 2003 erhobene Daten stützt. Erste Resultate weisen darauf hin, dass alle Zielgrössen erreicht werden konnten, wobei eine durchschnittliche Verbesserung um lediglich 3% erzielt wurde (bei einer tieferen Quote in der Beurteilung durch die Patienten) [7]. Vor rund einem Jahr kündigte die IHA die wohl bislang grösste Untersuchung zu P4P an, die sich auf ganz Kalifornien erstrecken und mit einem verbesserten Indikatorenset arbeiten soll. Eine weitere Referenzstudie ist in Massachusetts mit Daten von 2001 bis 2003 bei 5350 Leistungserbringenden der grössten fünf Health Plans durchgeführt und im Juli 2008 veröffentlicht worden [8]. Auch hier kommen die Autoren zu einem ernüchternden Schluss: «Overall P4P contracts were not associated with greater improvement in quality compared to a rising secular trend.»

Praktisch alle zitierten Autoren weisen darauf hin, dass die bisherigen Untersuchungen zwar tendenziell Hinweise für verbesserte Qualität unter P4P ergeben, dass jedoch Verbesserungen nicht signifikant nachgewiesen werden konnten. In den Studiendiskussionen werden dazu folgende Gründe genannt:

- Der Beobachtungszeitraum ist noch viel zu kurz für signifikante Aussagen.
- Die Indikatorensets, die den US-Studien zu P4P zugrunde gelegt wurden, waren qualitativ wie quantitativ dürftig.
- Die finanziellen Anreize sind in den USA sehr tief und übertreffen nur in wenigen Fällen 5% des ärztlichen Einkommens.
- Es gibt in den USA keine «before-after-studies» mit den gleichen Ärztekollektiven, da vor der Einführung von P4P die entsprechenden Qualitätsdaten nicht erfasst wurden.
- Die medizinische Qualität ist im Beobachtungszeitraum der Studien gesamthaft gestiegen; es lässt sich anhand der Studien aber kaum eruieren, welcher Teil dem P4P-Ansatz und welcher der allgemeinen Entwicklung in der Medizin zuzuschreiben ist.

So weit die USA. Im britischen NHS mit dem weit höheren qualitätsabhängigen Lohnanteil und einem gegenüber den USA wesentlich komplexeren Bemessungssystem müssten somit schlüssigere Resultate erwartet werden. Doch ist hier der Zeithorizont noch kürzer als bei den amerikanischen Health Plans. Die wohl meist-zitierte Longitudinalstudie ist Mitte 2007 im NEJM publiziert worden und untersuchte die Qualität in 42 Praxen zu den Zeitpunkten 1998, 2003 und 2005 (d.h. nach Einführung des New Contract im Jahr 2004) [9]. Auch hier sind die Ergebnisse (noch) wenig aussagefähig, doch zeigt sich tendenziell, dass für die Diagnosen Koronare Herzkrankheit, Diabetes mellitus und Asthma der Qualitätszuwachs zwischen 2003 und 2005 grösser war als im Zeitraum zwischen 1998 und 2003. Dies könnte ein Hinweis für die Wirksamkeit des New Contract sein. Nur zeigte sich in der gleichen Studie kein signifikanter Unterschied bei den Verbesserungsraten von zwei Indikatorensets, wovon eines dem P4P-Konzept unterstellt war, das andere nicht.

In seiner Analyse des NHS-Programms von 1998 bis 2008 zieht der «Nuffield Trust», eine der führenden unabhängigen Institutionen in der Gesundheitspolitik, den Schluss, dass seit der Reform die Qualität offensichtlich gesamthaft substantiell verbessert wurde, dass es aber nicht klar ersichtlich ist, welcher Teil davon auf die medizinischen Fortschritte, gesellschaftliche Verhaltensänderungen und auf die systemische Entwicklung der Versorgung zurückzuführen ist [10]. Dies würde bedeu-

ten, dass die qualitätsabhängigen Vergütungen teilweise für Qualitätsinitiativen entrichtet worden sind, die schon zuvor – durch die säkularen Entwicklungen in der medizinischen Versorgung – wirksam waren.

Fazit aus heutiger Sicht

Fassen wir zusammen:

- Studien gibt es kaum,
- die Beobachtungszeit ist noch viel zu kurz, um schlüssige Aussagen machen zu können,
- die verglichenen P4P-Programme sind in fast allen Parametern höchst verschieden,
- deren Resultate zeigen meist eine zumindest tendenziell positive Wirkung auf die Qualität durch P4P,
- die Anreize im Umfang von 5% sind viel zu klein,
- die meisten Indikatorensets bedürfen noch wesentlicher Verfeinerungen.
- Und: Qualität kostet – zumindest initial.

Dies ist das kurze Fazit, das sich aus den bisherigen Untersuchungen der qualitätsabhängigen Honorierung in der Medizin ziehen lässt. Von den erwähnten Argumenten für und wider das P4P-Prinzip kann somit kaum eines schlüssig entkräftet oder erhärtet werden. Erst mit wachsendem Zeithorizont werden aussagefähigere und in der Tendenz klarere Resultate verfügbar sein. Es ist jedoch zu erwarten, dass sich in den meisten Gesundheitssystemen – so auch im schweizerischen – eine Form von qualitätsabhängigen Leistungskomponenten durchsetzen wird. Dass für solche Programme schon vor der Einführung ein sorgfältiges Monitoring etabliert wird, muss vorausgesetzt werden; kostspielige Grosseperimente sind nicht zu empfehlen.

Literatur

- 1 Patrick DL (Governor). Patrick Administration Announces Non-Payment Policy for 28 Serious Reportable Events. Press Release Health and Human Services. www.mass.gov. 2008.
- 2 Epstein AM. Paying for Performance in the United States and Abroad. *NEJM*. 2006;355 (4).
- 3 Rosenthal MB et al. Pay for Performance in Commercial HMOs. *NEJM*. 2006;355 (8).
- 4 Doran T et al. Pay-for-Performance Programs in Family Practices in the United Kingdom. *NEJM* 2006;355 (4).
- 5 Fisher ES. Paying for Performance – Risks and Recommendations. *N Engl J Med*. 2006; 355 (2): 1845 ff.
- 6 Petersen LA et al. Does Pay for Performance improve the Quality of Health-Care? *Annals of Internal Medicine*. 2006; 145 (4).
- 7 Lempert L, Wilkins D. Integrated Healthcare Association. Press Release: www.iha.org/pressrel. 2006.
- 8 Pearson SD et al. The Impact of Pay-for-Performance on Health Care Quality in Massachusetts 2001–2003. *Health Affairs*. 2008;27 (4).
- 9 Campell et al. Quality of Care in England with the Introduction of Pay for Performance. *NEJM*. 2007;357 (2).
- 10 Leatherman S, Sutherland K. The Nuffield Trust. *The Quest for Quality: Refining the NHS Reforms*. 2008.

Korrespondenz:

Dr. med. et lic. oec. Kurt Hess
 Projektleiter im Gesundheitswesen
 Vizepräsident Stiftung EQUAM
 Grossmünsterplatz 1
 8001 Zürich
kurt.hess@hin.ch