

1 Pay for Performance – Aktueller Stand und Perspektiven

Matthias Schrappe

1.1 Einleitung

Pay for Performance (P4P, qualitätsorientierte Vergütung) ist eine meist sektoral eingesetzte Vergütungsform, die in Gesundheitssystemen mit fallpauschalierter Finanzierung zum Einsatz kommt. Das primäre Ziel von P4P-Programmen besteht darin, die durch DRG's oder periodenbezogene Fallpauschalen bedingten Fehlanreize (insbesondere Mengenausweitung) durch finanzielle Anreize zur Qualitätsverbesserung zumindest teilweise zu neutralisieren. P4P steht in der Regel nur für einen kleineren Teil der Vergütung, das dominierende Vergütungssystem wird nicht außer Kraft gesetzt.

In Deutschland ist P4P durch das Krankenhausstrukturgesetz (KHSG) vom 22.12.2015 für den stationären Sektor in die Sozialgesetzgebung eingeführt worden (vorher bestand lediglich eine Regelung für den ambulanten Bereich nach § 136a Abs. 4 a.F.). Entsprechend internationaler Vorbilder (z.B. USA: Schrappe 2014, S. 175ff.) ist diese Entwicklung in einen umfangreichen, auf Qualitätsverbesserung ausgerichteten Kontext der aktuellen Gesetzgebung eingebettet. In erster Linie handelt es sich um weitere wettbewerbliche Instrumente (Public Reporting mit Qualitätsbericht nach § 136b Abs. 6 SGB V), die Qualitätsorientierung der Versorgungsplanung (z.B. qualitätsorientierte Krankenhausplanung nach § 6 Abs. 1a KHG und § 136c SGB V), Zugangsindikator i.R. der Sicherstellungszuschläge n. § 17b KHG, selektives Kontrahieren

(Qualitätsverträge nach § 110a SGB V und Neufassung des § 140a zur Besonderen Versorgung), Stärkung der Patientenorientierung durch Neufassung des § 1 Abs. 1 KHG und § 136b Abs. 6 SGB V, Schwerpunktsetzung Hygiene und Patientensicherheit in den Detailregelungen des § 136a und § 136b SGB V, Relativierung der Überversorgung durch Regelungen zur Mengenproblematik (§ 10 und § 17b KHG, § 10 Abs. 13 KHEntgG, Neufassung der Regelungen zu den Chefarztverträgen n. § 135c SGB V und Zweitmeinungsverfahren nach § 27b SGB V), die Beauftragung zur Indikatorenentwicklung an das IQTiG nach § 137a und die Möglichkeit zur Evaluation entsprechender Strukturmaßnahmen durch den Innovationsfonds nach § 92a SGB V. Vor diesem Hintergrund soll kurz auf die Einordnung, die Voraussetzungen einer erfolgreichen Umsetzung und auf die eventuellen unerwünschten Effekte von P4P eingegangen werden.

1.2 Einordnung

Das geschilderte umfassende Vorgehen deutet auf eine langfristig angelegte Entwicklung. Wenngleich in § 1 Abs. 1 des Krankenhausgesetzes (KHG) die Erweiterung des Begriffes der „bedarfsgerechten Versorgung“ durch die Formulierung „qualitativ hochwertige, patienten- und bedarfsgerechte Versorgung“ zunächst auf die Krankenhausplanung zu verweisen scheint, macht die Gesetzesbegründung sehr deutlich, dass es um eine umfassende Orientierung in Richtung Qualitätsverbesserung geht, denn: „eine wirtschaftliche Sicherung der Krankenhäuser ist ohne eine Zielorientierung und Sicherstellung des allgemein Standards der Krankenhausversorgung mit Qualitätssicherungsgesichtspunkten nicht durchgreifend zu festigen und zu bewirken.“ Das Instrument P4P soll offensichtlich nicht als Einzelmaßnahme, sondern integriert in eine umfassende Stärkung des Qualitätsgedankens implementiert werden (z.B. Crossing the Quality Chasm, IOM 2001, s.a. Schrappe 2014, S. 331ff.).

Das deutsche Gesundheitssystem schließt sich damit der internationalen Entwicklung an. In den USA werden ab dem 1.10.2016 6% des Budgets der an der Medicare-Versorgung beteiligten Krankenhäuser über Qualitätsindikatoren verteilt. Es handelt sich hier um das sog. Value-Based Purchasing Program (VBP), eine Sonderform von P4P, das nicht nur die Qualität der Leistung, sondern die Effizienz zum Maßstab nimmt, weiterhin um das Hospital Readmission Reduction Program und das Hospital Acquired Condition Reduction Program (Kahn et al. 2015). In der ambulanten Versorgung wurde vor Kurzem der Physician Value Based Payment Modifier durch das MIPS (Merit-based Incentive Payment System) abgelöst (Rosenthal 2015). Ebenfalls im ambulanten Bereich werden in Großbritannien das Quality and Outcome Framework mit sehr guten, die chronischen Erkrankungen betreffenden

Indikatoren, und in Frankreich der Contrat d'amélioration des pratiques individuelles (CAPI) (Brunn u. Hassenteufel 2015) eingesetzt.

Die in Deutschland zu verspürende Ablehnung in der Selbstverwaltung, die in deutlichem Gegensatz zum erklärten Willen des Gesetzgebers und zu den internationalen Erfahrungen steht, ist bemerkenswert (z.B. Veit et al. 2012), vor allem wenn inadäquate Operationalisierungen von Begrifflichkeit und Evidenz Verwendung finden. Natürlich sind zahlreiche praktische Umsetzungsfragen zu lösen (s.u.), weiterhin ist die isolierte (!) Wirksamkeit von P4P erwartungsgemäß begrenzt (Ryan et al. 2015), trotzdem rechnet z.B. in den USA niemand damit, dass man die P4P-Programme zurückziehen und wieder zur reinen Mengensteuerung zurückkehren wird („Like it or not, hospital P4P is here to stay“, Ryan u. Blustein 2012). Die Erfolge dieser umfassenden Anstrengung sind nämlich nicht zu übersehen, so berichten die Centers of Disease Control (CDC) von einem deutlichen Rückgang der nosokomialen Infektionen in allen Gruppen (außer Katheter-bedingten Harnwegsinfektionen), teilweise um knapp die Hälfte (zentrale Venenkatheter-bezogene Infektion um 46%) (CDC 2015).

P4P geht theoretisch auf die Principal Agent-Theorie zurück, die auf das Gesundheitswesen angepasst wurde (z.B. Allen et al. 2014). Dieser Theorie liegt die Annahme einer Informationsasymmetrie zwischen der finanzierenden (Principal: Kostenträger) und der ausführenden Seite (Agent: Leistungserbringer) zugrunde. Durch finanzielle Anreize wird die Aufmerksamkeit des Agent auf Ziele gerichtet, die den Präferenzen des Principal entsprechen. Unter den Bedingungen des deutschen DRG-Systems, das auf die medizinische Homogenität der Fallgruppen verzichtet, kann man den P4P-Ansatz so interpretieren, dass sorgfältig ausgewählte kostenhomogene Fallgruppen durch Qualitätsparameter zusätzlich in qualitätshomogene Gruppen aufgespalten werden, die dann einer gesonderten Steuerung unterliegen. Hierdurch ergäbe sich in Deutschland z.B. die Möglichkeit, statt der sektoralen Optimierung der operativen prozeduralen Akutmedizin vermehrt das dominante Problem der chronischen Mehrfacherkrankungen in den Fokus zu rücken. Alternativen zur Weiterentwicklung des DRG-Systems bestehen nur in zweierlei Hinsicht:

- Behebung der Informationsasymmetrie durch Zusammenführung von finanzieller und medizinischer Verantwortung in populationsbezogenen Versorgungssystemen mit Managed Care-Charakter (s. Empfehlung des SVR 2009, Nr. 1179), und
- Übertragung der finanziellen Verantwortung für Qualität und Kosten auf die Wahlentscheidung des Patienten (diskutiert von David Cutler [2015]).

1.3 Voraussetzungen

Adäquate Definition

In Erweiterung der Definition des SVR (2008, Nr. 732) muss eine exakte Definition von P4P auf die Dualität von Qualitätsmonitoring durch definierte Indikatoren auf der einen Seite und die Kopplung an Vergütungsbestandteile auf der anderen Seite Bezug nehmen. Es besteht eine Analogie mit dem DRG-System, bei dem eine epidemiologische Systematik mit einer Vergütungssystematik verknüpft wird.



Pay for Performance (P4P) im Gesundheitswesen basiert auf dem Monitoring von Qualität 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 (Schrappe 2014, S. 193).

Zielorientierung

Die wichtigste inhaltliche Voraussetzung ist eine adäquate Zielorientierung der P4P-Einführung. Statt der Qualitätsdarstellung („Qualitätssicherung“), historisch auf die Einführung von nicht-adjustierten Fallpauschalen im Gesundheitsstrukturgesetz von 1993 zurückgehend, steht nun auch in Deutschland der Übergang zu einem Konzept der Qualitätsverbesserung (Quality Improvement) an, das den auf Organisationen bezogenen Qualitätsmanagement-Gedanken auf die Systemebene transferiert (Qualitätsverbesserung auf Systemebene). Bei diesem Verbesserungs- und Steuerungsansatz ist eine konkrete Zielbestimmung von Bedeutung, insbesondere welche Aspekte bei der Qualitätsverbesserung im Vordergrund stehen und mit welchen Indikatoren die Verbesserung zu fassen ist. Die bisherige akutmedizinisch-operative Ausrichtung der „Qualitätssicherung“, die keinen einzigen Indikator aus dem chronisch-konservativen Bereich beinhaltet (s. Schrappe 2014, S. 76), ist angesichts der Morbiditätsentwicklung (Dominanz chronischer Mehrfacherkrankungen) nicht mehr aktuell. Es müssen außerdem andere Qualitätsperspektiven eingenommen werden (z.B. Patientenperspektive mit Patient Reported Outcome Measures, außerdem Nutzenaspekte). Weiterhin sind neben leistungsbezogenen Qualitätsaspekten auch Strukturfragen zu adressieren, die z.B. Koordinations- und Integrationsdefizite betreffen. Insgesamt ergibt sich ein dreidimensionaler Orientierungsrahmen für die Planung und Evaluation von Maßnahmen der Qualitätsverbesserungen (s. Abb. 1).

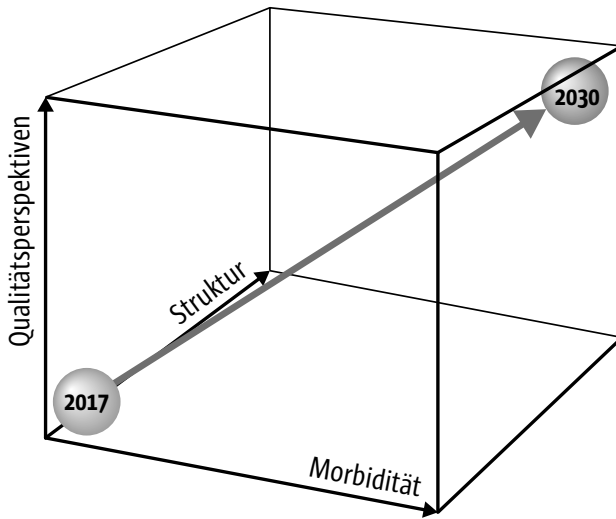


Abb. 1 Dreidimensionaler Orientierungsrahmen. Unter Qualitätsperspektiven versteht man u.a. die Anbieter-, Patienten- oder gesellschaftliche Perspektive, zur Morbiditätsdimension gehört auch Prävention. Die Strukturdimension umfasst Koordination und Integration (mod. n. Schrappe 2014, S. 70).

Ökonomische und verhaltenspsychologische Grundannahmen

Gerade wenn man Informationsasymmetrie und unvollständigen Marktzugang etc. als Kennzeichen des Gesundheitssystems diskutiert, dürfen bei der Einführung von P4P implizite Annahmen nicht außer Acht gelassen werden. Im Vordergrund steht die Annahme, dass Patienten und andere Nutzer des Gesundheitswesens sich rational verhalten, dass die Motivation der Leistungserbringer durch finanzielle Anreize verstärkt und nicht konterkariert wird, dass sich die verwendeten Qualitätsindikatoren umsetzbar und widerspruchsfrei darstellen und dass die Anreize stärker wirksam sind als alternative Verhaltensweisen (z.B. Leistungserbringer: Mengensteigerung; Kostenträger: Risikostrukturausgleich) (aus Schrappe 2017). Ein P4P-Programm muss also nicht nur von einer adäquaten Zielformulierung ausgehen, sondern auch so konfiguriert werden, dass angesichts der unterschiedlichen Interessen eine Chance zur Umsetzung besteht (s.u.).

1.4 Umsetzung: Qualitätsindikatoren

Entsprechend der Definition von P4P (s.o.) muss das Monitoring durch Qualitätsindikatoren getrennt von ökonomischen Umsetzungsproblemen betrachtet werden. Vier Problemkreise haben sich in der Diskussion in Deutschland herauskristallisiert, die sich gegenseitig negativ beeinflussen:

Lineares Konzept der Qualitätsmessung

Die Quantifizierung von Qualität ist nicht voraussetzungslos möglich (Schrappe 2014, S. 94). Es können vier Messverfahren unterschieden werden:

- generierende Verfahren wie CIRS zur Behebung „blinder Flecken“ (geringe Sensitivität, geringe Spezifität),
- quantitative, lineare Messungen (z.B. der Mortalität) mit Kompromiss zwischen Sensitivität und Spezifität,
- Monitoring durch Indikatoren (hohe Sensitivität, geringere Spezifität) und
- komplexe Messvorgänge mit höchstmöglicher Sensitivität und Spezifität zur wissenschaftlichen Evaluation.

Für den langfristigen Einsatz scheiden generierende Verfahren wegen ihrer geringen Sensitivität (ca. 1%) und komplexe Messvorgänge wegen des Messaufwandes aus. In Deutschland wird fast ausschließlich die lineare quantitative Messung eingesetzt, wobei der Zusammenhang zwischen Messwert (z.B. Mortalität) und Qualität der Leistungserbringung oft nicht hergestellt werden kann. Indikatoren sind dagegen speziell für das Monitoring entwickelt, sie sind hoch-sensitiv eingestellt. Indikatoren werden bezüglich ihres Vorhersageverhaltens von Qualitätsdefiziten validiert – alle negativen Ereignisse sollen detektiert werden (hohe Sensitivität), man lässt sich jedoch auf Kompromisse bei der Spezifität ein (JCAHO 1991). Ähnlich wie im Straßenverkehr (z.B. Geschwindigkeitsbegrenzung) wird nicht angenommen, dass das Ansprechen eines Indikators stets identisch mit schlechter Qualität sei, es erhält also auch ein Formel 1-Rennfahrer eine Verwarnung, obwohl er tatsächlich durch die besonders gute Beherrschung seines Fahrzeuges keine Gefährdung darstellt. Stattdessen nimmt man an, dass die positive Beeinflussung von Indikatoren im Zusammenwirken mit anderen Interventionen eine Verbesserung der Qualität zur Folge hat.

Die hohe Sensitivität, die von Indikatoren gefordert wird, zeigt übrigens die Problematik sogenannter „Exzellenz-Indikatoren“, die gesetzlich gefordert sind und mit denen Anreize für besonders gute Qualität verbunden werden. Sie unterscheiden sich vom gängigen Indikatorenkonzept grundlegend, denn bei ihnen ist eine hohe Spezifität gefordert (die Belohnung darf nicht fälschlicherweise zugesprochen werden). Die Forderung nach „Exzellenz-Indikatoren“ ist daher aus wissenschaftlicher Sicht sehr kritisch zu sehen.

Einseitige Fixierung auf Ergebnis-Indikatoren

Die international führenden, umfassenden Projekte zur Qualitätsverbesserung setzen fast ausnahmslos Prozessindikatoren ein (Ryan u. Blustein 2012).

Obwohl Ergebnisindikatoren durch den Anschein gestützt und im SGB V genannt werden, sind sie von einer Reihe maßgeblicher Nachteile begleitet:

- sie betreffen bereits eingetretene Ereignisse, während Prozessindikatoren Ereignisse vorhersagen (bad apple-Problematik),
- sie müssen im Ggs. zu Prozessindikatoren risikoadjustiert werden,
- im Ggs. zu Prozessindikatoren ist die Verantwortlichkeit oft nicht klar (z.B. nach Entlassung aus der stationären Behandlung),
- kleine Einrichtungen werden aus statistischen Gründen benachteiligt (höhere Häufigkeit von „Ausreißern“, Einzelereignisse sind nicht zu neutralisieren),
- sie sind wenig motivierend wegen der bad apple-Problematik, während Prozessindikatoren ein präventives Eingreifen möglich machen, und vor allem
- ist ihnen ein Mengenanreiz immanent, was besonders bei P4P-Programmen eine wichtige Rolle spielt (Attraktion leichter Fälle bei mengenmäßig „ausbaubaren“ Leistungen, dadurch Anstieg der die Ergebnisqualität),
- außerdem sind Risikoadjustierungsmodelle bei großen Anbietern bzw. Datenmengen leicht zu beeinflussen (Gaming).

Bei Prozessindikatoren sind jedoch ebenfalls zwei Aspekte zu berücksichtigen:

- sie sind nicht so stark mit den Ergebnissen korreliert, wie man es gerade bei stark EBM-abgesicherten Parametern meinen sollte. Es ist zwar durchaus ein (hoch-signifikanter) Effekt von Prozessindikatoren auf *Outcomes* nachweisbar, wie Werner und Bradlow (2006) bei immerhin 3.657 Krankenhäusern in den USA anhand der Daten aus dem *Hospital Compare* Programm nachweisen konnten, aber quantitativ enttäuschen die Ergebnisse dennoch. Die zu ihrer Evaluation verwendeten Studiendesigns, die den komplexen Bedingungen, in denen solche Prozessparameter wirken, nicht gerecht werden, stehen als Ursache in der Diskussion.
- ganz entgegen dem Augenschein sind insbesondere solche Prozessindikatoren, die nicht durch EBM bzw. Leitlinien abgesichert sind, bei denen also der Informationsvorsprung der „Experten vor Ort“ noch besteht, besonders wirksam, während bekannte EBM-abgesicherte Prozessindikatoren in P4P-Programmen keinen Effekt zeigen. Dieser irritierende Befund lässt sich aus der Principal Agent Theorie erklären. Bei Indikatoren, bei denen die Informationsasymmetrie aufgehoben ist, ist eine Einzelfallvergütung sehr viel effektiver.

Einseitige Fixierung auf Routine-Daten

Routinedaten sind insbesondere wegen ihres Sensitivitätsproblems schlecht geeignet, da sie in erster Linie die Vergütungs-bezogene Dokumentation widerspiegeln (z.B. Maass et al. 2015). Die Sensitivitäts-Problematik wird durch die kürzlich auf Routinedaten umgestellte Erhebung des Dekubitus unterstrichen (ältere Patienten: Underreporting – keine Vergütungsrelevanz; jüngere Patienten: sehr hohe Raten – hohe Vergütungsrelevanz). Weitergehend sind jedoch folgende Aspekte zu beachten:

- Routinedaten sind (ebenso wie das DRG-System sui generis) Prozeduren-lastig, sie fördern also die operativ-akutmedizinische Ausrichtung des Systems,
- sie bilden trotz aller „transssektoralen“ Versuche in erster Linie die sektorale Logik der Vergütungssysteme ab und fördern nicht die Integration,
- sie verstärken den Mengenanreiz des Systems (statt der Prävention),
- sie präjudizieren den Gebrauch von Ergebnis-Indikatoren (mit wiederum Routinedaten-gestützter Risikoadjustierung), und
- sie stärken den Anbieterbezug (da diese die Vergütung auslösen) und nicht den Patientenbezug.

Routinedaten behindern also die notwendige Neuausrichtung des Systems und stabilisieren die derzeitige akutmedizinisch-prozedurale Ausrichtung. Es gibt jedoch auch Indikatoren, die sehr gut mit Routinedaten zu erheben sind (z.B. Mindestmengen – aber auch hier: Gefahr der Mengenausweitung).

Mit Nachdruck ist darauf hinzuweisen, dass es seit über vierzig Jahren ein prominentes Beispiel gibt, bei dem man mit klinisch-epidemiologischen Falldefinitionen (nicht zu verwechseln mit klinischen Falldefinitionen) zu stabilen, international vergleichbaren Zahlen kommt: die Infektionsepidemiologie mit ihren Falldefinitionen der *Centers of Disease Control* (CDC). Die Erarbeitung von klinisch-epidemiologischen Falldefinitionen für die allgemeine Qualitätsdarlegung sollte mit hoher Priorität auch außerhalb der Infektionsepidemiologie vorangetrieben werden.

Patienten- und regionale Perspektive statt Anbieter-Perspektive einnehmen

Die gegenwärtig in Deutschland verwendeten Qualitätsparameter weisen (gerade wegen ihres Vergütungsbezugs) in erster Linie eine Anbieterperspektive auf. Stattdessen müssen Patient Reported Outcome Measures (PROMs) in den Vordergrund gerückt werden, dazu gehören alle Dinge, die der Patient selbst und nur selbst berichten kann (z.B. postoperative Schmerztherapie). Sie betreffen vornehmlich den Bereich der Prozessindikatoren (Koordination, Information, Kommunikation). Diese „PROMs“ stellen einen wichtigen Einstieg in das Thema unterschiedlicher Qualitätsperspektiven dar und sind

international bereits im Zusammenhang mit P4P-Programmen in Gebrauch (Ryan u. Blustein 2012). Entsprechende Entwicklungen durch das IQTIG nach den Bestimmungen des § 137 sind daher von großer Wichtigkeit. Gleiches gilt für die Area-Indikatoren, bei denen nicht einzelne Anbieter, sondern das Zusammenspiel der regionalen Versorgungseinrichtungen zum Gegenstand wird.

1.5 Umsetzung: ökonomische und verhaltenspsychologische Faktoren

Kongruenz von Aufgabe und Verantwortung

Analog zum allgemeinen Organisationsprinzip und entsprechend der sog. RUMBA-Regel („B“ für behaviourable) ist darauf zu achten, dass die P4P-Anreize für die Behandlung eingesetzt werden, für die die entsprechende Institution (oder Person) auch die Regelungsverantwortung hat. Ansätze der sog. transsektoralen Qualitätssicherung müssen auf die Einhaltung dieser Regel überprüft werden.

Kein Dual Use von Public Reporting (PR, Qualitätsbericht) und P4P

Zwar befinden sich die Hypothesen zur Wirkungsweise von PR und P4P stark in Bewegung (Übersicht Schrappe 2014, S. 181f. und 189ff.), aber vom gleichzeitigen Einsatz von Indikatoren für PR und P4P (dual use) ist dringend abzuraten. Dies gilt nicht nur, soweit Indikatoren z.B. in Public Reporting Programmen bereits im Einsatz waren und nicht mehr (auch nicht durch P4P) weiter verbesserbar sind (sog. ceiling-Effekt) – auch wenn hier die größte Gefahr in der derzeitigen Entwicklung in Deutschland besteht, dass nämlich die Indikatoren aus dem Qualitätsbericht mit finanziellen Mitteln aufgewertet werden (die absehbare Erfolglosigkeit würde den Skeptikern natürlich recht geben). Die eigentliche Begründung liegt jedoch in der Art der ökonomischen Wirkung: PR beeinflusst nicht das Verhalten der Patienten oder der Kostenträger, sondern wirkt über einen befürchteten Reputationsverlust direkt auf der Ebene der Krankenhäuser, und zwar vor allem in hoch-kompetitiven Regionen (SVR 2008, Nr. 685ff.). P4P zeigt in hoch-kompetitiven Bereichen bei Einsatz gleicher Indikatoren jedoch keine zusätzliche Wirkung mehr (stattdessen führt die Konkurrenz um Mengen), ist aber in anderen Regionen mit geringerer Konkurrenz seinerseits wirkungsvoller als PR. Es ist also dringend angeraten, für die PR- und P4P-Programme unterschiedliche Indikatoren-Sets einzusetzen, die für die spezifischen Einsatzbereiche validiert sind, und die zudem regelmäßig aktualisiert werden (um ein ceiling zu vermeiden).