Qualitätsorientierte Anreiz- und Vergütungssysteme in der stationären Versorgung

–

Entwicklung und Anwendung eines Kriterienkatalogs

vorgelegt von
Dipl.-Kffr.
Dörte Gesche Seeger
aus Berlin

von der Fakultät VII – Wirtschaft und Management
der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
– Dr.rer.oec. –

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Reinhard Busse
Erstgutachter: Prof. Dr. Thomas Mansky
Zweitgutachter: Prof. Dr. Klaus-Dirk Henke

Tag der wissenschaftlichen Aussprache: 4. November 2013

Berlin 2013

D 83

Danksagung

Zunächst möchte ich meinem Doktorvater für zahlreiche, konstruktive Diskussionen danken – diese Unterstützungen und Ratschläge haben maßgeblich zum Gelingen der Arbeit beigetragen. Ein weiterer Dank gilt dem Team des Fachgebiets “Strukturentwicklung und Qualitätsmanagement im Gesundheitswesen” für die außergewöhnliche nette Zusammenarbeit und Bereitschaft, jederzeit Themen (ob fachspezifisch oder nicht) zu diskutieren. Meinem Zweitgutachter gilt Dank für viele bereichernde Gespräche und Hilfestellungen. Auch den Interviewpartnern der im Projekt beteiligten Krankenkassen und Krankenhäusern danke ich herzlich für die spannenden Gespräche und vielen Anregungen.

Ein besonderer Dank gilt selbstverständlich den Personen, die mich nicht erst während der Dissertationszeit unterstützen – vor allem meine Familie und Freunde sorgen stets für Rat und Tat und Abwechslung.

Zusammenfassung

In dieser Arbeit wird ein in vielen Bereichen neuartiger, umfassender, praxisorientierter Kriterienkatalog zur Entwicklung und Bewertung von qualitätsorientierten Vergütungssystemen im Gesundheitssektor (oft als Pay-for-Performance, P4P, bezeichnet) entwickelt: Der Katalog kombiniert Anforderungen aus der Qualitätsmessung der Medizin mit solchen aus Anreizsystemen der Betriebswirtschaftslehre. Der Katalog bietet damit eine umfassende Sicht auf alle System- und Funktionselemente von qualitätsorientierten Vergütungssystemen und definiert gleichzeitig sehr detailliert und praktisch anwendbar die wesentlichen Beurteilungskriterien.

Anhand von sieben qualitativ-empirischen Fallstudien bei Leistungserbringern und Kostenträgern bezüglich der Umsetzung von P4P in IV-Verträgen mit Gewährleistungsklauseln wurde zudem einerseits iterativ der Kriterienkatalog angepasst, andererseits lassen sich folgende Aussagen zur praktischen Implementierung ableiten:

- In keinem der untersuchten IV-Verträge mit Gewährleistungsklauseln wurden die darin enthaltenen, qualitätsorientierten Vergütungselemente vollständig umgesetzt. Oft scheitert die Umsetzung der getroffenen Vereinbarungen bereits daran, dass Gewährleistungsfälle nicht sicher erkannt werden. Hier wären auch im bestehenden System Verbesserungen möglich bzw. sind teilweise geplant.
- Eine weitere Schwierigkeit der Umsetzung von P4P in IV-Verträgen ist die geringe Zahl von eingeschlossenen Fällen und somit auch daraus folgenden Gewährleistungsfällen. Dies führt zu geringer, strategischer Relevanz und behindert Investitionen in Prozessveränderungen und Software. Außerdem ist durch die unterschiedliche Gestaltung von Selektivverträgen sowohl eine standardisierte Umsetzung und Evaluierung der P4P-Elemente mit Skaleneffekten als auch ein Vergleich verschiedener Systeme kaum möglich.
- Gleichzeitig zeigen aber die meisten Befragten in den Fallstudien ein großes Interesse an P4P und gehen davon aus, dass diese Vergütungsform mittel- oder langfristig in Deutschland eingesetzt werden wird. Hierfür wird vor allem eine valide, aussagekräftige, manipulationssichere, akzeptierte Messung von Qualität sowie eine Vergütung, die sich überwiegend auf Bonus-Elemente konzentriert, erwartet.
- Außerhalb der untersuchten IV-Verträge haben einige der befragten Leistungserbringer und Kostenträger erste P4P-Elemete etabliert, bspw. in Qualitätsmanagementsystemen oder Ideen für Bonusverträge, die hier am Rande mit analysiert werden.

In der Arbeit werden außerdem erste Empfehlungen für eine mögliche Einführung eines bundesweiten P4P-Systems gegeben, nämlich die Umsetzung anhand eines kollektivvertraglichen Systems mit Fokus auf Ergebnisqualität, deren Kennzahlen mittels eines Gesamtscores vergütet werden. Es verbleibt die Erkenntnis, dass eine qualitätsorientierte Vergütung in Deutschland umsetzbar wäre und langfristig die Qualitätsorientierung im stationären Sektor verbessern könnte.

Abstract

In this thesis, a comprehensive, practice-oriented, in most parts innovative criteria catalogue for the design and evaluation of pay-for-performance-systems (P4P) in the inpatient sector has been developed: the catalogue combines requirements from quality measurement in medicine with those from incentive schemes in business economics. The catalogue thus offers a comprehensive view of all system and functional elements of quality-oriented compensation schemes and at the same time defines the essential evaluation criteria in a very detailed and practical way.

In seven qualitative-empirical case studies with providers and payors the implementation of P4P in integrated care contracts with warranty clauses was analyzed. This allowed on one hand the iterative adjustment of the criteria catalogue and on the other the following statements regarding their practical implementation:

- In none of the examined integrated care contracts with warranty clauses the quality-oriented compensation elements contained in these contracts were completely implemented. Often in the affected contracts the warranty claims are not recognized with enough certainty. In this matter, improvements would be possible in the existing system and in some cases are already planned.
- Another difficulty in implementing P4P in integrated care contracts is the low number of included cases and therefore the warranty claims resulting from these. This leads to less strategic relevance and makes investments in procedural changes and software more difficult. In addition, because each selective contract is designed differently, a standardized implementation and evaluation of P4P elements with scale effects as well as a comparison of different systems are almost impossible.
- But at the same time, most of the people interviewed in the case studies showed great interest in P4P and assume that this compensation form will be implemented in the medium or long term in Germany. For this, it is expected above all that there should be a valid, meaningful, accepted measurement of quality that is immune to manipulation as well as compensation that is focused primarily on bonus elements.
- Outside of the examined integrated care contracts, some of the providers and payors questioned have established first P4P elements, e.g., in quality management systems or ideas for bonus contracts.

In addition, the thesis gives first recommendations for a possible introduction of a nationwide P4P system in Germany, namely implementation based on a collective contract system focused on quality of results, whose key figures can be compensated based on a total score. It can be concluded that quality-oriented compensation in Germany would be viable and long-term quality orientation in the inpatient sector could improve.

Kurzverzeichnis

Danksagung I

Zusammenfassung II

Abstract III

Kurzverzeichnis IV

Inhaltsverzeichnis V

Abbildungsverzeichnis XI

Tabellenverzeichnis XIII

Abkürzungsverzeichnis XV

1. Einleitung 1

2. Qualität in stationärer Versorgung: Eingrenzung des Qualitätsbegriffs und Bedeutung von Qualität 11

3. Entwicklung des Qualitätsteils des Kriterienkatalogs für P4P-Systeme 43

4. Entwicklung des Anreizteils des Kriterienkatalogs für P4P-Systeme 89

5. Anwendung des entwickelten Katalogs bei P4P-Systemen 147

6. Diskussion der Erkenntnisse und weiterführende Anwendung 239

Resümee 257

Anhang: Interviewleitfäden i

Literaturverzeichnis xx

Inhaltsverzeichnis

Danksagung ..... **I**

Zusammenfassung ..... **II**

Abstract ..... **III**

Kurzverzeichnis ..... **IV**

Inhaltsverzeichnis ..... **V**

Abbildungsverzeichnis ..... **XI**

Tabellenverzeichnis ..... **XIII**

Abkürzungsverzeichnis ..... **XV**

1. Einleitung **1**

1.1. Diskussion der Problemstellung und Zielsetzung des Promotionsvorhabens 1

1.1.1. Problemstellung des Promotionsvorhabens 1

1.1.2. Zielsetzung: Entwicklung und Nutzung eines Kriterienkatalogs für P4P-Projekte 3

1.2. Zentrale Fragenstellungen und Gliederung 3

1.3. Beschreibung der methodischen Vorgehensweise 6

1.4. Literaturüberblick für theoretische Grundlagen 7

2. Qualität in stationärer Versorgung: Eingrenzung des Qualitätsbegriffs und Bedeutung von Qualität 11

2.1. Der Qualitätsbegriff – Definitionen, Dimensionen und Eingrenzungen 11

2.1.1. Diskussion verschiedener Qualitätsdefinitionen 11

2.1.2. Aufteilung des Qualitätsbegriffs nach Donabedian: Struktur-, Prozess- und Ergebnisqualität 14

2.1.3. Weitere Dimensionen und verschiedene Perspektiven der Qualität als Möglichkeit einer breiteren Definition 20

2.1.4. Eingrenzung der an stationärer Qualität beteiligten Akteure 22

2.1.5. Entwicklung einer „Working Definition“ 25

2.2. Bedeutung von Qualität im stationären Sektor 25

2.2.1. Qualitätsdiskussion im stationären Sektor: Entwicklung und derzeitiger Stand 25

2.2.2. Qualitätsmessung im stationären Sektor: historische Entwicklung und Darstellung verschiedener Indikatorensysteme 27

2.2.3. Qualitätsinitiativen im stationären Sektor: Gesetzliche Grundlagen und Überblick 34

2.2.4. Exkurs: Wahrnehmungsverzerrung in der Qualitätsdiskussion in Deutschland? 41

2.3. Zwischenfazit: Die Diskussion zeigt neue Sichtweisen auf Qualität und auf die verschiedenen Qualitätsinitiativen im stationären Sektor 42

3. Entwicklung des Qualitätsteils des Kriterienkatalogs für P4P-Systeme 43

3.1. Wichtigkeit der Operationalisierung stationärer Qualität durch qualitätsorientierte Vergütung 43

3.2. Beschreibung der Entwicklung des Qualitätsteils des Kriterienkatalogs 43

3.2.1. Quellen für die deduktive Entwicklung des Kriterienkatalogs 43

3.2.2. Aufbau des Kriterienkatalogs 44

3.2.3. Systemelemente des Qualitätsteils des Kriterienkatalogs 45

3.2.4. Kriterienkategorien des Qualitätsteils des Kriterienkatalogs 48

Kriterienkategorie 1: Aussagekraft 48

Kriterienkategorie 2: Aussagekraft /„Das Richtige messen“ 50

Kriterienkategorie 3: Akzeptanz 51

Kriterienkategorie 4: Wissenschaftlichkeit /„Richtig messen“ 51

Kriterienkategorie 5: Messaufwand und -sicherheit 52

3.2.5. Beschreibung der Skala für die Kriterienbewertung 52

3.3. Der 1. Teil des Kriterienkatalogs: Beschreibung und Diskussion der Kriterien des Qualitätsteils 53

Kriterium 1a: Abdeckung des Gesamtsystems 54

Kriterium 2a: Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet 55

Kriterium 2b: Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich 57

Kriterium 3a: Medizinische Bedeutung für das Behandlungsergebnis des Patienten 60

Kriterium 3b: Relevanz des Indikators für ein bestimmtes Fachgebiet 61

Kriterium 3c: Sektorbezug vs. sektorübergreifende Nutzung 63

Kriterium 4a: Medizinische Bedeutung in der medizinischen Fachliteratur 65

Kriterium 4b: Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften 66

Kriterium 4c: Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern . 67

Kriterium 4d: Verständlichkeit für Patienten und interessierte Öffentlichkeit 69

Kriterium 4e: Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit 71

Kriterium 5a: Reliabilität 72

Kriterium 5b: Validität 74

Kriterium 5c: Risikoadjustierung und Risikoadjustierbarkeit 75

Kriterium 6a: Objektiver Messaufwand und -komplexität 78

Kriterium 6b: Manipulationssicherheit / Objektivität der Messung 80

Kriterium 6c: Grad der Integration in automatische Prozesse / IT-Systeme 85

3.4. Zwischenfazit: Der Qualitätsteil des Kriterienkatalogs ist zu vielen Teilen neuartig und ermöglicht eine neue Blickweise auf P4P-Systeme 87

4. Entwicklung des Anreizteils des Kriterienkatalogs für P4P-Systeme 89

4.1. Grundlagen von Anreizsystemen 89

4.1.1. Theoretische Grundlagen von Anreizsystemen 89

4.1.2. Funktionsweise, Systemelemente und Gestaltungsdimensionen von Anreizsystemen 96

4.1.3. Theoretische Anforderungen an Anreizsysteme 99

4.1.4. Relevanz der theoretischen Grundlagen für Pay-for-Performance-Systeme 104

4.2. Theoretische Grundlagen von qualitätsorientierten Anreizsystemen in der stationären Versorgung (P4P) 104

4.2.1. Relevanz von Pay-for-Performance (P4P) 104

4.2.2. Definition und Begrifflichkeit von Pay-for-Performance 107

4.2.3. Ziele von P4P-Systemen 108

4.2.4. Funktionsweise von P4P-Systemen 110

4.2.5. Gestaltungsdimensionen von P4P-Systemen 110

4.3. Beschreibung der Entwicklung des Anreizteils des Kriterienkatalogs 113

4.3.1. Systemelemente des Anreizteils des Kriterienkatalogs 113

4.3.2. Kriterienkategorien des Anreizteils des Kriterienkatalogs 114

Kriterienkategorie 6: Wirksamkeit 114

Kriterienkategorie 7: Operationalisierbarkeit 114

Kriterienkategorie 8: Vergleichbarkeit 114

4.4. Der 2. Teil des Kriterienkatalogs: Beschreibung und Diskussion der Kriterien des Anreizteils 114

Kriterium 7a: Zielkonsistenz 115

Kriterium 7b: Zielkongruenz 116

Kriterium 7c: Reichweite 118

Kriterium 8a: Erreichbarkeit 120

Kriterium 8b: Transparenz 122

Kriterium 8c: Akzeptanz 123

Kriterium 8d: Langfristige Anreizwirkung der Ziele 124

Kriterium 9a: Gerechtigkeit 125

Kriterium 10a: Effizienz 126

Kriterium 10b: Effektivität 129

Kriterium 10c: Flexibilität 130

Kriterium 11a: Anteil der Vergütung 131

Kriterium 11b: Planbarkeit 133

Kriterium 12a: Nutzung zur Steuerung 135

Kriterium 12b: Aktualität 136

Kriterium 12c: Kontrollierbarkeit 138

Kriterium 12d: Praktische Umsetzung im Controlling 138

Kriterium 12e: Verursachungsgerechte Zuordnung und Beeinflussbarkeit 139

Kriterium 13a: Zeitverlauf 141

Kriterium 13b: Querschnittsvergleich/Ranking 142

Kriterium 13c: Justiziabilität 144

4.5. Zwischenfazit: Der neu entwickelte Kriterienkatalog beschreibt und bewertet P4P-Systeme umfassend und beinhaltet neuartige Elemente sowie eine praxisorientierte Strukturierung..... 145

5. Anwendung des entwickelten Katalogs bei P4P-Systemen 147

5.1. Umsetzungsbeispiele von P4P-Projekten 147

5.1.1. Entwicklung von P4P-Systemen – Schwerpunkt in den USA 147

5.1.2. Derzeitige Bewertungen von P4P-Systemen in der Literatur 148

5.1.3. Exkurs: Neueste Entwicklung von P4P in den USA 151

5.2. P4P-Ansätze in Deutschland: Gesetzliche Rahmenbedingungen und IV-Verträge mit Qualitätsgarantien 154

5.2.1. Rahmenbedingungen der stationären Vergütung..... 155

5.2.2. Verträge der Integrierten Versorgung - Beschreibung und Bestandsaufnahme..... 155

5.2.3. Besonderheit von qualitätsorientierter Vergütung in IV-Verträgen - Beschreibung und Gestaltungsoptionen 158

5.3. Methodische Ansätze für die Bewertung der Umsetzungsbeispiele 160

5.3.1. Methodische Argumentation für den Fallstudienansatz..... 160

5.3.2. Methodische Entscheidung für leitfadengestütztes Interview 162

5.3.3. Auswertungsverfahren nach Mayring 165

5.4. Beschreibung der Fallstudien: Derzeitige Umsetzung von P4P-Systemen in Deutschland 166

5.4.1. Fallstudie 1: Krankenkasse A 166

5.4.2. Fallstudie 2: Krankenkasse B 185

5.4.3. Fallstudie 3: Krankenkasse C 199

5.4.4. Fallstudie 4: Krankenhaus A 204

5.4.5. Exkurs: Auswertung des internen Qualitätsmanagementsystems anhand des Kriterienkatalogs.. 212

5.4.6. Fallstudie 5: Krankenhaus B..... 217

5.4.7. Exkurs: Auswertung des „Bonus-Vertrages“ anhand des Kriterienkatalogs 222

5.4.8. Fallstudie 6: Krankenhaus C 226

5.4.9. Fallstudie 7: Krankenhaus D 232

5.5. Bewertung der Nutzbarkeit des Kriterienkatalogs 237

5.6. Zwischenfazit: Die Anwendung des Kriterienkatalogs ermöglicht detaillierte Ergebnisse zur Umsetzung von P4P-Elementen in den Fallstudien 237

6. Diskussion der Erkenntnisse und weiterführende Anwendung 239

6.1. Bedeutung des Kriterienkatalogs: Durch den Kriterienkatalog können P4P-Systeme einheitlich und umfassend bewertet und entwickelt werden 239

6.2. Erkenntnisse der Fallstudien: In Deutschland werden P4P-Systeme in der stationären Versorgung bei den hier untersuchten Verträgen derzeit nicht vollständig umgesetzt 244

6.3. Schritte zu einer möglichen Umsetzung von P4P: Ausgewählte Empfehlungen für die Entwicklung eines P4P-Systems in Deutschland 249

Resümee 257

Anhang: Interviewleitfäden i

Literaturverzeichnis xx

Abbildungsverzeichnis

Abbildung 1: Graphische Darlegung der Gliederung 6

Abbildung 2: Darstellung des Literaturüberblicks für theoretische Grundlagen (Auszug) 7

Abbildung 3: Das magische Viereck der Gesundheitsökonomie 13

Abbildung 4: Anwendung von Strukturqualität in einem Vertrag der integrierten Versorgung 15

Abbildung 5: Zusammenhänge zwischen Struktur-, Prozess- und Ergebnisqualität 19

Abbildung 6: Akteure im Gesundheitswesen und Perspektiven der Qualität 24

Abbildung 7: Häufigste Todesursachen in den USA, 1998 und 2002.....26

Abbildung 8: Überblick über die Entwicklung ausgewählter Qualitätsinitiativen 36

Abbildung 9: Systemelemente des Kriterienkatalogs (Fokus auf Qualitätsteil) 45

Abbildung 10: Beispiel für Systemelemente anhand des G-IQI-Systems 48

Abbildung 11: Dichotome Bewertungsskala für Sektorbezug 64

Abbildung 12: Zusammenhang zwischen Verständlichkeit, Nutzen, Erfahrungen, Erkenntnisstand, wahrgenommener Relevanz und Akzeptanz 68

Abbildung 13: Bedürfnispyramide nach Maslow.....92

Abbildung 14: Relevante Principal-Agent-Beziehungen im stationären Sektor 96

Abbildung 15: Systemelemente des Kriterienkatalogs (Anreizsysteme) 98

Abbildung 16: Umsatzanteile eines Modell-IV-Vertrages mit Gewährleistungsklausel 133

Abbildung 17: Anzahl eingeschlossener Publikationen zu P4P im BQS-Gutachten.....149

Abbildung 18: Berechnungsbeispiels für den Indikator „AMI-7a - Fibrinolytic Therapy“.....153

Abbildung 19: Anzahl IV-Verträge in Deutschland, 2004-2011 157

Abbildung 20: Teilnehmer von und Ausgaben für IV-Verträge, Deutschland, 2008-2011 157

Abbildung 21: Wichtigkeit und Erfüllung einzelnen Aspekte für die Teilnahme an IV-Verträgen, Umfrageergebnisse von 84 Krankenkassen.....158

Abbildung 22: Möglichkeiten der Erkennung, Umsetzung und Messung von Gewährleistungsfällen in Fallstudie 1.....171

Abbildung 23: Systemelemente des Kriterienkatalogs.....240

Abbildung 24: Überblicksdarstellung des Kriterienkatalogs 243

Abbildung 25: Möglichkeiten der Erkennung, Umsetzung und Messung von Gewährleistungsfällen, Fallstudie 1, Krankenkasse A.....247

Tabellenverzeichnis

Tabelle 1: Anzahl gefundene Artikel bei Stichwortsuche in wissenschaftlichen Journals 27

Tabelle 2: Auszug stationärer Qualitätsindikatoren 30

Tabelle 3: Umsetzung von stationären Qualitätssystemen nach Kategorien (Auswahl) 36

Tabelle 4: Überblicksdarstellung des Kriterienkatalogs 54

Tabelle 5: Anforderungslisten im Überblick (Auswahl) 100

Tabelle 6: Verschiedene Zielsetzungen von P4P-Programmen 109

Tabelle 7: Beispielhafte Gestaltungsdimensionen für P4P-Programme (Auswahl) 111

Tabelle 8: Beschreibung einer Auswahl stationärer P4P-Projekte in den USA 148

Tabelle 9: Auswahl P4P-Projekte in Deutschland 159

Tabelle 10: Formen und Merkmale qualitativer Interviews 163

Tabelle 11: „SPSS-Methode“ der Leitfaden-Erstellung 164

Tabelle 12: Bewertungen Fallstudie 1, Systemelemente Gesamtsystem und Versorgungsbereich 175

Tabelle 13: Bewertungen Fallstudie 1, Systemelement Qualitätsindikator 175

Tabelle 14: Bewertungen Fallstudie 1, Systemelement Qualitätsmessung 177

Tabelle 15: Bewertungen Fallstudie 1, Systemelement Qualitätsziele 179

Tabelle 16: Bewertungen Fallstudie 1, Systemelement Vergütungsfunktion 181

Tabelle 17: Bewertungen Fallstudie 1, Systemelement Messergebnis 182

Tabelle 18: Bewertungen Fallstudie 2, Systemelemente Gesamtsystem und Versorgungsbereich 191

Tabelle 19: Bewertungen Fallstudie 2, Systemelement Qualitätsindikator 191

Tabelle 20: Bewertungen Fallstudie 2, Systemelement Qualitätsmessung 193

Tabelle 21: Bewertungen Fallstudie 2, Systemelement Qualitätsziele 194

Tabelle 22: Bewertungen Fallstudie 2, Systemelement Vergütungsfunktion 196

Tabelle 23: Bewertungen Fallstudie 2, Systemelement Messergebnis 197

Tabelle 24: Bewertungen Exkurs Fallstudie 4, Systemelemente Gesamtsystem und Versorgungsbereich 213

Tabelle 25: Bewertungen Exkurs Fallstudie 4, Systemelement Qualitätsindikator 213

Tabelle 26: Bewertungen Exkurs Fallstudie 4, Systemelement Messung.....214

Tabelle 27: Bewertungen Exkurs Fallstudie 4, Systemelement Ziele.....215

Tabelle 28: Bewertungen Exkurs Fallstudie 4, Systemelement Vergütungsfunktion 216

Tabelle 29: Bewertungen Exkurs Fallstudie 4, Systemelement Messergebnis.....217

Tabelle 30: Bewertungen Exkurs Fallstudie 5, Systemelement Gesamtsystem und Versorgungsbereich .222

Tabelle 31: Bewertungen Exkurs Fallstudie 5, Systemelement Qualitätsindikator 223

Tabelle 32: Bewertungen Exkurs Fallstudie 5, Systemelement Messung.....224

Tabelle 33: Bewertungen Exkurs Fallstudie 5, Systemelement Ziele.....224

Tabelle 34: Bewertungen Exkurs Fallstudie 5, Systemelement Vergütungsfunktion 225

Tabelle 35: Bewertungen Exkurs Fallstudie 5, Systemelement Messergebnis.....226

Tabelle 36: Bewertungen Fallstudie 6, Systemelement Ziele 231

Tabelle 37: Bewertungen Fallstudie 6, Systemelement Messergebnis 231

Tabelle 38: Bewertungen P4P-Modell Systemelemente Gesamtsystem, Versorgungsbereich, Qualitätsindikator, Messung.....255

Tabelle 39: Bewertungen P4P-Modell Systemelemente Ziele, Vergütungsfunktion, Messergebnis 256

Abkürzungsverzeichnis

AHRQ	Agency for Healthcare Research and Quality, Forschungsstelle, untersteht dem United States Department of Health and Human Services (HHS).
AOK	Allgemeine Ortskrankenkasse.
AQUA	Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen.
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.
BaG	Bundesamt für Gesundheit in der Schweiz.
BIP	Bruttoinlandsprodukt.
BQS	Bundesgeschäftsstelle Qualitätssicherung gGmbH.
CH-IQI	Swiss Inpatient Quality Indicators.
CMS	Centers for Medicare & Medicaid Services, für das Medicare Programm zuständige Gesundheitsamt der USA.
DALY	Disability-Adjusted-Life-Years, Lebensjahre durch Krankheitslast reduziert, behinderungsbereinigt.
DIN	Deutsches Institut für Normierung.
DMP	Disease management Program - systematisches Behandlungsprogramm für chronisch kranke Menschen.
DRG	Diagnoses-related Groups.
EFQM	European Foundation for Quality and Management, Qualitätsmanagementsystem.
G-BA	Gemeinsamer Bundesausschuss, höchstes Gremium der gemeinsamen Selbstverwaltung im Gesundheitswesen Deutschlands.
G-IQI	German Inpatient Quality Indicators.
GKV	Gesetzliche Krankenversicherung.
HHS	U.S. Department of Health and Human Services (HHS), US-Bundes- Gesundheitsministerium
HSMR	Hospital Standardised Mortality Ratios.
HVBP	Hospital Value-Based Purchasing.
InEK	Institut für das Entgeltsystem im Krankenhaus (InEK GmbH).

IQI	Inpatient Quality Indicators.
IQM	Initiative Qualitätsmedizin.
IV-Vertrag	Vertrag der Integrierten Versorgung nach §§ 140a-d SGB V
IOM	Institute of Medicine: unabhängige Prüfstelle im Gesundheitsbereich der US National Academy of Sciences.
ISO	International Organization for Standardization, internationale Vereinigung von Normungsorganisationen.
JCAHO	Joint Commission on Accreditation of Healthcare Organisations.
KHEntgG	Krankenhausentgeltgesetz, gemeinsam mit Krankenhausfinanzierungsgesetz relevant für Vergütung von Krankenhausleistungen.
KTQ	Kooperation für Transparenz und Qualität.
MDK	Medizinischer Dienst der Krankenversicherungen.
NHS	National Health Service (NHS), staatliches Gesundheitssystem in Großbritannien.
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung .
PDCA	Plan-Do-Check-Act (Deming-Zyklus).
P4P	Pay for Performance – Qualitätsorientierte Vergütungssysteme in der stationären Versorgung.
PHQID	CMS/Premier Hospital Quality Incentive Demonstration (PHQID).
PSI	Patient Safety Indicators, Indikatorenset der Patientensicherheit.
QALY	qualitätskorrigiertes Lebensjahr (quality adjusted life year).
QSR	Qualitätssicherung mit Routinedaten.
SGB V	Sozial-Gesetz-Buch, fünftes Buch.
SVR	Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen.
TK	Techniker Krankenkasse.
TQM	Total Quality Management.
WidO	Wissenschaftliches Institut der AOK.

1. Einleitung

1.1. Diskussion der Problemstellung und Zielsetzung des Promotionsvorhabens

1.1.1. Problemstellung des Promotionsvorhabens

> *„Überhaupt aber beruhen neun Zehntel unseres Glückes allein auf der Gesundheit. Mit ihr wird alles eine Quelle des Genusses“ (Schopenhauer 1851, S. 105)*

Die hohe Bedeutung von Gesundheit und somit von medizinischen Leistungen ist unumstritten – sowohl für die Einzelperson als auch für die Bevölkerung. Daher ist eine hohe medizinische Qualität im ureigensten Interesse einer jeden Gesellschaft. Seit Mitte des 19. Jahrhunderts wird das Thema Qualität in der stationären Versorgung diskutiert. Neben den Möglichkeiten von Qualitätsmessungen ist in den letzten Jahren vermehrt auf die Frage eingegangen worden, wie Qualität aktiv und auch unter Managementbeteiligung verbessert werden kann. Ein – aus der betriebs- und motivationswissenschaftlichen Literatur bekannter – Ansatz ist dabei die qualitätsorientierte Vergütung – d. h. finanzielle Anreize, die eine direkte Verknüpfung zwischen Qualität und Bezahlung schaffen. In Deutschland sind Aspekte der Krankenhausvergütung und -qualität derzeit – auch gesetzlich – eher getrennt (vgl. Leber et al. 2007). Bekannt geworden ist das Konzept der qualitätsorientierten Vergütung als „Pay-for-Performance (P4P)“.

Wie gut ist die Qualität der stationären Versorgung in Deutschland? Mit einer Lebenserwartung von 80 Jahren liegt Deutschland sowohl über dem europäischen als auch weltweiten Durchschnitt von 75 bzw. 68 Jahren (vgl. WHO 2010). Auch andere Indikatoren positionieren die deutsche Gesundheitsversorgung im oberen Mittelfeld (vgl. WHO 2006, S. 5ff.). Trotz einer guten Ausgangsposition kann die Qualität des Gesundheitssystems allerdings noch verbessert werden.

Wie kann Qualität im Krankenhaus gemessen werden? Welche Indikatoren und Messverfahren dazu geeignet sind, Qualität im Krankenhaus darzustellen, wird nach wie vor von Wissenschaftlern und Praktikern diskutiert. Es gibt eine Vielzahl von Indikatorensystemen, die in Deutschland und international verwendet werden (so u. a. AHRQ-Indikatoren, AQUA-Qualitätsberichte, G-IQI, QSR-Indikatoren, PSI). Auch ein einheitliches Verständnis des Qualitätsbegriffes gibt es nur bedingt. Zwar wird die Qualitätsdefinition des Institutes of Medicine (IOM) weitestgehend akzeptiert, gleichzeitig geht jedoch die Bewertung verschiedener Qualitätsdimensionen (bspw. nach Donabedian: Struktur-, Prozess- und Ergebnisqualität; vgl. Donabedian 1978) deutlich auseinander. In Deutschland wird dieses gerade bei der Diskussion von Mindestmengen (als Strukturqualitätsindikator) deutlich (vgl. Geraedts/de Cruppe 2006, S. 87; Khuri/Henderson 2005, S. 1222f.; Velasco-Garrido/Busse 2004 sowie Flintrop 2012, S. 11).

Wie kann Qualität im Krankenhaus umgesetzt werden? Der Schwerpunkt der Qualitätsdiskussion wird seit einigen Jahren unter anderem auf die Frage von qualitätsorientierter Vergütung oder Pay-for-Performance gelegt. Die Idee solcher Anreizsysteme basiert auf Konzepten der Motivations- und Organisationstheorie. Es wird versucht, Leistungserbringer (oder konkret Krankenhäuser) durch monetäre Anreize zur langfristigen Verbesserung der Behandlungsqualität zu motivieren (vgl. Amelung/Zahn 2009, S. 7ff.). Hierfür gibt es verschiedene Umsetzungsmöglichkeiten, die auch für den deutschen Markt bereits diskutiert werden (vgl. z. B. Mansky 2010a, S. 8f.). Angeregt wurde die Diskussion unter anderem auch durch die Arbeit von Porter und Teisberg aus dem Jahre 2006 „Redefining Health Care“ (vgl. Porter/Teisberg 2006). Hierin wird die Idee von „value-based purchasing/competition“ (VBP) im Gesundheitswesen aufgebracht, die Vorteile des freien Wettbewerbs aus anderen Industrien auf den Gesundheitsmarkt übertragen soll.

Welche P4P-Projekte gibt es bereits? Erste Pilotprojekte für P4P-Systeme wurden in den USA und in Großbritannien seit Anfang des 21. Jahrhunderts umgesetzt und sind bereits mehrfach bewertet worden – u. a. durch systematische Reviews (vgl. Rosenthal et al. 2005; SVR 2007; van Herck et al. 2010; Scheppach et al. 2011). Insgesamt zeigt sich, dass durch die Integration von Qualitätsaspekten in monetäre (oder auch nicht monetäre) Anreizsysteme die medizinische Qualität verbessert werden kann, wenn auch die Studienlage häufig schwierig ist. Wovon die Ausprägung dieser Verbesserungen abhängt, ist teilweise noch unklar – verschiedene Umsetzungsmodalitäten wie beispielsweise die Vergütungshöhe oder die Bemessungsgrundlage sind mögliche Gestaltungsfaktoren (vgl. Scheppach et al. 2011, S. 102ff.).

Welche Anforderungen muss somit ein erfolgreiches P4P-System erfüllen? Ein einheitliches Verständnis, welche Anforderungen und Kriterien ein solches System erfüllen sollte, ist derzeit nicht vorhanden. Demzufolge sind die verschiedenen Pilotprojekte nur teilweise vergleichbar, und neue Entwicklungen von Systemen können auf keine standardisierten Umsetzungsempfehlungen aufbauen.

Wie werden die Konzepte des P4Ps in Deutschland umgesetzt? Die Grundidee der qualitätsorientierten Vergütung im stationären Sektor kann derzeit im Rahmen von Verträgen der integrierten Versorgung umgesetzt werden. Dies findet unter anderem durch die Anwendung von Gewährleistungsklauseln statt, die eine qualitätsvariable Nicht-Erstattung ermöglichen. Inwieweit diese Garantien tatsächlich umgesetzt, gemessen und nachverfolgt werden, ist derzeit nur wenig evaluiert. Dies soll in diesem Promotionsvorhaben analysiert werden.

Problemstellung: Obschon P4P als „Trend“-Konzept sowohl umgesetzt als auch evaluiert und diskutiert wird, gibt es derzeit kein einheitliches Verständnis dafür, wie ein qualitätsorientiertes Vergütungssystem aufgebaut sein sollte. Ein detaillierter Anforderungskatalog für die Qualitäts- und Anreizaspekte eines solchen Systems fehlt. Doch nur bei einem einheitlichen Verständnis, welche Kriterien erfüllt sein müssen, kann eine standardisierte Bewertung und Entwicklung von P4P-Systemen stattfinden.

Gleichzeitig müssen die bereits umgesetzten Pilotprojekte evaluiert werden, um aus der derzeitigen Implementierung Schlüsse für künftige P4P-Projekte ziehen zu können.

1.1.2. Zielsetzung: Entwicklung und Nutzung eines Kriterienkatalogs für P4P-Projekte

Um die Wirksamkeit von P4P-Systemen zu erhöhen und auch eine breitere Implementierung zu ermöglichen, ist ein Kriterienkatalog sinnvoll, der die Anforderungen an ein qualitätsorientiertes Anreizsystem beschreibt und bewertet. Daher soll die vorliegende Arbeit einen solchen Katalog entwickeln – sowohl für die Anforderungen an die Qualitätsaspekte als auch für die Anreizsystematik.

Ziel des Dissertationsvorhabens ist außerdem, diesen Katalog bereits an ersten Umsetzungsbeispielen in Deutschland anzuwenden. Gerade in Deutschland können dadurch erste Ansätze von Malus-Systemen beispielsweise bei Verträgen der Integrierten Versorgung mit Gewährleistungsklauseln detailliert untersucht werden. Zweck dieser Anwendung ist neben einer detaillierten, standardisierten Bewertung von P4P-Ansätzen die Überprüfung des Kriterienkatalogs und seiner Anwendbarkeit.

Ein weiteres Ziel der Dissertation ist die Analyse und Evaluation von ersten praktischen Beispielen qualitätsorientierter Vergütung im deutschen stationären Sektor. Es wird untersucht, ob solche Vergütungssysteme innerhalb von Verträgen der integrierten Versorgung tatsächlich umgesetzt werden und inwieweit dadurch finanzielle Anreize entstehen und zu besseren Qualitätsergebnissen führen. Bei potenziellen Umsetzungsschwierigkeiten und -hindernissen soll die Anwendung des Kriterienkatalogs ermöglichen, die Implementierung zu analysieren und ggf. zu optimieren.

Die letzte Zielsetzung ist die Beschreibung von ersten Empfehlungen zur Einführung eines P4P-Systems, das stationäre Qualität bundesweit in Deutschland vergüten kann. Auch hierfür wird der Kriterienkatalog herangezogen.

1.2. Zentrale Fragenstellungen und Gliederung

Die relevanten Fragestellungen des Dissertationsvorhabens sind:

- Welche Bedeutung hat Qualität in der stationären Versorgung?
  - Welche Entwicklungen gibt es in der Qualitätsdiskussion und welche Bedeutung hat Qualität in der stationären Versorgung?
  - Wie kann der Qualitätsbegriff und die -messung definiert und operationalisiert werden?
- Welche Kriterien sollte ein qualitätsorientiertes Anreizsystem erfüllen?
  - Welche Elemente und Funktionen hat ein qualitätsorientiertes Anreizsystem?
  - Welche Anforderungen sollten der Qualitäts- und Anreizaspekt eines solchen Systems erfüllen?

- Empirischer Teil: Wie gut erfüllen bisherige Pilotprojekte von P4P diesen Kriterienkatalog?
  - Welche P4P-Projekte gibt es derzeit in Deutschland?
  - Wie stringent werden vorhandene P4P-Systeme bei beispielhaft ausgewählten Krankenkassen und Anbietern in Deutschland umgesetzt?
  - Wie gut wird der Kriterienkatalog von diesen Systemen erfüllt?
  - Wie konsequent wird das Management der P4P-Vereinbarungen in der betrieblichen Praxis umgesetzt?
  - Welche Verbesserungsmöglichkeiten bestehen für die P4P-Projekte?
  - Wie wird die Anwendbarkeit des Katalogs bewertet?
- Kann der Kriterienkatalog bei der Etablierung eines deutschen Pay-for-Performance-Systems Verwendung finden?
  - Wie könnte ein deutsches P4P-System aussehen (Modellcharakter)?
  - Welche Umsetzungsphasen und ggf. -hindernisse bestehen?

Zur Beantwortung dieser Fragestellungen wird die Promotionsarbeit in vier Blöcke mit insgesamt 6 Kapiteln unterteilt:

A. Erfassung, Zielstellung der Doktorarbeit: *Kapitel 1 (Einleitung) und Kapitel 2 (Qualität in stationärer Versorgung)*

Im ersten Block werden zunächst die Grundlagen für das Forschungsziel gelegt. Nach einer kurzen Literaturanalyse und einer Diskussion des methodischen Vorgehens werden im zweiten Kapitel die Grundlagen der stationären Qualität näher beleuchtet. Hierfür wird zunächst der Qualitätsbegriff diskutiert und eine „working definition“ entwickelt. Anschließend wird die Entwicklung der stationären Qualität dargelegt. Aspekte der Qualitätsmessung, -indikatoren, -managementsysteme und des Public Reporting kommen dabei zum Tragen.

B. Entwicklung des Kriterienkatalogs: *Kapitel 3 (Entwicklung des Qualitätsteils des Kriterienkatalogs für P4P-Systeme) und Kapitel 4 (Entwicklung des Anreizteils des Kriterienkatalogs für P4P-Systeme)*

Im zweiten Block wird der Kriterienkatalog entwickelt. Um die Aspekte der Qualität(smessung) und des Anreizsystems voneinander zu trennen, wird der Katalog in zwei Schritten erstellt: Zunächst die Kriterien mit Qualitätsbezug anschließend die Anforderungen an das Anreizsystem.

(1) Kriterien mit Qualitätsbezug: Es werden Anforderungen definiert, die für den ausgewählten Versorgungsbereich, die Qualitätsindikatoren und die -messung relevant sind. Neben einer Herleitung

aus existierender Literatur werden für alle Kriterien mögliche Umsetzungs- und Bewertungsverfahren im Sinne einer Operationalisierung definiert.

(2) Kriterien des Anreizsystems: Auch für die Anreizsystematik werden Anforderungen definiert – vor allem das Zielsystem, die Vergütungsfunktion und das Messergebnis sind hierbei entscheidend. Zu diesem Zweck werden zunächst die theoretischen Grundlagen der VWL, BWL und Motivationslehre sowie primär des Principal-Agent-Theorems diskutiert und für die Entwicklung angewendet.

C. Anwendung des Kriterienkatalogs auf Umsetzungsbeispiele*: Kapitel 5 (Anwendung des entwickelten Katalogs bei P4P-Systeme)*

Der entwickelte Kriterienkatalog soll im dritten Block auf seine Funktionsfähigkeit und Anwendbarkeit überprüft werden. Außerdem werden Erkenntnisse über den Stand von P4P-Vorhaben in den USA und Deutschland diskutiert. Der Katalog wird bei verschiedenen Fallstudien angewendet und zwar für IV-Verträge mit Qualitätsgarantien in Deutschland:

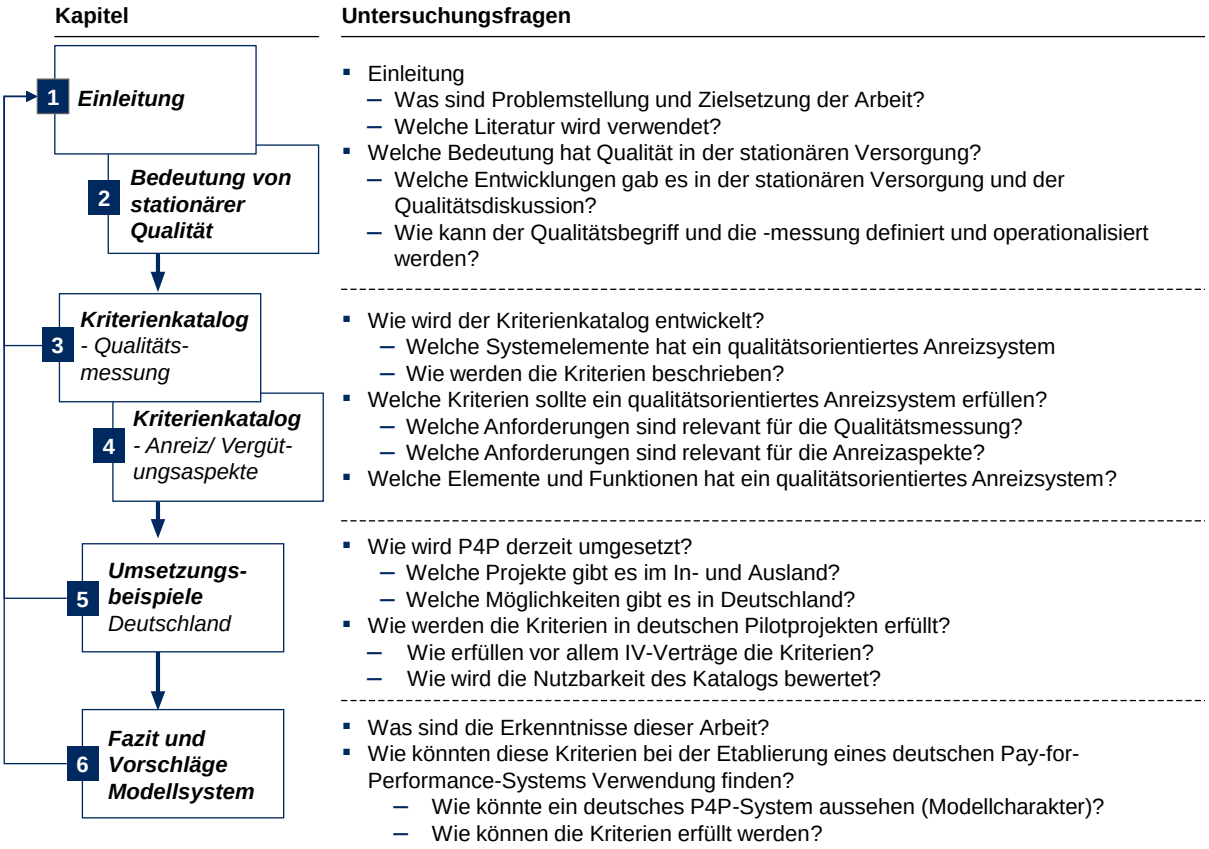
IV-Verträge mit Qualitätsgarantien: Da es nur wenige Pilotprojekte in Deutschland gibt, werden IV-Verträge mit Qualitätsgarantien ausgewählt, um den Kriterienkatalog anzuwenden. Diese setzen erste Aspekte von P4P-Systemen um und können daher anhand des Katalogs analysiert werden. Verschiedene Verträge werden nach Gesprächen mit den Beteiligten und Erhebung von Primärdaten mithilfe der Kriterien evaluiert. Anschließend wird ein Fazit zur Funktionsfähigkeit des Katalogs gezogen.

D. Zusammenfassung der Ergebnisse und Ausblick*: Kapitel 6 (Diskussion der Erkenntnisse und* weiterführende Anwendung*)*

In diesem Block wird zunächst ein zusammenfassendes Fazit der Forschungsergebnisse vor allem hinsichtlich Anwendbarkeit, Funktionsfähigkeit und Struktur des Katalogs diskutiert. Auch die Erkenntnisse der sieben Fallstudien werden diskutiert. Als letzter Schritt wird ein Ausblick auf eine mögliche Umsetzung eines bundesweiten übergreifenden P4P-Systems gegeben.

Diese Gliederung ist in Abbildung 1 graphisch dargelegt:

Abbildung 1: Graphische Darlegung der Gliederung



1.3. Beschreibung der methodischen Vorgehensweise

Insgesamt stellt die Arbeit eine deduktive Entwicklung eines Konstruktes mit Überprüfung anhand von Fallstudien dar. Der Ansatz ist positivistisch gewählt, da speziell die objektive Einschätzung und Bewertung der Umwelt mit möglichst wertfreien Mechanismen angestrebt wird (vgl. Blumberg et al. 2008, S. 20–21). Gleichzeitig wird die Möglichkeit einer auch subjektiven Wahrnehmung der Ergebnisse nicht negiert – gerade im Kapitel 5 werden Informationen durch qualitative Interviews mit Krankenhaus- und Krankenkassenmitarbeitern mit „realistischem“ Ansatz gewonnen (vgl. King/Horrocks 2010, S. 12).

Die Entwicklung des Kriterienkatalogs in den Kapiteln 3 und 4 entspricht einer deduktiven Herangehensweise. Dazu werden nach eingehender Literaturanalyse zunächst die verschiedenen Systemelemente eines Kriterienkatalogs definiert. Anschließend wird die für jedes Kriterium relevante Literatur untersucht, zusammengeführt und mit konkreten Anwendungsmöglichkeiten sowie Bewertungsoptionen versehen.

Nach der Beschreibung und ersten Bewertung des Kriterienkatalogs wird dieser auf Nutzbarkeit und Funktionalität hin überprüft. Hierfür werden deutsche Projekte analysiert. Als erstes werden wiederum mit einem qualitativen Ansatz Projekte von drei Krankenkassen und vier Krankenhäusern im Rahmen von IV-Verträgen anhand des Kriterienkatalogs bewertet. Insgesamt werden die Daten ex-post-facto, causal

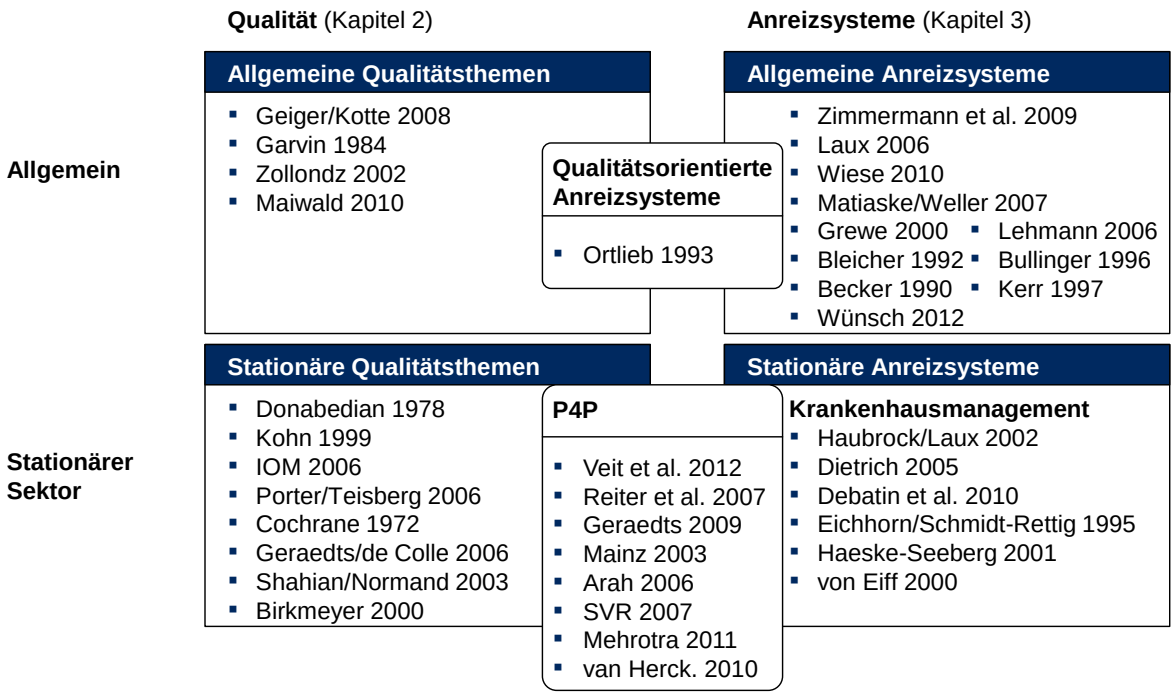
und cross-sectional analysiert (vgl. Blumberg et al. 2008, S. 196). Besonders für den empirischen Teil wird im Kapitel 5 erneut detailliert auf die verwendeten Methoden eingegangen, so bspw. der qualitative, multiple Fallstudienansatz (vgl. Wrona 2005, S. 5ff.; Meuser/Nagel 2009), das leitfadengestützte problemzentrierte Interview (vgl. Witzel 2000) und die Auswertungsmethode der qualitativen Inhaltsanalyse nach Mayring (vgl. Mayring 2010b).

Im letzten Kapitel wird eine kurze Ausschau auf einen prädiktiven Ansatz gegeben, um ein mögliches System in Deutschland zu beschreiben.

1.4. Literaturüberblick für theoretische Grundlagen

An dieser Stelle soll ein kurzer Überblick über eine Auswahl der vorhandenen und verwendeten Literatur gegeben werden. Für diese Dissertationsarbeit werden verschiedene Stränge der aktuellen Literatur zusammengebracht. Während für den ersten Teil des Anforderungskatalogs Literatur zum Thema Qualität und Qualitätsmanagement relevant ist, werden im zweiten Teil Arbeiten zu Anreizsystemen genutzt. Da im Kapitel 5 vor allem der Gesundheitssektor in Deutschland und deutsche Pilotprojekte untersucht werden, soll ein Fokus (soweit möglich) auf deutscher Literatur liegen. Für eine Übersicht der Literaturstränge und einen Auszug der verwendeten Arbeiten, siehe Abbildung 2.

Abbildung 2: Darstellung des Literaturüberblicks für theoretische Grundlagen (Auszug)



Eine Auseinandersetzung mit Literatur zur Qualität muss mit einer Begriffsdefinition beginnen, für die meist die Definition des ISO (International Organization for Standardisation) herangezogen wird – (vgl. DIN-EN-ISO 8402; Geiger/Kotte 2008). Weitere, allgemein gültige Differenzierungen sind in den Ansätzen

von Garvin zu finden (vgl. Garvin 1984). Zur Anwendung des Qualitätsbegriffs in Managementsystemen eignen sich unter anderem die Arbeiten von Zollondz 2002 und Maiwald 2010.

Qualität im stationären Sektor wird seit vielen Jahren gemessen, bewertet, verbessert und vor allem diskutiert. Für die Anfänge der Messung der Krankenhausqualität werden hierbei die Darstellung von Kelsey 2001; Chassin 1996 und Chassin/Loeb 2011 verwendet, die die Arbeiten von Florence Nightingale, Ernest Amory Codman und Abraham Flexner beschreiben. Für die Kategorisierung der verschiedenen Qualitätsdimensionen war vor allem Donabedian verantwortlich (vgl. Donabedian 1978). Die Verbesserung der medizinischen Ergebnisse ist ein zentrales Thema vieler klinischer Forschungsarbeiten. Gibt man das Stichwort „quality“ in der am meisten angewandten medizinisch wissenschaftlichen Suchmaschine PubMed ein, so werden mit Stand Mai 2012 582.069 Arbeiten als Treffer angezeigt. Dies spiegelt aber nur einen Teil der die Qualität betreffenden Arbeiten wider, da fast jede medizinische Therapiestudie eine Verbesserung der Behandlungsverfahren und damit auch der Qualität zum Ziel hat, auch wenn das Stichwort „quality“ nicht explizit erwähnt wird. Ein Meilenstein in der medizinisch-politischen Qualitätsdiskussion war die 1999 vom Institute of Medicine (IOM) publizierte Metaanalyse zu Fehlern in der stationären Versorgung. Im gesundheitsökonomischen Bereich stach in den letzten Jahren beispielsweise Michael Porter mit dem Thema „Value based purchasing“ hervor (vgl. Kohn et al. 1999; Institute of Medicine 2007; Porter/Teisberg 2006). Die neueste Arbeit von Michael Porter beschäftigt sich mit dem deutschen Gesundheitswesen und ist im März 2012 entstanden (Porter/Guth 2012).

Für die Entwicklung des hier vorgestellten Anforderungskatalogs für qualitätsorientierte Vergütung stehen in dieser Arbeit Fragestellungen der stationären Qualitätssicherung im Vordergrund. Zu diesem Zweck werden existierende praktische Ansätze besprochen, z. B. von AQUA, IQI, G-IQI, QSR, AHRQ, aber auch andere Konzepte betrachtet, wie z. B. Zertifizierungen, oder Qualitätsmanagementsysteme wie JCAHO, EFQM, KTQ und DIN-ISO 9000. Außerdem gibt es eine Vielzahl an Arbeiten und Studien, die sich mit der Wirksamkeit einzelner Qualitätsindikatoren oder den Qualitätsdimensionen auseinandersetzen (ein kleiner Auszug ist Luft et al. 1979; Birkmeyer 2000; Dudley et al. 2000; Haeske-Seeberg 2001, S. 113; Shahian/Normand 2003; Khuri/Henderson 2005 und Geraedts/de Cruppe 2006; McAuliffe 1978; Rubin 2001; Lowrie et al. 2001; Mant 2001; Bradley et al. 2006; Werner/Bradlow 2006; Fonarow et al. 2007; Mehta et al. 2008 und Krumholz et al. 2007).

Für die theoretischen Grundlagen des zweiten Teils des Katalogs wird zunächst die ökonomische Literatur zu Anreizsystemen und deren Anforderungen herangezogen. Diese sind u. a. für volkswissenschaftliche Anreize Coase 1937; Zimmermann et al. 2009; Zweifel et al. 2009 und Wiese 2010. In der betriebswissenschaftlichen Literatur (die hier im Vordergrund steht) gibt es mehrere Fachbereiche, die sich der Thematik der Anreizsysteme annehmen, so unter anderem die Unternehmensführung (z. B. Schanz 1991; Hungenberg 2006), Unternehmenspolitik und -organisation (z. B. Grewe 2006), Personalmanagement und Unternehmenssteuerung (z. B. Lindert 2001) aber auch Strategie (z. B. Becker 1990; Bleicher 1992) oder Qualitätsmanagement (z. B. Ortlieb 1993). Übergreifend gibt es zwei

Strömungen: die verhaltenstheoretischen Konzepte mit den wichtigsten Arbeiten von Maslow und Herzberg (auch Arbeiten von Deci/Ryan 1985; Frey 2005 und Matiaske/Weller 2006 finden Verwendung) und die Principal-Agent-Theorie. Letztere ist für diese Dissertationsarbeit besonders relevant – primär werden die Arbeiten von Petersen 1989; Mas-Colell et al. 1995; Staehle et al. 1999; Frey/Benz 2004, S. 22; Kräkel 2004; Schreyögg/von Werder 2004 und Lehmann 2006 genutzt. Eine Verknüpfung dieser Theorie mit den Verhältnissen im stationären Sektor schafft in erster Linie Ernst 1999. Einige deutsche Arbeiten zu Anreizsystemen und sogar zu den Anforderungen an solche Systeme sind für das Promotionsvorhaben von hoher Relevanz: Becker 1990; Seidel 1991; Bleicher 1992; Bullinger 1996; Kerr 1997; Grewe 2000; Lehmann 2006; Laux 2006; Wünsch 2012. Qualitätsorientierte Anreizsysteme wurden vor allem von Ortlieb 1993 analysiert.

Anreizsysteme der stationären Versorgung sind in Deutschland bislang nur teilweise erforscht. Während es eine Vielzahl von Arbeiten zu Krankenhausmanagement, -betriebslehre oder -personalwirtschaft gibt (z. B. Hoefert 1997; Schwartz 1997; Haeske-Seeberg 2001; Sturm 2002, Haubrock/Laux 2002; Dietrich 2005; Fleßa 2007; Debatin et al. 2010); werden Anreizsysteme und insbesondere das Zusammenwirken von Leistungserbringern und Kostenträgern dabei meist nicht betrachtet. Einige Arbeiten konzentrieren sich vorwiegend auf die Motivation von Mitarbeitern und Führungskräften (Eichhorn/Schmidt-Rettig 1995; Eiff 2000; Humburg 2001) – dies ist zwar generell von Interesse, die Fokussierung auf innerbetriebliche Aspekte ist für dieses Promotionsvorhaben dagegen weniger geeignet, daher werden Krankenhausmanagementaspekte nur sekundär betrachtet.

Für die Entwicklung des Anforderungskatalogs und die Diskussion der relevanten Kriterien werden verschiedene Literaturstränge herangezogen:

- Existierende Auflistungen von verschiedenen Anforderungen für die Entwicklung oder Bewertung von Qualitätsindikatoren[1], so z. B. in Deutschland von QUALIFY (so unter anderem in Reiter et al. 2007; Geraedts et al. 2002 und Zorn/Ollenschläger 1999; oder international Remus et al. 2004; Campbell 2002; JCAHO 2006; Mattke et al. 2006; Mainz 2003; Arah et al. 2006; Ten Asbroek et al. 2004; Vos et al. 2009 und Wollersheim et al. 2007)
- Bewertungsschemata für P4P-Systeme in – systematischen – Reviews (z. B. in Deutschland SVR 2007; Scheppach et al. 2011; Veit et al. 2012a[2], Schrappe/Gültekin 2010; Mehrotra et al. 2010a und Amelung/Zahn 2009 oder international Rosenthal et al. 2005; van Herck et al. 2010; Bufalino et al. 2007; Institute of Medicine 2007 und Kuhmerker/Hartman 2007)

[1] Einige Anforderungen auch in Arbeiten zur Bewertung von Indikatorensystemen und systemischen Reviews

[2] Hier u. a. auch ein „Algorithmus“ zur Überprüfung der Kompatibilität und Eignung von Indikatoren für verschiedene P4P-Systeme, deren einzelne Anforderungen jeweils bei den Kriterien erwähnt werden (vgl. Veit et al. 2012e, S. 42ff. sowie Veit et al. 2012f, S. 69ff.)

- Anforderungskataloge für Qualitätssysteme in anderen Bereichen (z. B. in der ambulanten Versorgung oder für Ärzteportale (so unter anderem Thomeczek 2002; Diel 2009))
- Anforderungen, die sich aus klinischen Praxisanwendungen entwickelt haben (vgl. u. a. Mansky 2010a).
- Anforderungskataloge an Anreizsysteme aus der betriebswissenschaftlichen Literatur (bspw. Becker 1990; Bleicher 1992; Seidel 1991; Bullinger 1996; Kerr 1997; Ortlieb 1993; Grewe 2000; Lehmann 2006; Laux 2006; Wünsch 2012)

Insgesamt zeigt sich, dass die deduktive Erarbeitung des Kriterienkatalogs prinzipiell auf zahlreiche Literatur und wissenschaftliche Arbeiten zurückgreifen kann. Gleichzeitig jedoch sind mehrere der Einzelkriterien und speziell deren Detaillierung sowie die Entwicklung eines umfassenden, detaillierten Katalogs als neuartig zu bezeichnen (vgl. hierfür auch Kapitel 3 und 4).

Für die Anwendung der Kriterienkataloge bei deutschen Projekten von qualitätsorientierter Vergütung werden nebst einiger Sekundärliteratur zu Projektbewertungen in den USA und in Deutschland (z. B. van Herck et al. 2010; Rosenthal et al. 2005; Mehrotra et al. 2008; Petersen et al. 2006; Bufalino 2006; SVR 2007; Scheppach et al. 2011; Schrappe/Gültekin 2010; Veit et al. 2012a; Amelung/Zahn 2009) primär eigene Primärquellen (u. a. Interviewtranskripte) genutzt, die in den jeweiligen Kapiteln näher beschrieben werden.

Für die Beschreibung der verwendeten Methodik wird nach einer Diskussion des qualitativen Ansatzes und im Speziellen der Verwendung qualitativer Fallstudien mit Arbeiten von Wrona 2005 und Yin 2003; vorwiegend auf die Entwicklung und Durchführung der Interviews eingegangen (King/Horrocks 2010; Witzel 2000; Lamnek 2005; Flick 2011; Marotzki 2003; Meuser/Nagel 2009; Helfferich 2005). Für die Auswertung der Interviews wird die qualitative Inhaltsanalyse nach Mayring herangezogen mit Literatur von unter anderem Mayring 2010b und Langer 2010.

2. Qualität in stationärer Versorgung: Eingrenzung des Qualitätsbegriffs und Bedeutung von Qualität

Zur Entwicklung des Kriterienkatalogs in der vorliegenden Arbeit ist es zunächst notwendig, ein Verständnis für Qualität in der stationären Versorgung zu entwickeln. Zu diesem Zweck müssen einige Fragen diskutiert werden:

- Was heißt „Qualität“ in der Medizin und wie wird der Begriff definiert?
- Welche Dimensionen der Qualität sind relevant und wie unterscheiden sich die Qualitätsperspektiven von verschiedenen Akteuren des stationären Sektors?
- Welche Entwicklungen gab es bisher in der Qualitätsmessung und der Nutzung von stationären Qualitätskennzahlen?

Für die erste Frage wird zunächst auf die vorherrschende Begriffsdefinition von „Qualität“ eingegangen. Diese Begriffsbedeutung hat eine lange Tradition in den Qualitätswissenschaften und wird auch von unter Akteuren im Gesundheitswesen viel diskutiert. Aus den vorangegangenen Erkenntnissen wird eine Arbeitsdefinition („working definition“) der Qualität für diese Arbeit erstellt. Dies soll ermöglichen, die Entwicklung des Kriterienkatalogs entlang der Arbeitsdefinition zu konkretisieren.

Anschließend werden verschiedene Aspekte der Qualität näher beleuchtet: Neben den Qualitätskategorien nach Donabedian sind dies verschiedene Dimensionen (z. B. nach Garvin) sowie Perspektiven von verschiedenen Akteuren im Gesundheitswesen. Die Diskussion dieser verschiedenen Qualitätsaspekte erlaubt eine weitere Eingrenzung des Forschungsthemas von diesem Dissertationsvorhaben.

Im letzten Schritt wird kurz auf die bisherigen Entwicklungen sowohl der Qualitätsdiskussion als auch der Qualitätsmessung eingegangen. Hierbei wird neben einem Blick in die USA speziell ein Schwerpunkt auf die deutschen Entwicklungen gelegt.

2.1. Der Qualitätsbegriff – Definitionen, Dimensionen und Eingrenzungen

2.1.1. Diskussion verschiedener Qualitätsdefinitionen

„Qualität ist das Anständige“ (Heuss 1951, S. 80)

Der Qualitätsbegriff wird seit vielen Jahren in der Wissenschaft und Wirtschaft diskutiert. Spätestens seit Einführung des „Qualitätsmanagements“ vor ca. 50 Jahren wird versucht, Qualität klarer abzugrenzen und zu definieren (vgl. Geiger/Kotte 2008, S. V). Eine mögliche Definition lautet:

> „Die Beschaffenheit einer betrachteten Einheit (z. B. eines Produkts) sollte nach deren Realisierung die Forderungen an diese Einheit erfüllen.“ (Geiger/Kotte 2008, S. 3).

Leichte Abwandlungen dieser allgemeinen Definition werden international und in Deutschland angewendet. So führt z. B. das ISO-Institut die Geigersche Kurzdefinition in seiner Begriffsnorm. Bei Zollondz wird Qualität definiert als „Realisierte Beschaffenheit einer Einheit bezüglich Qualitätsanforderungen an diese“ (Zollondz 2002, S. 152).

Geschichtlich reicht der Qualitätsbegriff in der Medizin weit zurück: Bereits im 18. Jahrhundert v. Chr. wurden erste Grundlagen medizinischer Qualität im „Codex Hammurapi“ formuliert, so bspw. Strafen bei Qualitätsmangel, Falldarstellungen und eine Gebührenordnung (vgl. Schott 2005). Im Hippokratischen Eid und in der seit 1948 genutzten „Genfer Deklaration des Weltärztebundes / Genfer Gelöbnis“ dagegen wird Qualität nicht explizit erwähnt, der Beruf soll mit „Gewissenhaftigkeit und Würde“ ausgeübt werden (vgl. Riha 2010; Weltärztebund 1994).

Der Qualitätsbegriff in der Medizin scheint aus Sicht der verschiedenen Akteure zunächst klar und konsensfähig zu sein: Eine gute Qualität hat eine Leistung dann, wenn sie nach dem Stand der medizinischen Wissenschaft das bestmögliche Ergebnis für den Patienten erbringt. Dazu gehört auch, dass die Behandlung selbst den Patienten nicht unangemessen schädigt („Primum non nocere“ – zuerst einmal nicht schaden, Grundsatz aus der hippokratischen Tradition). Dieser Qualitätsbegriff liegt letztlich auch der Verbesserung der Behandlungsverfahren in der Medizin zugrunde. Seit es eine wissenschaftliche Medizin gibt, ist die Verbesserung der Patientenergebnisse das zentrale Anliegen. Probleme ergeben sich immer dann, wenn es um die Präzisierung und Operationalisierung dieser Definition geht. Was sind dann die *Messgrößen*, die dieses Ergebnis am besten charakterisieren?

Den allgemeinen Qualitätsbegriff in der Medizin im oben genannten Sinne hat das „Committee to Design a Strategy for Quality Review and Assurance in Medicare“ nach Analyse von über 100 verschiedenen Definitionen in Form der ‚Versorgungsqualität’[1] wie folgt definiert:

> “A degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (Lohr 1990, S. 21).

In dieser Definition ist ein Zielaspekt vorhanden: die ‘erwünschten Gesundheitsergebnisse’. Diese erwünschten Ergebnisse können dabei eine Vielzahl von Dimensionen einnehmen: Die Verbesserung der Gesundheit sowohl von Einzelpersonen als auch der Gesamtgesellschaft sind mögliche Aspekte. Gleichzeitig geht es um die Vermeidung von Risiken bzw. der Verschlechterung der Gesundheit.

Die ‚Versorgungsqualität’ beinhaltet jedoch auch den Aspekt der effektiven medizinischen Versorgung. Nicht nur das Ziel (‚desired health outcomes’) sondern auch der Weg dahin (‚health services’) sind

1 Die vorliegende Arbeit betrachtet Versorgungsaspekte im stationären Sektor. Daher ist ein Fokus auf Versorgungsqualität naheliegend.

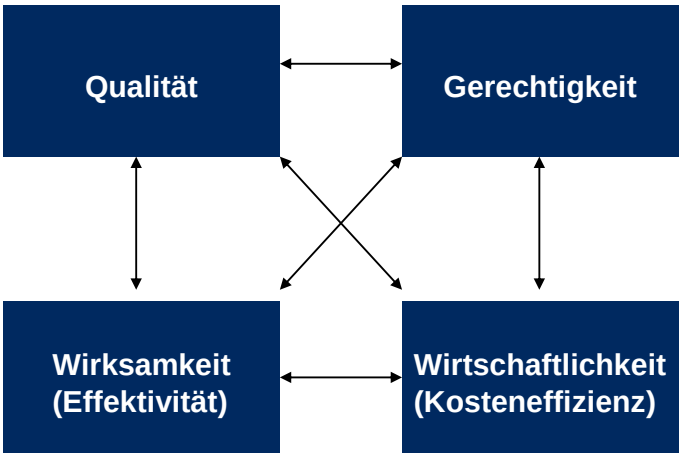
relevant. Insofern geht diese Definition über den rein medizinischen Qualitätsbegriff hinaus, schließt diesen aber ein.

Eine inadäquate Versorgung kann ebenfalls einen Qualitätsmangel darstellen, da sie entweder nicht zum gewünschten Ergebnis führt oder den Patienten unnützen Risiken aussetzt. Das Institute of Medicine beschreibt drei mögliche Probleme in der Versorgung (vgl. National Research Council 1997, S. 3):

1. Verwendung von unnötiger oder unpassender Versorgung
2. Nicht ausreichende Verwendung von benötigter, effektiver und passender Versorgung
3. Unzulänglichkeiten bei technischen und personellen Aspekten der Versorgung

Diese Problematik findet sich auch im Gutachten des Sachverständigenrats 2000/2001 Band III als Über-, Unter-, Fehlversorgung (vgl. Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen 2001). Dies zeigt, dass neben den medizinischen Qualitätsergebnissen auch die eingesetzten (finanziellen) Mittel bzw. Fehlallokationen von (begrenzten) Ressourcen in Betracht gezogen werden können. Bei Gesundheitsausgaben in Deutschland von 10,5 % des BIPs in 2008 (vgl. OECD 2011, S. 1) sind auch wirtschaftliche Aspekte entscheidend. In der vorliegenden Arbeit sollen jedoch diese Aspekte getrennt voneinander betrachtet werden. Auch in der Gesundheitsökonomie wird Qualität zwar getrennt von Effektivität und Effizienz diskutiert – die Korrelationen dieser Dimensionen werden aber im „magischen Viereck“ deutlich dargestellt. So spricht das „Magische Viereck der Gesundheitsökonomie“ neben den Elementen „Qualität“, „Gerechtigkeit“ und „Wirksamkeit (Effektivität)“ auch vom Aspekt der „Wirtschaftlichkeit (Effizienz)“:

Abbildung 3: Das magische Viereck der Gesundheitsökonomie



Quelle: Eigene Darstellung mit Anlehnung an Fedderwitz 2012

2.1.2. Aufteilung des Qualitätsbegriffs nach Donabedian: Struktur-, Prozess- und Ergebnisqualität

Die Weiterentwicklung des Qualitätsbegriffs wurde vom amerikanischen Forscher Avedis Donabedian vorangetrieben. Obschon Donabedian vorwiegend medizinische Belange erforschte (so z. B. in „The quality of medical care“ (vgl. Donabedian 1978), oder „Evaluating the quality of medical care“ (vgl. Donabedian 1966) oder „Methods for deriving criteria for assessing the quality of medical care“ (vgl. Donabedian 1980)), wurde seine Strukturierung in drei verschiedene Qualitätsdimension auch für andere Fachrichtungen relevant (vgl. Zollondz 2002, S. 147).

Die drei Qualitätsdimensionen von Donabedian sind (vgl. Donabedian 1988, S. 1745):

- Strukturqualität (appraisal of structure)
- Prozessqualität (assessment of process)
- Ergebnisqualität (assessment of outcome)

In Deutschland hat diese Aufteilung sogar in den SGB V Einzug erhalten: in § 137a Abs. 1 SGB V werden Mindestanforderungen an Struktur-, Prozess- und Ergebnisqualität erwähnt. Nach einer Erläuterung sollen hier die Vor- und Nachteile der Dimensionen diskutiert werden.

2.1.2.1. Strukturqualität nach Donabedian

Die *Strukturqualität* beschreibt die Rahmenbedingungen einer Organisation (bspw. Ausstattungen, Fähigkeiten oder technischen Ausrüstungen), die zur unternehmerischen Leistungserbringung notwendig sind. In der stationären Versorgung entspricht dies neben dem Vorhandensein spezifischer Fachabteilungen vor allem der Anzahl und fachlichen Qualifikationen des ärztlichen und nicht-ärztlichen Personals, sowie der räumlichen, infrastrukturellen und (medizin-) technischen Ausstattung. Besonders diese Qualitätsdimension wird vielfach als Voraussetzung für Vertragsbeziehungen genutzt, so z. B. in Abbildung 4 als Voraussetzung für einen Vertrag der integrierten Versorgung nach §140a SGB V:

Abbildung 4: Anwendung von Strukturqualität in einem Vertrag der integrierten Versorgung

Strukturqualität Krankenhaus

Für die stationäre Versorgung teilnehmender Versicherter aufgrund der Diagnose Asthma soll vorrangig in Krankenhäuser überwiesen werden, die die Inhalte der Anlage 9 Ziffer 1 bis 3 der Risikostruktur-Ausgleichsverordnung (RSAV) beachten und folgende Voraussetzungen erfüllen:

1. Fachliche Voraussetzungen

Mindestens eines der unten genannten Kriterien muss erfüllt werden:

- Zugelassene Weiterbildungsstätte Pneumologie mit mindestens 1 ständig erreichbaren Facharzt für Innere Medizin Schwerpunkt Pneumologie in Vollzeitbeschäftigung
- Abteilung für Innere Medizin mit mindestens 1 ständig erreichbaren Facharzt für Innere Medizin Schwerpunkt Pneumologie oder Facharzt für Innere Medizin mit der Berechtigung zum Führen der Teilgebietsbezeichnung „Lungen- und Bronchialheilkunde"

Für Kinder:

- Pädiatrische Abteilung mit mindestens einem ständig erreichbaren Facharzt für Kinder- und Jugendmedizin mit der Zusatzbezeichnung ‚Allergologie'

Das Krankenhaus hat darüber hinaus mindestens 2 intensivmedizinische Betten nach Krankenhausbedarfsplanung und die ständige Möglichkeit der konsiliarischen Einbeziehung eines Kardiologen und eines Radiologen vorzuweisen.

2. Organisatorische und apparative Voraussetzungen

Quelle: Kassenärztliche Vereinigung Schleswig-Holstein; Anhang 3

Die Elemente der Strukturqualität sind in der Regel leicht messbar, da konkrete Ausstattungsaspekte abgefragt werden können (vgl. Haeske-Seeberg 2001, S. 112). Gleichzeitig können die Kennzahlen durch externe Stellen (z. B. den Medizinischen Dienst der Krankenversicherungen MDK) überprüft werden – allerdings mit hohem Aufwand und nicht automatisiert. Eine hohe Verständlichkeit der Parameter für Patienten und für die Öffentlichkeit ermöglicht ebenfalls eine gute Anwendung der Strukturqualität – auch in der Außenvermarktung.

Allerdings muss hinterfragt werden, welche Korrelationen oder Kausalität zwischen der Strukturqualität und dem medizinischen Behandlungsergebnis besteht. So gibt es sicherlich einige Grundvoraussetzungen der Strukturqualität, die „notwendig" sind, um eine gute Behandlungsqualität zu gewährleisten (so z. B. das Vorhandensein medizinischer Geräte). Die Strukturqualität alleine ist in der Praxis jedoch noch keine hinreichende Determinante einer gleich hohen Ergebnisqualität, zumal auch die strukturellen Anforderungen in der Praxis nicht beliebig hoch gesetzt werden können. Auch bei im Rahmen der praktischen Möglichkeiten gleichen Strukturen kann es unterschiedliche Ergebnisse geben. Dies ist eine der Erkenntnisse aus der Studie „To Err is Human", die Ergebnisunterschiede in der klinischen Versorgung unter den in den USA vorgegebenen, für alle Kliniken gleichen strukturellen Anforderungen nachweist (vgl. Kohn et al. 1999). Diese Beziehungen werden in Abschnitt 2.1.2.4 näher diskutiert.

2.1.2.2. Prozessqualität nach Donabedian

Die *Prozessqualität* beschreibt die Handlungsabläufe und Aktivitäten der Organisation. Im stationären Kontext bedeutet dies zum Beispiel die Behandlungsabläufe auf einer Station oder gewisser administrativer Aktivitäten. Die Sicherung der Prozessqualität findet dabei häufig durch die Einhaltung von Leitlinien und Standards statt.

Vorteile der Prozessqualität liegen vor allem in der hohen Akzeptanz der Elemente für das medizinische Fachpersonal. Leitlinien bspw. – die von anerkannten Fachgesellschaften formuliert wurden – haben oft eine hohe Bedeutung für die alltägliche Arbeit im Krankenhaus und werden für die Ausbildung von Ärzten angewandt. Zusätzlich kann aufgrund einer Vielzahl von wissenschaftlichen Arbeiten von einer höheren Korrelation zwischen Prozess- und Ergebnisqualität ausgegangen werden, als dies bei der Strukturqualität der Fall ist (siehe Abschnitt 2.1.2.4).

Allerdings ist die Dokumentation und Messung der Prozessqualität zumeist aufwendiger als bei der Strukturqualität (vgl. Krumholz et al. 2007, S. 77f.): Die Abläufe und Prozesse müssen häufig von den Akteuren im Krankenhaus selbst dokumentiert werden. Somit ist auch eine externe Überprüfung schwierig, da bspw. eine leitliniengerechte Behandlung vielfach nur anhand von „Eigen-Dokumentation“ nachweisbar ist, im Gegensatz zum Vorhandensein eines medizinischen Geräts in der Strukturqualität.

Außerdem sind viele Kriterien einer guten Prozessqualität für den Patienten kaum verständlich – während eine Patientenaufklärung als gute Qualität gewertet werden kann, sind Aspekte wie „Vorhofsonden mit Amplitude ≥ 1,5 mV“ aus dem AQUA-Qualitätsbericht für Laien kaum einzuordnen (vgl. Haeske-Seeberg 2001, S. 113; AQUA-Institut GmbH 2011a, S. 36). Auch andere Aspekte wie bspw. die Validität der Prozessqualität werden diskutiert (vgl. unter anderem McAuliffe 1978, S. 907; Lowrie et al. 2001, S. 1917; Mant 2001, S. 475; Rubin 2001, S. 469; Peterson et al. 2006, S. 1912; Wang et al. 2009, S. 1411). Zusätzlich werden bei Prozessindikatoren eine komplizierte Aktualisierung auf neue Standards sowie eine mögliche Fokussierung im klinischen Alltag nur auf die gemessenen Prozesse als Nachteile erwähnt (vgl Krumholz et al. 2007, S. 76ff.).

2.1.2.3. Ergebnisqualität nach Donabedian

Die *Ergebnisqualität* entspricht der Differenz aus dem Eingangszustand und dem Ausgangszustand (vgl. Zollondz 2002, S. 147). Hiermit ist das eigentliche Behandlungsergebnis oder „Outcome“ gemeint. Im stationären Kontext bedeutet dies vor allem die gesundheitliche Verbesserung der Patienten und stellt somit den direktesten Indikator für Qualität einer Versorgungsleistung im eingangs genannten Sinne dar (vgl. Pira 2000, S. 79). Beispiele hierfür sind verschiedene Mortalitätsraten für schwerwiegende Krankheiten und Eingriffe. Ergebnisqualität wird manchmal auch mit den „5 Ds“ beschrieben: „death, disease, discomfort, disability and dissatisfaction“ (Mainz 2003, S. 526), wobei die ersten beiden „D“s – d. h. Mortalität und Morbidität – im Vordergrund stehen.

Während die Prozess- und Strukturqualität veränderbare Dimensionen darstellen, die die Qualität beeinflussen, setzt die Ergebnisqualität auf das Outcome, das am Ende der Aktivitätskette steht. Somit ist sie auch das zentrale Qualitätselement, das als Ziel einer „Qualitätsverbesserung“ tatsächlich verbessert werden soll. Da es kaum Möglichkeiten gibt, durch Qualitätsmaßnahmen die Ergebnisqualität direkt zu verbessern, finden aktive Veränderungen meist entlang von Prozessen und Strukturen statt, die wiederum die Ergebnisqualität nachhaltig verbessern.

Der Vorteil der Ergebnisqualität ist, dass diese Qualität das Endergebnis verschiedener Aktivitäten darstellt. Auch die zumeist hohe Akzeptanz solcher Indikatoren seitens des medizinischen Personals sowie der Patienten und der interessierten Öffentlichkeit sind zu erwähnen (vgl. Krumholz et al. 2007, S. 76ff.).

Auch und gerade bei Ergebnisqualitätsindikatoren muss gefragt werden, ob jeweils „Das Richtige“ gemessen wird. So ist sicherlich die mittel- und langfristige Sterblichkeitsrate (z. B. 5-Jahres-Überleben) bei onkologischen Erkrankungen ein wichtiges und relevantes Ergebnis für den Patienten, die direkte Beeinflussbarkeit durch die behandelnden Ärzte ist jedoch fraglich, da im Einzelnen Faktoren wie Schweregrad der Erkrankung oder Patientenhistorie mitentscheiden – eine Risikoadjustierung ist daher notwendig (vgl. u. a. Mansky 2008, S. 162; und Kriterium 5c). Gleichzeitig muss Ergebnisqualität verständlich definiert sein, um eine hohe Akzeptanz seitens der Patienten zu erreichen. Daher werden an die Messung von Ergebnisqualität hohe Anforderungen gestellt, die im nächsten Kapitel diskutiert werden.

In späteren Kapiteln wird diskutiert, dass die direkte „Leistungsorientierung“ eine der wichtigsten Anforderungen an Anreizsysteme ist (vgl. Becker 1990). Hiermit ist stets ein „Ergebnis“ (meist ein Betriebsergebnis) gemeint. Diese Leistungsorientierung entspricht in der Medizin am ehesten der Ergebnisqualität und nicht eventuellen Surrogatparametern. Erstere soll daher hier in den Mittelpunkt gerückt werden. Diese Relevanz des „outcome-measurements“ ist unter anderem auch bei Porter zu finden (vgl. Porter/Guth 2012). Gerade die Möglichkeit (finanzielle) Anreize direkt an das Behandlungsergebnis zu koppeln, kann – gemäß der Motivationstheorie – eine direkte Verbesserung erzeugen. Daher ist speziell die Ergebnisqualität für die Anwendung in P4P-Systemen geeignet (vgl. u. a. auch Veit et al. 2012e, S. 10ff.[1]).

2.1.2.4. Beziehungen zwischen Struktur-, Prozess- und Ergebnisqualität

Struktur-, Prozess- und Ergebnisqualität können einerseits als Dimensionen und andererseits auch als Phasen der „Dienstleistungserstellung“ verstanden werden:

[1] Sowohl die generelle Eignung als auch die notwendigen Anforderungen von der Nutzung von Ergebnisqualitätsparamtern in P4P werden diskutiert.

(1) sobald eine spezifische *Struktur* geschaffen ist,

(2) können *Prozesse* durchgeführt werden,

(3) die zu den gewünschten *Ergebnissen* führen können .

Die Aufteilung nach Prozessschritten wird nicht erst seit den Arbeiten von Donabedian diskutiert. Bereits Anfang der 90er diskutierte Joseph Moses Juran in seiner „Straßenkarte für Qualitätsplanung“ die Elemente „Eingabe“, „Aktivität“ und „Ausgabe“ (vgl. Juran 1991, S. 24f.). Hier ist ein ähnlicher Ansatz wie bei Donabedian erkennbar.

Doch welche Beziehungen bestehen genau zwischen Strukturen, Abläufen und Ergebnissen in der Medizin? Nachweisbare Korrelationen sind vor allem zwischen Prozess- und Ergebnisqualität feststellbar. So können Studien auch empirisch einen Zusammenhang zwischen Prozessindikatoren und dem Behandlungsergebnis sowie einige Vor- und Nachteile von den Indikatorenarten zeigen (so z. B. McAuliffe 1978; Rubin 2001; Mant 2001 – vor allem die Beziehung von Leitlinien zu Ergebnisparametern ist meistens relevant: Grimshaw/Russell 1993; Kopp 2011). Andere Studien zeigen jedoch, dass einige Prozessindikatoren kaum mit der Ergebnisqualität (wie Mortalität) korrelieren; insbesondere Einzelindikatoren – im Gegensatz zu zusammengesetzten „composite adherence scores“ – können häufig nur einen geringen Anteil des Behandlungsergebnisses erklären (vgl. u. a. Lowrie et al. 2001; Bradley et al. 2006; Peterson et al. 2006; Wang et al. 2009).

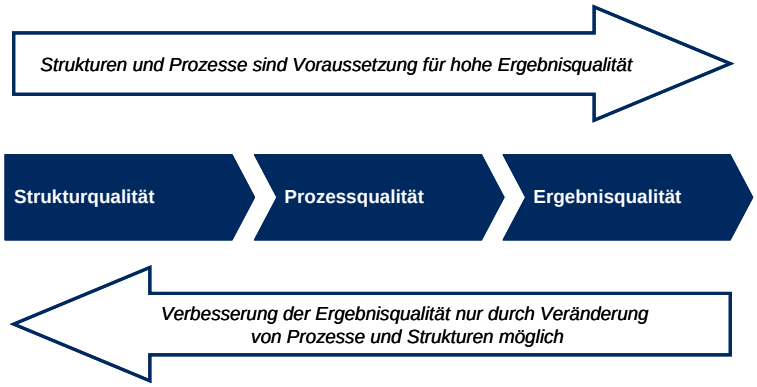
Von einigen Strukturindikatoren wurde ebenfalls eine hohe Korrelation zu Ergebnisparametern erwartet. Dies gilt vor allem für die Frage der Mindestmenge von Fallzahlen. Bei verschiedenen schweren Krankheitsbildern korrelieren die Fallzahlen invers mit der Sterblichkeitsrate (vgl. Luft et al. 1979; Birkmeyer 2000; Dudley et al. 2000;). Dieser Effekt kann entweder mit der Hypothese „practice makes perfect“ oder mit „selektivem Einweisen“ – d. h. der Auswahl des Krankenhauses nach der besseren Ergebnisqualität – erklärt werden (vgl. Luft et al. 1979). Auf jeden Fall ist der Zusammenhang, wenn gegeben, nur statistischer Art. Im Einzelfall heißt dies, dass auch bei nachweisbarem Zusammenhang ein Haus mit niedriger Fallzahl gut und eines mit hoher Fallzahl schlecht sein kann. Dies beeinträchtigt die Justiziabilität dieser Kennzahl. Darüber hinaus wird in den letzten Jahren der positive Effekt von Mindestmengen und Fallzahlvolumen auf die Ergebnisqualität immer öfter angezweifelt (so z. B. bei Shahian/Normand 2003; Khuri/Henderson 2005 und Geraedts/de Cruppe 2006). Aus diesen Gründen wurde in Deutschland 2011 beispielsweise die Mindestmengenregelung für den operativen Einsatz von Kniegelenk-Totalendoprothesen nach ablehnenden Gerichtsentscheidungen zunächst ausgesetzt (vgl. u. a. Flintrop 2012).

Doch gibt es auch eine „rückwärtige“ Beziehung von der Ergebnisqualität hin zur Prozess- und Strukturqualität? Zum einen kann hier die „Zuweiserhypothese“ erwähnt werden: Dabei wird davon ausgegangen, dass bei einer guten Ergebnisqualität die Patientenflüsse zu diesen „guten“

Krankenhäusern gelenkt werden (von zuweisenden Ärzten). Auf diese Weise steigt das Fallvolumen, wodurch mittelbar die Struktur- und Prozessqualität z. B. durch bessere Standardisierung der Abläufe und spezifische Investitionen steigen kann (vgl. Birkmeyer 2000 oder „selektive Einweisung“ bei Luft et al. 1979). Gemäß den oben diskutierten Mindestmengen-Mechanismen kann dann auch die Ergebnisqualität weiter steigen.

Doch eine weitere Relation von Ergebnis- hin zu Prozess- und Strukturqualität ist denkbar: Wenn die Ergebnisqualität der Behandlung verbessert werden soll, ist es für Praktiker oft entscheidend, die Struktur- und Prozessqualität aktiv zu verbessern. Durch Leitlinien bspw. ist die Verbesserung von Strukturen und Prozessen durch die beteiligten Akteure (d. h Ärzte und medizinisches Personal) umsetzbar. Besonders durch bspw. Peer-Reviews und PDCA-Zyklen können Qualitätsstandards für Ergebnisqualität in die Strukturen und Prozesse eines Krankenhauses implementiert werden (vgl. Krahwinkel et al. 2011; Rink 2011 und Helios Kliniken Gruppe 2005, S. 69ff. und Abschnitt 2.2.3.3). Somit kann die Ergebnisqualität operationalisiert werden, indem gewisse Standards für Struktur- und Prozessqualität definiert oder besser umgesetzt werden (siehe graphische Darstellung der Zusammenhänge in Abbildung 4).

Abbildung 5: Zusammenhänge zwischen Struktur-, Prozess- und Ergebnisqualität



Quelle: eigene Darstellung mit Anlehnung an Donabedian 1978

Mit welcher Dimension sollen sich nun Pay-for-Performance-Programme vorrangig beschäftigen? Ein System, das eine dauerhafte und kontinuierliche Verbesserung der Qualität ermöglichen soll, sollte vor allem Ergebnisqualität beinhalten. Nur so kann sichergestellt werden, dass das eigentliche Outcome der Behandlung, also die gesundheitliche Verbesserung, im Fokus steht. Aus ärztlicher Sicht kann es aber hilfreich sein, auch die Struktur- und Prozessqualität zu betrachten. Zumindest muss das jeweilige System ermöglichen, auf der Handlungsebene (d. h. bei Strukturen und Abläufen) Veränderungen so einzuführen, dass es zu einer Verbesserung der Ergebnisqualität führt.

2.1.3. Weitere Dimensionen und verschiedene Perspektiven der Qualität als Möglichkeit einer breiteren Definition

Neben den Dimensionen von Donabedian gibt es weitere Möglichkeiten, Qualität zu definieren. Eine für diese Arbeit aufschlussreiche Unterscheidung ist das partialanalytische Qualitätsverständnis von Garvin (vgl. Garvin 1984; Zollondz 2002, S. 145f.). Garvin beschreibt darin den Begriff der Qualität aus verschiedenen Perspektiven, die auf den medizinischen Kontext übertragbar sind. Für diese Arbeit sind folgende Qualitätsarten relevant:

Transcendent Quality (Absolute Qualität): Beschreibung der „Güte", also der bestmöglichen Ausprägung (vgl. Zollondz 2002, S. 145f.):

Hier ist vor allem die möglichst gute Zweckerfüllung gemeint (Industriebeispiel: CMA-Gütesiegel oder Hotelkategorien mit unterschiedlichen Qualitätsansprüchen).

Einen „Goldstandard" für die Definition einer Zielerreichung ist in der Medizin nicht allgemein akzeptiert. Nur im Rahmen von Leitlinien gibt es Ansätze, die einem Standard gleichkommen. Doch gerade dieses Verständnis von Qualität kann für die Nutzung von P4P-Systemen sehr hilfreich sein. Bei der Umsetzung von Qualitätsverbesserungsmaßnahmen ist eine Standardisierung von Abläufen und Infrastruktur sinnvoll. Doch auch ein „Goldstandard" für die Ergebnisqualität ist bedeutsam. Wie in den folgenden Kapiteln weiter differenziert wird, kann die Definition von konkreten Qualitätszielen und somit „Mindeststandards" oder „Goldstandards" ein Anreizsystem operationalisieren. Die Betrachtung von „absoluter" Qualität ist somit insbesondere bei Fragestellungen von definierten „Qualitätszielen" relevant.

User-based Quality (Qualität für den Kunden): Kundensubjektive Bewertung der Leistung (vgl. Zollondz 2002, S. 145f.):

Die kundenbasierte Qualität betrachtet die Kundenperspektive des Produktes oder der Dienstleistung. Dabei werden die nachgefragten und wahrgenommenen Kundenbedürfnisse bewertet und mit der eigentlichen Beschaffenheit verglichen. Auch hier ist demzufolge eine Zielerreichung (mit sehr spezifischen Ziel) von Bedeutung (vgl. Zollondz 2002, S. 146).

Diese Dimension der Qualität ist auch für den medizinischen Sektor sehr relevant. Patientenorientierung wird zunehmend wichtiger: so z. B. auf internetbasierten „Klinikführern", bei denen neben der Behandlungsqualität vermehrt auch Elemente der „Patientenzufriedenheit" Einzug gehalten haben. In verschiedenen Publikationen wird ebenfalls zunehmend die Patientenperspektive diskutiert – so beispielsweise in „Der patientenzentrierte Qualitätsbegriff und seine Implikationen" von Klemperer (vgl. Klemperer 1995, S. 189ff.). Hierbei wird vorwiegend zwischen medizinischer Leistung, pflegerischer Leistung und sogenannten Hotelleistungen unterschieden. Neuere Umfragen zeigen jedoch, dass die medizinische Qualität auch bei der Beurteilung durch Patienten immer mehr in den Vordergrund rückt.

Eine Umfrage der „Weißen Liste“ der Bertelsmann-Stiftung zeigt, dass Patienten sowohl bei leichten als auch schweren Erkrankungen als wichtigstes Auswahlkriterium eines Krankenhauses die „medizinischen Behandlungsergebnisse“ angeben (28 % bzw. 35 %) (vgl. Mansky 2013, S. 138ff.).

Eine zusätzliche Berücksichtigung der Dimension patientenbasierte Qualität kann im Rahmen eines P4P-Systems den Patientenfokus erhöhen und sicherstellen, dass Qualitätsverbesserungen auch und vor allem zugunsten der Patienten entwickelt werden. Daher ist diese Einteilung für diese Arbeit von Bedeutung.

Value-based Quality (Qualität als Wert): Wertorientierter Ansatz mit Preis-Leistungs-Relation (vgl. Zollondz 2002, S. 145f.):

Die fünfte Dimension nach Garvin ist besonders interessant. Hier wird deutlich, dass das Qualitätsverständnis explizit sehr breit ist und über die medizinischen Definitionen hinausgeht. Neben der „Qualitätsgröße“ wird nämlich zusätzlich der Kostenaspekt (sowohl monetär als auch Zeitaufwand) mit einbezogen. Garvin betrachtet diesen Aspekt vor allem aus der Kundenperspektive (d. h. wie der Preis des Gutes im Vergleich zur Qualität wahrgenommen wird). Qualität ist damit eine relative Größe und kann als „Wert“ eines Gutes oder einer Dienstleistung bezeichnet werden.

Dieser Aspekt ist für den medizinischen Kontext von hoher Relevanz – insbesondere die Arbeiten von Michael Porter fokussieren auf den „value“ von Behandlungen, der definiert ist als (vgl. Porter/Teisberg 2006, S. 4):

$$\frac{\text{Behandlungsergebnis (Outcome)}}{\text{Kosten zur Erreichung dieses Ergebnisses}}$$

Während dieses Verständnis in anderen Industrien als „Preisvorstellung“ des Konsumenten verbreitet ist, ist ein derartiger wertorientierter Ansatz im Gesundheitssektor neu. Im Grunde genommen handelt es sich hier um die Sicht des Einkäufers der Leistung und nicht des Patienten. Der Einkäufer (eine Krankenversicherung bzw. in den USA mittelbar der Betrieb, der diese Versicherung abschließt und bezahlt) ist besonders an diesem Preis-/Leistungsverhältnis interessiert. Für den Patienten steht aber in der Regel das Behandlungsergebnis selbst (d. h. hier nur der Zähler) im Vordergrund. Da die medizinischen Leistungen zwar von den Patienten in Anspruch genommen werden, aber nicht von diesen sondern von den Krankenversicherungen bezahlt werden, entsteht dabei das Risiko des „Moral-Hazard“ – d. h. der Abkoppelung von Leistungsinanspruchnahme und Zahlungsbereitschaft. (vgl. auch die Beschreibungen des „Principal-Agent-Prinzips“ in Abschnitt 4.1.1.2.2).

Daher ist ein „Preis-Leistung-Gedanke“ in der Medizin zunächst nicht vorrangig – zumindest nicht beim Endkonsumenten (d. h. dem Patienten). Die Krankenversicherer dagegen beschäftigen sich selbstverständlich auf sehr aktive und detaillierte Weise mit diesem Thema. Somit ist auch für die

Krankenhäuser ein Preis-Leistungsverhältnis von Bedeutung – meist jedoch fokussiert auf den Kosten-Mengen-Aspekt.

Insgesamt zeigt sich, dass die garvinsche Definition von Qualität breiter ist als z. B. die Dimensionen von Donabedian. Dies wird vor allem bei drei Aspekten deutlich:

- *Aspekt des „Goldstandards“:* Während die Zweckerfüllung auch bei der Definition vom Institute of Medicine (IOM) bereits erwähnt wird, findet sie im landläufigen Verständnis von Qualität bislang wenig Anklang. Dieses Element ermöglicht aber durch die Definition von „Qualitätszielen“ eine nachvollziehbare, rasche Operationalisierung des Qualitätsbegriffs in die tägliche Arbeit von Ärzten und medizinischem Personal.
- *Kosten-Nutzen-Verhältnis*: Wie bereits beschrieben, ist es von Bedeutung, diesen Aspekt bereits in der Definition von Qualität zu betrachten. So kann die notwendigerweise enge Verbindung von Qualität der Leistungserbringung und deren Kosten in den Fokus gerückt werden. Dies ermöglicht dann eine sehr zentrale Betrachtung der Ziele und Vergütungsfunktion von P4P-Systemen.
- *Verschiedene Perspektiven:* Durch die Aufteilung nach Garvin wird ersichtlich, dass der Begriff Qualität von verschiedenen Akteuren der medizinischen Gesundheitsversorgung unterschiedlich verstanden werden kann. Diese Perspektiven werden explizit (z. B. in der kundenorientierten Qualität) oder implizit (z. B. „wertigen“ Qualität) erwähnt.

2.1.4. Eingrenzung der an stationärer Qualität beteiligten Akteure

Das Gesundheitswesen in Deutschland ist geprägt von verschiedenen Akteuren, die trotz ihrer engen Zusammenarbeit unterschiedliche Perspektiven auf Qualität haben. Die für diese Arbeit relevanten Akteure sind (vgl. auch Henke 2011, S. 43ff.[1]):

- *Leistungserbringer*: Im Gesundheitswesen wird die medizinische Dienstleistung von einer Vielzahl von Anbietern erbracht. Da in den unterschiedlichen Sektoren (stationäre, ambulant, Rehabilitation, etc.) sehr verschiedene Fragestellungen relevant sind und die Akteure sich in den Organisationsformen und Anforderungen deutlich unterscheiden, kann sich diese Arbeit nur auf eine Art fokussieren: die akut-stationären Leistungserbringer. Besonders durch das DRG-System sind die stationären Vergütungssysteme einheitlich und die Dokumentationen der Leistungen ausführlich. Außerdem ermöglichen die größeren Unternehmen im stationären Bereich die Anwendung von organisationstheoretischen Modellen für Anreizsysteme. Zusätzlich wird Krankenhäusern eine

[1] Der Akteur der „Industrie“ wie bei Henke 2011, S. 43 wird hier aufgrund der geringeren Relevanz nicht betrachtet.

qualitativ wie quantitativ besondere Rolle im Gesundheitswesen zugesprochen (vgl. Henke/Göpffarth 2010, S. 35).

- Gerade für den Leistungserbringer ist die Versorgungs- und Behandlungsqualität ein relevantes Thema, das auch in den betrieblichen Alltag Einzug erhalten hat. Denn genau hier wird die Qualität beeinflusst und kann verbessert werden. Dadurch wird Qualität zu einer Managementaufgabe auf allen Ebenen des Krankenhauses. Im Rahmen des hohen Kostendrucks scheint manchmal jedoch ein Spannungsfeld zwischen Qualität und Effizienz zu entstehen, das nach Expertenmeinung allerdings nicht besteht – bessere Qualität senkt die Kosten bspw. durch Vermeidung von Komplikationen und stationären Wiederaufnahmen (vgl. u. a. Helios Kliniken Gruppe 2005, S. 69ff.).

- *Krankenversicherungen/Kostenträger*: Die „Zahler“ im deutschen Gesundheitssystem haben einen Versicherungsvertrag mit den Bürgern (in ihrer Rolle als Versicherte) und einen Versorgungs- oder Vergütungsvertrag mit den Leistungserbringern und vergüten die erbrachten Dienstleistungen gemäß den gesetzlichen Bestimmungen (vgl. Busse et al. 2010, S. 2–3). Aufgrund des hohen Anteils von Versicherten in gesetzlichen Krankenversicherungen (GKV) werden in der vorliegenden Arbeit vor allem diese Krankenkassen betrachtet. Um trotzdem eine Auswahl an Perspektiven zu haben, werden mehrere Kassenarten und -größen einbezogen.

- Für Kassen ist neben der Qualität auch das Wirtschaftlichkeitsgebot nach §12 Abs. 1 SGB V relevant. Dies muss aber wiederum keinen Widerspruch zu einer Qualitätsorientierung bedeuten. Speziell im Wettbewerb um Versicherte ist die Versorgungsqualität ein wichtiges Thema. Ob und inwieweit Qualität aber auch außerhalb eines „Marketing-Effekts“ verstanden wird, soll gerade im Kapitel 5 diskutiert werden.

- *Patienten/Versicherte*: Die Patienten oder Versicherten stellen im Gesundheitssystem den „Endkonsumenten“ dar. Sie interagieren sowohl mit der Kasse als auch mit dem Leistungserbringer und können als Zentralfiguren des Gesundheitsmarktes verstanden werden. Diese Gruppe erfährt für diese Arbeit keine Einschränkungen, für die operationelle Bewertung bspw. der Kriterien werden auch Patientenvertreter (z. B. Patientenverbände) vorgeschlagen.

  Aus Patientensicht ist die Behandlungsqualität natürlich von höchster Bedeutung. Neuere Patientenbefragungen zeugen vom großen Interesse bezüglich der medizinischen Behandlungsqualität als Kriterium bei der Krankenhauswahl. Außerdem befürworten 36 % der Befragten eine Form von qualitätsorientierter Vergütung, von diesen wünschen sich 65 % einen sehr hohen variablen, qualitätsorientierten Anteil; 37 % erwarten, dass Krankenhäuser für vermeidbare Komplikationen keine Vergütung erhalten sollen (vgl. Mansky 2013, S. 153ff.).

- *Regulierung:* Mit diesem Akteur sind die verschiedensten Entscheidungsträger gemeint, die bei der politischen und gesellschaftlichen Entwicklung des Gesundheitswesens Einfluss haben. Dies sind

neben dem Gesundheitsministerium auch der gemeinsame Bundesausschuss und die verschiedenen Einrichtungen der Selbstverwaltung (vgl. Raspe/Wissing 2010, S. 26).

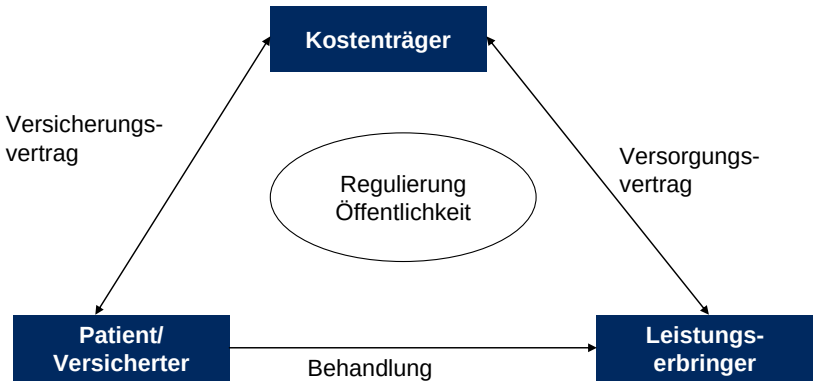
Für die Regulierung ist die medizinische Versorgungsqualität von hoher Bedeutung. Gerade für politische Entscheidungen sind möglichst faktenbasierte, medizinische Ergebnisse zu verschiedenen Optionen relevant. Gleichzeitig können jedoch auch Erkenntnisse genutzt werden, die eine mögliche Steuerung beim Leistungserbringer hervorbringen (so bspw. Aussagen zu Struktur- und Prozessqualität, wie bspw. Mindestmengen), um durch die Gesetzgebung Qualitätsstandards zu setzen, die die Versorgungsqualität langfristig verbessern.

- *Öffentlichkeit/Medien*: Der letzte Akteur für diese Arbeit wird aus verschiedenen Einzelakteuren zusammengefasst: Regulierung, interessierte Öffentlichkeit, Medien. Diese bilden zusammen einen Akteur mit Interesse an hoher Versorgungsqualität, der aber nur indirekt an der Leistungserbringung beteiligt ist.

  Aus Sicht der Öffentlichkeit sind vor allem Aspekte der „allgemeinen“ Versorgungsqualität entscheidend. Neben einer kontinuierlichen Verbesserung der „Volksgesundheit“ wird aber anhand der Berichterstattung auch deutlich, dass vor allem Themen zur Patientensicherheit (u. a. Sentinelereignisse) für die Gesellschaft von hohem Interesse sind. Allerdings entsteht durch die mediale Berichtserstattung manchmal Qualitätswahrnehmungen, die bei detaillierter Betrachtung kaum nachvollzogen werden können (vgl. „Exkurs: Wahrnehmungsverzerrung in der Qualitätsdiskussion in Deutschland?“).

Abbildung 6 zeigt die verschiedenen Akteure im Gesundheitswesen und ihre Beziehungen.

Abbildung 6: Akteure im Gesundheitswesen und Perspektiven der Qualität



Quelle: Eigene Darstellung in Anlehnung an Busse et al. 2010, S. 2

2.1.5. Entwicklung einer „Working Definition“

In den vorherigen Abschnitten wurden unterschiedliche Definitionen, Dimensionen und Akteure der Qualität diskutiert. Es wurden bereits einige Einschränkungen für die Arbeitsdefinition erwähnt. Im Kontext dieser Arbeit wird von „stationärer Versorgungsqualität“ gesprochen, die folgende Aspekte beinhaltet:

- Qualität als Struktur-, Prozess- und Ergebnisqualität nach Donabedian mit Fokus auf der Ergebnisqualität.
- Zusätzlich Qualität als „absolute“, „patienten-orientierte“ und „wert-orientierte“ Qualität nach Garvin, um Qualitätsziele sowie Wert-Aspekte und die Patientensicht zu integrieren.
- Fokus auf den stationären Sektor sowie gesetzliche Krankenkassen.

Diese Eingrenzungen sind vor allem für die Katalogentwicklung sowie für die Anwendung des Katalogs auf deutsche Beispiele entscheidend.

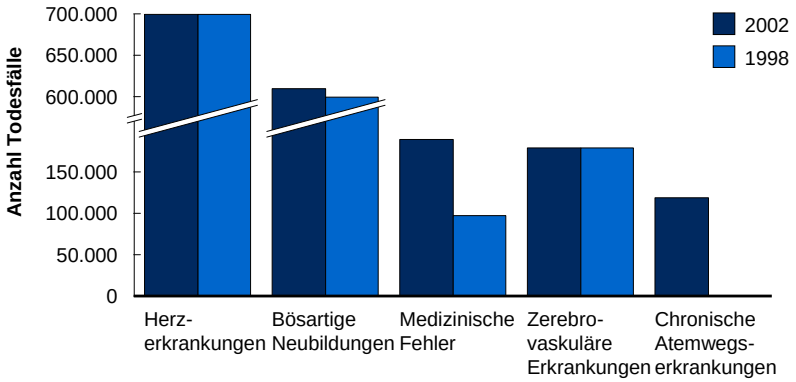
2.2. Bedeutung von Qualität im stationären Sektor

2.2.1. Qualitätsdiskussion im stationären Sektor: Entwicklung und derzeitiger Stand

Ein wichtiges Thema im 20. Jahrhundert war die evidenzbasierte Medizin, wie im wegweisenden Buch „Effectiveness and Efficiency: Random Reflections on Health Services“ dargestellt. Von Archie Cochrane in England seit 1972 vorangetrieben, wurde die Qualitätsdiskussion zum Teil durch diese Entwicklung geprägt (vgl. Cochrane 1972a; Cochrane 1972b). Evidenzbasierte Medizin ist „the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients“ (Sackett et al. 1996, S. 71) und ermöglicht dem praktizierenden Arzt, Ergebnisse der klinischen und patientenorientierten Forschung in die Therapieentscheidungen einzubeziehen und somit der Versorgung zu verbessern.

Zu Beginn des 21. Jahrhunderts kamen einige der wegweisenden Beiträge zur Qualitätsdiskussion aus den USA. So können zwei Berichte des IOMs als Meilensteine der Qualitätsdiskussion bezeichnet werden: „To Err is Human“ und „Crossing the Quality Chasm“ (Kohn et al. 1999; Institute of Medicine 2001). Der erste Bericht im Jahr 2000 konnte aufzeigen, dass jedes Jahr 98.000 Todesfälle in den USA auf medizinische Fehler („medical errors“) zurückzuführen sind. Dies führt zu zusätzlichen Kosten im Gesundheitssystem von 37-50 Mrd. Dollar pro Jahr. Diese Fehler sind somit unter den 10 häufigsten Todesursachen in den USA zu finden (siehe Abbildung 7).

Abbildung 7: Häufigste Todesursachen in den USA, 1998 und 2002



Quelle: Eigene Darstellung mit Daten von Porter/Teisberg 2006, S. 24-25

Die Feststellung der absoluten Höhe der Todesfälle in Krankenhäusern löste eine Diskussion in Fach- und öffentlichen Kreisen aus. In beiden Berichten entwickelt das IOM Empfehlungen, die neben neuen Dokumentationssystematiken vor allem auch eine qualitätsbezogene Vergütung vorschlagen (vgl. Institute of Medicine 2001).

Die Anknüpfung von leistungsorientierten Aspekten an die Vergütung wurde auch im nächsten Meilenstein der Qualitätsdiskussion vorgeschlagen: Porters „Redefining Health Care“ (vgl. Porter/Teisberg 2006; Porter/Guth 2012). Porter entwickelt in diesem Werk seine Vorstellung von „value-based competition/purchasing“ anhand derer eine Verbesserung sowohl der Qualitäts- als auch der Kostenbasis möglich sein soll. Diese Werke haben die Diskussion der Pay-for-Performance-Modelle in den USA intensiviert.

Auch in Deutschland gibt es vermehrt Diskussionen über die qualitätsorientierte Vergütung. Spätestens seit dem Bericht des Sachverständigenrats 2007 „Kooperation und Verantwortung“ ist auch hierzulande eine leistungsorientierte Vergütung im Rahmen von P4P-Systemen im Gespräch (vgl. SVR 2007). Die Umsetzungsmodalitäten werden ebenfalls diskutiert (vgl. Amelung/Zahn 2009; Schrappe/Gültekin 2010; Scheppach et al. 2011), und es wurden Vorschläge für den deutschen Markt erarbeitet (vgl. Mansky 2010a; Mansky/Seeger 2011).

Da für eine qualitätsorientierte Vergütung zunächst die Qualität gemessen werden muss, soll im Folgenden die Entwicklung von Qualitätsmessung diskutiert werden.

2.2.2. Qualitätsmessung im stationären Sektor: historische Entwicklung und Darstellung verschiedener Indikatorensysteme

2.2.2.1. Historische Entwicklung der Qualitätsmessung

„… quality can indeed be legitimately quantified, measured, traced, and trended – just like many other business variables and activities" (McIntyre et al. 2001, S. 8)

An die Frage „Was ist Qualität?" schließt sich nahtlos die Frage „Wie kann Qualität objektiv gemessen werden?" an. Bereits seit dem Codex Hammurapi, 1750 v. Chr. wird medizinische Qualität gemessen – und sogar in das Vergütungskonzept einbezogen (vgl. Schott 2005; Leber 2009). Hier wurde anhand von sehr einfachen Indikatoren medizinische Ergebnisqualität gemessen – meist der binäre Erfolg oder Misserfolg einer Behandlung.

Eine der ersten übergreifenden Qualitätsmessungen aus Krankenhaussicht wurde von Florence Nightingale angestoßen. Sie entwickelte 1859 die weltweit ersten Leistungstabellen von Krankenhäusern. Außerdem zeigte sie, dass hohe Sterblichkeitsraten – vor allem in großen Kliniken – vermeidbar waren (vgl. Nightingale 1863; Kelsey 2001; Maynard 2012). Diese Errungenschaften für die Messung der Krankenhausqualität und statistische Auswertungen wurden schon von Zeitgenossen geehrt: 1858 bereits wurde Nightingale in die Royal Statistical Society aufgenommen, unter anderem dank der statistischen Analysen in der Epidemiologie und im öffentlichen Gesundheitswesen (vgl. McDonald 1998).

Ende des 19. Jahrhunderts gab es noch weitere Arbeiten zum Thema der stationären Ergebnisqualität. So wurden bereits 1864 Sterblichkeitsraten nach Amputationen und anderen Krankheiten gemessen und verschiedene Krankenhäuser im städtischen und ländlichen Bereich verglichen (vgl. Wells 1864, S. 384). Die vermutete Ursache, dass die Sterblichkeit mit der Etage korreliert, auf der sich die Station befindet, konnte wissenschaftlich nicht bestätigt werden. Die akademische Beschäftigung mit der Messung von Behandlungsergebnissen ist der Beginn einer Vielzahl von wissenschaftlichen Arbeiten zum Thema Behandlungsergebnisse, Sterblichkeit und Qualität. Tabelle 1 zeigt einen kurzen Überblick über die Anzahl von Arbeiten unter den Suchergebnissen „outcome", „quality" und „mortality" in den wissenschaftlichen Journals „New England Journal of Medicine (NEJM)", „British Medical Journal (BMJ)", „The Lancet" und „The Journal of the American Medical Association (JAMA)" für das Jahr 2011.

Tabelle 1: Anzahl gefundene Artikel bei Stichwortsuche in wissenschaftlichen Journals

	Quality	**Outcome**	**Mortality**

The Lancet (all types of articles)	467	369 (161[1])	525 (106[2])
JAMA (research articles only)	105	83 (57)	56 (34)
BMJ (research articles only)	926	1336 (906)	772 (503)
NEJM (research articles only)	66	141 (48)	64 (12)
Summe	1.564	1.929 (1.172)	1.417 (655)

Quelle: Eigene Suche, Beschränkung auf Publikationen 2011

Die Tabelle zeigt, dass täglich mehr als vier Artikel zum Thema Qualität, Behandlungsergebnis und Mortalität allein in diesen Top-Journals veröffentlicht werden. Das verdeutlicht die hohe Bedeutung dieser Themen für die wissenschaftliche Arbeit.

Weitere Pioniere wie Abraham Flexner (mit Bemühungen um medizinische Bildung ca. 1910) und Ernest Amory Codman (Entwickler von Krankenhausstandards) sind ebenfalls zu erwähnen (vgl. Chassin/Loeb 2011). Ein bekanntes Zitat von Codman zeugt von den Herausforderungen bei der Umsetzung von Qualitätsmessungen:

> *„It is against the individual interests of the medical and surgical staffs of hospitals to follow up, compare, analyze and standardize all their results“ (Codman 1996, S. 9; zitiert u. a. von McIntyre et al. 2001, S. 9)*

Qualitätsmessungen fanden im 20. Jahrhundert vorwiegend im Rahmen der Prozess- und Strukturqualität statt. Dies wird vor allem durch die Entwicklungen von Zertifizierungen, Managementsystemen und Berichtswesen deutlich, die auf diese Aspekte fokussiert sind (siehe Abschnitt 2.2.3.3.). Da für die vorliegende Arbeit vor allem die Ergebnisqualität bedeutsam ist, soll die Entwicklung der Ergebnismessung kurz dargestellt werden.

Der Übergang von der Messung von Prozess- und Strukturqualität hin zur Messung und Darstellung von Ergebnisqualität anhand von Routinedaten vollzog sich am Anfang der 90er Jahre. Administrative Daten wurden bereits in den 1970er und 1980er Jahren für kleinräumige Analysen bzw. Ergebnisforschung genutzt (vgl. Wennberg/Gittelsohn 1973; Holcomb 2000; Zhan 2003). Später von der Agency for Health Care Research and Quality (AHRQ) weiterentwickelt, wurden Krankenhausabrechnungsdaten für Qualitätsmessungszwecke von der Expertengruppe um Dr. Iezzoni im „Complications Screening Programm“ 1992 erstmals systematisch angewandt (vgl. Iezzoni et al. 1992; Johantgen et al. 1998). Vor allem zwei Indikatorensets sind auch international bekannt: die Inpatient Quality Indicators (IQI) Sterblichkeitskennzahlen und die Patient Safety Indicators (PSI) (vgl. Mansky/Nimptsch 2010b, S. 145ff.; AHRQ 2006).

[1] Anzahl „Quality“ und „Outcome“

[2] Anzahl „Quality“, „Outcome“ und „Mortality”

In Deutschland hat die Anwendung von Routinedaten für die Ergebnismessung mit der DRG-Einführung 2004 an Bedeutung gewonnen. Bereits 2000 wurde ein Indikatorenset parallel zu den IQI-Indikatoren von den Helios-Kliniken entwickelt (vgl. Busse et al. 2009), das 2010 bereits von mehr als 400 Kliniken in Deutschland angewendet wird (vgl. Mansky et al. 2011). Angepasst an die landesspezifischen Bedingungen werden die Indikatoren auch in der Schweiz und (als nationales Berichtssystem) in Österreich genutzt (vgl. Bundesamt für Gesundheit [BAG] 2010; Fuchs 2010).

Eine weitere Entwicklung stellt die Einbeziehung von krankenhaus- und sektorübergreifenden Daten auf Kassenebene dar. Im Rahmen des QSR-Projektes (Qualitätssicherung der stationären Versorgung mit Routinedaten) können seit 2008 standardisiert Langzeitergebnisse analysiert werden (vgl. AOK-Bundesverband et al. 2007).

Eine weitere Entwicklung der Qualitätsmessung liegt in der Entstehung verschiedenster Register, so bspw. für Endoprothesen oder Tumorpatienten. Diese Sammlung klinischer Daten wird derzeit vor allem für die Versorgungsforschung angewandt (vgl. Wegscheider 2004; Bestehorn 2005). Da sie bislang kaum einer krankenhausvergleichenden Qualitätsdarstellung dienen oder für die Umsetzung von Qualitätsverbesserungsmaßnahmen im Klinikmanagement genutzt werden, sollen diese medizinischen Register hier nicht weiter betrachtet werden.

2.2.2.2. Darstellung verschiedener Indikatorensysteme

Eine Vielzahl unterschiedlicher Qualitätsindikatoren existiert in Deutschland und international. Qualitätsindikatoren werden verstanden als „Maße, deren Ausprägungen eine Unterscheidung zwischen guter und schlechter Qualität von Strukturen, Prozessen und/oder Ergebnissen der Versorgung ermöglichen sollen“ (Jäckel 2009b, S. 2). In der Literatur werden mehrere Einsatz- und Funktionsmöglichkeiten der Qualitätsindikatoren unterschieden. Neben Evaluation, Monitoring, Alarmfunktion und Entscheidungsunterstützung für den internen Gebrauch sind die externen Anwendungsmöglichkeiten in der Überwachungs-, Kontroll- und Rechenschaftsfunktion sowie der Information der Öffentlichkeit, Bestrafung, Leistungserbringervergleich, Zertifizierung und nicht zuletzt leistungsorientierten Vergütung zu finden (vgl. Geraedts 2009). Insgesamt bieten sie eine quantitative Grundlage für die Praxis, um Verbesserungen der klinischen Qualität und der Behandlungsprozesse zu erreichen (vgl. Mainz 2003, S. 524). Daher sind die Analyse und Diskussion von Qualitätsindikatoren ein wichtiger Trend im Gesundheitswesen – auch und vor allem wegen des Einsatzes in der Praxis und ggf. der Anwendung in qualitätsorientierten Anreiz- und Vergütungssystemen.

Qualitätsindikatoren können sowohl auf Routinedaten als auch auf Meldedaten basieren. Erste Datenbanken versuchen bereits einen Überblick über die Indikatoren verschiedener Quellen zu schaffen, so bspw. der „Quinth-Thesaurus“ des GKV-Spitzenverbandes (http://quinth.gkv-spitzenverband.de/content/index.php). Ein Auszug der existierenden Systeme ist in Tabelle 2 dargestellt:

Tabelle 2: Auszug stationärer Qualitätsindikatoren

Im Folgenden diskutierte Indikatorensysteme	Weitere Indikatorensysteme
IQI (Inpatient Quality Indicators der AHRQ)	Pediatric Quality Indicators (*PDIs*) der AHRQ
PSI (Patient Safety Indicator)	Prevention Quality Indictors (*PQIs*) der AHRQ
G-IQI (German Inpatient Quality Indicator)	*OECD*-Indikatoren
A-IQI (Austrian Inpatient Quality Indicators)	Qualitätsindikatoren für Kirchliche Krankenhäuser – *QKK*
CH-IQI (Swiss Inpatient Quality Indicators)	*WHO*-PATH
QSR (Qualitätssicherung mit Routinedaten)	*COMPAQH* der Haute Autorité de Santé (Frankreich)[1]
AQUA (Indikatoren der externen Qualitätssicherung)	Danish National Indicator Project
	Leap Frog Leaps (USA)

IQI / PSI: Die Agency for Healthcare Research and Quality (AHRQ) bietet vier verschiedene Indikatorensysteme an, die alle auf Routinedaten basieren (vgl. Bott 2009):

1. Inpatient Quality Indicators (IQIs),
2. Patient Safety Indicators (PSIs),
3. Pediatric Quality Indicators (PDIs) *[nicht näher betrachtet],*
4. Prevention Quality Indictors (PQIs) *[nicht näher betrachtet].*

Die Indikatoren können neben der Anwendung zur Qualitätsverbesserung und zum Public Reporting auch für P4P-Zwecke genutzt werden. Allen Indikatoren ist gemein, dass sie auf Routinedaten basieren, ergebnisorientiert sind, stationäre Qualität messen, auf Wissenschaftlichkeit geprüft wurden und von der Öffentlichkeit genutzt werden können (vgl. Hussey et al. 2007, S. 6).

Das Inpatient Quality Indicators-System (*IQIs*) besteht aus 32 Indikatoren, die risikoadjustierte Sterblichkeitsraten für 7 konservative und 8 operative Krankheitsbilder, 11 Anwendungs- bzw. Verbreitungsraten von Eingriffen (utlilization rates) und diverse Fallzahlindikatoren (volume) abdecken (vgl. Hussey et al. 2007, S. 4f.). Die Indikatoren werden national und international von vielen Initiativen genutzt und/oder weiterentwickelt (so z. B. den OECD-Indikatoren des HCQI-Projektes – Health Care Quality Indicators) und haben daher eine sehr hohe Relevanz.

[1] Vgl. Grenier-Sennelier et al. 2005

Die Patient Safety Indicators (*PSI*) betrachten potenziell vermeidbare Komplikationen und unerwünschte Ereignisse (adverse events) wie bspw. die Eingriffsverwechslung bei Operationen. Hierfür wurden 27 Indikatoren entwickelt (vgl. Hussey et al. 2007, S. 5). Auch in Deutschland gewinnt das Thema der Patientensicherheit immer mehr an Bedeutung – das Aktionsbündnis Patientensicherheit (APS) entwickelt und veröffentlicht hierfür verschiedene Handlungsempfehlungen und Checklisten, eigene Indikatoren oder Messungen der Patientensicherheit gibt es allerdings nicht (vgl. Lessing 2009, S. 610f.).

G-IQI, A-IQI, CH-IQI: Das verbreitetste, europäische Indikatorensystem wurde parallel zu den AHRQ IQI entwickelt. Zunächst entwickelt und genutzt von der HELIOS Klinikgruppe, verwenden derzeit nach Expertenschätzung ca. 25% aller deutschen Krankenhäuser freiwillig G-IQI-Indikatoren (500 Krankenhäuser), davon 100 Häuser im Rahmen der Initiative Qualitätsmedizin (IQ^M) und schätzungsweise 900 insgesamt in Deutschland, der Schweiz und Österreich (vgl. Nimptsch/Mansky 2012a, S. 1456; Mansky et al. 2011, S. A-5). Die Indikatoren basieren auf Routinedaten und beschränken sich (methodisch notgedrungen) auf den Krankenhausaufenthalt. Ein Fokus liegt auf der Standardized Mortality Ratio (SMR), die die tatsächliche Sterblichkeitsrate mit der aufgrund der bundesweiten Mittelwerte erwarteten Sterblichkeit ins Verhältnis setzt. Die Sterblichkeitsindikatoren beziehen sich auf Krankheitsbilder, bei denen die Sterblichkeit prinzipiell beeinflusst werden kann (vgl. Mansky et al. 2011, S. A-9). Durch die Auswahl der Indikatoren kann diese Messmethode insgesamt 36 % aller Todesfälle in Deutschland abbilden und stellt damit den „umfassendsten Indikatorensatz zur Kontrolle und Verbesserung eines wesentlichen Behandlungsergebnisses, nämlich der beeinflussbaren Krankenhaussterblichkeit“ (Mansky et al. 2011, S. A-11) dar. Eine Variante der G-IQI-Indikatoren 3.1 wird in der Initiative Qualitätsmedizin (IQ^M) im Rahmen eines Qualitätsmanagementsystems mit trägerübergreifenden Peer-Reviews genutzt (vgl. Zacher 2011; Rink 2011).

Für die Anwendung in Österreich und in der Schweiz wurden die G-IQI-Indikatoren entsprechend angepasst. Gerade die verpflichtende Veröffentlichung der Messergebnisse der CH-IQI-Indikatoren aller Krankenhäuser in der Schweiz ist eine bemerkenswerte Entwicklung für Transparenz durch einheitliche Indikatorensysteme (vgl. Zahnd 2011).

QSR: Eine weitere wichtige Entwicklung in Deutschland stellen die QSR-Indikatoren dar. Diese Indikatoren entstammen aus der seit 2002 bestehenden Zusammenarbeit der HELIOS Kliniken Gruppe, FEISA (Forschungs- und Entwicklungsinstitut für das Sozial- und Gesundheitswesen Sachsen-Anhalt) sowie des WidO (Wissenschaftliches Institut der AOK) und dem AOK-Bundesverband. Durch die Nutzung von Langzeitdaten kann mit diesen Indikatoren über den Krankenhausaufenthalt hinaus Ergebnisqualität gemessen und analysiert werden – so werden z. B. Sterblichkeitsraten nicht nur im Krankenhaus, sondern auch nach 30 Tagen, 90 Tagen und einem Jahr evaluiert (vgl. AOK-Bundesverband et al. 2007, S. 13). In der ersten Version von QSR wurden zehn Leistungsbereiche (Tracer) definiert, die eine hohe medizinische, epidemiologische und ökonomische Relevanz haben. Diese sind (vgl. AOK-Bundesverband et al. 2007, S. 12f.):

- Herzinsuffizienz
- Herzinfarkt
- Hirninfarkt oder intrazerebrale Blutung
- Kolon- bzw. Rektum-Operation bei kolorektalem Karzinom
- Appendektomie
- Implantation einer Hüftgelenks-Endoprothese bei Coxarthrose
- Implantation einer Hüftgelenks-Endoprothese oder Osteosynthese bei Hüftfraktur
- Implantation einer Kniegelenks-Totalendoprothese
- Gallenblasenentfernung bei Gallensteinen

Auch für die QSR-Indikatoren werden unter anderem die „SMRs“ angewandt. Einige Ergebnisse werden über den AOK-Krankenhausnavigator veröffentlicht (http://weisse-liste.krankenhaus.aok.de/). Gerade die krankenhausübergreifende Messung und Auswertung bietet dabei verschiedenste Möglichkeiten neuartiger, auch fall- und sektorübergreifender Aussagen zur Versorgungsqualität, sowohl bezogen auf einzelne Kliniken in den Klinikberichten, als auch auf Bundesebene im bundesweiten QSR-Bericht (AOK-Bundesverband et al. 2007). In einer Vielzahl anderer Publikationen werden weitere künftige Anwendungsmöglichkeiten aufgezeigt (z. B. Malzahn et al. 2010; Heller 2010).

AQUA-Indikatoren: Auf der Grundlage des §137a SGB V wird die gesetzliche Qualitätssicherung seit 2010 vom AQUA-Institut umgesetzt. Die zuvor seit 2001 von der Vorgängerinstitution BQS (Bundesgeschäftsstelle Qualitätssicherung) angewandten Indikatoren umfassen derzeit 30 Leistungsbereiche und rund 400 Indikatoren (vgl. AQUA-Institut GmbH 2011a, S. 3). Neben einigen Ergebnisindikatoren (bspw. 1-Jahres-Überlebensraten nach Transplantationen oder Organverletzung bei Hysterektomie) werden vor allem Prozessindikatoren verwendet (bspw. Indikation zur PCI oder Anwesenheit eines Pädiaters bei der Geburt von Frühgeborenen) – auch Mengenangaben werden dargelegt (vgl. AQUA-Institut GmbH 2011a). Derzeit arbeitet das AQUA Institut an einer Weiterentwicklung vorhandener Indikatoren und im Auftrag des GB-A an Neuentwicklungen für weitere Leistungsbereiche. Der Schwerpunkt soll dabei auf sektorübergreifenden Indikatoren liegen. Derzeitiger Projektstand ist bspw. Probebetriebe zu den Themen Kataraktoperation und Konisation seit dem 1. Juli 2012[1]. Die Veröffentlichung der Ergebnisse im strukturierten Qualitätsbericht sowie der Prozess der Datenvalidierung werden im nächsten Abschnitt diskutiert.

Insgesamt zeigt sich, dass verschiedene Indikatorensysteme bestehen. Wie die Akzeptanz der G-IQI Indikatoren seitens vieler deutscher Kliniken sowie ihr Einsatz in der Schweiz und Österreich und parallel

[1] Vgl. https://www.sqg.de/entwicklung/projektstand/index.html, zuletzt überprüft 23.01.2013

auch die Entwicklungen in den USA zeigen, finden Indikatorensysteme, die auf administrativen oder Routinedaten beruhen, zunehmend Zuspruch. Folgende Vorteile der Routinedatensysteme zeichnen sich ab (vgl. u. a. Mansky 2008, S. 149ff.):

- Mit der zunehmenden Verfügbarkeit medizinischer Daten im Abrechnungsbereich sind früher nicht mögliche medizinische Auswertungen umsetzbar
- Die Daten sind vollständig hinsichtlich der abgerechneten Fälle
- Die Zusammenführung dieser Daten erlaubt auch Langzeitauswertungen, die über eine Datenerhebung nicht mit vertretbarem Aufwand möglich wären (vgl. Nimptsch/Mansky 2012b, S. 885ff.)
- Aufgrund der Vollständigkeit und der erfolgenden Kontrollen im Abrechnungsprozess ist die Manipulationssicherheit der Daten höher als bei separaten Erhebungen nur für Zwecke der Qualitätsmessung
- Die Daten sind im Managementbereich bei den Leistungserbringern, bei den Krankenkassen, aber auch (über das Statistische Bundesamt) im wissenschaftlichen Bereich verfügbar
- Sie stehen zumindest beim Leistungserbringer zeitnah zur Verfügung
- Es gibt keinen Zusatzaufwand in der Datenerfassung
- Nach anfänglicher Skepsis gibt es eine steigende Akzeptanz in der Ärzteschaft und im Management, da das Kosten-Nutzen-Verhältnis positiv gesehen wird
- Durch Verbesserung der medizinischen Klassifikationen (ICD, OPS, Zusatzkennzeichen wie „present on admission / (POA)“) stehen zunehmend wichtige Sachverhalte für die Auswertung zur Verfügung

Im wissenschaftlichen Bereich konnte oft gezeigt werden, dass die administrativen Daten hinsichtlich der in diesem Rahmen verfügbaren Informationen in vielen Fällen gleichwertig zu klinischen Datenerhebungen eingesetzt werden können (vgl. u. a. Heller et al. 2008, S. 171ff.; Mansky 2008, S. 154). Andererseits ist unbestreitbar, dass viele klinische Detailinformationen in den Abrechnungsdaten fehlen und trotz weiteren Ausbaus der Klassifikationen, der aber aufgrund des Abrechnungsaufwandes limitiert sein wird, sicher auch in Zukunft nicht komplett zur Verfügung stehen werden. Insofern zeichnet sich eine sich ergänzende Koexistenz beider Systeme ab: Die administrativen Daten ermöglichen rasche und sogar auf nationaler Ebene komplette Datenauswertungen wesentlicher Sachverhalte. Spezialerhebungen können und müssen diese ggf. ergänzen, wenn zusätzliche Details gewünscht oder erforderlich sind. Die Grenze wird sich fortlaufend weiter verschieben, da mit zunehmender Verbreitung elektronischer Krankenakten immer mehr Daten als „Routinedaten“ verfügbar werden. Auf die zunehmende Nutzung solcher Daten z. B. bei CMS und im Rahmen von P4P wird später eingegangen.

2.2.3. Qualitätsinitiativen im stationären Sektor: Gesetzliche Grundlagen und Überblick

2.2.3.1. Gesetzliche Grundlagen zur Qualitätssicherung im stationären Sektor

Eine Vielzahl von Gesetzesvorschriften beschäftigt sich mit der stationären Qualität. Krankenhäuser sind nach dem SGB V und den sich daraus ergebenden Vereinbarungen explizit zur Qualitätssicherung verpflichtet. Mittelbar ergibt sich auch vertragsrechtlich (durch den Behandlungs- bzw. Versorgungsvertrag zwischen Leistungserbringer, Patient und Krankenkasse) die Verpflichtung zu hoher Qualität. Zunächst besteht zwischen Patienten und Ärzten ein Behandlungsvertrag, der nach herrschender Meinung einen Dienstvertrag nach §§ 611ff. BGB darstellt (vgl. Griebau 2011, S. 607). Aufgrund der Komplexität des menschlichen Organismus kann ein Heilerfolg zunächst nicht garantiert werden, so dass es sich nicht um einen Werkvertrag nach §§ 631ff. BGB handelt. Ob es auch zwischen dem Arzt und der GKV-Patienten (auch im Rahmen einer Krankenhausbehandlung) zu einem Dienstvertrag kommt, ist umstritten (vgl. Griebau 2011, S. 609).

Eine Vielzahl von Abhandlungen beschäftigt sich mit der Arzthaftung, dies kann und soll hier nicht abschließend betrachtet werden. Wenn der Arzt allerdings im Rahmen des Behandlungsvertrages gegen seine Sorgfaltspflichten verstößt, so können sich daraus Schadensersatzansprüche für den Patienten ergeben – sowohl ein Verschulden als auch eine Kausalität muss hierfür vom Patienten nachgewiesen werden (vgl. Kaiser 2008, S. 601ff.). Relevant ist vor allem, dass abweichende Garantieregelungen zwischen den Vertragspartnern (vor allem auch zwischen dem Kostenträger und dem Krankenhaus) möglich sind. So ist es bspw. in selektivvertraglichen Verträgen der Integrierten Versorgung möglich, Gewährleistungsregelungen oder Garantieübernahmen zu vereinbaren – allerdings verursacht dies für den Leistungserbringer ein erhöhtes Risiko, da „der Eintritt der Gewährleistungspflicht bzw. des Garantiefalls auf vielen Ursachen beruhen kann, die sich dem Einflussbereich des Leistungserbringer entziehen“ (Möller et al. 2008, S. 549). Außerdem entstehen für den Leistungserbringer Versicherungsschutzlücken, da der Haftpflichtschutz vertraglich vereinbarte Ansprüche, die über die gesetzliche Verpflichtung hinausgehen, ausschließt (vgl. Möller et al. 2008, S. 549). Vertragspartner können zusätzlich die Form und Höhe der Vergütung vereinbaren (vgl. Möller et al. 2008, S. 551). Nach § 140b Abs. 3 S. 3 SGB V sind die Vertragspartner in IV-Verträgen außerdem an den Mindeststandard der gesetzlich verpflichtenden Qualitätssicherung gebunden[1], die im Folgenden diskutiert wird.

Über die ärztlichen Berufsordnungen der Länder sind Ärzte dazu verpflichtet, an Qualitätssicherungsmaßnahmen teilzunehmen, ohne explizit auf konkrete Methoden einzugehen (vgl. Bundesärztekammer 2011, § 5). Außerdem sind sie „zur gewissenhaften Versorgung mit geeigneten Untersuchungs- und Behandlungsmethoden“ verpflichtet (Bundesärztekammer 2011, § 11).

[1] Eine Bindung an die Qualitätsvorgaben nach §§ 135ff. SGB V besteht nicht (vgl. Möller et al. 2008, S. 550)

Für diese Arbeit ist aber besonders die explizite Qualitätssicherung nach den Vorschriften des SGB V von Relevanz. Zunächst sind alle nach § 108 SGB V zugelassenen Krankenhäuser zur Teilnahme an der externen, vergleichenden Qualitätssicherung verpflichtet. §135a SGB V verpflichtet die Leistungserbringer zu „Sicherung und Weiterentwicklung der Qualität der von ihnen erbrachten Leistungen“, insbesondere zur Beteiligung an „einrichtungsübergreifenden Maßnahmen der Qualitätssicherung“ mit dem Ziel der Ergebnisqualitätsverbesserung und zur Einführung und Entwicklung eines einrichtungsinternen Qualitätsmanagements (§ 135a SGB V). Konkrete Umsetzungsvorschriften bezüglich der Wahl eines Qualitätsmanagementsystems werden nicht gestellt.

Die Paragraphen 137, 137a, 137b und 137c regeln die verpflichtende Qualitätssicherung in der stationären Versorgung. Während § 137 SGB V u. a. die Zuständigkeiten des Gemeinsamen Bundesausschusses (G-BA) als oberstes Beschlussgremium der gemeinsamen Selbstverwaltung von Ärzten, Krankenhäusern und Krankenkassen beschreibt, werden in § 137a SGB V die Aufgaben aufgelistet, mit denen derzeit AQUA beauftragt ist. Im Rahmen der „einrichtungsübergreifenden Qualitätssicherung“ nach § 115b Abs. 1, § 116b Abs. 3 S. 3, § 137 Abs. 1 und § 137f. Abs. 2 Nr. 2“ (§ 137a Art. 1 SGB V) sind dies sowohl die Indikatorenentwicklung für Messung und Darstellung der Versorgungsqualität, die Dokumentation und die Durchführung der einrichtungsübergreifenden Qualitätssicherung als auch die Veröffentlichung der Ergebnisse (§ 137a Art. 2 SGB V). Die Paragraphen § 137b und § 137c beschreiben weitere Aufgaben des G-BAs wie bspw. die Bewertung von Untersuchungs- und Behandlungsmethoden und die Erstellung eines Berichts über den Stand der Qualitätssicherung. Dadurch entstehen mittelbare Auswirkungen für die Qualitätssicherung in Krankenhäusern (§ 137b und c SGB V).

Auch andere gesetzliche Regelungen verpflichten zu Berücksichtigung von Qualitätsaspekten, wie bspw. die Paragraphen zum ambulanten Operieren (§ 115b SGB V) oder zu den Verträgen der integrierten Versorgung (§ 140a-b SGB V). Gerade letztere werden später detailliert betrachtet.

Insgesamt zeigt sich, dass eine Vielzahl von Regelungen zur stationären Qualitätssicherung bestehen, ohne die Auswahl der Maßnahmen zur Qualitätsverbesserung zu sehr einzuschränken. Diese Qualitätsmaßnahmen werden im nächsten Abschnitt beschrieben und diskutiert.

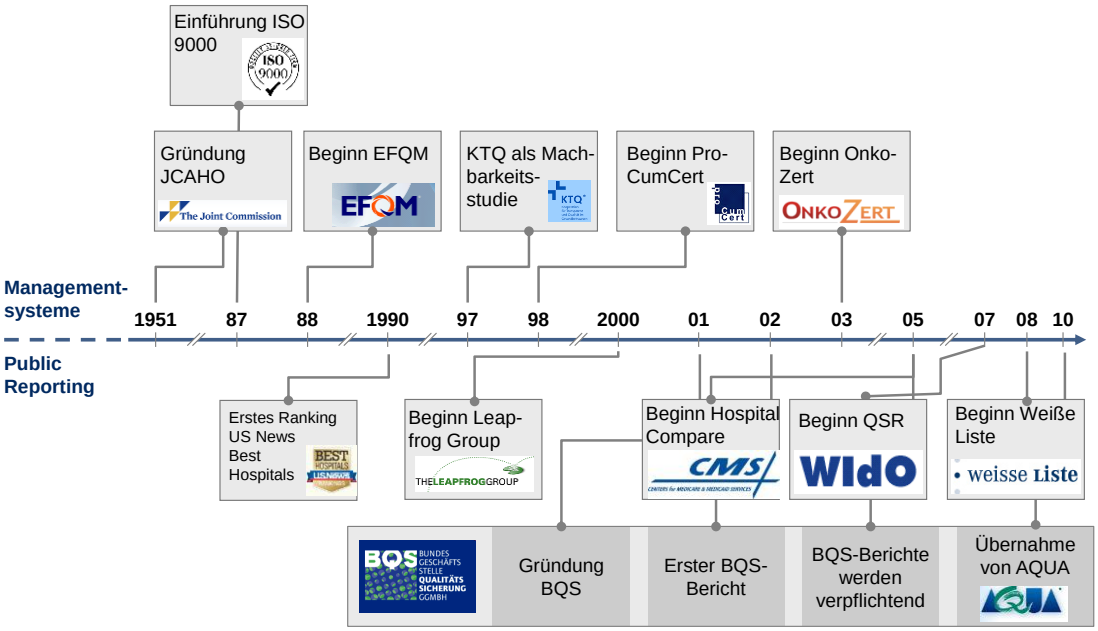
2.2.3.2. Überblick und Kategorisierung der Qualitätsinitiativen

Neben P4P sind auch Managementsysteme und Public Reporting als Umsetzungssysteme für Qualitätsverbesserungen in dieser Arbeit zu diskutieren. Ohne innerbetriebliche Qualitätsmanagementsysteme – egal in welcher Form – kann P4P keine Verbesserungswirkungen entfalten. Vom Public Reporting, das oft als immanenter Teil von P4P gesehen wird (vgl. Emmert 2008, S. 75; SVR 2007, S. 167; Kuhmerker/Hartman 2007), wird eine ergänzende Wirkung auf den Patienten

erwartet (vgl. u. a. Lindenauer et al. 2007; Chien et al. 2007). Diese würde die finanzielle Wirkung von P4P verstärken.

Es gibt neben der gesetzlichen Qualitätssicherung eine Vielzahl von Initiativen in Deutschland und auch international, die verschiedene Aspekte wie Zertifizierungen, Managementsysteme und Public Reporting umfassen. Abbildung 8 und Tabelle 3 geben eine Übersicht über die beiden Kategorien Qualitätsmanagementsysteme und Public Reporting.

Abbildung 8: Überblick über die Entwicklung ausgewählter Qualitätsinitiativen



Quelle: Eigene Darstellung mit Daten aus Internetrecherche und vorherigen Abschnitten

Tabelle 3: Umsetzung von stationären Qualitätssystemen nach Kategorien (Auswahl)

Managementsysteme	Public Reporting
JCAHO	Qualitätsberichte
DIN EN ISO 9000	Internetbasierte „Klinikführer“
EFQM	Qualitätsrankings
TQM	
KTQ	

Es ist festzuhalten, dass die zuvor angesprochenen Indikatorensysteme auch Anwendung im Rahmen von Managementsystemen oder Public Reporting finden, so bspw. die G-IQI Indikatoren die anhand von Peer-

Reviews und einem PDCA-Zyklus ins Management integriert werden oder die Veröffentlichung von QSR-Berichten (vgl. z. B. Busse et al. 2009; AOK-Bundesverband et al. 2007). Gleichzeitig werden auch Ergebnisse von Zertifizierungen von Managementsystemen veröffentlicht (bspw. JCAHO und KTQ). Eine überschneidungsfreie Zuordnung der verschiedenen Initiativen zu diesen Kategorien ist daher nicht möglich.

2.2.3.3. Qualitätsmanagementsysteme und -zertifizierungen

Qualitätsmanagementsysteme wurden Mitte des 20. Jahrhunderts erstmals entwickelt und fanden ihren vorläufigen Endpunkt in der Entwicklung des „Total Quality Management (TQM)“ (vgl. Gucanin 2003) und somit der vollständigen Ausrichtung aller Prozesse auf Qualitätsaspekte (vgl. Fleßa 2007, S. 244). Vorher waren bereits Konzepte wie „Kaizen“ (vgl. Imai 1993) – als Grundkonzept aller Managementmethoden – oder „Lean Management“ relevant (vgl. z. B. Haubrock 2002, S. 120f.). Die Anwendung in der Medizin, insbesondere im stationären Sektor fand erst ziemlich spät statt und wurde vor allem von einer strengeren Gesetzgebung Ende der 1990er Jahre forciert (vgl. Fleßa 2007, S. 245).

In Deutschland stehen vor allem die Managementsysteme nach DIN ISO 9000 und der European Foundation for Quality Management (EFQM) im Vordergrund (vgl. Hildebrand 2005). Diese Systeme fordern die Umsetzung eines Managementsystems (z. B. mit den Kernelementen des PDCA- oder Deming-Zyklusses – „Plan-Do-Check-Act“). Dies soll für das Krankenhaus zu einer kontinuierlichen Beschäftigung mit Qualität führen und zu der Möglichkeit, das Thema im Unternehmen und der Unternehmenskultur zu etablieren.

Ein weiteres konkretes Beispiel für ein etabliertes Qualitätsmanagementsystem basierend auf Ergebnisqualität ist das System der Helios Kliniken Gruppe. Basierend auf den G-IQI Indikatoren wird hier die Qualität durch ein Peer-Review-Verfahren und klare Vorgaben eines Qualitätsmanagementsystems und Zielvereinbarungen kontinuierlich verbessert (vgl. Mansky/Nimptsch 2010a; Mansky/Nimptsch 2010b; Helios Kliniken Gruppe 2005, S. 69ff.; Busse et al. 2009). Durch das System können vor allem die Qualitätsergebnisse von Kliniken, die vorher unterdurchschnittliche Ergebnisse hatten, deutlich verbessert werden. Auch innerbetriebliche finanzielle Anreize im Rahmen von Zielvereinbarungen werden hier etabliert.

Eine Vielzahl der zuvor erwähnten Qualitätsinitiativen beschäftigt sich mit der externen Bewertung von Qualitätsmanagementsystemen durch Zertifizierungen oder Akkreditierungen. Hier sollen die wichtigsten Beispiele kurz dargestellt werden: DIN EN ISO 9000ff, JCAHO, EFQM und KTQ.

DIN EN ISO 9000ff.: Die Zertifizierung nach der International Standardization Organisation (ISO) bzw. des Deutschen Instituts für Normung (DIN) kann als einer der formalisiertesten Systeme angesehen werden. Die Hauptelemente stellen dabei das Qualitätshandbuch (mit dokumentierten, standardisierten Aufbau- und Ablauforganisationen, QM-Beauftragten und -Prozessen) und externe und interne Audits dar (vgl.

Pfitzinger 2011; Fleßa 2007, S. 251). Obschon DIN-Normen als solche einen hohen gesellschaftlichen Bekanntheitsgrad und eine prinzipiell hohe Akzeptanz haben, sind im Bereich der Medizin einige Nachteile zu vermerken: So ist die Entwicklung des QM-Handbuches oft technokratisch und zeigt vielfach aus Sicht des medizinischen Personal wenig Praxisbezug – außerdem ist die Zertifizierung aufgrund der hohen Kosten und des großen Zeitaufwandes für ein gesamtes Krankenhaus kaum möglich und wird daher fast nur in Teilbereichen durchgeführt (vgl. Fleßa 2007, S. 251f.).

JCAHO: Durch die Gründung der Joint Comission on Accreditation of Healthcare Organizations (JCAHO) 1951 wurden die – bereits von der American College of Surgeons (ACS) 1918 entwickelten – Mindeststandards für Krankenhäuser als Zertifizierungen umgesetzt (vgl. McIntyre et al. 2001). Die Akkreditierung durch die JCAHO und der jährliche Bericht bleibt auch heutzutage wichtiger Bestandteil der Qualitätsdiskussion in den USA – primär weil sie Voraussetzung für Zahlungen der größten Versicherungen (Medicare und Medicaid) ist und somit rund 80 % aller Krankenhäuser in den USA akkreditiert sind (vgl. Fleßa 2007, S. 252). Während in Deutschland der Nachweis eines einrichtungsinternen Qualitätsmanagement gesetzlich verpflichtend ist (vgl. auch vorherigen Abschnitt)[1], gibt es in den USA keine solche gesetzliche Regelung. Dagegen wirkt die Verpflichtung seitens Medicare/Medicaid (als größter „Zahler“ in den USA), eine JCAHO Zertifizierung vorzuweisen, als hoher (finanzieller) Anreiz für die Nutzung eines Qualitätssicherungssystems, der in Deutschland aus der gesetzlichen Regelung entsteht. Die JCAHO Qualitätsindikatoren fokussieren sich vor allem auf Prozessqualität (vgl. JCAHO 2010), doch auch Infektionsraten oder Mortalitäten werden gemessen. Die Ergebnisse werden in einer Datenbank gesammelt und veröffentlicht (u. a. http://www.qualitycheck.org).

EFQM: Die Zertifizierung der European Foundation of Quality Management (EFQM) befasst sich im Gegensatz zum DIN EN ISO 9000ff.-Verfahren mit dem gesamten Krankenhaus. Neben der Überprüfung der Prozesse im Unternehmen beschäftigt sich das EFQM-Modell zu 50 % mit Outcomeparametern, wie Ergebnissen, gesellschaftlicher Verantwortung oder Mitarbeiterzufriedenheit (vgl. Fleßa 2007, S. 255). Auch wenn hierbei nicht explizit von Ergebnisqualität gesprochen wird, ist die klare Ergebnisorientierung ein Vorteil gegenüber anderen Systemen. Allerdings wird für die Ergebnisqualität keine standardisierte Messung vorgegeben. Die Ergebnisorientierung bleibt somit in der Regel betriebsintern und ermöglicht folglich nicht zwangsläufig einen Vergleich mit anderen Leistungserbringern.

KTQ: 1997 wurde KTQ (Kooperation für Transparenz und Qualität) als Machbarkeitsstudie begonnen, 2001 wurde es offiziell eingeführt. Sowohl Qualitätsverbesserung als auch Transparenzerhöhung werden explizit angestrebt, es stehen wiederum Prozesse und Strukturen im Rahmen eines PDCA-Zyklus im Vordergrund (vgl. Fleßa 2007, S. 258). Außerdem werden spezielle KTQ-Qualitätsberichte veröffentlicht.

[1] Welche Managementsysteme umgesetzt werden, wird nicht vorgeschrieben, so dass auch einige deutsche Kliniken die internationale Varianten der JCAHO Zertifizierung nutzen (bspw. DRK-Kliniken).

Die Qualitätsberichte werden allerdings als „anekdotisch“ bezeichnet und beinhalten weder Ergebnisinformationen noch aussagekräftige Informationen zu Prozessen oder Strukturen (vgl. Hildebrand 2005, S. 40–41).

Insgesamt lässt sich feststellen, dass derzeit noch nicht abschätzbar ist, welches Qualitätsmanagementsystem und welche Zertifikate sich langfristig am Markt durchsetzen werden, da die Initiativen viele Vor- und Nachteile umfassen (vgl. Emmert 2008, S. 299).

2.2.3.4. Public Reporting: Umsetzung von stationärer Qualität durch Qualitätsberichtserstattung

Public Reporting – d. h. die Veröffentlichung von Informationen zur stationären Qualität – wurde zwar in den USA „pioniert“, doch auch in Deutschland gehören Qualitätsberichte und vergleichende Internetplattformen mittlerweile zur Tagesordnung. Die wichtigsten Ansätze sind außerdem gesetzlich verankert (vgl. Emmert 2008, S. 259).

Zunächst einmal ist festzuhalten, dass die Veröffentlichungen von Qualitätsergebnissen von Ärzten prinzipiell als vergleichende Werbung gesehen werden könnte, die im stationären Sektor nur unter bestimmten Voraussetzungen gestattet ist (vgl. nach Bundesärztekammer 2011, §§ 27 Abs. 3; Emmert 2008, S. 260f.). Gesetzlich vorgegeben ist die *externe vergleichende Qualitätssicherung* nach § 137 SGB V. Seit 2009 durchgeführt durch das AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH (vorher durch die BQS) werden deutschlandweit verpflichtend Qualitätsdaten von 1.800 Krankenhäusern anhand von 30 Leistungsbereichen und 400 Qualitätsindikatoren übermittelt und ausgewertet (vgl. AQUA-Institut GmbH 2011a). Die bundesweiten Ergebnisse werden u. a. in den Qualitätsreports von AQUA zusammengefasst. Der Bundesbericht dient zusammen mit den vom AQUA Institut den Kliniken zur Verfügung gestellten Auswertungen „in erster Linie dem internen Vergleich von Kliniken untereinander und soll so die Förderung und Verbesserung der Behandlungsqualität unterstützen“ (AQUA-Institut GmbH 2011a, S. 3). Insgesamt werden ca. 17 % aller stationär behandelten Patienten erfasst. Die Indikatoren umfassen sowohl Prozess- als auch Struktur- und Ergebnisqualität – allerdings liegt ein Fokus auf Prozessqualität. Außerdem basieren die Messungen vor allem auf der Meldung durch die Leistungserbringer – die Ergebnisse werden aber auf Plausibilität, Vollständigkeit und Vollzähligkeit geprüft (vgl. Gemeinsamer Bundesausschuss 01.01.2012, § 9). Bei auffälligen Ergebnissen werden die Krankenhäuser dem „strukturierten Dialog“ unterzogen, um ggf. durch die Einbeziehung von Fachexperten die interne Qualitätsverbesserung zu fördern. Manipulationsfreiheit kann aber trotzdem nicht garantiert werden. Einige Vergleiche zeigen bereits deutliche Unterschiede der Qualitätsmessung anhand von Routinedaten und den Daten des AQUA-Berichts (vgl. bspw. Heller/Schnell 2007, S. 1983f.; Jakob 2010).

Ein weiterer Aspekt der externen stationären Qualitätssicherung nach § 137 Abs. 3 Satz 1 Nr. 4 SGB V ist die verpflichtende zweijährliche Erstellung und Veröffentlichung eines *„strukturierten Qualitätsberichts"* für alle Krankenhäuser, erstmals für das Berichtsjahr 2004 (vgl. Emmert 2008, S. 351). Ziele dieses Berichtes sind (vgl. Gemeinsamer Bundesausschuss 23.06.2011, § 1):

1. Information und Entscheidungshilfe für alle interessierten Personen, insbesondere im Vorfeld einer Krankenhausbehandlung,
2. Orientierungshilfe bei der Einweisung und Weiterbetreuung von Patienten und Patientinnen insbesondere für Leistungserbringer in der ambulanten Versorgung und für Krankenkassen,
3. Möglichkeit für ein Krankenhaus, seine Leistungen nach Art, Anzahl und Qualität nach außen transparent und sichtbar darzustellen.

In den Berichten müssen nach derzeitigem Stand alle Indikatoren der Kategorien 1 (Verpflichtende Veröffentlichung empfohlen) und Kategorien 2 (Verpflichtende Veröffentlichung empfohlen, Erläuterung und/oder leichte Anpassung notwendig) aus der externen Qualitätssicherung (also den „AQUA-Indikatoren) veröffentlicht werden. Außerdem werden Struktur- und Leistungsdaten des Krankenhauses und der Fachabteilungen, sowie Themen des Qualitätsmanagements (so u. a. Qualitätspolitik und Qualitätsziele) integriert (vgl. Gemeinsamer Bundesausschuss 23.06.2011).

Ein weiterer Trend in der Qualitätsdiskussion sind vergleichende Darstellungen und Rankings der Qualität von Krankenhäusern im Internet. Einige amerikanische Beispiele sind Leap Frog (http://www.leapfroggroup.org/), US News Best Hospitals (http://health.usnews.com/best-hospitals), Daten der JCAHO (http://www.qualitycheck.org) und Medicare/CMS (http://www.hospitalcompare.hhs.gov). Diese Darstellungen erlauben einen Vergleich von Krankenhäusern und eine regionale und ggf. auch krankheitsspezifische Suche. Die genutzten Indikatoren variieren dabei stark – von Aussagen zu Akkreditierung (JCAHO) über einen Fokus auf Patientensicherheit (Leap Frog) bis zu gewichteten Scores und Rankings verschiedener Dimensionen (US News) (vgl. Hill et al. 1997; Comarow 2006).

Auch in Deutschland gibt es bereits mehrere vergleichende Internetdarstellungen von Leistungserbringern – sowohl für den ambulanten Sektor im Rahmen von Arztportalen (vgl. Emmert et al. 2009; Hügler 2010) als auch für Krankenhäuser (vgl. Geraedts 2006; Petersen et al. 2007; Norbert/Martin 2008; Van Emmerich/Metzinger 2010). Einige dieser Klinikführer sind:

- Weiße Liste (Bertelsmann Stiftung) (www.weiße-liste.de)
- Klinik-Lotse (www.klinik-lotse.de)
- Kliniken Rhein-Ruhr (www.kliniken-rhein-ruhr.de)

- TK-Klinikführer (www.tk-online.de)
- Klinikbewertungen (www.klinikbewertungen.de)

Im stationären Sektor wird diese Entwicklung sogar als „Wildwuchs“ bezeichnet (vgl. Petersen et al. 2007, S. 1112). Die Portale bewerten und ranken die Krankenhäuser anhand teilweise sehr unterschiedlichen Daten und Scorings. Der Nutzen der Portale wird aufgrund fehlender Patientenorientierung hinterfragt (vgl. Petersen et al. 2007), auch wenn 22 % der Befragten einer repräsentativen Umfrage im Gesundheitsmonitor der Bertelsmann Stiftung solche Angebote nutzen wollen (vgl. Geraedts 2006, S. 159). Gleichzeitig gehen 47 % der Befragten davon aus, dass nicht ausreichende Informationen für eine Krankenhauswahl zur Verfügung stehen und 86 % haben noch nie das Internet für Informationen zur Krankenhausqualität genutzt (vgl. Mansky 2013, S. 145ff.).

2.2.4. Exkurs: Wahrnehmungsverzerrung in der Qualitätsdiskussion in Deutschland?

Bestimmte Qualitätsprobleme werden in der öffentlichen Diskussion nicht oder nur verzerrt wahrgenommen. Dies soll an drei Beispielen diskutiert werden:

Beispiel 1: Cholezystektomie: Die Todesfallrate bei der einfachen Cholezystektomie (ohne Tumorfälle) ist beispielsweise in Deutschland signifikant höher als in den USA oder der Schweiz (vgl. Nimptsch/Mansky 2012a, S. 1456). Dennoch zählt dieser Indikator nach Einstufung der Bundesfachgruppen nicht zu den Indikatoren „mit besonderem Handlungsbedarf“, seit 2010 auch sonst kein Indikator des Leistungsbereiches Cholezystektomie (bis 2009 noch der Indikator „QI 4a: Eingriffsspezifische Komplikationen (Okklusion oder Durchtrennung des DHC)“) (vgl. AQUA-Institut GmbH 2011a, S. 16). In den USA und Kanada wird dieser Indikator dagegen zusammen mit seltenen Todesfällen bei anderen Eingriffen mit niedrigem Risiko als wichtiges Merkmal der Patientensicherheit angesehen (vgl. AHRQ 2006; Barua/Esmail 2011, S. 47).

Beispiel 2: Patientensicherheit vs. andere Qualitätsthemen: Während in der Presse häufig über Themen der Patientensicherheit gesprochen wird – die ohne Zweifel eine hohe Bedeutung haben – werden Qualitätsverbesserungen bei als „unspektakulär“ wahrgenommenen Krankheiten (wie Herzinfarkt oder Schlaganfall) nur selten diskutiert. Dabei gibt es für diese Krankheiten (Herzinfarkt, Herzinsuffizienz, Pneumonie, Schlaganfall) in Deutschland insgesamt 116.458 Krankenhaustodesfälle bei 1.149.890 Fällen (d. h. ca. 10 %; vgl. Mansky et al. 2011, S. A-11). Eine Reduktion der Sterblichkeit um 5 % (relativ) ist nach vorliegenden Daten möglich, d. h. 5.253 Todesfälle wären hier vermeidbar (vgl. Busse et al. 2009). Das Risiko einer „wrong-side-surgery“, als Beispiel eines typischen Patientensicherheitsproblems, beläuft sich dagegen auf 1:112.994 Operationen (vgl. Kwaan 2006, S. 353) und liegt in Deutschland damit vermutlich in der Größenordnung von ca. 100 Fällen (oder weniger). Das Beispiel bedeutet nicht, dass die Vermeidung der wrong-side-surgery unwichtig wäre. Es zeigt aber im Kontrast zu den

Reduktionsmöglichkeiten von Todesfallraten bei häufigen Krankheiten, dass die relative Gewichtung in der öffentlichen Wahrnehmung verzerrt ist.

Beispiel 3: Thrombolyse bei Schlaganfall: Beim Schlaganfall wird die Thrombolyse als wichtige Maßnahme zur Qualitätsverbesserung dargestellt. Derzeit werden aus verschiedenen Gründen nur ca. 6 % aller Schlaganfallpatienten lysiert (vgl. Mansky et al. 2011, S. A-33). Wenn die Lyse die Todesfallrate um 10 % senken könnte, würde sie die genannte Gesamtsterblichkeit aufgrund von Schlaganfall um 0,6 % reduzieren können. Die tatsächlichen Verbesserungen, die HELIOS oder IQ^{M} mit vielen anderen Maßnahmen erzielen, liegen weit darüber (vgl. Busse et al. 2009).

Lösungsvorschlag: Insgesamt zeigt sich die Notwendigkeit, die öffentliche Wahrnehmung von Qualitätsproblemen stärker zu objektivieren. Derzeit zählen eher Emotionen als Zahlen. Eine verzerrte Wahrnehmung der Probleme führt nicht unbedingt zu einer falschen Allokation von Mitteln.

2.3. Zwischenfazit: Die Diskussion zeigt neue Sichtweisen auf Qualität und auf die verschiedenen Qualitätsinitiativen im stationären Sektor

Erste Erkenntnisse für diese Promotionsarbeit lassen sich wie folgt zusammenfassen:

- Qualität als Begriff und Thematik wird in verschiedensten Fachrichtungen detailliert und vielseitig diskutiert und bewertet.
- Es existieren mehrere Begriffsdefinitionen, die sich auf verschiedene Aspekte fokussieren – sowohl im Bereich der Gesundheitsversorgung als auch in anderen Sektoren.
- Aus diesen Abhandlungen zum Qualitätsthema lassen sich neben den üblichen Dimensionen und Beschreibungen (so bspw. nach Donabedian mit Struktur-, Prozess- und Ergebnisqualität) weitere Aspekte herausarbeiten, die für diese Arbeit von hoher Relevanz sind, so bspw. die Frage nach Qualität als „Wert“, „Kundenorientierung“ oder auch die Perspektiven verschiedener Akteure im Gesundheitswesen.
- Zusammenfassend ergibt sich hieraus eine „working definition“ für Qualität, die verschiedene Eingrenzungen der vorliegenden Arbeit (so unter anderem anhand der Akteure) beinhalten.
- Anschließend konnte ein Überblick über die Bedeutung von Qualität speziell im stationären Sektor gegeben werden. Aus der Beschreibung der historischen Entwicklung der Qualitätsdiskussion generell, vor allem aber der Qualitätsmessung und der Anwendung und Umsetzung von Qualität u. a. in Managementsystemen, Zertifikaten und Berichten wurde deutlich, dass nicht nur eine hohe Relevanz von Qualität im stationären Sektor besteht. Sowohl im Klinikalltag als auch in verschiedenen Qualitätsinitiativen konnte das Thema weiterentwickelt werden. Gleichzeitig gibt es weiterhin Defizite, die unter anderem durch qualitätsorientierte Vergütung adressiert werden könnten.

- Der Exkurs zeigt eine unterschiedlich wahrgenommene Relevanz von Qualitätsthemen bei verschiedenen Akteuren im deutschen stationären Sektor und zeigt erste Lösungsvorschläge auf.

3. Entwicklung des Qualitätsteils des Kriterienkatalogs für P4P-Systeme

3.1. Wichtigkeit der Operationalisierung stationärer Qualität durch qualitätsorientierte Vergütung

Damit der vorher diskutierte Qualitätsbegriff auch in der Praxis umgesetzt werden kann, ist besonders eine hohe Operationalisierbarkeit oder Anwendbarkeit bedeutsam. Neben verschiedenen Anreizen zur Umsetzung sind hierfür im ersten Schritt verschiedene Anforderungen an die Definition von Qualitätsindikatoren und die Umsetzung der Qualitätsmessung notwendig. Die Qualitätsindikatoren und Messung sollen auch für die Verwendung in einem qualitätsorientierten Anreizsystem geeignet sein und somit die Inzentivierung für nachhaltige Qualitätsverbesserungen unterstützen.

Qualitätsmesssysteme existieren in der stationären Versorgung bereits, so dass Anforderungen hieran deduktiv erarbeitet werden können. Auch für Anreizsysteme gibt es einige Vorstellungen zu Kriterien. Qualitätsorientierte Anreizsysteme wurden zwar in ihrer Gesamtheit bereits verschiedentlich in Reviews bewertet, ein einheitlicher Anforderungskatalog besteht jedoch nicht. Aus diesem Grund ist es hilfreich, beide Aspekte (d. h. Qualität und Anreize) zunächst einzeln und dann gemeinsam zu analysieren. Eine getrennte Betrachtung der Anreizaspekte ermöglicht außerdem eine klare Fokussierung auf die Qualitätsaspekte und deren Verständnis.

3.2. Beschreibung der Entwicklung des Qualitätsteils des Kriterienkatalogs

3.2.1. Quellen für die deduktive Entwicklung des Kriterienkatalogs

Die Entwicklung des Kriterienkatalogs findet überwiegend deduktiv statt. Neben einer eingehenden Literaturrecherche wurden auch neue Aspekte, die aus dem Anwendungsgebiet der Anreizsysteme entstammen, hinzugefügt.

Anhand der Literaturrecherche wurden vier relevante Quellenarten definiert:

- Existierende Auflistungen von verschiedenen Anforderungen für die Entwicklung oder Bewertung von Qualitätsindikatoren[1], so z. B. in Deutschland von QUALIFY (unter anderem in Reiter et al. 2007; Geraedts et al. 2002 und Zorn/Ollenschläger 1999; oder international Remus et al. 2004; Campbell

[1] Einige Anforderungen auch in Arbeiten zur Bewertung von Indikatorensystemen und systemischen Reviews

2002; JCAHO 2006; Mattke et al. 2006; Mainz 2003; Arah et al. 2006; Ten Asbroek et al. 2004; Vos et al. 2009 und Wollersheim et al. 2007)

- Bewertungsschemata für P4P-Systeme in – systematischen – Reviews (so z. B. in Deutschland SVR 2007; Scheppach et al. 2011; Veit et al. 2012a[1], Schrappe/Gültekin 2010; Mehrotra et al. 2010a und Amelung/Zahn 2009 oder international Rosenthal et al. 2005; van Herck et al. 2010; Bufalino et al. 2007; Institute of Medicine 2007 und Kuhmerker/Hartman 2007[2])
- Anforderungskataloge für Qualitätssysteme in anderen Bereichen (so z. B. in der ambulanten Versorgung oder für Ärzteportale; unter anderem Thomeczek 2002; Diel 2009)
- Anforderungen, die sich aus klinischen Praxisanwendungen entwickelt haben (vgl. u. a. Mansky 2010a).

Mittels dieser und weiterer Literaturquellen sowie eigener weiterführender Überlegungen wurde der Kriterienkatalog entwickelt. Anschließend wurde er anhand der empirischen Ergebnisse sowie der qualitativen Interviews iterativ angepasst und verbessert.

3.2.2. Aufbau des Kriterienkatalogs

Für eine bessere Verständlichkeit ist der Kriterienkatalog anhand verschiedener Bausteine aufgebaut, nämlich:

a. *Systemelemente* – Die höchste Ebene des Anforderungskatalogs besteht aus den „Systemelementen“. Diese beschreiben die Kernelemente der Qualität bzw. Anreize, aus denen der Prozess der Qualitätsmessung bzw. der Festlegung von Anreizen besteht. Es handelt sich hierbei nicht um Anforderungen oder Kriterien sondern um Elemente des qualitätsorientierten Anreizsystems.

b. *Kriterienkategorien* – Die zweite Ebene besteht aus „Kriterienkategorien“, die jeweils den Systemelementen zugeordnet sind. Diese Ebene ist aus einer Subsummierung und Kategorisierung der Einzelkriterien entstanden und ist auch in anderen Katalogen (z. B. QUALIFY vgl. Reiter et al. 2008) zu finden.

c. *Kriterien* – Die Ebene der „Einzelkriterien“ oder „Kriterien“ ist die detaillierteste Ebene des Kataloges. Hier werden für die einzelnen Systemelemente und Kriterienkategorien konkrete Anforderungen beschrieben. Diese werden unter anderem mit Hilfe von Anwendungsbeispielen und Bewertungsoptionen näher detailliert.

---

[1] Hier u. a. auch ein „Algorithmus“ zur Überprüfung der Kompatibilität und Eignung von Indikatoren für verschiedene P4P-Systeme, deren einzelne Anforderungen jeweils bei den Kriterien erwähnt werden (vgl. Veit et al. 2012e, S. 42ff. sowie Veit et al. 2012f, S. 69ff.)

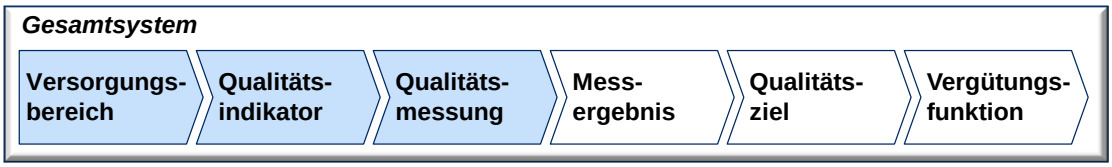
[2] In der ersten Entwicklungsphase wurden außerdem erste Ideen bei Horvath 2008 gefunden.

3.2.3. Systemelemente des Qualitätsteils des Kriterienkatalogs

Die Elemente von Anreizsystemen werden in der betriebswirtschaftlichen Literatur unterschiedlich definiert. Den meisten Darstellungen sind zwei Elemente (nämlich die „Bezugsgröße“ und die „Vergütungsfunktion“) gemein. Hier ist zunächst die sogenannte „Bezugsgröße“ als einer der wichtigsten Parameter des Systems (vgl. Becker 1990, S. 29) relevant. Diese Größe beschreibt das Element, anhand dessen die variable Vergütung gestaltet wird. In betrieblichen Anreizsystemen können dies z. B. finanzielle oder operative Erfolgsfaktoren sein (vgl. Becker 1990).

Für ein qualitätsorientiertes Anreizsystem besteht die Bezugsgröße – und somit der „Erfolg“ – aus Qualitätsaspekten. Um eine konkrete Bezugsgröße oder sogar ein zu erreichendes Ziel zu definieren, muss die stationäre Qualität somit messbar sein und auch tatsächlich gemessen werden. Ein Messergebnis für medizinische Qualität wiederum bedingt einen Qualitätsindikator und einen Messvorgang. Zusätzlich muss für die Auswahl eines Qualitätsindikators zunächst ein Versorgungsbereich definiert werden, auf den sich der Indikator bezieht. Auf diese drei „Qualitätselemente“ – Versorgungsbereich (Krankheitsbild und/oder Eingriff), Qualitätsindikator und Messung – wird im Folgenden eingegangen und die relevanten Anforderungen für diese Aspekte des qualitätsorientierten Anreizsystems diskutiert.

Abbildung 9: Systemelemente des Kriterienkatalogs (Fokus auf Qualitätsteil)



3.2.3.1. Systemelement 1 – Gesamtsystem

Das erste Systemelement besteht aus der übergreifende Ebene des Gesamtsystems. Das Gesamtsystem (d. h. ein System der Qualitätsmessung und -vergütung) besteht zunächst einmal aus der Summe aller weiteren Systemelemente – z. B. einem Messsystem, das mehrere Versorgungsbereiche und Indikatoren beinhaltet und anhand eines „Gesamtscores“ vergütet. Neben den Anforderungen, die sich für die einzelnen Systemelemente ergeben, soll diese Ebene ermöglichen, Anforderungen, die sich nicht aus der Summe aller Teile ergeben, darzustellen. So können Defizite, die einzelnen Indikatoren bspw. haben, in der Summe der Messergebnisse revidiert werden. Für das Gesamtsystem sind vor allem die Aussagekraft des Indikatorensystems bzw. deren Abdeckungsgrad relevant.

3.2.3.2. Systemelement 2 – Versorgungsbereich

Existierende Ansätze von qualitätsorientierten Anreizsystemen beziehen sich derzeit zumeist auf einen oder nur wenige Versorgungsbereiche bzw. Indikationen oder Krankheiten. Dies wird zu Teil auch als „Auswahl des Versorgungsaspekten“ bezeichnet (vgl. Blumenstock 2009, S. 12) und soll hier „Versorgungsbereich“ genannt werden.

Auch bei weitgreifenden Systemen werden für jeweilige Versorgungsbereiche spezifische Qualitätsindikatoren entwickelt (siehe auch Beschreibung der gängigsten Indikatorensysteme im vorherigen Kapitel). Systeme, die beispielsweise anhand von übergreifenden Indikatoren wie den „Hospital Standardized Mortality Ratios (HSMR)“ qualitätsspezifisch vergüten, existieren nur sehr vereinzelt. Solche krankenhausweiten Qualitätsindikatoren werden sehr kontrovers diskutiert und zeigen Schwierigkeiten in Validität, Aussagekraft und Nutzbarkeit für die Bewertung von Qualität (vgl. Shahian et al. 2010; Lilford/Pronovost 2010 und Black 2010). Daher wird im Folgenden zunächst davon ausgegangen, dass bei der Entwicklung eines qualitätsorientierten Anreizsystems die Auswahl des betroffenen Versorgungsbereiches ein relevantes Systemelement darstellt.

3.2.3.3. Systemelement 3 – Qualitätsindikator

Ein weiteres relevantes Steuerungsinstrument ist der Qualitätsindikator (vgl. Geraedts 2009). Um die Qualität vergüten zu können, muss zunächst definiert werden, anhand welches Indikators die Qualität gemessen werden soll.

Qualitätsindikatoren haben eine hohe Bedeutung für die Versorgungsqualität und stellen eine Form der „Klinischen Messgrößen“ dar (vgl. Jäckel 2009a, S. 8). Für Qualitätsindikatoren wurden bereits von verschiedenen Stellen Anforderungen und Mindestkriterien definiert und diskutiert (vgl. u. a. Jäckel 2009a; Reiter et al. 2007). In Deutschland ist hier vor allem das „QUALIFY“-Instrument der BQS zu erwähnen. Dieses Instrument bewertet methodisch die Fähigkeit von Qualitätsindikatoren, zuverlässige Aussagen zu Qualitätsunterschieden in der Versorgung zu tätigen (vgl. Lüngen/Rath 2011, S. 39). Wie bereits kurz erwähnt war dies eine der Hauptquellen für die Entwicklung des hier vorliegenden Katalogs. Die übernommenen und ggf. angepassten Kriterien werden in den folgenden Kapiteln näher diskutiert.

3.2.3.4. Systemelement 4 – Qualitätsmessung

Das vierte Systemelement der Qualität des entwickelten Kriterienkatalog besteht aus der Qualitätsmessung. Für die Anwendung der Qualitätsindikatoren muss der Messvorgang genau definiert und beschrieben werden. Einige Handbücher für Qualitätsindikatoren und deren Messung geben dies bereits vor (vgl. z. B. Mansky et al. 2011; AQUA-Institut GmbH 2010, S. 19ff.; AHRQ 2006).

Für das Systemelement der Qualitätsmessung ist dabei entscheidend, dass sowohl der Vorgang der Datenerhebung als auch der Datenanalyse und ggf. Ergebnisberechnung betrachtet wird. Die Tätigkeiten

der Datenerhebung und -analyse werden häufig bei unterschiedlichen Akteuren durchgeführt (nämlich die Datenerhebung in der Regel vom Leistungserbringer, die Datenanalyse dagegen beispielsweise von der SQG oder von der Krankenkasse). Die Kriterien und Anforderungen an die Qualitätsmessungen müssen diese Tatsache widerspiegeln – so ist bspw. das Kriterium der Manipulationssicherheit besonders für den Leistungserbringer relevant, die Frage des objektiven Messaufwands dagegen findet vermutlich meist beim Kostenträger Anwendung.

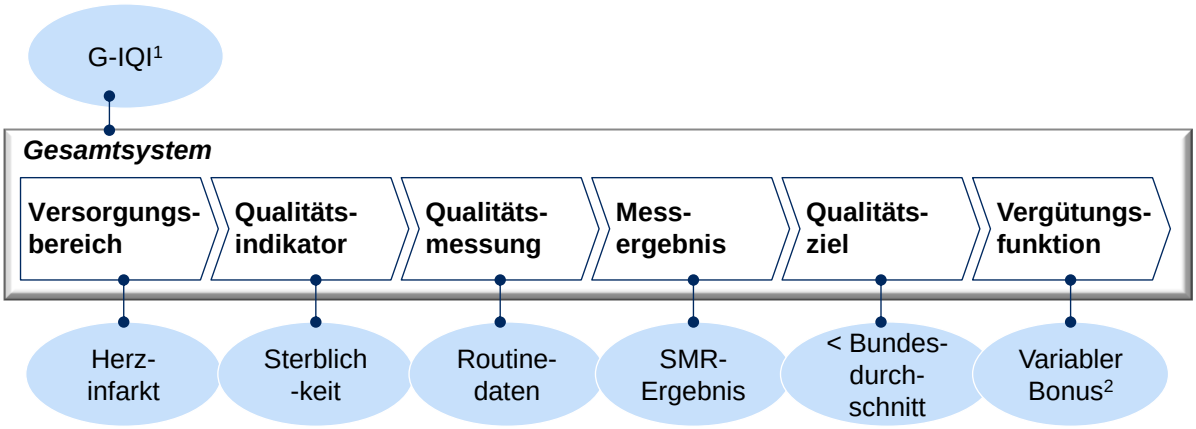
Für die Qualitätsmessung werden ferner klare Indikatordefinitionen benötigt. Solch eine detaillierte Definition der Indikatoren als Grundlage für die Datenanalyse und -berechnung findet sich beispielsweise bei G-IQI. Im Definitionshandbuch der Version 3.1 werden alle Indikatoren beschrieben und die technischen Details der Messung bzw. Berechnung dargelegt (vgl. Mansky et al. 2011). Die Messverfahren der externen, sektorübergreifenden Qualitätsmessung oder der AHRQ-Indikatorensysteme sind ebenfalls detailliert beschrieben (vgl. AQUA-Institut GmbH 2010, S. 19ff.; AHRQ 2006).

Auch für die Qualitätsmessung wurden bereits in der Literatur Kriterien formuliert. Diese finden sich jedoch meist in den Anforderungen für die Qualitätsindikatoren wieder (so z. B. in der detaillierten Beschreibung bei „QUALIFY“, Reiter et al. 2007). Damit der Kriterienkatalog für alle Akteure anwendbar und möglichst praktikabel ist, sollen die zwei Aspekte der Qualitätsindikatoren und -messung hier getrennt betrachtet werden. Dies ist besonders deswegen sinnvoll, weil die Entwicklung von Qualitätsindikatoren und die Definition insbesondere aber die praktische Umsetzung der Qualitätsmessung vorwiegend von unterschiedlichen Stellen durchgeführt wird. Ein umfassender Kriterienkatalog sollte daher beide Elemente diskutieren.

Sofern einzelne Kriterien sowohl für die Qualitätsindikatoren als auch die -messung relevant sind, wird dies im Folgenden beschrieben. Ein Beispiel ist hierfür die Validität: Wenn ein Qualitätsindikator so definiert wird, dass er valide Ergebnisse hervorbringen kann, muss auch die Qualitätsmessung – die ja dann das eigentlich Messergebnis ermöglicht – ebenso valide gestaltet werden.

Abbildung 10 zeigt beispielhaft die Beschreibung von Systemelementen mittels des G-IQI Messsystems inklusive einer innerbetrieblichen Umsetzung anhand einer Zielvereinbarung.

Abbildung 10: Beispiel für Systemelemente anhand des G-IQI-Systems



1 Vergütungssystem von Helios als Beispiel
2 Bspw. in Zielvereinbarung

3.2.4. Kriterienkategorien des Qualitätsteils des Kriterienkatalogs

Nach der Unterteilung in vier relevante Systemelemente für die Qualitätsaspekte (Gesamtsystem, Versorgungsbereich, Qualitätsindikator und Qualitätsmessung) wurde für den Kriterienkatalog eine weitere Kategorisierung vorgenommen: Kriterienkategorien und Einzelkriterien. Bei 17 Kriterien allein für die Qualitätsperspektive und weiteren 21 für den Anreizteil ermöglicht dies eine bessere Übersichtlichkeit und Verständlichkeit.

Bei einigen der existierenden Anforderungskataloge ist eine ähnliche Unterteilung zu finden, so z. B. mit „Kategorien“ und „Methodische Gütekriterien“ (vgl. Reiter et al. 2007). Bei QUALIFY wird auf der obersten Ebene mit den drei Kategorien „Relevanz, Wissenschaftlichkeit und Praktikabilität“ gearbeitet (vgl. Jäckel 2009a, S. 10).

Im dem hier neu zu entwickelnden Katalog wurden die Kriterienkategorien entsprechend der jeweiligen Systemelemente definiert. Methodisch wurden dabei zunächst die Einzelkriterien definiert und anschließend zu Kategorien zusammengefasst. Für die bessere Verständlichkeit sollen hier allerdings zunächst die Kriterienkategorien und anschließend die Einzelkriterien beschrieben werden.

Kriterienkategorie 1: Aussagekraft

Systemelement: Gesamtsystem / Versorgungsbereich

Gerade für die Auswahl des Versorgungsbereichs ist es von großer Bedeutung, eine möglichst hohe Aussagekraft der gewählten Krankheitsgebiete oder Behandlungen zu erreichen. Um ein stichhaltiges System zu entwickeln, sollten Versorgungsbereiche ausgewählt werden, die innerhalb eines medizinischen Fachgebietes entweder besonders häufige oder aber medizinisch oder ökonomisch besonders wichtige Leistungen beschreiben. Alle Versorgungsbereiche und Qualitätsindikatoren

zusammen sollten auch für das Gesamtsystem eine möglichst hohe Aussagekraft im Sinne eines hohen Abdeckungsgrades erreichen.

Alternativ zur Auswahl einzelner Versorgungsbereiche gibt es in der medizinischen Qualitätsanalyse auch umfassende Ansätze unter Einsatz von Gesamtindikatoren. Einer davon ist die sogenannte „Hospital Standardized Mortality Ratio (HSMR)“. Diese misst im Rahmen eines Top-Down-Ansatzes die Krankenhausqualität anhand eines singulären Indikators. Hierfür wird die gesamte Sterblichkeitsrate eines Leistungserbringers gemessen und adjustiert – der HSMR misst somit, ob die gemessene Sterblichkeitsrate eines Krankenhauses dem Erwartungswert entspricht, ähnlich wie die bereits diskutierte SMR. Differenzen zwischen den beiden Indikatoren ergeben sich aus dem Hauptansatz – Top-Down vs. Bottom-up: Während die SMRs für einzelne Krankheitsbilder „von unten“ rechnen, bewerten HSMRs die Sterblichkeitsrate für ein gesamtes Krankenhaus „von oben“, um anschließend unter Nutzung verschiedener Daten zu adjustieren, so z. B. (vgl. Knorr und Weyermann 2011):

1. Alter (in Jahren)
2. Geschlecht (männlich / weiblich)
3. Aufnahmekategorie (elektiv / Notfall)
4. Verlegungsstatus (zuverlegt: ja / nein)
5. Verweildauerkategorie (1 / 2 / 3-9 /10-15 / 16-21 / 22-365 Tage)
6. Relevante Diagnosegruppe (I20 Angina pectoris)
7. Komorbiditätsbeurteilung nach Charlson-Index (0 / 1 / 2)

Von Sir Brian Jarman am Imperial College 1999 in London entwickelt, werden HSMR in mehreren Ländern angewandt, so unter anderem bei Dr. Foster oder vom Canadian Institute for Health Information (CIHI). Zwar gibt es mit dem „HSMR“ einen Indikator, der die Qualität eines gesamten Krankenhauses darstellen soll, allerdings wird gerade die Aussagekraft dieses indikationsübergreifenden Indikators aufgrund von verschiedenen wissenschaftlichen Schwierigkeiten hinterfragt (vgl. unter anderem Lilford/Pronovost 2010; Bottle et al. 2011).

Die Schwierigkeiten der HSMR kann mit einem Vergleich anderer Industriequalitätskennzahlen erklärt werden. Die Medizin steht insgesamt vor ähnlichen Problemen wie andere Branchen mit stark differenziertem Produktspektrum: So würden Kunden eine Gesamtbewertung von Miele für nicht relevant empfinden, nur die Bewertung der einzelnen Modelle von Staubsaugern oder Backöfen können bei der Kaufentscheidung Ausschlag geben. Daher wird die Produktqualität von Stiftung Warentest z. B. nur für Einzelprodukte dargestellt, Bewertungen einer gesamten Produktlinie oder eines

Herstellerunternehmens sind nicht relevant, da die Produkte zu unterschiedlich sind[1]. In einem Krankenhaus werden auch sehr unterschiedliche Dienstleistungen angeboten – die Sterblichkeit (als möglicher Qualitätsindikator) schwerwiegender Krankheitsbilder in der Onkologie kann mit solchen der Augenheilkunde nicht verglichen werden. Ein Krankenhaus daher „Top-Down“ anhand der übergreifenden (wenn auch adjustierten) Sterblichkeit zu bewerten, führt zu Schwierigkeiten hinsichtlich der Wissenschaftlichkeit aber auch der Verständlichkeit und Aussagekraft.

Kriterienkategorie 2: Aussagekraft /„Das Richtige messen“

Systemelement: Qualitätsindikator

Das Kriterium der Aussagekraft gilt auch für den Qualitätsindikator. Die für das qualitätsorientierte Anreizsystem verwendeten Indikatoren sollen das betrachtete Krankheitsbild möglichst genau und hinsichtlich der gemessenen Qualitätskriterien homogen beschreiben, so dass für das Anreizsysteme „das Richtige“ gemessen wird. Gleichzeitig sollen die Indikatoren eine möglichst hohe Aussagekraft, wenn möglich hinsichtlich des Ergebnisses oder aber beispielsweise auch hinsichtlich wichtiger Prozessschritte haben.

Dies erfordert in der Medizin wenn möglich „risikohomogene“ Gruppen bzw. eine ausreichende Risikostratifizierung – bspw. durch die Bildung von einer recht risikohomogenen Gruppen bei „Cholezystektomie bei Gallensteinen ohne Tumorfälle, Anteil laparoskopischer OPs“ (vgl. Mansky et al. 2011, S. A-44). Soweit diese nicht in ausreichendem Maße zu bilden sind oder die Gruppe zu klein wird, muss eine geeignete Risikoadjustierung möglich sein. Gleichzeitig wird versucht, die Indikatorgruppen so umfassend wie möglich zu halten.

Diese Kriterienkategorie ist in anderer Gestaltung auch in einigen der bestehenden Anforderungskataloge vorzufinden. In QUALIFY beispielsweise wird es mit „Relevanz“ beschrieben (vgl. Reiter et al. 2007). Auch im „Health Care Quality Indicators Project (HCQI)“ der OECD findet sich dieses Kriterium wieder (vgl. Mattke et al. 2006). Hier wird Relevanz wie folgt definiert: „The impact of the construct captured by a quality measure on persons’ mortality, morbidity, and functioning and in some cases the cost of care.” (Mattke et al. 2006, S. 2).

Für den hier entwickelten Kriterienkatalog besteht die Kriterienkategorie „Aussagekraft” aus Einzelkriterien, von denen zwei unter die Definition von „Relevanz“ einzuordnen sind. Hinzu kommt die Möglichkeit eines Sektorbezuges. Somit ist die Kriterienkategorie der „Aussagekraft“ der Qualitätsindikatoren weiter gefasst als die reinen Anforderungen an die „Relevanz“.

[1] Vgl. zum Beispiel Stiftung Warentest, Test „Staubsauger“ vom 29.03.2012 oder „Backöfen“ 30.08.2012

Kriterienkategorie 3: Akzeptanz

Systemelement: Qualitätsindikator

Um eine erfolgreiche Umsetzung zu ermöglichen ist eine hohe Akzeptanz bei den beteiligten Akteuren notwendig. Sowohl für die „aktiven“ Umsetzer des Anreizsystems (d. h. vor allem Kostenträger und Leistungserbringer inklusive deren Arbeitsebene) als auch die „passiv“ Beteiligten (also z. B. Patienten und Öffentlichkeit) ist dies von großer Bedeutung.

Diese Kriterienkategorie wird nur in einigen bislang existierenden Katalogen in unterschiedlichen Aspekten adressiert: Zum einen wird beispielsweise bei QUALIFY die „Verständlichkeit und Interpretierbarkeit für Patienten und interessierte Öffentlichkeit“ unter dem Gütekriterium der „Praktikabilität“ zugeordnet (vgl. Reiter et al. 2007). Auch für die hier genannte Kriterienkategorie ist eine hohe Verständlichkeit des Indikators von Bedeutung. Doch zur Erreichung einer hohen Akzeptanz für alle Beteiligten sollte zusätzlich die wahrgenommene Bedeutung des Indikators möglichst hoch sein. Gerade der Aspekt der „subjektiven Wahrnehmung“ ist hierbei relevant. Auch bei anderen Auflistungen sind Thematiken wie „Konsens über Bedeutung der relevanten Versorgungsaspekte“ (Geraedts et al. 2002, S. 94 oder auch bei Campbell 2002, S. 360). Ferner wird die Akzeptanz von P4P als Gesamtsystem bei Veit diskutiert (vgl. Veit et al. 2012e, S. 61ff.). Hier werden als wichtige Faktoren bspw. Verständlichkeit des Programms, medizinische Relevanz, Bedeutsamkeit des Anreizes und Gerechtigkeit erwähnt – diese Aspekte werden in den Einzelkriterien diskutiert.

Dass neben den eigentlichen Kriterien für eine hohe Akzeptanz vor allem die Integration der beteiligten Gruppen in die Entwicklung der Indikatoren und des Gesamtsystems entscheidend ist, ist speziell für die Frage der Umsetzung und Implementierung eines qualitätsorientierten Anreizsystems von Bedeutung. Es ist kein eigenständiges Kriterium, kann sich aber mittelbar auf das Akzeptanzkriterium auswirken.

Kriterienkategorie 4: Wissenschaftlichkeit /„Richtig messen“

Systemelement: Qualitätsmessung (zum Teil auch Qualitätsindikator)

Neben der Kriterienkategorie 2 „Das Richtige messen“ ist es ebenso entscheidend, „richtig“ zu messen, also den Messprozess sowie die Spezifizierung der Qualitätsindikatoren unter wissenschaftlichen Aspekten korrekt durchzuführen.

Die „Wissenschaftlichkeit“ ist somit ein Kriterienkategorie, das gleich zwei Systemelemente betreffen kann – den Qualitätsindikator und die Qualitätsmessung. Die Qualitätsmessung und deren Ergebnisse sollten beispielsweise eine hohe Reliabilität aufweisen sollten, und ein Qualitätsindikator sollte durch die Adjustierung von möglichen Einflussgrößen risikoadjustierbar sein.

Diese „Gütekriterien der empirischen Sozialforschung“ sind allgemein verwendete Kriterien zur Bewertung der „Wissenschaftlichkeit“ (vgl. unter anderem Albers 2006). Diese Kriterienkategorie findet sich in ähnlicher Form u. a. auch bei QUALIFY wieder (vgl. Reiter et al. 2008). Unterschiede ergeben sich vor allem in den Einzelkriterien, die in den einzelnen Abschnitten diskutiert werden.

Kriterienkategorie 5: Messaufwand und -sicherheit

Systemelement: Qualitätsmessung

Die letzte Kriterienkategorie ist der Messaufwand und die Messsicherheit. Während der Aufwand für den Messprozess eine der Kernfragen für die Umsetzbarkeit in den klinischen und betrieblichen Alltag darstellt (bei QUALIFY als „Praktikabilität“ bezeichnet – allerdings mit anderen Kriterien versehen, vgl. Reiter et al. 2008), ergibt sich aus der Messsicherheit vor allem eine Konsequenz für die Nutzbarkeit des Messergebnisses bspw. auch zu Management- oder Gerichtsbarkeitszwecken (Details hierzu im Kapitel 4).

Für den objektiven Messaufwand ist vorwiegend der Datenerfassungsprozess entscheidend. Auch die Einbindbarkeit in bestehende IT-Prozesse sowie ggf. die Mehrfachnutzung existierender Datenbestände sind hierfür relevant. Aus diesem Grunde ist für dieses Kriterium beispielsweise die Nutzung von Routinedaten (englisch „administrative data“) von Bedeutung (vgl. bspw. Wennberg et al. 1987; Mitchell et al. 1994; Hannan et al. 1997; Mansky 2009).

Für die Messsicherheit ist gerade im Rahmen eines Vergütungssystem ferner die Manipulationsresistenz äußerst entscheidend. Wichtig dafür sind der Erfassungsprozess, die Objektivität der Messung und die Überprüfbarkeit der Daten.

3.2.5. Beschreibung der Skala für die Kriterienbewertung

Um die Kriterien möglichst standardisiert bewerten zu können, muss ein detaillierter Evaluierungsprozess mit dem Ziel etabliert werden, die Kriterien möglichst objektiv und quantitativ zu bewerten. Hierzu kann besonders die Entwicklung einer Bewertungsskala dienen. Einige Einzelkriterien, die bereits quantitativ sind, sind dazu besonders geeignet (z. B. ökonomische Bedeutung). Langfristig wäre eine einheitliche (Intervall-)Skala und eventuell sogar international anerkannte, einheitliche Grenzwerte für die verschiedenen Kriterien wünschenswert (z. B. einen Mindestwert auf der Bewertungsskala bezüglich der ökonomischen Relevanz zur Entwicklung eines P4P-Systems). Eine solche quantitative Skalenbewertung würde langfristig sogar die Erstellung eines „gewichteten Gesamtscores“ zur Evaluierung ermöglichen und somit einen transparenten, einheitlichen Vergleich verschiedener Systeme – auch eine direkte Verknüpfung zur qualitätsorientierten Anpassung der Vergütung wäre auf dieser Grundlage erstrebenswert (vgl. auch Mansky 2010a).

Aufgrund der noch recht neuartigen Entwicklung eines umfassenden Anforderungskatalogs an P4P-Systeme jedoch, ist eine solche Bewertung anhand einer Skala derzeit nicht möglich. Die derzeit bestehenden Auflistungen nutzen meist die Delphi-Befragung eines Expertenpanels (vgl. Häder 2002), um die Bewertung durchzuführen (vgl. Mattke et al. 2006, S. 2 und Reiter et al. 2008). Auch für die vorliegende Kriterienbewertung wird eine standardisierte Expertenbefragung anhand der Delphi-Methode mit einer quantitativen Informationsgrundlage vorgeschlagen. Außerdem wird aufgrund einer fehlenden Intervallskala oder eines festgesetzten Grenzwertes zunächst die Ordinalskala von QUALIFY (vgl. Reiter et al. 2007) empfohlen:

Bewertungsstufen

1 = trifft nicht zu

2 = trifft eher nicht zu

3 = trifft eher zu

4 = trifft zu

Die Bewertung findet anhand von recherchierten Daten zu den Versorgungsbereichen durch qualitative Einschätzung statt und wird im Einzelnen detailliert diskutiert[1].

3.3. Der 1. Teil des Kriterienkatalogs: Beschreibung und Diskussion der Kriterien des Qualitätsteils

Tabelle 4 zeigt alle Kriterien des Katalogs im Überblick – in diesem Kapitel werden zunächst nur die Kriterien des Qualitätssystems 1a-6d beschrieben, die restlichen Kriterien des Anreizsystems werden in Kapitel 4 erläutert.

[1] Für die Anwendung des Katalogs in den Kapitel 5 werden ebenso zunächst möglichst detailliert quantitative Daten gesammelt. Außerdem werden qualitative Anhaltspunkte in den Experteninterviews genutzt. Die eigentliche Bewertung findet anschließend durch die Autorin statt und wird durch die Diskussion mit Experten und in einer „Forschungswerkstatt“ überprüft.

Tabelle 4: Überblicksdarstellung des Kriterienkatalogs

Systemelement	Kriterienkategorie	#	Einzelkriterium
Gesamtsystem	Aussagekraft	1a	Abdeckung des Gesamtsystems
Versorgungs-bereich	Aussagekraft	2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet
		2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich
Qualitäts-indikator	Aussagekraft/ "Das Richtige messen"	3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten
		3b	Relevanz des Indikators für ein bestimmtes Fachgebiet
		3c	Sektorbezug vs. sektorübergreifende Nutzung
	Akzeptanz	4a	Medizinische Bedeutung in der medizinischen Fachliteratur
		4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften
		4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern
		4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit
		4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit
Messung	Wissenschaftlichkeit/ "Richtig messen"	5a	Reliabilität
		5b	Validität
		5c	Risikoadjustierung und Risikoadjustierbarkeit
	Messaufwand und Messsicherheit	6a	Objektiver Erhebungsaufwand und -komplexität
		6b	Manipulationssicherheit / Objektivität der Messung
		6c	Grad der Integration in automatische Prozesse / IT-Systeme
Ziele	Wirksamkeit/ Zielgerichtetheit	7a	Zielkonsistenz
		7b	Zielkongruenz
		7c	Reichweite
	Operationalisier-barkeit	8a	Erreichbarkeit
		8b	Transparenz
		8c	Akzeptanz
		8d	Langfristige Anreizwirkung der Ziele
	Vergleichbarkeit	9a	Gerechtigkeit
Vergütungs-funktion	Wirksamkeit/ Zielgerichtetheit	10a	Effizienz
		10b	Effektivität
		10c	Flexibilität
	Operationalisier-barkeit	11a	Anteil der Vergütung
		11b	Planbarkeit
Messergebnis	Operationalisier-barkeit	12a	Nutzung zur Steuerung
		12b	Aktualität
		12c	Kontrollierbarkeit
		12d	Praktische Umsetzung im Controlling
		12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit
	Vergleichbarkeit	13a	Zeitverlauf
		13b	Querschnittsvergleich/Ranking
		13c	Justiziabilität

Kriterium 1a: Abdeckung des Gesamtsystems

Systemelement: Gesamtsystem

Kriterienkategorie: Aussagekraft

Beschreibung: Das erste Kriterium beschreibt die Aussagekraft des Gesamtsystems. Es soll bewertet werden, wie groß die Reichweite und Abdeckung des Gesamtsystems ist. Hierunter wird vor allem das Indikatorensystem verstanden, d. h. die Abdeckung der Summe aller Indikatoren und deren Messergebnisse.

Begründung für die Auswahl des Kriteriums: Nicht nur die Einzelindikatoren sind für die Bewertung eines qualitätsorientierten Vergütungssystem relevant, sondern auch das Zusammenspiel und die Summe aller Indikatoren.

Bezug zu existierenden Systemen: Während die Aussagekraft von Indikatoren in mehreren Auflistungen relevant ist (bspw. „Bedeutung des mit dem Qualitätsindikators erfassten Qualitätsmerkmals […]“ bei Reiter et al. 2007, S. 5; aber auch Mattke et al. 2006, S. 2; siehe vor allem Kriterium 3a). Allerdings wird hier nicht das Indikatorensystem als Gesamtsystem betrachtet. Nur bei Amelung werden „ganzheitliche Vergütungssysteme“ angesprochen, aber keine detaillierten Anforderungen definiert (vgl. Amelung/Zahn 2009, S. 17–18). Inwieweit alle Indikatoren zusammen einen besonders hohen Abdeckungsgrad bspw. aller Krankenhaus- oder Todesfälle erreichen, wird bislang noch nicht als Anforderung definiert, einige etablierte Indikatorensysteme weisen aber spezifisch auf diesen Aspekt hin (vgl. Mansky et al. 2011, S. A-11).

Anwendungsdefinition und -beispiele: Für dieses Kriterium können einige, recht gut quantifizierbare Bewertungsaspekte herangezogen werden:

- Anteil durch das Messsystem (vor allem die Indikatoren) abgedeckte Krankenhausfallzahlen und Todesfälle …
  - ... pro Leistungserbringer, bspw. für Maximalversorger oder Fachkliniken
  - ... pro Fachgebiet (siehe auch Kriterium 3b)
  - … deutschlandweit (vgl. bspw. Mansky et al. 2011, S. A-10f.)
- Anzahl der inkludierten Fachgebiete und der in den Fachgebieten relevanten Indikatoren (bspw. auch die Nutzung von wichtigen Sentinelereignissen, vgl. Kriterium 3a)

Skala und Bewertung: Es liegen bislang keine einheitlichen Grenzwerte vor. Daher wird empfohlen, mit einer Ordinalskala (bspw. im Rahmen eine Delphi-Befragung, vgl. Reiter et al. 2007, S. 42f.) zu bewerten. Langfristig jedoch kann dieses Kriterium sehr gut objektiviert werden, so ist eine Intervallskala mit ggfs. sogar internationalen Grenzwerten denkbar und erstrebenswert.

Informationsquellen: Zur Bewertung dieses Kriterium muss zunächst die Beschreibung des Indikator- und Vergütungssystems herangezogen werden. Anschließend können verschiedene Statistiken über die Häufigkeit von spezifischen Fallzahlen und Todesfällen genutzt werden (vgl. bspw. Spindler 2010, S. 350ff.; Mansky et al. 2011, S. A-11f.).

Kriterium 2a: Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet

Systemelement: Versorgungsbereich

Kriterienkategorie: Aussagekraft

Beschreibung: Das zweite Kriterium ist die medizinische Relevanz des Versorgungsbereichs, d. h. inwieweit das Krankheitsbild oder die Behandlung, die für den/die Qualitätsindikator(en) ausgewählt

wurde, bedeutsam sind – entweder für das ganze Versorgungssystem oder zumindest das Fachgebiet. Hierbei geht es noch nicht um den Qualitätsindikator selbst (siehe Kriterien 3a/b), sondern zunächst um das betrachtete Krankheitsbild bzw. den Eingriff (bspw. der Herzinfarkt als wichtiges Krankheitsbild einer Kardiologie).

Begründung für die Auswahl des Kriteriums: Als zu betrachtende Leistungsbereiche kommen entweder Krankheitsentitäten in Frage, die hinreichend häufig sind, oder solche Krankheitsbilder bzw. Operationen, die unabhängig von der Häufigkeit als sehr wesentlich in einem Fachgebiet wahrgenommen werden. Während die Häufigkeit objektivierbar ist, entspricht die Wichtigkeit einer subjektiven Wahrnehmung, wobei aber innerhalb eines Fachgebiets in der Regel wenige Differenzen in deren Wahrnehmung bestehen (z. B. Ösophagusresektionen in der Viszeralchirurgie werden fast immer als wichtig eingestuft, vgl. Bartels 2009, S. 780f.). Eine erfolgreiche Auswahl nach den genannten Kriterien führt in der Regel zu hoher Akzeptanz innerhalb des Fachgebietes aber auch bei den Patienten. Die Auswahl häufiger Leistungsbereiche stellt darüber hinaus sicher, dass ein möglichst hoher Anteil eines Fachgebiets abgedeckt werden kann.

Bezug zu existierenden Systemen: Am besten kann dieser Aspekt mit dem Begriff der „health importance“ verglichen werden, den das IOM beschreibt als „measure [that] should include the prevalence of the health condition to which it applies and the seriousness of the health outcome affected” (Hurtado et al. 2001, S. 81). Während das IOM hier auch die Fragen von Verständlichkeit und Beeinflussbarkeit mit subsummieren, soll es bei diesem Kriterium vor allem um die medizinische Bedeutung der Indikation gehen. Andere Auflistungen mit ähnlichen Anforderungen sind JCAHO 2006; Wollersheim et al. 2007, S. 16 und Remus et al. 2004, S. 8.

Anwendungsdefinition und -beispiele: Während bei der Relevanz des Indikators vor allem zu beachten ist, dass „wesentliche Aspekte der Lebensqualität, Morbidität oder Mortalität“ (vgl. Reiter et al. 2007) erfasst sind, können für die Auswahl des Versorgungsbereiches folgende Aspekte bedeutsam sein:

- Häufigkeit der Indikation im Versorgungsbereich (vor allem „Volkskrankheiten“ wie Magen-Darm-Grippe im ambulanten Bereich als einer der häufigsten Indikationen in Deutschland (vgl. Spindler 2010, S. 376)) oder kolorektale Karzinome als wichtige und häufige Ursache für Darmoperationen im stationären Bereich (bspw. nach G-IQI Definition 44,5%, vgl. Nimptsch/Mansky 2012a, S. 1453).
- Bekannte oder vermutete Versorgungsprobleme (beispielsweise eine hohe Fallzahl von unerwünschten Ereignissen z. B. im internationalen Vergleich, beispielsweise Revisionen bei Hüft-TEPS (fast 24.948 im Jahr 2010), die evtl. zum Teil vermeidbar wären (vgl. AQUA-Institut GmbH 2011a, S. 135)). Hier gibt es Anhaltspunkte dafür, dass die Revisionsraten in Deutschland zu hoch sind. Auch große Unterschiede in der „tatsächlichen“ und „möglichen“ Versorgungsqualität fallen hierunter (vgl. Hurtado et al. 2001, S. 83).

- Hohe Variabilität in der Versorgungsqualität (d. h. große Unterschiede der Qualität zwischen Leistungserbringern oder Regionen).
- Indikation mit besonderer Wichtigkeit bei Versorgungsqualität (Schwere der Gesundheitsschädigung bei Nicht-Erreichen des Qualitätsziels bzw. Schweregrad eines unerwünschten Ereignisfalles, beispielsweise Herzinfarkt oder andere Herzerkrankungen mit hoher Sterblichkeitsrate; Langzeitfolgen in der Frühgeborenenbetreuung), bei denen möglicherweise erhebliche Qualitätsverbesserungen zu erzielen sein könnten.
- Seltene, aber spezialisierte Verfahren, die im Fachgebiet als sogenannte „Königsdisziplin“ verstanden werden (so z. B. Ösophaguschirurgie, vgl. Mansky et al. 2011, S. A-9).

Diese Liste enthält Elemente der Definition des QUALIFY Kriteriums „Bedeutung des mit dem Qualitätsindikator erfassten Qualitätsmerkmals für die Patienten und das Versorgungssystem“ (vgl. Reiter et al. 2007) sowie Aspekte des IOM (vgl. Hurtado et al. 2001, S. 83ff.) und Ergebnisse aus Experteninterviews und weitergehenden Überlegungen.

Skala und Bewertung: Langfristig sollte das Kriterium der medizinischen Bedeutung anhand einer Intervallskala semiquantitativ bewertet werden. Außerdem wären zumindest für die häufigen Leistungsbereiche einheitliche (vielleicht sogar international-einheitliche) Grenzwerte für die Bewertung wünschenswert (d. h. eindeutige Definitionen bspw. ab welcher Häufigkeit einer Indikation der Versorgungsbereich von hoher Bedeutung ist). Da die Unterpunkte einzeln quantitativ bewertbar sind, wäre bei ausreichender Datenlage und vor allem festgelegten Grenzwerten sogar eine „gewichtete Scorebewertung“ der Unterpunkte möglich. Derzeit jedoch wird die Bewertung im Rahmen einer Delphi-Befragung mit der Ordinalskala nach QUALIFY empfohlen (vgl. Reiter et al. 2007).

Informationsquellen: Zur Erstellung einer ausreichenden Datenbasis zur Bewertung des Kriteriums stehen eine Vielzahl von Quellen offen: Neben bspw. Diagnose- und/oder Fallpauschalenstatistiken (vgl. Spindler 2010) oder Mortalitätszahlen können dies auch wissenschaftliche Publikationen zu evidenzbasierter Medizin sein – auch die Daten der externen Qualitätssicherung stehen zur Verfügung (vgl. AQUA-Institut GmbH 2011a).

Kriterium 2b: Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich

Systemelement: Versorgungsbereich

Kriterienkategorie: Aussagekraft

Beschreibung: Dieses Kriterium beschreibt die wirtschaftliche Relevanz des ausgewählten Versorgungsbereichs für die Gesamtversorgung oder einen Fachbereich, lässt sich anhand vorhandener Statistiken messen und ist damit objektivierbar.

Begründung: Jede Qualitätsdiskussion hat auch ökonomische Aspekte. Da die Qualität der Versorgung die Folgekosten eines Krankheitsbildes oder einer Behandlung beeinflussen kann (vgl. Malzahn et al. 2010, S. 131ff.), stellt sich die Frage einerseits nach der ökonomischen Bedeutung eines Versorgungsbereiches, andererseits aber insbesondere auch nach der potenziellen Beeinflussbarkeit der Ausgaben durch Qualitätsverbesserung. Für qualitätsorientierte Vergütungssysteme bedeutet dies, dass Systeme, die einen Versorgungsbereich mit hohen und stärker qualitätsabhängigen Kosten betreffen, vorzuziehen sind. Gleichzeitig hängt die Aussagekraft des Versorgungsbereiches auch nach gesellschaftlichem Verständnis neben der medizinischen Relevanz von der ökonomischen Bedeutung ab. Während die absolute ökonomische Bedeutung eines Versorgungsbereiches zumindest für eine Behandlungsepisode anhand vorhandener Statistiken messbar ist, sind die Folgekosten bzw. auch der Anteil der qualitätsvariablen Kosten oft nur schätzbar. Nur in Einzelfällen liegen hierzu bisher empirische Daten vor (vgl. u. a. Malzahn et al. 2010).

Die Wirtschaftlichkeitsbetrachtung ist auch hier eine Form der „Kosten-Effektivitäts-Analysen“, die gerade vor dem Hintergrund von Opportunitätskosten relevant sind (vgl. Büscher/Gerber 2010, S. 63).

Bezug zu existierenden Systemen: Die gesundheitsökonomische Evaluation ist ein eigenes wissenschaftliches Teilgebiet mit einer Vielzahl von Methoden und Ansätzen (vgl. unter anderem Zweifel et al. 2009; Schöffski 2007a; Greiner 2007; Büscher/Gerber 2010; Busse 2006; siehe auch Arbeiten zum „Gesundheitssatellitenkonto“ von Henke, vgl. Henke 2010). Hierbei werden für das Studiendesign verschiedene Kosten- und Nutzenaspekte berücksichtigt (vgl. Schöffski 2007b, S. 65). Während für den „Nutzen“ häufig „Quality-Adjusted-Life-Years (QUALY)“ oder „Disability-Adjusted-Life-Years (DALY)“ herangezogen werden, sind für die Kostenseite indirekte, direkte und intangible, externe und zukünftige Kosten relevant (vgl. Büscher/Gerber 2010). Verschiedene Modelle stehen dabei zur Evaluierung zur Verfügung: von reinen Kostenstudien (Krankheitskostenstudien – cost-of-illness-studie [COI]) bis hin zu Kosten-Nutzen-Analysen im engeren Sinne (cost-benefit analysis [CBA]) (vgl. Büscher/Gerber 2010). Für die hier zur Debatte stehende Auswahl eines adäquaten Versorgungsbereiches, der für das qualitätsorientierte Vergütungssystem genutzt wird, ist vor allem die Krankheitskostenstudie relevant (vgl. Henke et al. 1997, S. 123ff.). Hierbei werden die direkten und indirekten Kosten einer Krankheit als Ganzes betrachtet, um deren „gesellschaftliche Bedeutung“ (vgl. Schöffski 2007b, S. 67) bzw. in diesem Fall auch die Einsparpotenziale zu ermitteln. Der Fokus liegt hierbei auf den volkswirtschaftlichen Ausgaben für Krankheiten und nicht auf der Auswahl alternativer Behandlungsmethoden (vgl. Schöffski 2007b, S. 67).

Der ökonomische Aspekt der Bedeutung des Krankheitsbildes wird in den genannten Kriterienkatalogen derzeit nicht explizit als eigenes Kriterium betrachtet. Die Kostenaspekte werden z. B. bei QUALIFY im Kriterium „Bedeutung des mit dem Qualitätsindikator erfassten Qualitätsmerkmals für die Patienten und das Versorgungssystem“ als „hohe Kosten“ erwähnt, aber nicht weiter erläutert (vgl. Reiter et al. 2007). Möglicherweise hängt die bisherige geringere Berücksichtigung der wirtschaftlichen Aspekte damit

zusammen, dass die Kostensenkungswirkungen von Qualitätsverbesserungen noch zu wenig analysiert wurden. Datenlage und Auswertungsmöglichkeiten haben sich hier verbessert (vgl. z. B. AOK-Bundesverband et al. 2007, S. 11ff.), so dass die Bedeutung dieses Kriteriums vermutlich zunehmen wird.

Anwendungsdefinition und -beispiele: Das Kriterium der ökonomischen Bedeutung kann z. B. anhand der folgenden Aspekte definiert werden:

- Gesamtkosten der Erkrankung innerhalb einer Periode für das Gesundheitssystem (vgl. Schöffski 2007b, S. 69; z. B. „Revisionen oder Ersatz des Hüftgelenks“ mit einem „Erlösvolumen“ von ca. 960 Mio. EUR in 2008 und somit 2 % Anteil am Gesamterlösvolumen (vgl. Spindler 2010, S. 377))
- Kosten eines neu diagnostizierten Patienten über seine Restlebenszeit (vgl. Schöffski 2007b, S. 69; bspw. die aktuelle Diskussion ob die „Erhöhungshäufigkeit von Pflegestufen“ bei der Bewertung von Hüft-Endoprothesen-Implantationen hinzugezogen werden soll (vgl. Malzahn et al. 2010)).
- Gesamtkosten von potenziell vermeidbaren Qualitätsproblemen (bspw. von Revisionsoperationen nach Hüft-TEPs mit ca. 176 Mio. EUR oder 0,26 % der stationären Ausgaben in 2009 (vgl. Mansky/Seeger 2011, S. 201)). Während die Höhe dieser potenziell vermeidbaren Qualitätsprobleme messbar ist, kann die mögliche Reduktion dagegen nur geschätzt werden.

Skala und Bewertung: Während die Gesamtkosten von Patienten, Erkrankungen und potenziell vermeidbaren Qualitätsproblemen objektiv messbar sind, müssen die möglichen Reduktionen geschätzt werden. Die ökonomische Bedeutung der Indikation ist einerseits anhand der hier genannten Krankheitskostenanalyse objektiv messbar, so dass quantitative Relationen der Versorgungsbereiche offengelegt werden (vgl. Schöffski 2007b, S. 67). Da jedoch die möglichen Verbesserungen eingeschätzt werden müssen, ist die Definition eines Grenzwertes schwierig. Daher sollte zunächst die Ordinalskala von QUALIFY (vgl. Reiter et al. 2007) genutzt werden. Basis für diese Bewertung kann z. B. ein Expertpanel mit den bereits angesprochenen Daten und Evaluationen sein.

Informationsquellen: Wie auch schon im vorherigen Kriterium stehen hier eine Vielzahl von Quellen als Datenbasis zur Verfügung, die bei Schöffski 2007 beschrieben werden: Z. B. im „Top-down-Ansatz“ mit verschiedenen volkswirtschaftlichen Daten (z. B. Morbiditäts- und Mortalitätsstatistiken, Statistiken über durchschnittliche Krankenhausaufenthalte und Arbeitsunfähigkeitstage) und „Bottom-up“ anhand von Behandlungsrichtlinien von Fachgesellschaften bewertet mit repräsentativen Entgelten (z. B. Tagessätzen, Punktwerten oder Fallpauschalen für Einzelbehandlungen) (vgl. Schöffski 2007b, S. 67). Zusätzlich können noch die Kostenstatistiken der Kostenträger (neben den Krankenkassen auch z. B. die Pflegeversicherung) oder der Leistungserbringer (auch hier neben Krankenhäusern Daten weiterer Leistungserbringer im ambulanten, Reha- oder Pflegesektor) zur Verfügung stehen. Gerade einrichtungsübergreifende Ansätze wie bspw. von QSR können für dieses Kriterium einen großen Mehrwert liefern.

Kriterium 3a: Medizinische Bedeutung für das Behandlungsergebnis des Patienten

Systemelement: Qualitätsindikator

Kriterienkategorie: Aussagekraft/„Das Richtige messen“

Beschreibung: Für dieses Kriterium soll die medizinische Bedeutung des Indikators für das Behandlungsergebnis des Patienten, insbesondere aber für die Beeinflussbarkeit und Verbesserbarkeit des Behandlungsergebnisses beleuchtet werden.

Begründung: Ein qualitätsorientiertes Anreizsystem soll „Das Richtige“ messen und anreizen, (d. h. bspw. weniger Schmerzen, „saved lives“ oder Erhöhung der „QALYs“) – dies bedeutet erstens, dass der Qualitätsindikator vor allem verbesserbare Messergebnisse liefert. Wichtige Sachverhalte, die durch Qualitätsmaßnahmen nicht beeinflussbar sind, sind hier also nicht relevant (z. B. Sterblichkeit in Endphase von nicht-therapiebaren Krebserkrankungen). Zweitens sollten die Ergebnisse aus Patientensicht quantifizierbar sein, bspw. durch die Darstellung in „saved lives“ oder „QALYs“ (siehe auch Kriterium 4d). Drittens sollte ebenso die Korrelation zwischen der Erfüllung des Indikators und der Verbesserung des Behandlungsergebnisses bzw. des Krankheitsbildes hoch sein – dies ist besonders bei Ergebnisindikatoren relevant.

Bezug zu existierenden Systemen: In der medizinischen Literatur sowie den existierenden Bewertungssystemen wird dieses Kriterium ähnlich diskutiert (vgl. Reiter et al. 2008; Mainz 2003) – so unter anderem mit dem Kriterium „Bedeutung des mit dem Qualitätsindikator erfassten Qualitätsmerkmals für die Patienten und das Versorgungssystem“, „Indikatorevidenz“ und „Nutzen“ (vgl. Reiter et al. 2007, S. 5f.; auch bei Mainz: „indicators should be based on the best available evidence“ (Mainz 2003, S. 524), aber auch JCAHO 2006 und Remus et al. 2004). Gleichzeitig steht die Bedeutung dieses Kriteriums in der Literatur außer Frage (vgl. ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) 2001). Nur wenn ein Indikator auch tatsächlich das eigentliche Behandlungsergebnis wiedergibt, kann „Das Richtige“ gemessen werden und der Indikator eine hohe Aussagekraft aufweisen.

Die medizinische Bedeutung eines Indikators für das Behandlungsergebnis ist von verschiedenen Faktoren abhängig. Eingängig ist zunächst, dass Ergebnisindikatoren eine höhere Bedeutung haben können als Prozess- oder Strukturparameter (vgl. auch Abschnitt 2.1.2.4). Doch auch einige Prozessindikatoren können eine hohe medizinische Bedeutung haben, solange die Korrelation nachgewiesen ist oder werden kann. Je höher die Korrelation zwischen dem Indikator und dem Behandlungsergebnis, desto größer die medizinische Bedeutung. Auch Michael Porter beschreibt die Wichtigkeit der Messung von „value“ und „outcome“ im Gesundheitswesen mit einem klaren Fokus auf Ergebnismessung: „… value depends on results, not inputs, value in health care is measured by the outcomes achieved, not the volume of services delivered; […] process measurement and improvement are important tactics but are no substitutes for measuring outcomes and costs.” (Porter 2010, S. 2477).

Anwendungsdefinition und -beispiele: Zu bewerten sind beispielsweise:

- Mögliche Verbesserungen der Messergebnisse des Qualitätsindikators
- Nachweis und Stärke der Korrelation zwischen Indikator und Behandlungsergebnis anhand von klinischen Studien oder Studien der Epidemiologie und Versorgungsforschung
- Nutzung des Indikators durch anerkannte Fachgesellschaften oder etablierte Indikatorensysteme für die Qualitätsmessung (wie bspw. IQI)
- Nutzung des Indikators in der Wissenschaft zur Bewertung verschiedener Behandlungsmethoden
- Bei Fehlen solcher Studien, sind Ergebnisindikatoren den Struktur- und Prozessindikatoren vorzuziehen und ggf. Expertenmeinungen hinzuzuziehen

Meist besteht innerhalb von Fachgesellschaften Konsens bezüglich der medizinischen Bedeutung: Dass bspw. die Sterblichkeit nach Herzinfarkt ein relevanter Indikator ist, wird kaum angezweifelt. Andere Indikatoren hingegen –z. B. Erhebung eines histologischen Befundes nach Cholezystektomie (vgl. AQUA-Institut GmbH 2011a, S. 16) – werden von einigen Experten hinterfragt.

Skala und Bewertung: Die medizinische Bedeutung für das Behandlungsergebnis kann im Einzelfall recht objektiv anhand von wissenschaftlichen Studien bewertet werden. Eine Vergleichbarkeit verschiedener Indikatoren für verschiedene Behandlungen ist damit jedoch meist nicht gegeben. Daher kann hier langfristig von einem klaren Grenzwert nicht ausgegangen werden. Wie bereits bei den vorherigen Kriterien sollten hier die Bewertungsstufen mit Hilfe des Delphi-Verfahrens eingeordnet werden (vgl. Reiter et al. 2007).

Informationsquellen: Im Fokus der Informationsquellen sollte hier wissenschaftliche Literatur, d. h. wenn möglich randomisierte, doppelblinde klinische Studien mit hohem JADAD-Score (vgl. Jadad et al. 1996) stehen. Auch Studien der Epidemiologie oder Versorgungsforschung können mit einbezogen werden. Ebenso sollten die Ergebnisse der Entwicklung von etablierten Indikatorensystemen betrachtet werden. Die Ergebnisse sollten von einem Expertengremium analysiert und somit nach Fachdiskurs die Bewertung zur „medizinischen Bedeutung für das Behandlungsergebnis“ festgelegt werden.

Kriterium 3b: Relevanz des Indikators für ein bestimmtes Fachgebiet

Systemelement: Qualitätsindikator

Kriterienkategorie: Aussagekraft/„Das Richtige messen“

Beschreibung: Hier steht nicht das Behandlungsergebnis im Vordergrund sondern die Bedeutung für ein medizinisches Fachgebiet oder einen bestimmten Versorgungsbereich des Fachgebiets.

Begründung: Qualitätsindikatoren sollen selbstverständlich das Behandlungsergebnis für Patienten verbessern. Doch ebenso sollen die Indikatoren die medizinische Leistung von Fachabteilungen beim Leistungserbringer oder sogar einzelnen Ärzten darstellen. Um dies ausdrücken zu können, muss der Indikator für das gesamte Fachgebiet oder zumindest einen Bereich des Fachgebiets möglichst relevant und aussagekräftig sein. Gerade zur Nutzung des Katalogs für qualitätsorientierte Vergütungssysteme sind die Einbeziehung der Beteiligten und die Relevanz für die Akteure entscheidend.

Bezug zu existierenden Systemen: In den genannten Auflistungen ist die Bedeutung für ein Fachgebiet bisher nicht explizit genannt. Bei QUALIFY bspw. wird zwar die Einbeziehung von Informationen aus dem Indikator bei der Entscheidungsfindung von Ärzten erwähnt (vgl. Reiter et al. 2007, S. 6; aber auch Veit et al. 2012e, S. 41), die Fachbereiche sind aber nicht spezifisch aufgeführt. Selbstverständlich wird im Entwicklungsprozess verschiedentlich die Einbeziehung der späteren Nutzer beachtet (vgl. Blumenstock 2009), die Relevanz des Indikators für diese Nutzergruppen wird aber nicht ausdrücklich betrachtet. Daher kann dieses Kriterium wahrscheinlich als recht neuartig angesehen werden.

Anwendungsdefinition und -beispiele: Der Indikator wird hier anhand von folgenden Aspekten bewertet

- Anteil der Fälle innerhalb eines Fachgebiets, die durch den Indikator abgedeckt sind (so bspw. „Sterblichkeit nach Herzinfarkt"; X% Fälle einer Kardiologie sind üblicherweise Herzinfarkte, Y% aller Todesfälle einer Kardiologie betreffen Sterbefälle nach Herzinfarkt, dadurch lässt sich insgesamt eine hohe Bedeutung ableiten) – diese Aspekte haben auch bei der Entwicklung von G-IQI eine Rolle gespielt (vgl. Mansky et al. 2011)
- Aber auch: „wahrgenommene" Relevanz des Indikators als „Königsklasse" des Fachgebiets (bspw. Ösophagus-Operationen, die trotz geringer Fallzahl als wichtige Disziplin der Viszeralchirurgie gesehen und deren Indikatoren als relevant betrachtet werden, vgl. Mansky et al. 2011, S. A-9)
- Expertenmeinungen zur Relevanz des Indikators für das Fachgebiet (vor allem aus den Fachgesellschaften)

Eine weitere mögliche Bewertung könnte anhand der Kerngebiete der „Koryphäen" des Fachgebietes stattfinden. Diese können als repräsentative Vertreter des Fachgebietes angesehen werden, so dass häufig eine höhere Bedeutung zugewiesen wird. Ob dies objektiv gerechtfertigt ist, muss im Einzelfall evaluiert werden – rein praktisch sollte die Bedeutung zumindest betrachtet werden.

Skala und Bewertung: Auch bei diesem (neuen) Indikator sollte wie zuvor das Delphi-Verfahren unter besonderer Einbeziehung der Fachgesellschaften durchgeführt werden.

Informationsquellen: Als Informationsquellen können hier vor allem Daten der Fachabteilung von Leistungserbringern sowie Expertenmeinung von Fachgesellschaften dienen, aber auch Bundesstatistiken und wissenschaftliche Literatur zu bspw. Behandlungsmethoden.

Kriterium 3c: Sektorbezug vs. sektorübergreifende Nutzung

Systemelement: Qualitätsindikator

Kriterienkategorie: Aussagekraft/„Das Richtige messen“

Beschreibung: Dieses Kriterium analysiert, ob ein Indikator die Möglichkeit dazu bietet, auch sektorübergreifend genutzt zu werden, oder er lediglich einen klaren Bezug zu einem bestimmten Versorgungssektor ausweist.

Begründung: Obschon in dieser Arbeit vor allem der stationäre Sektor im Vordergrund steht, sind für den gesamten Behandlungsverlauf eines Patienten alle Sektoren relevant. Beginnend mit dem ambulanten Sektor und weiterführend mit Rehabilitations- und Pflegemaßnahmen gehören viele Akteure zur Gesamtbehandlung. Nur wenn die Aktivitäten aller Sektoren zusammengefügt werden, kann von einem gesamten Behandlungsergebnis gesprochen werden, das für den Patienten den höchsten Mehrwert bringt. Auf diesen Aspekt wird bereits bei Porter/Teisberg hingewiesen (vgl. Porter/Teisberg 2006; Porter/Guth 2012). Um eine sektorübergreifende Evaluation zu ermöglichen, müssen Qualitätsindikatoren auch in der Lage sein, die Aktivitäten aus verschiedenen Sektoren darzustellen. Die sektorübergreifende Sicht ist für die Patienten auf jeden Fall vorteilhaft, weil für diese das gesamte Behandlungsergebnis wichtig ist. Dennoch kann es sein, dass für die praktische Anwendung auch einzelne, stark sektorbezogene Indikatoren wichtig sind.

Bezug zu existierenden Systemen: Wie bereits beschrieben, betrachten einige der existierenden Ansätze von Kriterienkatalogen bestimmte Sektoren (z. B. Thomeczek 2002 für den ambulanten Sektor). Diese Kataloge schließen eine sektorübergreifenden Ansatz derzeit nicht explizit als Kriterium mit ein, auch wenn sie in verschiedenen Sektoren nutzbar sein sollen (vgl. Reiter et al. 2007, S. 3; Döbler 2010). Auch eine Vielzahl der Indikatorensysteme sind sektoral ausgerichtet (G-IQI, IQI, PSI, AQUIK). Die Ansätze der OECD dagegen (HCQI und die dafür entwickelten Kriterien) versuchen, die gesamte Gesundheitsleistung eines Landes zu betrachten. Sektorübergreifende Indikatoren sind hier implizit relevant (vgl. Mattke et al. 2006; Arah et al. 2006; vgl. auch die neuen Ansätze von sektorübergreifender Qualitätssicherung in Deutschland (SQG) die sich zurzeit in der Entwicklungsphase befinden[1]). Diesen Aspekt ausdrücklich in den Kriterienkatalog mit aufzunehmen stellt daher eine Erweiterung dar, die zu einer differenzierten Auswahl zwischen sektorübergreifenden und sektorspezifischen Indikatoren genutzt werden kann.

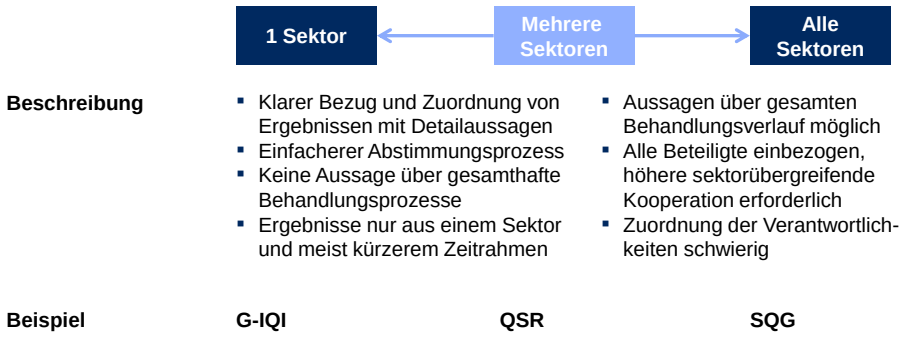
Anwendungsdefinition und -beispiele: Für dieses Kriterium soll überprüft werden, ob ein Indikator für eine sektorübergreifende Nutzung angewandt werden könnte bzw. bereits wird oder er spezifisch nur einen Sektor betrachtet. Die Indikatoren können nach dem Sektorbezug und der Anzahl der betroffenen Sektoren bewertet werden:

[1] Vgl. zum Projektstand http://www.sqg.de/entwicklung/projektstand/index.html, (zuletzt geprüft am 19.01.2013)

- Sektorbezug des Indikators ist unklar und nicht in der Definition enthalten (praktisch kaum relevant)
- Indikator bezieht sich auf einen Sektor und ist dementsprechend definiert (bspw. G-IQI)
- Indikator bezieht die Daten mehrere Sektoren mit ein (bspw. QSR – übergreifend über mehrere Sektoren werden jetzt schon Patientendaten seitens der Krankenkassen analysiert und bewertet, die Ergebnisse werden derzeit noch vor allem auf den stationären Sektor angewendet, vgl. Heller et al. 2008, S. 171ff.)
- Indikator wird in einem Sektor genutzt kann aber auf mehrere Sektoren ausgeweitet werden (z. B. aufgrund der Orientierung auf Behandlungsergebnisse, die nicht auf einen Sektor begrenzt sind)
- Indikator wird für zwei oder mehrere Sektoren angewandt

Skala und Bewertung: Für die Bewertung wird hier eine dichotome Skala vorgeschlagen. Für die Evaluierung ist die Sicht des jeweiligen Anwenders des Kriterienkatalogs relevant, somit kann keine klare Einteilung entlang einer Intervallskala durchgeführt werden – sowohl eine sektorübergreifende als auch eine sektorbezogene Systemausrichtung kann im Einzelfall erwünscht und sinnvoll sein. Daher sollte die Bewertung entlang der vorliegenden dichotomen Skala vollzogen werden:

Abbildung 11: Dichotome Bewertungsskala für Sektorbezug



Wenn P4P-Systeme entlang dieser Skala bewertet werden, muss der Anwender hier vorneweg entscheiden, ob nur ein oder mehrere Sektoren einbezogen sein sollen. Aus Patientensicht ist eine möglichst große, sektorübergreifende Ausrichtung zunächst stets von Vorteil – so können Behandlungswege in mehreren Sektoren aufeinander abgestimmt und in ihrer Gesamtheit betrachtet werden. Für die einzelnen Akteure oder den Gesetzesgeber allerdings kann es teilweise sinnvoller sein, nur einen Sektor zu betrachten – so kann auf die unterschiedlichen Rahmenbedingungen besser eingegangen und die Zuständigkeiten bestimmt werden. Zwar ist eine sektorübergreifende Betrachtung dann nicht möglich, vor allem realpolitisch jedoch können so Systeme zeitnäher diskutiert, umgesetzt und bewertet werden. Auch für diesen recht neuartigen Indikator sollte zunächst ein Delphi-Verfahren mit Fachexperten durchgeführt werden, um die Ansätze zu pilotieren.

Informationsquellen: Für die Bewertung des Sektorbezuges sind vor allem die Definition und der Anwendungsbereich des Indikators sowie die verwendete Datenbasis entscheidend. Zusätzlich können auch Expertengespräche zur Einschätzung einer möglichen Ausweitbarkeit des Sektorbezuges dienen.

Kriterium 4a: Medizinische Bedeutung in der medizinischen Fachliteratur

Systemelement: Qualitätsindikator

Kriterienkategorie: Akzeptanz

Beschreibung: Dieses Kriterium betrachtet die wissenschaftliche Bedeutung von Ergebnismessungen, die als Qualitätsindikatoren genutzt werden. In der medizinischen Fachliteratur werden häufig Outcome-Kriterien verwendet, die in diesen Studien nicht als Qualitätsindikatoren, sondern als Messkriterien oder Endpunkte für die Studienergebnisse verstanden werden. Es ist naheliegend, dass Ergebnisindikatoren, die auf solchen Messwerten beruhen, bei den Fachbeteiligten auf eine prinzipiell hohe Akzeptanz stoßen sollten. Ergebnisindikatoren dieser Art lagen beispielsweise der Entwicklung der G-IQI Indikatoren zugrunde. Die allgemeine Definition von Qualitätsindikatoren schließt diese Indikatorensichtweise mit ein (vgl. Jäckel 2009b, S. 2).

Begründung: Die Akzeptanz der Indikatoren hat einen ähnlichen Stellenwert für die Umsetzung eines P4P-Systems wie die zuvor genannten Relevanzkriterien der Indikatoren. Nur wenn die Fachebene die in der Qualitätssicherung entworfenen Indikatoren auch als für ihr Fach bedeutsam und nutzbringend ansieht, können die Indikatoren überhaupt außerhalb der Qualitätszirkel eine Wirkung entfalten. Die Bedeutung für ein Fachgebiet zeigt sich in diesem Fall in der wissenschaftlichen Anwendung der Kriterien[1].

Bezug zu existierenden Systemen: Die Bedeutung von Verständlichkeit und Akzeptanz von Qualitätsindikatoren wird in einer Vielzahl von Kriteriensammlungen diskutiert (u. a. Reiter et al. 2007, S. 24ff.). Die medizinische Fachliteratur als Quelle für die Einschätzung der wahrgenommen Relevanz wird hierbei nicht erwähnt. Auch in anderer Literatur wird diese Anforderung derzeit nach bestem Wissen nicht diskutiert. Daher kann dieses Kriterium als recht neuartig bezeichnet werden.

Anwendungsdefinition und -beispiele: Eine Einschätzung dieses Kriteriums kann anhand von Literaturstudien und in Expertengesprächen erfolgen. Es sind folgende Ansatzpunkte geeignet:

- Analyse der Nutzungshäufigkeit eines Indikators für die Messung von Behandlungsergebnissen (z. B. Sterblichkeit bei herzchirurgischen Eingriffen) in der Fachliteratur – d. h. Überprüfung, wie oft ein

[1] Auch in den später dargestellten Interviews wird erwähnt, dass für die beteiligten Chefärzte die Argumentation einer Verwendung der Indikatoren in der eigenen Fachliteratur häufig sehr wirksam ist.

Indikator in wissenschaftlichen Studien angewandt wird, um bspw. unterschiedliche Behandlungsmethoden zu bewerten. Bei neueren Verfahren auch Trends in der Nutzung entsprechender Outcome-Messungen.

- Subjektive Einschätzung von Experten der betroffenen Fachbereiche

Skala und Bewertung: Das Kriterium lässt sich anhand von Literaturübersichten in gewissem Maße objektivieren. Da Außenstehenden eine vollständige Bewertung kaum möglich sein wird, sollte die Skala aber im Rahmen einer Expertenbewertung qualitativ gestaltet sein (vgl. auch vorherige Kriterien).

Informationsquellen: Für eine Bewertung dieses Kriteriums ist die Recherche in der medizinischen Fachliteratur die Hauptquelle. Zum einen über existierende Reviews aber auch durch die Analyse Primärliteratur (u. a. aus PubMed) kann die wissenschaftliche Bedeutung analysiert werden. Ggf. können auch qualitative Experteninterviews genutzt werden, um die subjektiven Einschätzungen von Fachleuten zu analysieren.

Kriterium 4b: Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften

Systemelement: Qualitätsindikator

Kriterienkategorie: Akzeptanz

Beschreibung: Das Kriterium bewertet die Akzeptanz der Indikatoren von medizinischen Fachgesellschaften und kann als Zwischenebene zwischen der medizinischen Literatur und dem medizinischen Personal gesehen werden.

Begründung: Medizinische Fachgesellschaften stellen einen Zusammenschluss von wissenschaftlich und praktisch-klinisch arbeitenden Medizinern dar, die nach Fachrichtung organisiert sind. In Deutschland hat fast jede Fachrichtung eine eigene Fachgesellschaft, 163 von diesen sind wiederum im „AWMF - Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.“ organisiert (vgl. www.awmf.org, zuletzt geprüft 14.01.2013). Diese haben für die politischen, gesellschaftlichen Diskussionen aber auch für alltägliche, klinische Empfehlungen eine hohe Relevanz und genießen meist ein hohes Ansehen unter Medizinern. Sobald eine Fachgesellschaft Qualitätsindikatoren akzeptiert, ist mit einer besseren Akzeptanz auch beim medizinischen Personal zu rechnen[1].

Bezug zu existierenden Systemen: Während die Verständlichkeit und Akzeptanz von Qualitätsindikatoren, wie bereits erwähnt, in Kriteriensammlungen Beachtung findet, ist die wahrgenommene Relevanz für die

[1] Dies wird auch in den später dargestellten Interviews diskutiert.

Fachgesellschaften nach bestem Wissen ein recht neuartiges Kriterium[1] - in den betrachteten Auflistungen wird diese Anforderung nicht genannt.

Anwendungsdefinition und -beispiele: Dieses Kriterium kann zum einen durch die Auswertung bestehender Literatur von den Fachgesellschaften (bspw. klinische Studien, Kommentare, Empfehlungen), zum anderen aber auch durch spezifische Interviews und ggfs. sogar Umfragen bewertet werden. Folgende Ansatzpunkte können hierfür hilfreich sein:

- Analyse der Nutzung(shäufigkeit) eines Indikators von der oder als Empfehlung der Fachgesellschaft, in denen der Indikator zur Messung des Behandlungsergebnisses genutzt wird
- Einschätzung von medizinischen Fachgesellschaften bezüglich des Qualitätsindikators (bspw. in Kommentaren und Empfehlungen) in existierender Literatur
- Ergebnisse von Experteninterviews oder spezifischen Umfragen zur Einschätzung zu den Indikatoren

Skala und Bewertung: Dieses Kriterium ist nur teilweise quantitativ messbar. Daher ist hier eine qualitative Skala notwendig und ein Expertenpanel für die Bewertung dieses neuartigen Kriteriums sinnvoll.

Informationsquellen: Für dieses Kriterium können zunächst die Publikationen und Kommentare von Fachgesellschaften genutzt werden, in denen Qualitätsindikatoren beschrieben werden. Zusätzlich können qualitative Experteninterviews als Informationsquelle dienen, langfristig sind sogar größer angelegte Umfragen denkbar.

Kriterium 4c: Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern

Systemelement: Qualitätsindikator

Kriterienkategorie: Akzeptanz

Beschreibung: Das Kriterium bewertet die Akzeptanz der Indikatoren bei den medizinischen Endanwendern (je nach Indikator Ärzte, Pflegekräfte und andere Beteiligte).

Begründung für die Auswahl des Kriteriums: Gerade für die Anwendung in einem qualitätsorientierten Vergütungssystem ist die Akzeptanz der Personen, die direkt an der Behandlung beteiligt und damit für die Behandlungsqualität primär verantwortlich sind, ausschlaggebend. Die Akzeptanz dieser Gruppe hängt von der wahrgenommenen Relevanz, dem Nutzen aber auch dem Messaufwand ab. Nur wenn der Nutzen des Indikators als höher empfunden wird als der Aufwand, kann der Indikator langfristig

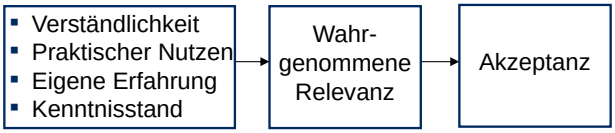
[1] Bei Veit wird zwar die Rolle der Fachgesellschaften diskutiert hier aber vor allem der Fokus auf Leitlinien gelegt – die Anforderung von „wahrgenommener medizinischer Relevanz“ wird nicht betrachtet (vgl. Veit et al. 2012e, S. 41).

umgesetzt werden. Nicht zu vergessen ist hierbei, dass der wahrgenommene Nutzen und die Relevanz zwar einerseits Voraussetzung für die Akzeptanz sein sollten, andererseits aber vom Informations- und Wissensstand des medizinischen Personals abhängen. Durch Weiterbildung bzw. Einbindung in den Entwicklungsprozess kann daher die Akzeptanz im Laufe des Verfahrens deutlich verbessert werden.

Bezug zu existierenden Systemen: Wie bereits erwähnt, wird die Thematik der Verständlichkeit und Interpretierbarkeit in mehreren Literaturstellen behandelt. Bei Reiter werden beispielsweise neben zwischen Patienten/interessierter Öffentlichkeit auch Ärzte/Pflegende betrachtet (vgl. Reiter et al. 2007). In anderen Sammlungen findet eine solche Unterteilung nicht statt (vgl. u. a. McGlynn 1998, S. 470; Mattke et al. 2006, S. 2ff.). Allerdings wird bei Campbell explizit auf die Akzeptanz der Indikatoren für die Personengruppen eingegangen, die beurteilt werden sollen (vgl. Campbell 2002, S. 361; siehe auch Amelung/Zahn 2009, S. 25).

Bei dem hier genannten Kriterium soll es nicht nur um die Verständlichkeit sondern direkt um die Einschätzung des medizinischen Personals gehen, nämlich die subjektiv wahrgenommene Relevanz. Um diese Einschätzung zu ermöglichen, ist Verständlichkeit zwar wichtig, aber nicht der einzige Faktor – auch der Kenntnisstand sowie ggf. die eigenen Erfahrungen des medizinischen Personals sind entscheidend. Ferner muss ein Nutzen im praktischen Qualitätsmanagement gegeben sein, d. h. die Messung und der Einsatz des Indikators sollten Verbesserungen ermöglichen und somit auch einen Nutzen für den Patienten haben. Für die Akzeptanz ist die wahrgenommene Relevanz ein notwendiges Kriterium. Daher kann die „wahrgenommene Relevanz“ als übergreifendes Kriterium zwischen der Akzeptanz und einzelnen Faktoren wie Verständlichkeit, Detaillierungsgrad der Definition, Kenntnisstand, Nutzen und ähnliches verstanden werden (siehe auch Abbildung 12).

Abbildung 12: Zusammenhang zwischen Verständlichkeit, Nutzen, Erfahrungen, Erkenntnisstand, wahrgenommener Relevanz und Akzeptanz



Anwendungsdefinition und -beispiele: Verschiedene Aspekte können bei der Anwendung des Kriteriums sinnvoll sein. Da diese Aspekte eher auf eine qualitative Einschätzung hinauslaufen, sind sie hier exemplarisch als Fragen formuliert, die beispielsweise in einer Umfrage oder in qualitativen Interviews an medizinische Endanwender gestellt werden können – und zwar sowohl des ärztlichen als auch des pflegerischen Dienstes, sowie an weitere Beteiligte, wie medizinisch-technische Assistenten oder medizinische Dokumentationsassistenten. Gerade wenn der Indikator in Pilotprojekten bereits umgesetzt wird, können die Einschätzungen des beteiligten Personals einen hohen Mehrwert darstellen:

- Wie schätzen Sie die medizinische Bedeutung des Indikators für das Behandlungsergebnis ein?
  - Kennen Sie wissenschaftliche Erkenntnisse zum Indikator?
  - Welche persönlichen Erfahrungen haben Sie mit dem Indikator?
  - Verwenden Sie den Indikator bereits zur Abschätzung Ihrer Behandlungsprozesse?
- Wie schätzen Sie den Nutzen des Indikators für den Patienten ein?
- Wie schätzen Sie den Nutzen des Indikators für die eigene Fachabteilung ein?
  - Welche Mitarbeitergruppen nutzen den Indikator (bspw. auch Pflegekräfte)?
  - Werden die Indikatorergebnisse im Alltag diskutiert und zur Verbesserung eingesetzt (z. B. M&M-Konferenzen)?
- Überwiegt der Mehrnutzen den Mehraufwand für die Dokumentation und Nachverfolgung des Indikators?

Skala und Bewertung: Die Bewertung des Kriteriums ist nur anhand von qualitativen Recherchen und Analysen möglich. Daher ist auch eine qualitative Skala notwendig. Auch ein Expertenpanel für die Bewertung dieses Kriteriums ist sinnvoll – gerade für dieses Kriterium ist die Einbindung von Vertretern des medizinischen Personals entscheidend. Bei bereits länger im Einsatz befindlichen Indikatoren sind repräsentative Umfragen unter medizinischem Personal erstrebenswert, die eine bessere Einschätzung ermöglichen können.

Informationsquellen: Neben den bereits erwähnten Umfragen und der Einbindung von Vertretern des medizinischen Personals stehen weitere Informationsquellen zur Verfügung, so z. B. Aussagen von Vertreterverbänden in der Presse oder detaillierte, qualitative Tiefeninterviews, die geführt werden können. Quantitative Daten sind für dieses Kriterium vermutlich (noch) nicht vorhanden.

Kriterium 4d: Verständlichkeit für Patienten und interessierte Öffentlichkeit

Systemelement: Qualitätsindikator

Kriterienkategorie: Akzeptanz

Beschreibung: Ein weiteres Akzeptanzkriterium ist die Verständlichkeit des Indikators für Patienten und interessierte Öffentlichkeit.

Begründung für die Auswahl des Kriteriums: Indikatoren, die sich auch an die Patienten und die Öffentlichkeit richten sollen, müssen für diese prinzipiell verständlich sein. Dies gilt daher vor allem für Indikatoren, die zur Veröffentlichung in Qualitätsberichten gedacht sind oder im Rahmen von Qualitäts-Rankings eingesetzt werden sollen (beispielsweise „Public Reportings“, vgl. Emmert 2008, S. 259ff.;

Geraedts et al. 2009, S. 232). Hier ist zwischen der Verständlichkeit für die Patienten und der zuvor genannten für die medizinischen Endanwender zu differenzieren. Es wird zwar aus Transparenzgründen vielfach verlangt, dass alle fachlichen Indikatoren auch veröffentlicht werden. Dennoch ist festzuhalten, dass nicht jeder fachlich sinnvolle bzw. evtl. für P4P eingesetzte Indikator außerhalb der beteiligten Fachebene zwangsläufig veröffentlicht werden muss.

Wenn Indikatoren für medizinische Laien veröffentlicht werden sollen, hängt angesichts der in der Regel unvermeidbaren hohen fachlichen Komplexität medizinischer Indikatoren die Verständlichkeit für die Patienten von sehr guten Erklärungen ggf. mit konkreten Beispielen ab. Der Inhalt des Indikators bleibt davon unbeeinflusst.

Bezug zu existierenden Systemen: Wie bereits erwähnt, ist die Verständlichkeit des Indikators in mehreren Arbeiten ein wichtiges Thema. In den meisten der Kriterienauflistungen (z. B. McGlynn 1998; ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) 2001; Geraedts et al. 2002, S. 94; JCAHO 2006) wird allerdings hinsichtlich der Verständlichkeit nicht zwischen verschiedenen Zielgruppen unterschieden (anders bei Reiter et al. 2007, S. 24ff.; hier wird speziell auf Patienten und die interessierte Öffentlichkeit verwiesen). Doch der medizinische und wissenschaftliche Kenntnisstand sowie die praktischen Erfahrungen sind bei den verschiedenen Gruppen so unterschiedlich, dass die Kommunikation der Indikatoren nicht die gleiche sein kann. Alles in allem ist dieses Kriterium nicht neu. Da häufig davon ausgegangen wird, dass „Public Reporting“ ein notwendiger Teilaspekt von P4P-Systemen ist (vgl. Emmert 2008, S. 73ff.), ist diese Anforderung für die Anwendung in P4P-Systemen von hoher Relevanz.

Anwendungsdefinition und -beispiele: Dieses Kriterium kann wie folgt bewertet werden:

- Bewertung des Indikators durch Vertreterverbände (bspw. Patientenverband)
- Bewertung durch „Laienpanels“, d. h. sozusagen eine Expertengruppe, die in diesem Fall aus medizinischen Laien zusammengesetzt wäre
- Soweit Portalanbieter Nutzungsstatistiken verfügbar machen, kann die Häufigkeit der Nutzung eines Indikators in öffentlich zugänglichen Suchportalen einen indirekten Hinweis darauf geben, ob der Indikator verstanden (und auch für relevant gehalten) wird. Es ist hier allerdings zu beachten, dass solche Portale nicht nur von Laien, sondern auch von Fachleuten (z. B. Wettbewerber) genutzt werden, was diese Statistiken verfälschen kann.
- Geeignet konzipierte Umfragen können ebenfalls zur Bewertung genutzt werden.

Skala und Bewertung: Auch diese Bewertung muss zu großen Teilen qualitativ erfolgen. Nur die Ergebnisse von repräsentativen Umfragen (wie bspw. die Umfragen der Bertelsmann Stiftung im Rahmen des Gesundheitsmonitors, Böcken et al. 2013; Mansky 2013, S. 136ff.) könnten eine Quantifizierung ermöglichen. Derzeit ist aber eine qualitative Skala sinnvoll, die durch ein Expertenpanel bewertet.

Informationsquellen: Besonders für dieses Kriterium sind möglichst repräsentative Ergebnisse beispielsweise anhand von Umfragen wünschenswert. Auch Aussagen von Vertreterverbänden können hilfreich sein. Insgesamt ist es wünschenswert, die Informationsbasis bezüglich der Verständlichkeit in Deutschland zu erweitern – auch um für das nächste Kriterium ausreichende Erkenntnisse zu erhalten.

Kriterium 4e: Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit

Systemelement: Qualitätsindikator

Kriterienkategorie: Akzeptanz

Beschreibung: Die wahrgenommene medizinische Bedeutung durch die (interessierte) Öffentlichkeit ist neben der Verständlichkeit des Indikators für diese Zielgruppe entscheidend, um die notwendige Akzeptanz für die Umsetzung zu erlangen.

Begründung: Wie schon beim medizinischen Personal ist auch für die interessierte Öffentlichkeit die Verständlichkeit nur der erste Schritt zur Akzeptanz und damit einer erfolgreichen Umsetzung. Auch Öffentlichkeit, Patienten und politische Entscheidungsträger sind daran beteiligt, wenn ein qualitätsorientiertes Anreizsystem implementiert werden soll. Eine hohe wahrgenommene medizinische Relevanz eines Indikators kann die Umsetzung deutlich erleichtern. Es ist allerdings anzumerken, dass die wahrgenommene Relevanz durch die Öffentlichkeit ein Ergänzungskriterium ist. Einzelne Indikatoren in P4P-Systemen könnten sich bei Konsens der Experten auch auf Fachkriterien beziehen, die aus Sicht der Öffentlichkeit weniger relevant sind. Eine öffentliche Akzeptanz erleichtert hingegen die Einführung eines solchen Systems, und es dürfte kaum ein System denkbar sein, das bei allen Indikatoren auf die öffentliche Akzeptanz verzichtet.

Bezug zu existierenden Systemen: Die Herleitung des Kriteriums der Verständlichkeit wurde bereits beim vorherigen Kriterium erwähnt. Doch wie bereits bei der wahrgenommenen Relevanz des medizinischen Personals wird die wahrgenommene Relevanz in der Öffentlichkeit bislang nur von einer Auflistung indirekt thematisiert: Geraedts spricht von notwendigen Konsens bei allen Verantwortlichen – ob hierbei auch die Öffentlichkeit gemeint ist, wird nicht explizit erwähnt (vgl. Geraedts et al. 2002, S. 94).

Anwendungsdefinition und -beispiele: Im Gegensatz zur Anwendung des Kriteriums bezüglich des medizinischen Personals ist hier die breite Öffentlichkeit beteiligt. Die Einbeziehung anhand von Vertretern oder gar repräsentativen Umfragen ist schwieriger. Für die Anwendung des Kriteriums werden daher neben möglichen Fragen im Rahmen von Umfragen oder Fokusgruppeninterviews weitere Aspekte hinzugenommen, die die Relevanz für die breite Öffentlichkeit verbessern und bewerten können, so z. B.:

- Möglichkeit der Nutzung der Indikatorergebnisse für übergreifende, gesellschaftlich relevante Aussagen und Entscheidungen

- Objektivierbarkeit/Quantifizierbarkeit der Messergebnisse: Indikator und Ergebnisverbesserungen werden in quantifizierten, möglichst aggregierten Größen dargestellt:
  - Anzahl geretteter Leben (saved lives),
  - Anzahl verlängerter Lebensjahre,
  - Anzahl betroffener Patienten deutschlandweit oder regional,
  - Anzahl verhinderter Krankenhaustage, Anzahl verhinderter Wiederaufnahmen
- Gleichzeitig ist weiterhin eine subjektive Bewertung in der Öffentlichkeit bezüglich der wahrgenommenen Relevanz entscheidend (z. B. Aussage in den Medien zur Relevanz des Indikators)

Skala und Bewertung: Auch dieses Kriterium sollte zunächst anhand einer qualitativen Skala und mittels eines Expertenpanels bewertet werden. Hier sollten Vertreter der Öffentlichkeit (beispielweise Vertreter der Selbsthilfegruppen, zufällig ausgewählte Laien/Patienten, ggf. auch institutionelle Patientenvertreter) beteiligt werden. Gleichzeitig kann durch eine Quantifizierung der Indikatorergebnisse eine objektivere Aussage getätigt werden.

Informationsquellen: Für dieses Kriterium sind bislang kaum Informationen über die wahrgenommene Relevanz vorhanden. Langfristig sind Umfragen und Detailinterviews erstrebenswert. Für die Quantifizierung der Indikatorergebnisse als „Stellvertreter“ für die wahrgenommene Bedeutung können klinische Studien, Studien der Versorgungsforschung und ggf. statistische Daten über Prävalenz zur Hochrechnung genutzt werden.

Kriterium 5a: Reliabilität

Systemelement: Messung

Kriterienkategorie: Wissenschaftlichkeit /„Richtig messen“

Die nächsten drei Kriterien beschreiben die drei Hauptgütekriterien der empirischen Sozialforschung und betrachten die Qualitätsmessungen. Allerdings sind diese Gütekriterien sowohl in den existierenden Kriteriensammlungen (vgl. z. B. Reiter et al. 2007; Mattke et al. 2006) als auch in der Wissenschaftstheorie (vgl. unter anderem Himme 2006; Albers 2006; Schnell et al. 2005) vielfach diskutiert. Da sie für den Katalog entscheidende Kriterien darstellen, werden sie hier kurz dargestellt, ohne auf umfassendere Bereiche der Wissenschaftstheorie einzugehen.

Beschreibung: Die Reliabilität bezeichnet die Zuverlässigkeit einer Messung. Hiermit wird die Reproduzierbarkeit der Messergebnisse bei wiederholten Messvorgängen bezeichnet.

Begründung: Auch für ein qualitätsorientiertes Anreizsystem ist die wissenschaftliche Verlässlichkeit der Qualitätsmessungen unerlässlich. Nur bei zuverlässig messbaren Ergebnissen der Indikatorenmessungen

können Qualitätsziele gesetzt und akzeptiert werden. Gerade die Möglichkeit, die Ergebnisse zu wiederholen/reproduzieren ist für ein System wichtig, das auch von verschiedenen Stellen überprüfbar sein soll.

Bezug zu existierenden Systemen: Die Reliabilität wird in den meisten der Kriteriensammlungen explizit erwähnt (vgl. u. a. Campbell 2002, S. 361; Geraedts et al. 2002, S. 94; Mainz 2003, S. 524; Remus et al. 2004, S. 8; JCAHO 2006, S. 2ff.; Wollersheim et al. 2007, S. 16). Bei QUALIFY wurde sogar ein eigenes Instrument entwickelt, dass die Reliabilität auch ohne Mehrfacherhebung abschätzen soll (vgl. Reiter et al. 2007). Insgesamt ist die Reliabilität notwendig für einen Kriterienkatalog, neuartig ist dieses Kriterium nicht.

Anwendungsdefinition und -beispiele: In der Wissenschaftstheorie werden zwei Arten der Testung zur Messung der Reliabilität unterschieden (vgl. Schnell et al. 2005, S. 151ff.). Allerdings hängt es von der Art der Qualitätsindikatoren ab, inwieweit diese Art von Testung anwendbar ist. Bei z. B. teilweise auf subjektiver Einschätzung der behandelnden Ärzte beruhenden Indikatoren kann die Reliabilität ein wichtiges Kriterium sein (Beispiel: Einschätzung des Behandlungserfolges bei psychischen Erkrankungen). Bei technisch eindeutigen, gleichbleibend definierten Indikatoren ist sie bei sorgfältiger Durchführung dagegen definitionsgemäß gegeben. Dieses ist zum Beispiel bei der Auszählung des Anteils der Herzinfarkte, die mit Acetylsalicylsäure-Gabe entlassen werden, der Fall. Hier kommt es nur auf die verlässliche Durchführung der Messung an (Validität). Ähnliches gilt beispielsweise auch bei Qualitätsindikatoren, die über eine logisch eindeutige, EDV-basierte Definition aus administrativen Daten gemessen werden (Beispiel G-IQI). Hier muss die Richtigkeit des Computerprogrammes sichergestellt werden (Validität).

Nur die hinsichtlich der Reliabilität zu untersuchenden Indikatoren müssten nach folgenden Verfahren getestet werden:

- *„Test-Retest-Methode“* (Messung zu verschiedenen Zeitpunkten, vom gleichen Messinstrument): Diese Methode setzt wiederholte Messungen am gleichen Kollektiv voraus. Sie ist zwar aus der physiologischen oder pharmazeutischen Forschung bekannt, ein Einsatz bei Qualitätsmessinstrumenten ist vor allem im Rahmen von methodologischen Studien oder psychologischen Effekten von Behandlungsarten gebräuchlich (vgl. Hjermstad et al. 1995, S. 1249f.; Downs/Black 1998, S. 381ff.; Backhaus et al. 2002, S. 737f.). In einem laufenden, qualitätsorientierten Anreizsystem, das die Behandlungsqualität langfristig und kontinuierlich verbessern soll, ist diese Methode nicht mehr einsetzbar, da hier systematische Veränderungen stattfinden können und die Reliabilität auf diese Weise nicht festgestellt werden kann.

- *„Parallel-Test-Methode“* (Messung zur gleichen Zeit von zwei Messinstrumenten): Für die Anwendung bei der Qualitätsmessung eines P4P-Systems scheint diese Methode zunächst besser geeignet. Zur Bewertung von Qualitätsmesssystemen wurde die Methodik bereits eingesetzt.

Beispielsweise wurde anhand einer Patientenstichprobe in Massachusetts die Reliabilität der HSMR-Messung auf diese Weise bestimmt (vgl. Shahian et al. 2010, S. 2531f.). Sollte dies nicht möglich sein, könnte zunächst auf die Bewertungsmethode nach QUALIFY zurückgegriffen werden (vgl. Reiter et al. 2007, S. 13).

Skala und Bewertung: Die Durchführung der o. g. Tests ergibt Messzahlen für die Übereinstimmung der Messungen. Deren Bewertung sollte anhand eines Expertenpanels erfolgen, ggf. sogar auf Grundlage vergleichbarer Messinstrumente im Rahmen einer „Parallel-Test-Methode".

Informationsquellen: Neben der Möglichkeit, die Reliabilität der Messergebnisse selbst zu überprüfen, soll hier auch erwähnt werden, dass es bereits eine Vielzahl von Indikatoren und Messungen gibt, deren Reliabilität von verschiedenen Arbeiten aufgezeigt wurde (vgl. u. a. Williams et al. 2006a, S. 246ff.). Neben der Auswertung der Messungen und Daten stehen als Informationsquelle daher wissenschaftliche Literatur und Auswertungen zur Verfügung.

Kriterium 5b: Validität

Systemelement: Messung

Kriterienkategorie: Wissenschaftlichkeit /„Richtig messen"

Beschreibung: Mit der Validität wird die Gültigkeit einer Messung beschrieben. Es wird demnach gefragt, ob das gemessen wird, was gemessen werden soll (vgl. Himme 2006).

Begründung: Besonders bei einem qualitätsorientierten Anreizsystem ist eine hohe Validität notwendig. Sonst kann die notwendige Akzeptanz nicht erreicht werden.

Bezug zu existierenden Systemen: Wie auch schon bei der Reliabilität beinhalten die meisten Sammlungen das Kriterium der Validität – teilweise mit unterschiedlichen Definitionen oder Ausprägungen (vgl. Zorn/Ollenschläger 1999; Campbell 2002, S. 361; Geraedts et al. 2002, S. 94; Mainz 2003, S. 524; Remus et al. 2004, S. 8; Arah et al. 2006; JCAHO 2006, S. 2ff.; Mattke et al. 2006; Reiter et al. 2007; Wollersheim et al. 2007, S. 16). Dieses Kriterium ist daher nicht neuartig, aber notwendig.

Anwendungsdefinition und -beispiele: In der Literatur werden für die Validität bei nichtexperimentellen Studien zwischen drei Arten unterschieden und verschiedene Anwendungsmöglichkeiten angegeben (vgl. Himme 2006, S. 491ff.):

- *Inhaltsvalidität*: Hier ist gemeint, inwieweit die theoretische Beschreibung des Indikators auch das Gleiche wie die Messergebnisse wiedergibt – d. h. bspw. ob die Messung der Sterblichkeitsrate bei Herzinfarkt auch tatsächlich diese Sterblichkeit misst. Voraussetzung für die Anwendung des Kriteriums ist daher eine präzise Definition des Indikators und der Messung (vgl. Himme 2006, S.

491ff.). Auch bei QUALIFY wird die Validität vor allem als Inhaltsvalidität verstanden (vgl. Reiter et al. 2007).

- *Kriteriumsvalidität*: Wenn eine Messung eine hohe Kriteriumsvalidität nachweisen kann, so korreliert sie stark mit der Messung eines anderen „Außenkriteriums“ (vgl. Himme 2006, S. 491ff.). Beispielsweise könnte ein Prozessqualitätsindikator als „prognostizierendes“ Kriterium für einen Ergebnisindikator herangezogen werden. Dieses Kriterium muss aber auch valide und reliabel sein. Die Auswahl eines solchen Kriteriums gestaltet sich häufig als sehr kompliziert und soll hier nicht weiter betrachtet werden.

- *Konstruktvalidität*: Diese Validität bezeichnet inwieweit die Messung mit dem eigentlichen Konstrukt übereinstimmt und nicht durch weitere Faktoren beeinflusst wird (vgl. Himme 2006, S. 493ff.). In der Praxis kann sich beispielsweise die Frage ergeben, ob die Sterblichkeit beim Herzinfarkt tatsächlich ein Qualitätskriterium ist. Da aber Qualität kein klares, monofaktorielles Konstrukt ist und es keinen von den Messkriterien unabhängigen Qualitätsbegriff als „Goldstandard“ gibt, kann diese Art von Validität oft nicht angewandt werden. Unter der impliziten Annahme, dass es einen Goldstandard gäbe, könnte auch die Konstruktvalidität gemessen werden. In einer Publikation von Bradley et al. (Bradley et al. 2006) wird beispielsweise erkennbar, dass die „hospital mortality“ seitens der Autoren implizit als „Goldstandard“ verwendet wird, um die Konstruktvalidität von Indikatoren der Prozessqualität zu beurteilen.

Im Rahmen eines qualitätsorientierten Anreizsystem bietet sich vor allem an, die Kriterien auf Inhaltsvalidität zu überprüfen.

Skala und Bewertung: Wie auch schon in der wissenschaftstheoretischen Literatur vorgeschlagen, sollte auch hier ein Expertenpanel über die Validität einer Messung entscheiden (vgl. auch Himme 2006).

Informationsquellen: Gerade für die Validität ist die präzise Definition des Indikators und der Messung von großer Bedeutung und dient als wichtigste Informationsquelle. Wie bereits bei der Reliabilität sind auch hier zunächst weitere Studien heranzuziehen, die bereits die Validität gewisser Indikatoren untersucht haben (so bspw. von AHRQ 2006).

Kriterium 5c: Risikoadjustierung und Risikoadjustierbarkeit

Systemelement: Messung

Kriterienkategorie: Wissenschaftlichkeit /„Richtig messen“

Beschreibung: Die Adjustierung von Messungen und Indikatoren bezüglich möglicherweise „verfälschender“ besonderer Risiken von einzelnen Leistungserbringern oder Fachabteilungen ist ein viel thematisierter Aspekt, der in diesem Zusammenhang diskutiert werden soll.

Begründung: Gerade bei der Verwendung von Ergebnisqualität ist es entscheidend, dass die qualitätsorientierte Vergütung an Elementen gebunden ist, die möglichst nur von den beteiligten behandelnden Personen nicht aber vom Ausgangszustand des Patienten abhängen. Nur wenn das Behandlungsergebnis vom medizinischen Personal beeinflusst werden kann, ist ein Anreizsystem sinnvoll. Um lediglich den beeinflussbaren Teil des Behandlungsergebnisses zu messen, müssen Unterschiede aufgrund von speziellen, vorbestehenden Charakteristika der Patientengruppen herausgerechnet werden. Dazu dient die Risikoadjustierung. Eine Risikoadjustierung hat somit als Ziel einen „fairen Krankenhausvergleich“ (Heller 2008, S. 3).

In der Praxis erweist sich das Vorgehen als schwierig. Einerseits ist es schwer, alle denkbaren Faktoren, die das Ergebnis beeinflussen könnten, zu identifizieren und im nächsten Schritt auch sicher zu messen. Andererseits ist darauf zu achten, dass Komplikationen infolge schlechter Behandlungsqualität nicht ihrerseits als Risikofaktoren benannt und dann rechnerisch „wegadjustiert“ werden. Dies ist insofern manchmal schwierig, weil Erkrankungen, die in einem Kontext vorbestehende Risikofaktoren sein können, in einem anderen Kontext als Komplikationen gewertet werden müssen. Ein Beispiel wäre die Lungenentzündung (Pneumonie): Sie kann einerseits bei Krankenhausaufnahme bestehen, d. h. außerhalb des Krankenhauses erworben sein. Andererseits kann sie als Komplikation im Krankenhaus auftreten.

Bezug zu existierenden Systemen: Risikoadjustierungen sind nicht nur in der Medizin bekannt. Auch bspw. in der Erfolgsrechnung der Finanzbranche werden verschiedene finanzielle Risiken eingerechnet und adjustiert. Auch in der Gesundheitsbranche bezieht sich der Begriff der „Risikoadjustierung“ nicht allein auf Anpassungen im Rahmen von Qualitätsmessung. In Krankenhausbetriebsvergleichen werden ebenfalls Risikoadjustierungen durchgeführt, um im Controlling-Kontext „nur Patienten mit ähnlichem Risikoprofil bzw. nur Abteilungen/Krankenhäuser mit ähnlichen Patientenzusammensetzungen („Patientenmix“) zu vergleichen“ (Schneeweiss/Sangha 2000, S. 133). Auch für die Zuweisung von Geldmitteln (z. B. Gesundheitsfond in Deutschland oder Medicare in den USA) werden verschiedene Adjustierungen genutzt (z. B. im Rahmen des Morbi-RSA), die in amerikanischen Arbeiten als „risk adjustment“ bezeichnet werden (vgl. z. B. Newhouse et al. 1997).

Für die Qualitätsmessung sind Risikoadjustierungen ein entscheidendes Thema. Bei der Messung von Ergebnisqualität sind diese Anpassungen besonders wichtig (vgl. DeLong et al. 1997). Die existierenden Indikatorensysteme, die Ergebnisqualität messen, beinhalten alle eine Form der Risikoadjustierung. Bereits seit mehr als 50 Jahren werden verschiedene Methoden angewandt, um die unterschiedlichen Ausgangssituationen des Patientenguts adäquat anzupassen (vgl. Moehring/Anderson 2011). Insgesamt soll bei den Methoden versucht werden, Adjustierungen durchzuführen, die genau die Elemente der Ergebnisqualität aussortieren, die im nicht-beeinflussbaren Risiko der Beteiligten liegen – gleichzeitig sollen die beeinflussbaren Qualitätsunterschiede herausgearbeitet werden. Hierfür existiert eine Vielzahl von untersuchten Adjustierungsmöglichkeiten (vgl. DeLong et al. 1997; Flämig et al. 2008; Iezzoni 1997;

Moehring/Anderson 2011; Newhouse 1998; Elixhauser et al. 1998, S. 10f.). In der gesetzlichen Qualitätssicherung werden ebenso verschiedene Varianten der Risikoadjustierung angewandt: 36 % der Qualitätsindikatoren wurden 2005 durch Stratifizierung adjustiert, 16% durch risikostandardisierte Fallkonstellationen, 6 % durch einen additiven Score und 1 % durch logistische Regression – allerdings wurden auch 23 % nicht adjustiert, für weitere 18 % war keine Risikoadjustierung erforderlich (vgl. BQS-Institut 2007). Insgesamt gibt es Indikatoren, die nicht adjustiert werden müssen, da bspw. klare Zielwerte in der medizinischen Literatur zu finden sind (vgl. Mansky et al. 2011, S. A-14). Auch das Bilden von risikohomogenen Gruppen kann eine sinnvolle Anpassung darstellen.

Die Gewichtung zwischen einer erwünschten, darzustellenden Differenzierung und eines „ungerechten“, nicht vermeid- oder beeinflussbaren Risikos der Leistungserbringer führt oft zu Diskussionen. Und auch die „Güte“ der Elemente, die zur Risikoadjustierung genutzt werden, müssen untersucht und diskutiert werden. Wenn beispielsweise ein Indikator und seine Messung valide, reliabel und weitestgehend manipulationssicher sind, die Elemente der Risikoadjustierung diese Kriterien aber nicht erfüllen, so muss es zu verzerrten Ergebnissen kommen. Daher muss die Risikoadjustierung die gleichen (vor allem wissenschaftlichen) Kriterien wie die Indikatoren erfüllen.

Allgemeiner Konsens besteht meist beim Einsatz von Alter und Geschlecht zur Risikoadjustierung bzw. Risikostratifizierung (vgl. Iezzoni 2003; Bundesamt für Gesundheit [BAG] 2007, S. 19). Die Korrelation vieler Ergebnisqualitätsindikatoren mit dem Alter der jeweiligen Patientengruppe sind gut belegt und für alle beteiligten Akteure klar verständlich (vgl. u. a. Bundesamt für Gesundheit [BAG] 2007, S. 19). Außerdem ist die Adjustierung gut möglich, da die benötigten Informationen in den Daten der jeweiligen Leistungserbringer manipulationssicher zu finden sind. Nebendiagnosen dagegen werden nicht immer vollständig und vollzählig erfasst bzw. kodiert und können daher nur in geringerem Maße als manipulationssicher betrachtet werden. Bei weiteren Elementen muss stets untersucht werden, ob die möglichen Risikofaktoren nicht vermeidbare Komplikationen darstellen, die eben genau dargestellt und nicht etwa „weg-adjustiert“ werden sollen. Die Akzeptanz dieser Methoden muss wiederum durch die Einbeziehung vom medizinischen Personal erreicht werden. Auch weitere Themen der Risikoadjustierung müssen bedacht werden, so bspw. verschiedene Arten von „Risiko-Bias“ - unter anderem die „constant risk fallacy“ (vgl. Nicholl 2007).

Die meisten der betrachteten Kriteriensammlungen beinhalten die Diskussion von Risikoadjustierung (vgl. u. a. Geraedts et al. 2002, S. 94; Remus et al. 2004, S. 8; JCAHO 2006, S. 3). Auch dies zeigt die hohe Bedeutung, die einer Adjustierung beigemessen wird. Da dieses Kriterium notwendig aber nicht neuartig ist, soll für die Anwendungsdefinition und die Bewertung besonders auf das Kriterium bei QUALIFY zurückgegriffen werden.

Anwendungsdefinition und -beispiele: In Anlehnung an die Bewertungsmechanismen des Kriteriums „Risikoadjustierung“ bei QUALIFY (vgl. Reiter et al. 2007) werden hier folgende Aspekte vorgeschlagen:

- Elemente, die nichts über die Qualität des Leistungserbringer aussagen, werden adjustiert, d. h:
  - Elemente, die die Behandlungsqualität verzerren
  - Elemente, die vom Leistungserbringer nicht beeinflussbar sind
  - Die Elemente der Risikoadjustierung sind – ebenso wie die Messung der Indikatoren – valide und reliabel.
- Qualitätsunterschiede, die von unterschiedlichen Leistungen des Leistungserbringer abhängen, werden weiterhin dargestellt, d. h.:
  - Elemente, die eine Aussage über die tatsächliche Qualität des Leistungserbringers ermöglichen, bleiben differenzierbar
  - Im Behandlungsprozess erworbene Komplikationen werden nicht weg-adjustiert, sondern bleiben als Unterschiede sichtbar.
- Die Risikoadjustierung ist transparent dargestellt, so dass für die interessierte Öffentlichkeit erkennbar bleibt, wie die Ergebnisse interpretiert werden können.

Skala und Bewertung: Auch für dieses Kriterium wird zunächst eine qualitative Bewertungsskala für die Bewertung durch ein Expertenpanel vorgeschlagen. Dabei sind auch die Ergebnisse von anderen Bewertungen existierender Indikatorensysteme relevant. Die Qualität zusätzlicher Elemente in der Risikoadjustierung kann anhand der ROC-Kurve gemessen und beurteilt werden (vgl. Shaw et al. 2002, S. 1105f.).

Informationsquellen: Zur Bewertung des Kriteriums ist zunächst die Beschreibung der Risikoadjustierungsverfahren des Qualitätsindikators relevant. Die Ergebnisse von evidenzbasierter Medizin und systematischen Reviews können dabei genutzt werden. Außerdem sollte die Akzeptanz der Risikoadjustierung bei angewandten Indikatoren in der Praxis analysiert werden.

Kriterium 6a: Objektiver Messaufwand und -komplexität

Systemelement: Messung

Kriterienkategorie: Messaufwand und -sicherheit

Beschreibung: Die nächste Kriterienkategorie der Qualitätsmessung betrifft den Messaufwand und die Messkomplexität. Das Einzelkriterium soll objektiv messen, mit wie viel zeitlichem und Ressourcenaufwand die Erhebung und Messung einhergeht.

Begründung: Um die Messung des relevanten Qualitätsindikators realistisch umsetzen zu können, ist eine gewisse Praktikabilität eine der Grundbedingungen. Gerade der bei der Erhebung und Messung der

Qualität entstehende Aufwand ist dabei ein Thema, das von Praktikern sehr oft besprochen wird. Ein hoher Aufwand in der Dokumentation oder Datenerhebung ist ein häufiger Beschwerdegrund – gerade für medizinisches Personal, da die administrative Belastung als sehr hoch empfunden wird. Neben der subjektiven Wahrnehmung (die bereits im Kriterienkategorie der Akzeptanz besprochen wurde) sind für eine sachliche Diskussion über die Kosten-Nutzen-Fragen von Qualitätsmessungen bei der Qualitätsmessung vor allem die objektive Komplexität und der Aufwand entscheidend.

Bezug zu existierenden Systemen: Der Aufwand und die Komplexität der Qualitätsmessung werden in verschiedenen Kriteriensammlungen erwähnt (vgl. u. a. Remus et al. 2004, S. 8; JCAHO 2006, S. 4; Wollersheim et al. 2007, S. 16). Bei QUALIFY wird beim übergreifenden Kriterium der „Praktikabilität“ der „Erhebungsaufwand“ diskutiert – auch die „Datenverfügbarkeit“ geht in eine ähnliche Richtung (vgl. Reiter et al. 2007). Bei den Kriterien des Australian Council on Healthcare Standards (vgl. Collopy et al. 2000, S. 38f.) wird zudem definiert, dass eine „einfache“ Datenerhebung notwendig ist. Allerdings werden diese Kriterien zumeist mittels qualitativer Einschätzungen bewertet. Doch gerade eine möglichst hohe Quantifizierbarkeit in der Bewertung des Kriteriums ist nötig, um die Diskussion von Dokumentations- und Erhebungsaufwand zu objektivieren. Daher ist die nähere Auseinandersetzung mit einer Operationalisierung dieses Kriteriums notwendig.

Anwendungsdefinition und -beispiele: Um Messaufwand und -komplexität darzustellen und zu bewerten, sollten verschiedene Aspekte betrachtet werden. Neben der Unterscheidung von Messaufwand und -komplexität ist hierbei auf die Datenerhebung versus Datenauswertung und -analyse zu achten.

- <u>Aufwand der Qualitätsmessung</u>
- Datenerhebung (bspw. Dokumentation von klinischen Daten für Qualitätsindikatoren durch Leistungserbringer):
  - Mehraufwand durch die Dokumentation und/oder Datenerfassung, die für den Indikator und die Qualitätsmessung entstehen (in Minuten oder Vollzeitäquivalenten bzw. in Euro für Arbeitszeit)
  - Weitere Mehrkosten durch notwendige Investitionen oder laufende Ausgaben in Dokumentationssysteme (bspw. im KISS-System)
- Datenauswertung (bspw. im Rahmen von qualitätsorientierten Vergütungssystemen zum Vergleich verschiedener Leistungserbringer durchgeführt von Krankenkassen oder übergreifenden Institutionen):
  - Mehraufwand durch die Auswertung der Daten (in Minuten oder Vollzeitäquivalenten bzw. in Euro für die Arbeitszeit)
  - Weitere Mehrkosten durch Investitionen oder laufende Ausgaben für Auswertungsprogramme oder -software

- Weitere Mehrkosten und -aufwand, die für das Gesamtsystem (d. h. das System bestehend aus Leistungserbringer, Krankenkassen, etc.) entstehen (bspw. durch Datenüberprüfungen von Dritten, Zusammenführungen verschiedenen Datenbasen)
- Aufsummierung des entstehenden Mehraufwandes anhand von monetären Bewertungen (d. h. Mehraufwand für Leistungserbringer, Krankenkassen und das Gesamtsystem)

Zusätzlich sollte auch die Komplexität bewertet werden, die eine ergänzende Dimension zum Aufwand darstellt und ein wichtiges Thema bei der Akzeptanz der Qualitätsmessung ist.

- Komplexität der Qualitätsmessung
- Anzahl von Arbeitsschritten zur Umsetzung der Qualitätsmessung
- Anzahl von beteiligten Organisationen (bspw. Leistungserbringer und Krankenkasse), Abteilungen und Personen

Skala und Bewertung: Zunächst einmal kann abhängig von der Art der genutzten Datenbasis bereits evaluiert werden, ob es überhaupt einen Mehraufwand für die Messung gibt – so entfällt bspw. bei der Nutzung von administrativen Daten die Datenerhebung durch zusätzliche Dokumentation. Da dieses Kriterium neuartige Aspekte und Bewertungsdimensionen aufweist, ist trotz des prinzipiell objektivierbaren Ansatzes aber zunächst eine qualitative Bewertung und Einschätzung im Rahmen von Expertenpanels notwendig. Langfristig könnte eine quantifizierte Skala aufgestellt werden und eine Bewertung mit Grenzwerten stattfinden.

Informationsquellen: Nach eingehender Recherche existieren derzeit keine Untersuchungen des Erhebungs- und Auswertungsaufwandes in Deutschland, nur Aussagen zum vergleichbar geringerem Erhebungsaufwand bei Qualitätsmessung durch Routinedaten sind vorhanden (vgl. Heller et al. 2008, S. 172; Mansky 2010a, S. 7f.; Thölen et al. 2010, S. 937). Daher müssten die Informationen aus qualitativen Interviews oder empirischen Erhebungen genutzt werden. Auch die Ergebnisse dieser Dissertation und die Analyse von anderen Projekten können Anwendung finden.

Kriterium 6b: Manipulationssicherheit / Objektivität der Messung

Systemelement: Messung

Kriterienkategorie: Messaufwand und -sicherheit

Beschreibung: Neben dem Aufwand und der Komplexität der Qualitätsmessung ist ein weiteres entscheidendes Kriterium die Manipulationssicherheit und damit Objektivität der Messung.

Begründung: Möglichst manipulationsresistente Messungen der Qualität sind eine der Grundvoraussetzungen für ein funktionsfähiges qualitätsorientiertes Anreizsystem. Nur wenn die

Messung und damit das Messergebnis möglichst frei von bewussten und unbewussten Verzerrungen und Beeinflussungen sind, können Anreize geschaffen werden, die eine bessere Behandlungsqualität inzentivieren. Unerwünschtes Verhalten – wie beispielsweise eine Beeinflussung der Messung – darf die Ergebnisse und somit die Anreize und ggf. Vergütungen nicht verändern.

Bezug zu existierenden Systemen: Die Notwendigkeit für eine Manipulationssicherheit der Messung ergibt sich unter anderem aus den wissenschaftlichen Gütekriterien an Messungen. Neben der Reliabilität und Validität ist hierbei die Objektivität eine Kriterienkategorie (vgl. Himme 2006, S. 383ff.). Die Objektivität ist vorhanden, wenn „verschiedene Personen, die die Messungen unabhängig voneinander durchführen, zu den gleichen Messergebnissen gelangen (vgl. Himme 2009, S. 485). Hierbei wird zwischen „Durchführungs-“, „Auswertungs-“ und „Interpretationsobjektivität“ unterschieden (vgl. Brühl/Buch 2006, S. 10f.). Während die Auswertungs- und Interpretationsobjektivität vor allem ein technisches Element der Qualitätsmessung darstellen, kann die Durchführungsobjektivität bereits bei der Datenerhebung ihren Anfang finden.

Konkret bedeutet dies, dass neben der Unabhängigkeit der Personen, die die Messung durchführen, auch entscheidend ist, dass die Dokumentation der für die Qualitätsbeurteilung relevanten Sachverhalte möglichst wenig beeinflussbar von den beteiligten Personen ist. Da die Datendokumentation durch medizinisches Personal durchgeführt werden muss, und eine medizinische Dokumentation keine vollständige Objektivität gewährleisten kann, ist eine vollkommene Manipulationsfreiheit nicht erreichbar.

Die Beeinflussbarkeit der Datengrundlage kann vor allem durch folgende Aspekte adressiert werden:

1. *Richtige Anreize setzen/Fehlanreize vermeiden*: Wenn bspw. die Dokumentation von stationären Daten Abrechnungszwecken dient, so bestehen Anreize, z. B. auch Komplikationen vollständig zu dokumentieren, da diese im DRG-System, das „complications and comorbidities“ als schweregraderhöhend bewertet (vgl. InEK 2011, Anhang C, 1057), häufig die Vergütung erhöhen. Eine Vereinheitlichung von Abrechnungs- und Qualitätsmesssystem schafft daher eine Balance der Anreizwirkungen. Bei separaten Bewertungssystemen bestünde in der Qualitätsmessung ein Anreiz, Komplikationen nicht zu erfassen. Dieser wird bei der Verknüpfung mit dem Vergütungssystem zum Teil konterkariert. Eine Nichtdokumentation von prinzipiell die Vergütung beeinflussenden Komplikationen würde zwar die Qualitätsdarstellung verbessern, müsste aber gleichzeitig mit Nichtvergütung dieser Komplikationen „erkauft“ werden. Ein Fehlverhalten der Leistungserbringer würde somit für sich bereits einen Vergütungsabschlag beinhalten. Bei wirksam konzipierten P4P-Systemen wäre aber denkbar, dass Fehlanreize entstehen, da der mittelbare „P4P-Gewinn“ größer ist als der Abschlag durch Fehldokumentation. Da beispielsweise die administrativen Daten in deutschen Krankenhäusern zu Abrechnungszwecken

erhoben werden, sind sie manipulationsresistenter als separate Datenerhebungen für die Qualitätssicherung[1].

2. *Datenbasis und -qualität überprüfen*: Neben den richtigen Anreizen muss ein Mechanismus etabliert werden, der die Datenbasis überprüfen kann. Sowohl die Vollständigkeit und Vollzähligkeit als auch die Richtigkeit des Datensatzes sollte so gut wie möglich überprüfbar sein. So sind die Routinedaten deutscher Krankenhäuser zum einen durch Plausibilitäts-, Vollständig- und Vollzähligkeitsprüfungen sowohl auf Seiten der Leistungserbringer selbst, als auch (mit anderem Ziel) der Krankenkassen vergleichsweise gut überprüft. In Zweifelsfällen sind ferner MDK-Prüfungen möglich und werden in der Praxis häufig auch durchgeführt (2,5 Mio. Prüfungen der Krankenhausleistungen in 2011, ca. 11% alle Fälle in 2007, vgl. MDK 2012; Blum et al., S. 111). Insgesamt ergibt sich daraus ein dreistufiges Prüfverfahren für diese Daten. Gleichzeitig zeigen in der Presse diskutierte Abrechnungsskandale, dass bei falscher Dokumentation die negativen Konsequenzen für alle Beteiligte sehr hoch sein können (vgl. u. a. Transplantationsskandal 2013). Die Überprüfung und damit Datenqualität der Routinedaten sind demnach hoch. Auch für die Daten der gesetzlichen Qualitätssicherung gibt es Prüfverfahren, die u. a. Vollständigkeit, Vollzähligkeit, Widerspruchsfreiheit prüfen sollen (u. a. im „Konzept zur Sicherung der Dokumentationsqualität“, vgl. AQUA-Institut GmbH 2011b, S. 20ff.)

3. *Gesetzliche Regelungen*: In Deutschland ist die bewusste Manipulation von Daten im Krankenhaus als Ordnungswidrigkeit untersagt. Ob sie strafrechtlich verfolgbar ist, wird zumindest diskutiert - im Zusammenhang mit dem „Organspendeskandal“ in Göttingen wurde dies als „Strafbarkeitslücke“ definiert und die Justiz dazu aufgefordert, zu prüfen, ob das Strafrecht verschärft werden müsse (vgl. Transplantationsskandal 2013). Gleichzeitig ist die Verpflichtung zur Dokumentation bei Ärzten u. a. im ärztlichen Standesrecht definiert (vgl. §10 Musterberufsordnung, Transplantationsskandal 2013). In den USA dagegen werden Manipulationen auch strafrechtlich verfolgt (vgl. False Claims Act, 18 U.S.C. § 287 und False Statements Act, 18 U.S.C. §1001 aus den Federal Criminal Statutes). Ahndungen nach dem Zivilrecht sind ebenso auch auf persönlicher Ebene möglich (vgl. Civil False Claims Act, 31 U.S.C. §3729(a)). Gleichzeitig gibt es in den USA eine Vielzahl von Kunstfehlerprozessen (bei bis zu 20 % der Ärzte einer Fachrichtung in einem Jahr), die auch wissenschaftlich evaluiert werden (vgl. Jena et al. 2011).

4. *Weitere Maßnahmen*: Beispiele für eine Reduktion von Manipulationen in der Qualitätsmessung; „Death from all causes“, „all-cause mortality“: Es besteht wegen der verbreiteten Multimorbidität älterer Patienten allgemein nicht nur in der Qualitätssicherung das Problem, ob

[1] Vermeidung von Fehlanreizen u. a. auch erwähnt bei Veit et al. 2012e, S. 73

bestimmte Todesfälle einer bestimmten Ursache zuzuordnen sind. Diese Unsicherheit hat dazu geführt, dass in der Ergebnisbeurteilung häufig alle Todesfälle unabhängig vom konkreten Zusammenhang mit dem zu untersuchenden Auslöser betrachtet werden (vgl. u. a. AHRQ 2006; Black 2002, S. 167f.). Dies ist beispielsweise eine Maßnahme zur Reduktion des Manipulationspotenzials.

In den existierenden Katalogen wird trotz der offenkundig bestehenden Fehlanreize die Manipulationssicherheit nur teilweise angesprochen. Während QUALIFY davon spricht, ob die Richtigkeit, Vollständigkeit und Vollzähligkeit der Daten überprüft werden kann (vgl. Reiter et al. 2007), wird beim Australian Council for Health Care von der Unbeeinflussbarkeit der Qualitätsbeurteilung durch subjektive Bewertung gesprochen (vgl. Collopy et al. 2000). Der Beginn der Diskussion von Manipulationssicherheit in Deutschland ist v. a. bei Mansky zu finden (vgl. u. a. Mansky 2008, S. 155f.) – so auch die Diskussion von Manipulationsresistenz im Zusammenhang mit P4P (vgl. Mansky 2010a, S. 6ff.). Beim BQS-Gutachten wird auf Fehlanreize und Möglichkeiten für „fehlerhafte Prozesse“ und mögliche Nebenwirkungen von P4P eingegangen (vgl. Veit et al. 2012e, S. 22ff., 35ff.). Allerdings liegt der Fokus hier auf den Anreizen, auf die Auswahl von Messsystemen wird nur kurz eingegangen. Auf die Manipulationssicherheit der Messung wird bei nur einigen Arbeiten aus den Wirtschaftswissenschaften kurz hingewiesen, so u. a. bei Laux und Wünsch (vgl. Wünsch 2012, S. 28; Laux 2006, S. 90).

Anwendungsdefinition und -beispiele: Für die Bewertung sollte zunächst definiert werden, welche Arten der Manipulation vermieden oder reduziert werden sollen.

- *Bewusste Manipulation*: Bei etablierten, finanziellen Anreizen kann medizinisches Personal die Motivation haben, bewusst unwahr zu dokumentieren. Neben einer „Falschdokumentation“ ist hierbei vor allem das Weglassen von Information (z. B. Nicht-Codierung von Komplikationen) oder das Auslassen ganzer Datensätze (z. B. „Gaming“, vgl. Scheppach et al. 2011; „Regelverletzung: motiviert“, „Regelverletzung aus Eigeninteresse“, „Selektionen der Projektteilnehmer“, „Selektion der Patienten“, vgl. Veit et al. 2012e, S. 22ff., 35ff.) zu erwähnen. Speziell wenn die Dokumentation zur Veröffentlichung von Qualitätsergebnissen dient, kann es zu Verzerrungen kommen – z. B. bei der Darstellung von Komplikationen. Bei der gesetzlichen Qualitätssicherung wird z. B. ab einer Unterdokumentation von <95 % wird der Leistungserbringer mit Abschlagzahlungen pro nicht dokumentiertem Fall bestraft (vgl. § 8 Abs. 4 KHEntgG i.V.m. § 137 Abs. 1 Satz 2 SGB V) – die Komplikationsraten liegen aber oft unter 5 %. Das selektive Weglassen einzelner Komplikationsfälle fällt daher nicht unbedingt auf. Die tatsächlich beobachteten Meldefehler und Dokumentationsunterschiede sind zum Teil erheblich (vgl. u. a. AOK-Bundesverband et al. 2007, S. 209; Maass et al. 2011, S. 409; Heller/Schnell 2007). In einem P4P-System muss aufgrund der verstärkten Anreize mit einer Zunahme der Manipulationsraten gerechnet werden, so dass der Manipulationsresistenz eine besondere Bedeutung zukommt.

- *Unbewusste Manipulation*: Doch auch eine ungewollte „Manipulation“ ist aufgrund der Komplexität und des hohen Dokumentationsaufwands möglich. Vor allem Meldebiases können hierunter fallen. So zeigt sich bspw., dass bei Registern unbeabsichtigt bevorzugt bestimmte Patienten gemeldet werden, so dass es zu einer Ergebnisverzerrung kommt (oder „selection bias“ vgl. u. a. Appelros et al. 2003, S. 145; Tu et al. 2004, S. 1414)[1]. Dabei ist darauf zu achten, Überprüfungsmechanismen einzuführen, die bereits die dokumentierende Person „warnt“. Ungewollte Manipulationen können beispielsweise entstehen, wenn ein behandelnder Arzt einen Todesfall bei multimorbiden, älteren Patienten nicht auf ein zu meldendes Krankheitsbild bezieht und daher nicht meldet. Flüchtigkeitsfehler sind ebenso denkbar (vgl. Veit et al. 2012e, S. 22).

Insgesamt sollte bewertet werden, wie manipulationssicher die Datenerhebung für bestimmte Qualitätsindikatoren ist. Insbesondere sollten folgende Punkte Beachtung finden:

- *Beeinflussbarkeit der Datenbasis bezüglich des konkreten Indikators*: Wie stark kann die Datenbasis von den Personen beeinflusst werden, die von den (finanziellen) Anreizen betroffen sind? Beispiele:
  - leicht beeinflussbar: Meldung von Komplikationen in Melderegistern – Weglassen täuscht Verbesserung vor
  - weniger leicht beeinflussbar: Todesfälle in Routinedaten – „Weglassen“ des Falles ist hier nicht möglich, aber durch manipulative Kodierung könnte ein Todesfall in eine nicht zum Indikator gehörende Krankheitsgruppe „verschoben“ werden
  - schwer beeinflussbar: Reoperationen in Langzeitnachverfolgungen (z. B. bei QSR) – die erstbehandelnde Klinik hat keinen Einfluss, wenn die Operation in einer anderen Klinik erfolgt. Eine Manipulation bei Durchführung in der eigenen Klinik wäre prinzipiell möglich, würde bei Verknüpfung mit den Abrechnungsdaten aber die Nichtabrechnung der Reoperation bewirken und wäre somit automatisch mit einer „Strafzahlung“ verbunden.
- *Anreize der Dokumentation*: Gibt es sinnvolle „Gegenanreize“ um eine hohe Datenqualität zu gewährleisten?
- *Spezifizierbarkeit der Erfassungsqualität*: Kann die Qualität der Datenbasis, Datenerfassung und Qualitätsmessung definiert und spezifiziert werden (z. B. durch die Nutzung bereits überprüfter Datengrundlagen, wie Routinedaten)?
- *Überprüfbarkeit der Vollständig- und Vollzähligkeit und Plausibilität der Daten*: Können die Daten von externen Dritten überprüft werden bzw. werden sie bereits überprüft?

[1] Andere Studien zeigen außerdem die Unterschiede der Datenergebnisse von Registern und Routinedatenmessungen (vgl. Pajunen et al. 2005, S. 132; Koek et al. 2007, S. 755)

- *Juristische Maßnahmen*: Kann die vermutete „Strafbarkeitslücke“ für Datenveränderungen in Deutschland geschlossen werden (z. B. nach Vorbild des „Code of Federal Regulation“ in den USA) und kann dadurch außerdem die persönliche Verantwortung von Ärzten beim Leistungserbringer erhöht werden?

Skala und Bewertung: Da es sich um eine erweiterte Anforderung handelt, müssen Skala und Bewertungsverfahren zunächst qualitativ und mit Hilfe von Expertengruppen durchgeführt werden.

Informationsquellen: Neben der Beschreibung der Qualitätsmessung und der Definition des Qualitätsindikators können vor allem verschiedene Analysen und Meinungen zur Erfassungsqualität verschiedener, existierender Datengrundlagen genutzt werden. Wenn es sich um neue Datendokumentationen handelt, sollte neben Experteneinschätzungen auch die Meinung der beteiligten Personen (vor allem derer, die die Dokumentation durchführen und beurteilen oder überprüfen) eingeholt werden.

Kriterium 6c: Grad der Integration in automatische Prozesse / IT-Systeme

Systemelement: Messung

Kriterienkategorie: Messaufwand und -sicherheit

Beschreibung: Ein weiterer Aspekt des Messaufwandes kann in der praktischen Umsetzung eines möglichst aufwandsarmen Messsystems gesehen werden. Hierfür ist die Möglichkeit einer Integration in automatische Prozesse und IT-Systeme entscheidend.

Begründung: Die Relevanz eines möglichst geringen Aufwandes der Qualitätsmessung wurde bereits im Kriterium 6a besprochen. Eines der wichtigsten Aspekte von Messaufwand ist die Möglichkeit, die Qualitätsmessung und -auswertung in existierende, automatische Systeme und IT-Prozesse zu integrieren. Dies kann eine aufwandsarme Umsetzung der Erfassungs- und Messmechanismen ermöglichen, die durch die hohe Standardisierung neben dem geringen Aufwand eine hohe Messstabilität mit sich bringen kann. Dieser Aspekt ist für alle Beteiligten der Datenerfassung, Messung und Auswertung relevant, besonders jedoch für die Krankenhäuser und Krankenkassen.

IT-Systeme, die die Qualität für ein qualitätsorientiertes Vergütungssystem evaluieren sollen, sollten sowohl beim Leistungserbringer als auch bei den Krankenkassen in die etablierten EDV-Systeme integriert werden können. Die Integrierbarkeit des Qualitätsindikators, die eine programmierbare, algorithmisch eindeutige Definition voraussetzt, ist dafür entscheidend. Gerade für die Fallstudien aus dem deutschen Markt kann die Frage nach der Integrierbarkeit der Messung genutzt werden, um die Umsetzbarkeit abzuschätzen und auf Schwierigkeiten hinzuweisen, die auf für diesen Zweck unzureichenden vertraglichen Vereinbarungen beruhen. Erste Initiativen versuchen, möglichst einheitliche, standardisierte IT-Systeme bspw. für Selektivverträge zu nutzen, dies stellt sich jedoch als recht schwierig

heraus (vgl. Krüger-Brand/Osterloh 2011). Gleichzeitig konnte gezeigt werden, dass durch die Umsetzung von P4P-Systemen die strategische Relevanz von IT-Implementierung deutlich ansteigen kann (vgl. Williams et al. 2006b).

Bezug zu existierenden Systemen: Die bisherigen Kriterienkataloge beinhalten nach bestem Wissen das Kriterium der Integrierbarkeit in IT- und automatische Prozesse nur indirekt – nur das JCAHO beschreibt die Anforderung, dass die ein spezifischer Algorithmus für die Qualitätsmessung entwickelt werden sollte, auf die Integration in IT wird nicht eingegangen (vgl. JCAHO 2006, S. 2). Beim SVR wird zwar auf die finanzielle Einbeziehung der Umsetzung und u. a. EDV-Ausstattung eingegangen, die Integrierbarkeit wird nicht erwähnt (vgl. SVR 2007, S. 209). Zwar wird auf den Messaufwand mehrfach eingegangen (wie bereits erwähnt), weitere Aspekte der Integration, wie beispielsweise die technische Umsetzbarkeit, ggf. höhere Akzeptanz und größere Stabilität der Ergebnisse werden nicht erwähnt. Daher kann das Kriterium als recht neuartig betrachtet werden.

Anwendungsdefinition und -beispiele: Ob ein Messsystem in den vorhandenen IT-Systemen verankert ist, kann bei einem bereits angewandten System gut untersucht werden. Vor allem der Grad der Integration ist dabei entscheidend (neben dem bereits im vorangegangen Kriterium definierten Aufwand und Komplexität). Folgende Aspekte sind relevant:

- Mögliche Funktionen des integrierten Messsystems:
  - Automatische Erkennung der Versicherten und Fälle, die prinzipiell einer in Frage kommenden Vereinbarung (z. B. IV-Vertrag) unterliegen und für die Qualitätsverfolgung relevant sind (z. B. Erkennung von Versicherten mit bestimmten Krankheiten oder Eingriffen auf ICD oder OPS-Basis anhand der Abrechnungsdaten).
  - Automatische Überprüfung der Ergebnisse auf Vollständigkeit, Vollzähligkeit, Plausibilität (und damit Aufdecken von Manipulation) [vor allem relevant für die Nutzung bei Krankenkassen; z. B. Überwachung der Fälle mit einem vertragsrelevanten Eingriff, die außerhalb der Vereinbarung abgerechnet wurden]
  - Bei einzelfallbezogenen Vereinbarungen: Automatische Erkennung von Qualitätsabweichungen, die eine Vergütungs- oder Abrechnungswirksamkeit haben können und Definition des Vorgehens bei solchen Fällen. Dazu würde bspw. die automatische Erkennung (potenzieller) Gewährleistungsfälle mit standardisiertem Workflow gehören.
  - Bei kennzahlenbezogenen Vereinbarungen: Automatische Berechnung der Kennzahlen mit zugehöriger automatischer Erkennung der Indikatorfälle.
  - Generell ist eine vollständig IT-basierte, algorithmische Messung ohne Annahmen oder subjektive Einschätzungen der Bearbeiter anzustreben.

- Zur Vereinbarung passende, adäquate Auswertungsmöglichkeiten, die für das Management sowohl definierte Messergebnisse als auch Analyseergebnisse zu Controllingzwecken regelmäßig verfügbar machen.

➢ Gewünschte Eigenschaften der integrierten IT-Unterstützung:

- Einfache Bedienung und Verständlichkeit für die Sachbearbeiter
- Flexibilität bei notwendiger Anpassung an neue Gegebenheiten der Messung oder der Datenbasis (Beispiel: Aktualisierungen der Kodiersysteme oder Ausweitung der Verträge)
- Möglichst effiziente Integration in das bestehende IT-System (bspw. nur eine, intuitive Bedienungsmaske)

Sofern noch keine IT-Integration etabliert ist und sich die technische Umsetzung noch in der Entwicklung befindet, müssen weitere Aspekte betrachtet werden:

➢ Klare Definition der notwendigen Schritte für die Qualitätsmessung (so bspw. konkrete Beschreibung der Datenbasis, Auswahlparameter, möglichen Abgrenzungen von nicht relevanten Datensätzen)

➢ Überprüfung, inwieweit die genannten Funktionsmöglichkeiten und Eigenschaften im IT-System umsetzbar sind

➢ Einschätzung der Investitionskosten und laufenden Aufwendungen für die IT-Integration (inklusive der Personalkosten für Projekte zur Entwicklung des Systems)

Skala und Bewertung: Da es sich hierbei um ein recht neuartiges Kriterium handelt, sind wiederum eine qualitative Skala und Bewertungsrunden mit Hilfe von Experten sinnvoll. Langfristig jedoch, kann sicherlich eine objektive Bewertung angestrebt werden, die eine konkrete Skala hinsichtlich Umsetzbarkeit und Kosten mit sich bringt.

Informationsquellen: Vor allem die tatsächliche Umsetzung bei den Kassen- und Krankenhausvertretern sollte hier als Bewertungsgrundlage dienen. Ergebnisse wie die dieser Arbeit, die die praktische Implementierung untersuchen, sind ein wesentlicher Schritt. Sofern diese nicht existieren, müssen Experteninterviews durchgeführt werden.

3.4. Zwischenfazit: Der Qualitätsteil des Kriterienkatalogs ist zu vielen Teilen neuartig und ermöglicht eine neue Blickweise auf P4P-Systeme

Erste Erkenntnisse aus dem Qualitätsteil des Kriterienkatalogs lassen sich wie folgt zusammenfassen:

➢ Als Grundlage werden die Notwendigkeit und die Ziele eines qualitätsorientierten Kriterienkatalogs dargestellt. Außerdem werden die bereits existierenden Auflistungen von Anforderungen an

Qualitätsindikatoren kurz dargelegt. Eine nähere Betrachtung findet im Rahmen der einzelnen Kriterien statt.

- Durch die Strukturierung des Kriterienkatalogs anhand von Systemelementen wird verdeutlicht, wie ein qualitätsorientiertes Vergütungssystem funktioniert, und dass für die unterschiedlichen Elemente auch jeweils spezielle Anforderungen notwendig sind. Dies kann bereits als erste Neuerung gegenüber den bereits bestehenden Auflistungen verstanden werden, die sich nicht an den konkreten Elementen eines Vergütungssystems orientieren.

- Ein weiterer, neuartiger Aspekt ist die Konzeption des Katalogs nicht nur für Qualitätsindikatoren als solche, sondern speziell für die Verwendung für ein Vergütungssystem. Dies bringt neue Anforderungen und vor allem Anpassungen mit sich, die einen großen Mehrwert für das System darstellen können. Bislang wird eine solche am Anwendungszweck orientierte Konzeption nur im kürzlich erschienen Gutachten der BQS berücksichtigt (vgl. Veit et al. 2012e, S. 22ff.).

- Ebenso sind bereits verschiedene Kriterienkategorien neuartig, so beispielsweise der explizite Fokus auf die Akzeptanz des Systems seitens der Fachgesellschaften und der anderen Akteure im Gesundheitswesen.

- Insbesondere auf der Ebene der Einzelkriterien reicht der vorliegende Katalog über frühere Kriterienkataloge hinaus und definiert vielfältige neue Aspekte. Dazu gehören z. B. die Aufteilung verschiedener Aspekte der Relevanz des Indikators und des Versorgungsbereiches, die Detaillierungsaspekte der Manipulationssicherheit oder der Grad der Integration in IT-Systeme.

- Durch die Beschreibung der Kriterien anhand von Herleitung, Begründung aber auch Anwendungsdefinitionen und -beispielen sowie möglichen Skalen und Bewertungen ist ein in der Praxis besser anwendbarer Katalog entstanden, der direkt existierende Systeme bewerten kann, aber auch Hilfestellungen zur Entwicklung eines neuen qualitätsorientierten Vergütungssystems liefert.

- Gleichzeitig kann der erste Teil des Katalogs aufgrund der übergreifenden Struktur als so vollständig wie möglich hinsichtlich des hier konzipierten Anwendungszweckes bezeichnet werden.

4. Entwicklung des Anreizteils des Kriterienkatalogs für P4P-Systeme

4.1. Grundlagen von Anreizsystemen

4.1.1. Theoretische Grundlagen von Anreizsystemen

Anreizsysteme zur zielgerichteten Motivation von Organisationen oder Mitarbeitern sind seit langem bekannt und gebräuchlich. Zielgerichtete Anreize sind ein wichtiger Gestaltungsfaktor zur Erreichung von erwünschtem Verhalten. Meist wird hierunter das Verhalten von Mitarbeitern oder Führungskräften verstanden – d. h. das innerbetriebliche Umfeld. Doch der Begriff des Anreizsystems kann auch außerbetrieblich verstanden werden. So können das Verhalten und die Verhaltensänderung von ganzen Organisationen im Mittelpunkt der Betrachtung stehen – d. h. eine volkswirtschaftliche Sicht.

Diese Arbeit betrachtet zwar ein qualitätsorientiertes Anreizsystem, das die Qualität im gesamten stationären Sektor verbessern soll. Die eigentlichen Qualitätsverbesserungen finden aber auch hier in den Organisationen – vor allem beim Leistungserbringer – statt. Daher fokussiert sich diese Arbeit vor allem auf die Auswirkungen des Anreizsystems auf betrieblicher Ebene – sowohl bei Krankenkassen als auch bei Leistungserbringern. Es sollen hier neben einem kurzen Abriss der volkswirtschaftlichen Theorien vor allem die betriebswirtschaftlichen Anforderungen an Anreizsysteme zur Diskussion kommen. Theorien, die in der mikroökonomischen und betriebswissenschaftlichen Literatur zum Tragen kommen, sind für diese Arbeit besonders relevant.

4.1.1.1. Die volkswirtschaftliche Sicht

Volkswirtschaftlich muss ein Anreizsystem als Motivation für Organisationen in einem Gesamtsystem und für das Gesamtsystem verstanden werden, das Verhalten zu verändern. Gerade wenn „externe Effekte“ vorliegen, kann eine solche „Lenkung“ des Marktes erwünscht sein. Unter „externen Effekten“ werden Güter verstanden, für die bspw. „kein Markt mit Knappheitspreisen besteht“ (Zimmermann et al. 2009, S. 455). Häufigstes Beispiel dafür ist, in der „sauberen Umwelt als öffentliches Gut“ zu sehen (vgl. Zimmermann et al. 2009, S. 455ff.) – wenn ein Unternehmen bspw. CO_2 ausstößt, so verändert es das öffentliche Gut, „saubere Umwelt“ hat aber keinerlei negative Effekte auf das Betriebsergebnis zu erwarten. Daher wird dieser Effekt als „extern“ bezeichnet.

Auch die Versorgungsqualität einer Bevölkerung kann als ein externer Effekt angesehen werden: Die Behandlungsqualität hat für ein Krankenhaus keinen direkten wirtschaftlichen Effekt, da derzeit kaum qualitätsbedingte Veränderungen der Vergütung und somit des Umsatzes der Leistungserbringer stattfinden. Allenfalls könnte damit gerechnet werden, dass ein Anteil der Patienten zu Kliniken mit besseren Qualitätsergebnissen wandern wird. Dies setzt aber voraus, dass die Ergebnisse für den Patienten hinreichend transparent sind. Auch Porter/Teisberg stellen daher fest, dass die

Marktmechanismen für die Gesundheitsbranche nicht ausreichen (vgl. Porter/Teisberg 2006, S. 3ff.). Die „negativen externen Effekte“, die eine schlechte Behandlungsqualität und damit ein nicht ausreichendes Gesundheitsergebnis auslösen, haben bei den Leistungserbringern kaum eine direkte Konsequenz, die sich etwa im Betriebsergebnis widerspiegeln könnte. Es ist derzeit demnach in der Regel nicht im betriebswirtschaftlichen Interesse eines Krankenhauses, eine möglichst gute Behandlungsqualität anzubieten[1].

Ähnlich verhält es sich bei Umweltverschmutzungen: Auch hier hat ein Unternehmen keine direkten, negativen Konsequenzen bspw. bei hohen CO_2-Emissionen zu befürchten. Daher wird durch verschiedene Instrumente versucht, die externen Effekte für das Unternehmen zu internalisieren und eine direkte Beziehung zum Betriebsergebnis herzustellen. Für diese sogenannte „Internalisierung externer Effekte“ gibt es verschiedene Ansätze. Zunächst kann zwischen finanzpolitischen und nicht-finanzpolitischen Instrumenten unterschieden werden (vgl. Zimmermann et al. 2009, S. 463). Die finanzpolitischen Instrumente (wie bspw. die Pigou-Steuer oder Coase-Theorien; vgl. Coase 1937; Pigou 1924; Wiese 2010, S. 400ff.) werden als Theorien der Finanzwissenschaft behandelt und haben nur eine geringe Anwendung in der Umweltpolitik. Einige nicht-finanzpolitische Instrumente allerdings – die auch in bisherigen Ansätzen des stationären Systems gefunden werden können – werden hier diskutiert:

Bei den nicht-finanzpolitischen Instrumenten sind zunächst Ver- und Gebote zu erwähnen. Dieses Ordnungsrecht wird von Ökonomen als nicht effizient bezeichnet (vgl. Zimmermann et al. 2009); gleichzeitig ist eine ordnungspolitische „Anweisung“ hoher Behandlungsqualität im vorliegenden Kontext nur schwer anwendbar, da ja nicht nur ein „Grenzwert“ der Qualität – der vermutlich recht niedrig ausfallen würde – erreicht werden soll, sondern auch eine kontinuierliche Qualitätsverbesserung angestrebt wird.

Von besonderem Interesse sind die ökonomischen Anreizinstrumente. Hierbei werden Anreize gesetzt, um die „erwünschten Verhaltensänderungen in einer ökonomisch effizienten Weise“ (Zimmermann et al. 2009, S. 464) zu erreichen – dies entspricht auch der Zielsetzung eines P4P-Systems, wie es in dieser Arbeit verstanden wird. Ein mögliches, nicht-finanzpolitisches Instrument ist dabei die Verursacherhaftung, die versucht, durch drohende Haftungsklagen von „Geschädigten“ Verhaltungsänderungen einer Organisation zu erreichen. Im stationären Kontext ist dies bspw. mit einer Beweislastumkehr für Krankenhäuser im Falle von Kunstfehlerklagen gleichzusetzen.

Das bekannteste, ökonomische, nicht-finanzpolitische Anreizinstrument hinsichtlich CO_2 Emissionen ist der Handel mit Emissionsrechten. Hier werden die externen Effekte dadurch internalisiert, dass der negative, externe Effekt der Emission einen Markt-/Verhandlungspreis erhält. Wer mehr externe Effekte

[1] Behandlungsqualität dagegen als „öffentliches Gut“ zu bezeichnen ist schwieriger: Weder eine Nicht-Rivalität noch das Ausschlussprinzip liegen vor.

„produzieren“ möchte, kann sich dieses Recht am Markt erkaufen (vgl. Wicke/Blenk 1993). Eine Art von Zertifikatehandel wurde auch bereits für den Gesundheitssektor ins Spiel gebracht, nämlich für die Zuweisung von OP-Leistungen – allerdings wurde die Idee bereits wieder verworfen (vgl. Zertifikatehandel 2012).

P4P-Systeme im derzeit angewandten Sinne sind mit diesen Theorien in Einzelaspekten vergleichbar. Elemente von Emissionsrechten und Verursacherhaftung sind ebenso wiederzufinden wie Grundzüge von Coase (vgl. Wiese 2010, S. 400ff.) – nämlich Verhandlungen zwischen Schädigern (Leistungserbringer) und Geschädigten (Krankenkassen) bspw. in P4P-Systemen in den USA. Die Grundidee der „Internalisierung externer Effekte“ zeigt allerdings das Kernproblem im stationären Sektor auf, nämlich die bislang nicht betriebswirtschaftlich relevante Stellung der Behandlungsqualität von Krankenhäusern.

4.1.1.2. Die betriebswirtschaftliche Sicht

Für diese Arbeit soll vor allem auf betriebswirtschaftliche Theorien Bezug genommen werden. Daher sind Aspekte zu diskutierten, die die betriebliche Zusammenarbeit zwischen Organisationen und vor allem die innerbetrieblichen Anreize beschreiben. Gerade einer der Kernansätze von Anreizsystemen (die Principal-Agent-Theorie, u. a. genutzt in der BWL und Mikroökonomie) ist in der Lage, auch das Zusammenspiel von Krankenkassen und Leistungserbringer zu beschreiben[1]. Auch sehr detaillierte Ansätze wie die „Prospect Theorie“ existieren und gehen über die Ideen des „Homo Oeconomicus“ hinaus (vgl. z. B. Wünsch) – insgesamt ist „Verhaltensökonomik“ oder „behavioural economics“ ein wichtiger, neuer Zweig der Wirtschaftswissenschaften[2].

Unter Anreizsystemen werden betriebliche Systeme verstanden, die Mitarbeiter oder Führungskräfte zu einem gezielten Verhalten motivieren sollen, oder die „Summe aller bewusst gestalteten Arbeitsbedingungen, die bestimme Verhaltensweisen (durch positive Anreize, Belohnungen etc.) verstärken, die Wahrscheinlichkeit des Auftretens anderer dagegen mindern (negative Anreize, Sanktionen)“ (Wild 1973, S. 47; zitiert u. a. von Schmeisser 2008, S. 65). Zwei Hauptansätze für die theoretischen Grundlagen von Anreizsystemen existieren: Verhaltenstheoretische und vertragstheoretische Grundlagen (vgl. Hungenberg 2006). Hier soll zunächst die Verhaltenstheorie angesprochen werden um anschließend detailliert auf die Principal-Agent-Theorie (vertragstheoretisch) einzugehen.

---

[1] Verschiedene Fachbereiche befassen sich mit der Thematik: Unternehmensführung (z. B. Schanz 1991; Hungenberg 2006), Unternehmenspolitik und -organisation (z. B. Grewe 2000); Personalmanagement und Unternehmenssteuerung (z. B. Lindert 2001) aber auch Strategie (z. B. Bleicher 1992; Becker 1990) oder Qualitätsmanagement (z. B. Ortlieb 1993).

[2] Diese Ansätze, die unter anderem in der Finanzwissenschaft genutzt werden, gehen über den Anwendungszweck dieser Arbeit hinaus, daher wird ein Fokus auf Anforderungen an Anreizsysteme wie im Folgenden diskutiert gelegt.

4.1.1.2.1. Verhaltenstheoretische Ansätze

Die engste, betriebliche Sicht auf Anreizsysteme ist in der Arbeitsmotivationstheorie zu finden. Es werden die Verhaltensweisen von Mitarbeitern und deren mögliche Motivation und Beeinflussbarkeit analysiert (vgl. Ortlieb 1993, S. 67). Die zwei bekanntesten Theorien sind dabei die Bedürfnispyramide von Maslow und die Zweifaktorentheorie von Herzberg.

Bedürfnispyramide nach Maslow: Maslow beschreibt fünf Ebenen der Motivgruppen (vgl. Abbildung 13). Diese Bedürfnisklassen haben eine hierarchische Beziehung. Erst wenn bspw. die physiologischen Grundbedürfnisse wie Hunger befriedigt sind, können weitere Bedürfnisse erfüllt werden (vgl. Rosenstiel 1980, S. 275ff.). Trotz vielfacher Kritik an Maslows Bedüfnispyramide (vgl. z. B. Wiswede 1980) ist sie eine der bekanntesten Theorien der Motivationsforschung, zur Erklärung von Anreizsystemen sind sie allerdings nur bedingt geeignet (vgl. Ortlieb 1993, S. 68). Die Bedürfnispyramide wird teilweise auch für die Beschreibung ganzer Gesellschaften und dem Wandel der Präferenzen genutzt (vgl. Henke et al. 2003, S. 643).

Abbildung 13: Bedürfnispyramide nach Maslow

Quelle: Darstellung gemäß Maslow 1943

Zweifaktorentheorie nach Herzberg: Die Zweifaktorentheorie nach Herzberg dagegen bietet erste Erklärungsansätze für Anforderungen von Anreizsystemen (vgl. Ortlieb 1993, S. 69). Herzberg bezieht sich in seinen Arbeiten explizit auf die betriebliche Motivation von Mitarbeiten und beschreibt zwei Faktoren: Die motivierenden Faktoren („Motivatoren“) und die Hygienefaktoren (vgl. Herzberg 1968). Hygienefaktoren stellen solche Elemente dar, die erfüllt sein müssen, um Unzufriedenheit zur verhindern, gleichzeitig aber nicht motivieren können (wie z. B. Arbeitsbedingungen). Motivatoren dagegen (wie z. B.

Anerkennung oder Verantwortung) können zu einer besseren Leistung und höheren Zufriedenheit führen. Gehalt nimmt hierbei eine Zwischenstellung ein, da es weder der „unabdingbaren Arbeitsumgebung“ noch den Motivationsfaktoren eindeutig zugeordnet werden kann (für Anreizsysteme bedeutet dies, dass eine „gute Arbeitshygiene“ zunächst eine Grundvoraussetzung dafür ist, erfolgreiche Motivation erreichen zu können). Ortlieb geht sogar noch darüber hinaus: „Zur Förderung der Qualitätsmotivation sollten demnach vor allem intrinsische Anreizsysteme eingesetzt werden.“ (Ortlieb 1993, S. 69)[1]. In einem materiellen/finanziellen Anreiz- bzw. Vergütungssystem müssen spezifischen Anforderungen erfüllt werden, um ein intrinsisches Anreizsystem zu erreichen. Diese Anforderungen stehen allerdings für ein finanziell-orientiertes Vergütungssystem im Sinne von P4P nicht im Vordergrund, da ein qualitätsorientiertes Vergütungssystem zunächst eindeutig eine extrinsische Motivation darstellt. Daher soll als nächstes die Frage geklärt werden, in welchem Verhältnis intrinsische zu extrinsischer Motivation stehen kann.

„Crowding-Out-Effekt“: Intrinsische Motivation beschreibt die Situation, in der ein Mitarbeiter von der Tätigkeit selbst motiviert wird – d. h. „von innen heraus“. Extrinsisch dagegen wird eine Motivation dann bezeichnet, wenn die Leistung(ssteigerung) im direkten Zusammenhang mit (externen) Anreizen (Belohnung oder Strafe) steht. Seit etwa Mitte der 80er Jahre wird in der Wissenschaft die Hypothese diskutiert, ob externe Anreizsysteme einen negativen Effekt auf die intrinsische Motivation haben können (erstmals von Edward Deci (u. a. Deci/Ryan 1985) und dann vorangetrieben von Bruno Frey (u. a. Frey 1992; Frey 2002). Diese als „Crowding-Out“ oder „Verdrängungshypothese“ bekannte Hypothese beschreibt den Effekt, dass finanzielle Belohnungen (als ein wichtiger, externer Anreiz) die intrinsische Motivation nicht bestärken sondern „verdrängen“ und dadurch langfristig eine niedrigere Leistung erreicht wird, als ohne die Belohnung. Verschiedene empirische Studien werden hierfür herangezogen (vgl. Matiaske/Weller 2006). Als Konsequenz empfiehlt Frey daher, Manager „wie Beamte“ zu vergüten – z. B. mit hohem Fixanteil der Vergütung (vgl. Frey 2005).

Bei Annahme dieser Hypothese wäre ein wirksames qualitätsorientiertes, finanzielles Anreiz- bzw. Vergütungssystem wie P4P nicht möglich. Der Grundsatz – „höhere Vergütung fördert bessere Qualität“ – wäre nicht haltbar – weder im betriebswirtschaftlichen Bereich noch bei der Internalisierung externer Effekte (vgl. Matiaske/Weller 2006). Weitere empirische und theoretische Arbeiten jedoch können zeigen, dass finanzielle Anreize sehr wohl positiv motivieren können, ohne intrinsische Motivation zu verdrängen (vgl. u. a. Le Grand 2003; Matiaske/Weller 2006). So wird davon ausgegangen, dass eine materielle Vergütung als „anerkannte Währung um Engagement zu honorieren [gilt]. Überschreiten jedoch die materiellen Anreize einen bestimmten Schwellenwert, tritt der von Frey postulierte Crowding-Out-Effekt ein.“ (Matiaske/Weller 2006, S. 129–130). Dies zeigt, dass im Rahmen einer „Rational-Choice-

[1] Vgl. hierfür die nicht eindeutige Zuordnung von „Gehalt“ in der Zwei-Faktoren-Theorie von Herzberg, wobei Hygienefaktoren vor allem die extrinsische, die Motivatoren vor allem die intrinsische Motivation betreffen (vgl. Ortlieb 1993, S. 69)

Theorie“ mit einem positiven Effekt von finanziellen Anreizsystemen gerechnet werden kann (vgl. Matiaske/Weller 2006). Allerdings sind sowohl die Anreizhöhe als auch weitere Faktoren wie bspw. Anerkennung zu beachten.

Gerade im Kontext von medizinischen Berufen sind der Berufsethos und somit die intrinsische Motivation sehr hoch, so dass auch hier von der Möglichkeit des „Crowding Out“ gesprochen wird – unter anderem weil durch finanzielle Anreize die gefühlte Verantwortung für ein motiviertes, ethisches Handeln an externe Stellen abgegeben werden kann (vgl. Wynia 2009, S. 886f.). Eine mögliche Gefahr von P4P-Systemen wäre somit, dass Ärzte nur noch wegen der Bezahlung handeln – nicht aufgrund von berufsethischer Überzeugung (vgl. Roland 2006, S. 417f.). Studien zum Vergleich von intrinsischer vs. extrinsischer Motivation gibt es im Gesundheitswesen nur wenige, erste Ergebnisse einer qualitativen Studie in Großbritannien zeigen, dass finanzielle Anreizsysteme die intrinsische Motivation nicht gefährden, Krankenschwestern allerdings besorgter bezüglich der zukünftigen Entwicklung sind (vgl. McDonald et al. 2007, S. 1361f.)[1]. Gleichzeitig kann auf der Systemebene gezeigt werden, dass das Verhalten durch finanzielle Anreize verändert werden kann, bspw. werden bei der „Fee-for-Service“-Vergütung deutlich mehr Behandlungen durchgeführt als bei fixen Budgets (vgl. Marshall/Harrison 2005, S. 4; Gosden et al. 2001, S. 44ff.) Um einen möglichst hohen Motivationseffekt von finanziellen Anreizen zu erzielen, wird unter anderem vorgeschlagen, einen ausreichenden Vergütungsanteil zu gewährleisten und außerdem die Ziele des qualitätsorientierten Vergütungssystem mit den intrinsischen Zielen der medizinischen Berufe zu vereinen (vgl. Marshall/Harrison 2005, S. 5; Mehrotra et al. 2010b, S. 502).

4.1.1.2.2. Principal-Agent-Theorie / Vertragstheorien

> *„Ein guter Hirt ist bereit, für seine Schafe zu sterben. Jemand, dem die Schafe nicht selbst gehören, ist kein richtiger Hirt. Darum läßt er sie im Stich, wenn er den Wolf kommen sieht […]. Wer die Schafe nur gegen Lohn hütet, läuft davon; denn die Schafe sind ihm gleichgültig.“ (Johannes 10, 11-12; Die Bibel; nach Lehmann 2006, S. 3).*

Die zweite Strömung für die Beschreibung von Anreizsystemen ist vor allem in der Principal-Agent-Theorie zu finden (vgl. z. B. Petersen 1989; Lehmann 2006). Diese Theorie ist ein Teilgebiet der „Neuen Institutionenökonomik“ (vgl. u. a. Kräkel 2004) und findet sowohl in der Mikroökonomik (vgl. u. a. Mas-Colell et al. 1995) als auch in der betriebswissenschaftlichen Lehre Anwendung, vor allem im Bereich Management (z. B. Staehle et al. 1999) und Organisation/Unternehmensführung (vgl. Schreyögg/von Werder 2004). Grundlagen sind hierbei die Vertragstheorien von Coase (vgl. Coase 1937).

[1] Aber auch Studien existieren, die Sorgen von Ärzten bezüglich bspw. Datenvalidität oder ausreichenden Forbildung und (technischer) Unterstützung innerhalb P4P-Projekten zeigen (vgl. Beckman 2006, S. 192; Locke/Srinivasan 2008, S. 22f.).

Die Principal-Agent-Theorie geht von einer vertraglichen Beziehung zwischen zwei Parteien aus (Principal = Auftraggeber (z. B. Arbeitgeber, Zulieferer) und Agent = Beauftragter (z. B. Arbeitnehmer, Abnehmer)), die nicht den gleichen Zugang zu Informationen und Eigentumsverhältnisse haben. Hierdurch entsteht eine Informationsasymmetrie, wodurch der Principal benachteiligt ist, gleichzeitig aber ein Delegationsrecht im hierarchischen Kontext hat (vgl. z. B. Kräkel 2004). Anders als in der Verhaltenstheorie wird hierbei von rational handelnden Personen ausgegangen.

Zwei Aspekte sind besonders wichtig für die Prinzipal-Agent-Theorie: „Hidden Information“ und „Hidden Action“ (oder „Moral Hazard“). Gerade die „Hidden Action“ – d. h. eine für den Principal nicht beobachtbare Handlung vom Agenten, um seinen eigenen Nutzen zu optimieren – erklärt den Sinn und die Notwendigkeit für Anreizsysteme. Frey bspw. schreibt dazu: „Anreizsysteme sind nötig, weil in aller Regel nicht davon ausgegangen werden kann, dass sich Mitarbeitende ohne entsprechende organisatorische Maßnahmen im Interesse einer Organisation verhalten“ (Frey/Benz 2004, S. 22). Aus dem Dilemma der Informations- und Zielasymmetrie lassen sich einige Anforderungen für Anreizsysteme ableiten, so bspw. die Notwendigkeit einer hohen Transparenz und möglichst großen Zielkongruenz.

Doch wie kann der Principal-Agent-Ansatz für ein qualitätsorientiertes Anreizsystem in der stationären Versorgung Anwendung finden? Wer ist Principal, wer ist Agent? Innerhalb eines Krankenhauses kommt es bereits zu einer Principal-Agent-Beziehung zwischen dem Arbeitgeber (Krankenhausträger) und den Führungskräften (z. B. Chefärzten) (vgl. Ernst 1999, S. 11ff.). Auch auf der nächsten hierarchischen Ebene (Führungskräfte-Chefärzte zu Mitarbeitern-Ober/Assistenzärzte) kommt es zu den typischen Charakteristika des Principal-Agent-Ansatzes (vgl. Litzkow 2008, S. 20). Ein innerbetriebliches Anreizsystem sollte auf diese Aspekte eingehen.

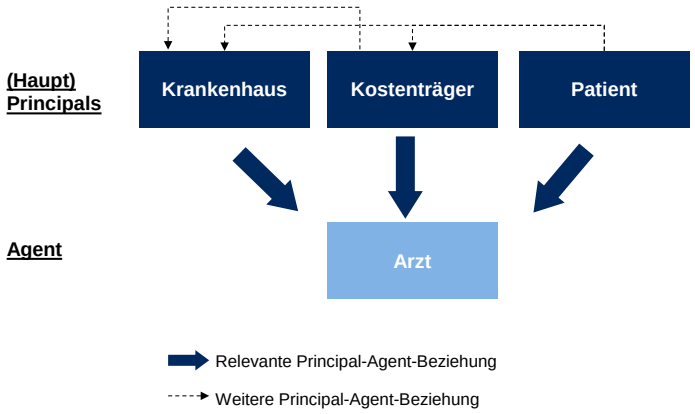
Außerdem ist es offensichtlich, dass eine Principal-Agent-Beziehung zwischen dem Arzt als Agent und dem Patienten als Principal besteht (vgl. hierzu z. B. Kräkel 2004 oder Schreyögg 2010, S. 152). Der Arzt hat zwar deutlich mehr Informationen, speziell im Rahmen der hohen Komplexität der Medizin, gleichzeitig wird er jedoch vom Patienten als „Dienstleister“ zur Behandlung beauftragt. Da es allerdings (im deutschen Gesundheitssystem) zu keiner direkten Vergütung zwischen diesen beiden Parteien kommt, ist die Übertragung auf ein Anreizsystem kaum möglich.

Daher muss der „Bezahler“, nämlich die Krankenkassen, mit in die Principal-Agent-Beziehung aufgenommen werden[1]. Blomqvist beispielsweise beschreibt die Rolle des Arztes als „double-agent“, nämlich zwischen Arzt (Agent) und Patient (Principal) sowie Arzt (Agent) und Krankenversicherung (Principal) (vgl. Blomqvist 1991, S. 412). Doch auch dies entspricht nicht der direkten monetären Beziehung im stationären Sektor: Die (gesetzliche) Krankenversicherung entlohnt nicht direkt den Arzt,

[1] Auch die Definition des Gesetzgebers als „Principal“ ist möglich (vgl. z. B. Schwartz 1997), für die hier diskutierten Anreizsysteme aber nicht zweckmäßig

sondern das Krankenhaus – und auch ein qualitätsorientiertes Vergütungssystem im Sinne von P4P würde vermutlich zunächst das Krankenhaus (Agent) als Adressaten für Anreize sehen. Somit kommt es zunächst zu einer Principal-Agent-Beziehung zwischen Krankenkasse und Krankenhaus, zwischen Krankenhaus und Klinikarzt (d. h. mittelbar auch zwischen Krankenkasse und Klinikarzt) und zwischen Patient und Klinikarzt. Ein Klinikarzt hat demzufolge eine „dreifache Agentenrolle“ (vgl. Picot et al. 2002, S. 85; Ernst 1999, S. 11ff.; Litzkow 2008, S. 21; vgl. für die verschiedenen Beziehungen auch Abbildung 14). In Deutschland allerdings ist diese Beziehung mittelbar, wohingegen in den USA die Patienten auch eine direktere Vertrags- und Vergütungsbeziehung zu Klinikärzten innehat.

Abbildung 14: Relevante Principal-Agent-Beziehungen im stationären Sektor



Quelle: Eigene Darstellung mit Anlehnung an Litzkow 2008, S. 21

Ein qualitätsorientiertes Vergütungssystem im Sinne von P4P, das Anreize für die Kliniken und Klinikärzte schaffen soll und sich auf Geldflüsse von Krankenkassen und die Behandlungsqualität von Patienten beruft, basiert auf einem komplexen System von Principal-Agent-Beziehungen.

4.1.2. Funktionsweise, Systemelemente und Gestaltungsdimensionen von Anreizsystemen

4.1.2.1. Arten und Funktionsweise von Anreizsystemen

In der Betriebswirtschaftslehre werden vor allem zwei Arten von Anreizsystemen unterschieden, die Becker dargestellt hat. Er spricht von Anreizsystemen im weiteren Sinne (i.w.S.) und solche im engeren Sinne (i.e.S.). Während Anreizsysteme i.w.S. das gesamte Führungssystem umfassen, sind unter dem Anreizsystem i.e.S. die zentralen Anreizgrundsätze im Rahmen des Motivationskonzeptes der Mitarbeiter zu verstehen (vgl. Becker 1990, S. 8). Im Folgenden soll vor allem auf das Anreizsystem i.e.S. eingegangen werden, da es sich hierbei um die steuerbaren Anreize handelt, die einer hohen, unternehmerischen Gestaltbarkeit unterliegen (vgl. Lehmann 2006, S. 17).

Für solche Anreizsysteme i.e.S. gibt es eine Vielzahl von Umsetzungsvarianten, von denen zwei angesprochen werden, da sie erste Einblicke in die Funktionsweise von Anreizsystemen und deren Grundelemente liefern. So wird bei Becker zwischen langfristigen, strategisch-orientierten und dualen Anreizsystemen unterschieden (vgl. Becker 1990, S. 36ff.). Langfristige Anreizsystemen können dabei bspw. marktindexiert (z. B. „Stock Option Plans“) oder leistungsbezogen (so bspw. „Performance Share Plans“) sein. Um eine möglichst große Auswahl an operativen sowie strategischen Anreizen vereinen zu können, fokussiert sich die vorliegende Arbeit auf ein duales Anreizsystem (vgl. Becker 1990, S. 67ff.).

Die zweite hier diskutierte Umsetzungsvariante ist „Management by objective“, ein von Peter Drucker entwickeltes Zielvereinbarungskonzept (vgl. Drucker 1956). Hierbei handelt es sich um eine Managementtheorie, anhand derer die „zielorientierte Beeinflussung aller Teilsysteme einer Organisation“ (Müller 1999, S. 271) beschrieben wird – vor allem die Steuerung von Führungskräften ist relevant. Diese Theorie verweist auf die Wichtigkeit von Zielen in einem Managementsystem, die eine richtungsweisende Funktion einnehmen können und daher ein erfolgreicheres Handeln versprechen (vgl. Müller 1999, S. 274). Diese Theorie findet sich vor allem in Zielvereinbarungskonzepten wieder, die klar definierte Ziele für jeden einzelnen Mitarbeiter definieren und auf Grundlage dieser Vereinbarung variabel vergüten. Eines der Grundelemente von Anreizsystemen ist daher in einem „Zielansatz“ zu finden (vgl. u. a. Becker 1990, S. 19).

4.1.2.2. Systemelemente von Anreizsystemen

Einige dieser Elemente geben die Funktionsweise von Anreizsystemen wieder und entsprechen den bereits diskutierten „Systemelementen“. Der Kern der Funktionsweise von Anreizsystemen ist im „Anreiz-Beitrags-Konzept“ oder „SOR (Stimulus-Organism-Response)“ zu finden (vgl. u. a. Seidel 1991, S. 172ff.). Dieses neobehavioristische Paradigma geht davon aus, dass ein gewisser Stimulus nach der Verarbeitung innerhalb bspw. eines Motivationsprozesses zum veränderten Verhalten führen kann (vgl. Seidel 1991, S. 172ff.). Somit sind der Anreiz (z. B. die variable Vergütung) und die Verhaltensänderung sowie deren Ergebnis Systemelemente des Anreizsystems.

Auch kann die Struktur eines Anreiz-Managementsystems mit dem eines Managementprozesses verglichen werden. So beschreibt Lindert die Struktur von Anreizsystemen mit vier Teilelementen (vgl. Lindert 2001, S. 136ff.).

1. *Leistungsdefinition*: Ziele des Anreizsystems müssen definiert und anschließend kommuniziert werden.

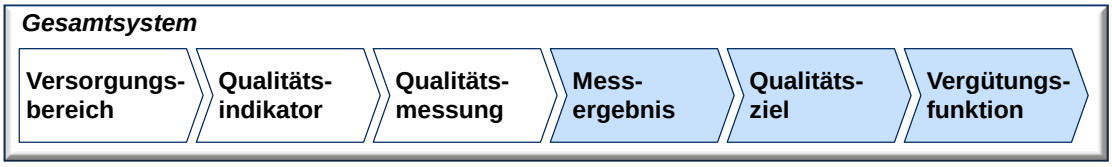
2. *Unterstützende Elemente*: Arbeitsbedingungen, Trainings etc. müssen etabliert werden, um die Zielerreichung zu fördern und zu ermöglichen.

3. *Leistungsmessung und -beurteilung*: Die vorher definierten Ziele werden gemessen und die Ergebnisse analysiert. Hierzu bedarf es zusätzlich eines Messindikators und einer Messbeschreibung.

4. *Leistungsbelohnung*: Hier werden den Leistungsergebnissen die Anreize und Entlohnungen zugeordnet – also eine Vergütungsfunktion definiert.

Diese Struktur entspricht großteils den Systemelementen, anhand derer bereits der erste Teil des Anforderungskataloges aufgebaut wurde (vgl. Abschnitt 3.2.3.) – siehe nochmals Abbildung 15.

Abbildung 15: Systemelemente des Kriterienkatalogs (Anreizsysteme)



Neben den Qualitätselementen (Versorgungsbereich, Qualitätsindikator und Qualitätsmessung) kommen nun noch ein festzusetzendes *Ziel* sowie die anhand des *Messergebnisses* umzusetzende eigentliche *Vergütung*(sfunktion) hinzu

4.1.2.3. Gestaltungsdimensionen von Anreizsystemen

Es gibt weitere Elemente (so genannte Gestaltungsdimensionen) die ein Anreizsystem definieren, so beispielsweise bei Grewe (vgl. Grewe 2000, S. 18ff.) und bei Bleicher (vgl. Bleicher 1992, S. 36). Bleicher unterscheidet bspw. zwischen Beteiligtenkreisen, Lang- oder Kurzfristigkeit des Systems, Flexibilität der Vergütung und immaterieller oder materieller Vergütung (vgl. Bleicher 1992, S. 36). Weitere Dimensionen beinhalten bspw. die Ausschüttungsperiode oder die Art der Leistungsbewertung (vgl. Becker 1990, S. 117) sowie verschiedene Arten von Entgeltsystemen (z. B. Verhältnis der Belohnung zur Risikohöhe, Einheitlichkeit des Entgeltsystems oder Fundierung der Gesamtquote, vgl. Kerr 1982).

Auch in der wissenschaftlichen Literatur zu Pay-for-Performance-Systemen wird (gerade in Deutschland) vermehrt auf das Thema Gestaltungsdimensionen eingegangen, so z. B. die Einbeziehung der Ärzte und anderer Berufsgruppen oder die Beachtung regionaler Besonderheiten. Auch eine kontinuierliche Evaluation und eine Kombination von finanziellen und nicht-finanziellen Anreizen werden als relevant diskutiert. Auch verschiedene sehr operative Aspekte wie ein guter Datenzugang oder unterstützende EDV-Ausstattungen sind für P4P-Systeme entscheidend. Neuerdings werden zusätzlich Themen wie „Ceiling-Effekte“, relevante Vergütungshöhe und mögliche Differenzen der Anreizwirkungen bei verschiedenen Krankheitsbildern) diskutiert (vgl. bspw. SVR 2007 und Schrappe/Gültekin 2011)[1].

[1] Vgl. auch Abschnitt 4.2.5.

Ziel dieser Arbeit ist es allerdings nicht, die verschiedenen Gestaltungsparameter zu evaluieren. Vielmehr werden die *Anforderungen* an das P4P-System diskutiert. Im empirischen Teil werden aber implizit auch Gestaltungsdimensionen der zu untersuchenden, existierenden P4P-Systeme analysiert.

4.1.3. Theoretische Anforderungen an Anreizsysteme

Es gibt eine Vielzahl von allgemeinen (nicht speziell auf P4P-Systeme bezogenen) Anforderungslisten an Anreizsysteme. Die folgende Auflistung zeigt die Anforderungen aus den wichtigsten Arbeiten im Überblick[1]:

[1] Bei Wünsch wird zwischen „Funktionalen Anforderungen“ und „Gestaltungsanforderungen“ unterschieden (vgl. Wünsch 2012, S. 26ff.), da in dieser Arbeit eine „optimale“ oder „pareto-effiziente“ Entlohnungsfunktion noch nicht angestrebt wird, soll hier der Fokus auf den Gestaltungsanforderungen liegen.

Tabelle 5: Anforderungslisten im Überblick (Auswahl)

#	Becker	Grewe	Lehmann	Bullinger	Kerr	Bleicher	Ortlieb	Wünsch
1	leistungs-orientiert	Leistungs-orientierung	Leistungs-orientierung					
2	transparent	Transparenz	Transparenz	verständlich (Transparenz von Aufbau und Struktur, ...)	Transparenz	Transparenz		Intersubjektive Überprüfbar-keit
3	flexibel	Flexibilität	Flexibilität			Flexibilität		Flexibilität
4	motivierend			motivations-förderlich		Motivation		
	belohnend				Belohnung/ Bestrafung	Belohnungs-wirksam		
5	anforderungs-gerecht	Gerechtigkeit	Gerechtigkeit	anforderungs-gerecht	Gerechtigkeit	Gerechtigkeit		Angemessen-heit
	marktgerecht			qualifikations-gerecht				
6			Akzeptanz	anpassungs-fähig				
7	wirtschaftlich möglich	Wirtschaftlich-keit	Wirtschaftlich-keit	wirtschaftlich	Effizienz	Wirtschaftlich-keit	Verbesserung Wettbewerbs-fähigkeit	Effizienz
	ergebnis-fördernd			produktivitäts-förderlich	Wirksamkeit			
8							steuerbar, organisierbar, kontrollierbar	
9			Anreizkompa-tibilität		Anspruchs-berechtigung			
			Anreizvalenz					
10			Rechtssicher-heit					
	individual- und gruppenweise		*Risiko-reduktion*		*Verfügbarkeit*	*Differenz-ierung*		*Manipulations-sicherheit*
					finanzielle und nicht-finanzielle Anteile	*Integration*		*Controllability*
					Reversibilität			
					Leistungs-messung			

Quelle: Eigene Darstellung mit Daten aus Becker 1990, S. 19; Bullinger 1996, S. 213f.; Kerr 1997, S. VIII; Bleicher 1992, S. 19f.; Grewe 2000, S. 16ff.; Lehmann 2006, S. 16ff.; Ortlieb 1993, S. 64ff.; Wünsch 2012, S. 28ff.

Es zeigt sich, dass eine große Einigkeit bezüglich der wichtigsten Anforderungen besteht (siehe farblich gekennzeichnete Anforderungen in Tabelle 5). Die Anforderungen mit der höchsten Übereinstimmung und Relevanz für P4P werden diskutiert, da sich auch in dem hier entwickelten Kriterienkatalog genutzt werden.

1. Leistungsorientierung: Bei Becker liegt das „Schwergewicht der Kriteriendiskussion“ auf der Leistungsorientierung (vgl. Becker 1990, S. 19). Kritisiert wird hier häufig eine einseitige Konzentration auf Unternehmensergebnisse (vgl. Grewe 2000, S. 16). Neben der messbaren Zielerreichung des jeweiligen Leistungsergebnisses kann hierbei das Leistungsverhalten und die Leistungsbedingung relevant sein (vgl. Lehmann 2006, S. 18). Somit müssen die Ziele zumindest teilweise an die Arbeitsbedingungen sowie individuelle Fähigkeiten angepasst werden. Gerade für P4P ist entscheidend, was genau „Leistung“ oder

„Performance“ bedeuten kann: diese Qualitätsorientierung wurde im Kapitel 3 bereits detailliert diskutiert. Für die weiteren Systemelemente des Anreizsystems sind verschiedene Subthemen der Leistungsorientierung entscheidend, so z. B. Erreichbarkeit der Zielsetzung, Flexibilität und Beeinflussbarkeit.

2. Transparenz: Alle Elemente des Anreizsystems (Bewertungsprozess, Belohnungen/Sanktionen, Veränderungen im Zeitverlauf) müssen für die beteiligten Personen (vor allem für die Person, der der Anreiz gilt) vergleichbar, nachvollziehbar und durchschaubar sein (vgl. Grewe 2000, S. 17; Becker 1990, S. 24; siehe auch „intersubjektive Überprüfbarkeit“ oder „Controllability“ bei Wünsch 2012, S. 28f.). Nur so kann der eigentliche Motivationseffekt eintreten, wenn der Anreiz vom Mitarbeiter erkannt wird. Außerdem ist nur dann eine hohe Akzeptanz und Gerechtigkeitsempfinden für das System erreichbar (vgl. Lehmann 2006, S. 18). Aufgrund der Principal-Agent-Beziehung zwischen Arzt, Krankenhaus, Krankenkasse und Patient ist für ein P4P-System im stationären Sektor die Transparenz für verschiedene Akteure von besonderer Bedeutung. Gerade auch Fragestellungen der Veröffentlichung von Qualitätsergebnissen sind hier relevant ebenso wie die Verständlichkeit der Qualitätsindikatoren.

3. Flexibilität: Ein langfristig erfolgreiches Anreizsystem muss sich zwangsläufig an Veränderungen im Unternehmensumfeld, der Arbeitsbedingungen, der Mitarbeiterbedürfnisse und der anderen Dimensionen anpassen können (vgl. Lehmann 2006, S. 17; Becker 1990, S. 24; Grewe 2000, S. 17; Wünsch 2012, S. 30f.). Auch müssen die unterschiedlichen Grundbedingungen von verschiedenen Beteiligten beachtet werden. Nur wenn diese Unterschiede berücksichtigt werden, können die Anreize als gerecht empfunden und akzeptiert sowie eine erfolgreiche Motivationswirkung erreicht werden. Speziell im stationären Sektor ist die Flexibilität ein wichtiges Thema, da bspw. medizinische Prozesse und Anforderungen aufgrund von Innovationen einem ständigen Änderungsprozess unterworfen sind.

4. Motivationswirkung: Zu Verhaltensänderungen zu motivieren, ist Sinn und Zweck eines Anreizsystems. Becker und Bullinger erwähnen diesen Aspekt explizit als Anforderung. Hierbei müssen für eine langfristige Anreizwirkung z. B. auch Mitarbeiterbedürfnisse und -entwicklungen beachtet werden (vgl. Becker 1990, S. 24; Bullinger 1996, S. 213f.). Unter anderem ist ein Gerechtigkeitsempfinden und Akzeptanz notwendig, um eine solche Motivationswirkung zu erreichen.

5. Gerechtigkeit: Es wird zwischen interner und externer Gerechtigkeit unterschieden, die ein betriebliches Anreizsystem erfüllen muss (vgl. u. a. Grewe 2000, S. 17; Lehmann 2006, S. 17; Wälchli 1995, S. 169f.; siehe auch „Angemessenheit“ bei Wünsch 2012, S. 29). Die interne (d. h. betriebsinterne) Gerechtigkeit beinhaltet Aspekte der Anforderungsgerechtigkeit (gerechte Belohnung relativ zur Verantwortung der Aufgabe), der Leistungsgerechtigkeit (gerechte Belohnung relativ zur Leistung) und der Sozialgerechtigkeit (gerechte Belohnung relativ zum sozialen Kontext des Mitarbeiters, siehe auch Becker 1990, S. 19). Diese Aspekte werden vom Mitarbeiter mit Kollegen verglichen und können so zu einem subjektiven Gerechtigkeitsempfinden führen. Die externe Gerechtigkeit vergleicht die Anreize und

Belohnungen im Kontext zu anderen Unternehmen. Diese Anforderung ist entscheidend, um langfristig andauernde Akzeptanz bei den Beteiligten zu erlangen und eine Motivationswirkung zu erreichen bzw. eine Demotivation zu verhindern, gleichzeitig ist sie aufgrund der Subjektivität des Kriteriums schwer zu erreichen.

***6. Akzeptanz*:** Die Anforderung der Akzeptanz wurde schon bei einigen der anderen Kriterien erwähnt – so kann ein gerechtes, transparentes, flexibles Anreizsystem zu einer hohen Akzeptanz bei den Beteiligten führen. Gleichzeitig kann die Akzeptanz von der individuellen Risikoneigung abhängen und durch Kommunikationsmaßnahmen (und damit steigender Transparenz) verbessert werden (vgl. Lehmann 2006, S. 16). Im ersten Teil des Katalogs waren die Kriterien „Medizinische Bedeutung in der medizinischen Fachliteratur“, „Wahrgenommene medizinische Bedeutung auf der Ebene der Fachgesellschaften“, „Wahrgenommene medizinische Relevanz/Nutzen von medizinischen Endanwendern“, „Verständlichkeit für Patienten und interessierte Öffentlichkeit“, „Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit“ zu „Akzeptanz des Qualitätsindikators“ zusammengefasst worden. Doch nicht nur das Systemelement des „Qualitätsindikators“ muss von den Beteiligten akzeptiert werden, auch die Anreizaspekte müssen dieser Anforderung entsprechen. Eine hohe Akzeptanz des Gesamtsystems betrifft dabei nicht nur die Mitarbeiter (oder in unserem Falle Ärzte und Pflegekräfte) sondern alle Beteiligten (also u. a. ebenso Krankenkassen und Patienten). Auf Systemebene wäre hierfür auch eine enge Einbindung dieser Personenkreise im Rahmen der Entwicklung wünschenswert.

7. Wirtschaftlichkeit/Effizienz: Aus betriebswirtschaftlicher Sicht ist die Wirtschaftlichkeit des Anreizsystems eine notwendige Bedingung für eine erfolgreiche, langfristige Implementierung und Anwendung. Das bedeutet zum einen, dass die Gestaltung und Höhe der Anreize wirtschaftlichen Aspekten unterliegen (vgl. Becker 1990, S. 24; Bleicher 1992, S. 20; Wünsch 2012, S. 29). Gleichzeitig müssen die Kosten der Implementierung und des Betreibens des Systems (bspw. IT-Programmierung, Messung und Evaluierung der Leistung, ggf. Training und andere Begleitprojekte) betrachtet werden (vgl. u. a. Grewe 2000, S. 18). Jedoch verursachen Anreizsysteme nicht nur Kosten sondern auch einen wirtschaftlichen Zusatznutzen. Gerade erfolgsorientierte Anreizsysteme sollen zu einer Verbesserung der Gewinnsituation des Unternehmens führen und positive Effekte/Betriebsergebnisse generieren (vgl. Becker 1990, S. 24). Doch auch qualitätsorientierte Anreizsysteme müssen Wirtschaftlichkeits-gesichtspunkten genügen (vgl. Ortlieb 1993, S. 64). Insgesamt sollte ein Kosten-Nutzen-Vergleich positiv ausfallen, jedoch ist die genaue Zuordnung der positiven Effekte auf das Anreizsystem häufig schwer, da sie nur einen von vielen Einflussfaktoren darstellt (vgl. Lehmann 2006, S. 29). Außerdem muss zwischen der Wirtschaftlichkeit für verschiedene Akteure unterschieden werden, bspw. hinsichtlich der Auswirkungen für den Leistungserbringer, die Krankenkasse oder das Gesamtsystem (vgl. hierzu erste Modellrechnungen bei Mansky/Seeger 2011 und Seeger 2011).

8. Steuerbarkeit/Organisierbarkeit/Kontrollierbarkeit: Konkrete Anforderungen an die Operationalisierbarkeit sind nur in einer der hier vorgestellten Kriterienlisten zu finden. Nur in den Grundanforderungen von qualitätsorientierten Anreizsystemen nach Ortlieb werden Steuerbarkeit, Organisierbarkeit und Kontrollierbarkeit erwähnt (vgl. Ortlieb 1993, S. 64f.). Gerade bei Qualität sind adäquate Managementsysteme und somit Umsetzungsaspekte häufig eines der Kernthemen (vgl. Vielzahl von Literatur über Qualitätsmanagement, z. B. Zollondz 2002; Andernach 2006; Kronsteiner 2010; Maiwald 2010 und natürlich Ortlieb 1993 – auch in der Gesundheitsbranche, z. B. Haubrock 2002; Fleßa 2007; Ollenschläger/Thomeczek 2002; ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) 2010; Haeske-Seeberg 2001; Mansky/Nimptsch 2010a; Thomeczek 2002; Thüsing 2005; Eberlein-Gonska 2011). Ein noch so gut durchdachtes Anreizsystem, das die bis hier diskutierten Anforderungen erfüllt, kann nur bei einer reibungslosen Implementierung im Unternehmen und in die betrieblichen Prozesse funktionieren. Daher sind Operationalisierbarkeit und Umsetzbarkeit wichtige Anforderungen an Anreizsysteme. Dies beinhaltet Kriterien, die verschiedene Systemelemente betreffen können, so beispielsweise die zeitliche und betriebsübergreifende Vergleichbarkeit der Messergebnisse, Planbarkeit und Aktualität der Vergütungsfunktion sowie Kontrollierbarkeit der Qualitätsziele.

9. Anreizkompatibilität/Anreizvalenz: Um steuerbar zu sein und gleichzeitig eine motivierende Wirkung zu haben, müssen die Anreize und vor allem die Zielsetzungen gewisse Anforderungen erfüllen. Um Steuerbarkeit zu erfüllen, ist das Kriterium der Kompatibilität oder Kongruenz der Ziele und Anreize relevant (vgl. Lehmann 2006, S. 17). Das bedeutet, dass die individuellen Anreize auch an die gesamtbetrieblichen Ziele angepasst werden und gleichgerichtet sein müssen (vgl. Bleicher 1992, S. 24; Becker beschreibt dies im Kriterium der Individual- und Gruppenanwendbarkeit - Becker 1990, S. 25). Gleichzeitig dürfen die Anreize aber den individuellen Zielen der betroffenen Mitarbeiter nicht widersprechen. Dies gründet sich in der Anforderung der Anreizvalenz – nur bei einer Konsistenz zum individuellen Zielsystem kann ein motivierender Effekt entstehen (vgl. Lehmann 2006, S. 17; Bullinger 1996, S. 213f.). Insgesamt zeigt sich, dass das Zielsystem der Anreize zielgerichtet und wirksam gestaltet sein muss.

10. Rechtssicherheit der Messergebnisse: Die letzte hier zu besprechende Anforderung an Anreizsysteme ist die Rechtssicherheit. Bei Lehmann wird dies nur unter dem Aspekt der Einhaltung gesetzlicher Vorschriften gesehen (vgl. Lehmann 2006, S. 18). Doch dieses Kriterium kann weitere Elemente beinhalten, so z. B. eine Möglichkeit bzw. eventuell auch die Gefahr der Nutzung von Messergebnissen des Anreizsystems im Rahmen von Rechtsverfahren (bspw. bei Kunstfehlerprozessen) oder bspw. auch die Justiziabilität von Messergebnissen bei Auseinandersetzungen zwischen Krankenkassen und Leistungserbringern.

4.1.4. Relevanz der theoretischen Grundlagen für Pay-for-Performance-Systeme

Aus der vorangegangenen Diskussion der theoretischen Grundlagen von Anreizsystemen in der Wirtschaftslehre lassen sich Erkenntnisse für die vorliegende Arbeit ableiten. Diese haben eine hohe Relevanz für die Überlegungen zu qualitätsorientierten Vergütungssystemen in der stationären Versorgung. Die wichtigsten sind:

- Die relevante Beziehung zwischen den Beteiligten an einem P4P-System entspricht dem Principal-Agent-Konzept. Hierbei ist vor allem die Beziehung zwischen dem Krankenhaus und der Krankenversicherung entscheidend. Es ist jedoch auch auf die Ebene der Führungskräfte/Chefärzte anwendbar.
- Insgesamt sollte P4P als ein Anreizsystem im betrieblichen Sinne verstanden werden, so dass sowohl die entsprechenden Gestaltungsdimensionen als auch Anforderungen von Anreizsystemen angewendet werden können.
- Weitere theoretische Konzepte (so z. B. die Verhaltenstheorien) bieten Erkenntnisse zur Funktionsweise von P4P und erste Diskussionsansätze über mögliche Fallstricke (wie bspw. der mögliche „Verdrängungseffekt“ von externem Anreiz gegen intrinsische Motivation).
- Bezüglich der Funktionsweise und der Gestaltungsdimensionen sind zunächst die Systemelemente des (Qualitäts)ziels und des Messergebnisses sowie der Vergütungsfunktion als Gesamtanreizaspekt zu betrachten. Anhand dieser Elemente kann dieser Kriterienkatalog strukturiert werden.
- Mittels der Darstellung und Gegenüberstellung der existierenden Anforderungssammlungen aus der Betriebswirtschaftslehre können erste Kriterien identifiziert werden, die im zu entwickelnden Katalog Anwendung finden werden. Diese sind vor allem: Transparenz, Flexibilität, Gerechtigkeit, Akzeptanz, Motivationswirkung, Operationalisierbarkeit, Anreizvalenz und -kompatibilität und Rechtssicherheit. Weitere Kriterien sind allerdings notwendig, um ein P4P-System abschließend zu beschreiben.

4.2. Theoretische Grundlagen von qualitätsorientierten Anreizsystemen in der stationären Versorgung (P4P)

4.2.1. Relevanz von Pay-for-Performance (P4P)

Es wurde bereits die historische Entwicklung von Qualitätsmessung und -management dargestellt. Auch in den wegweisenden Arbeiten wie „To Err Is Human: Building a Safer Health Care System“ als auch „Crossing the Quality Chasm“ vom IOM wurde die Relevanz von P4P erwähnt. Im zweiten Werk wurden

neue Forderungen an das Gesundheitssystem gestellt, so unter anderem: „Build in stronger incentives for quality enhancement“ (Institute of Medicine 2001, S. 182). Eine Fokussierung auf qualitätsorientierte Vergütung wird somit vorgeschlagen, neben einer Intensivierung der Transparenz und der IT-Nutzung (vgl. auch Emmert 2008, S. 77).

Nun ergibt sich die Frage, warum gerade ein qualitätsorientiertes Vergütungssystem sinnvoll sein soll, um Qualitätsverbesserungen zu erzielen und Qualitätsprobleme zu lösen. Auf zwei Schwierigkeiten ist dabei besonders einzugehen: Defizite im derzeitigen stationären Vergütungssystem (DRG) in Bezug auf Anreize und die Schwierigkeiten der Informationsasymmetrie im stationären Sektor.

4.2.1.1. Anreizdefizite des DRG-Systems

Ein Vergütungssystem kann und soll auch immer die richtigen Anreize für erwünschtes Verhalten setzen, dies wurde bereits in den vorherigen Abschnitten klar. Unter dieser Prämisse ist auch das derzeitige stationäre Entgeltsystem – DRG-System – entwickelt worden. Heutzutage wird zumeist von zwei hauptsächlichen Zielsetzungen des DRG-Systems als auslösendes Momentum für die Entwicklung ausgegangen, nämlich (vgl. Geissler et al. 2011a, S. 10):

1. Transparenzverbesserung der in Krankenhäusern erbrachten Leistungen
2. Anreizsetzung für eine effiziente Ressourcenverwendung in Krankenhäusern

Zusätzlich kann daraus eine Qualitätsverbesserung oder -sicherung abgeleitet werden (vgl. Geissler et al. 2011a, S. 10). Einer der „Ziehväter“ des DRG-Systems (Robert B. Fetter) beschreibt ebenfalls die zunächst intendierte Anwendungsmöglichkeit in Qualitätssicherungsprogrammen, die jedoch in der eigentlichen Implementierung dann zumeist ausschließlich auf Krankenhausentgelt- und -finanzierungssysteme (ohne direkten Bezug auf Qualität) stattfand (vgl. Fetter 1999, S. 101). 1999 (also 18 Jahre nach der ersten Einführung von DRGs) propagiert er die Nutzung von DRGs in „total quality management“-Programmen.

Das stationäre Vergütungssystem weist nur geringe Einbindung von Qualitätsorientierung auf, so dass auch nur geringe Verbesserungsanreize für Ärzte entstehen (vgl. u. a. Emmert 2008, S. 76). Das deutsche DRG-System beinhaltet einige Aspekte, die eher eine Qualitätsverschlechterung inzentivieren, so bspw. die zusätzliche Vergütung für Beatmungsstunden oder die hohe Zahlungsrelevanz von (auch im Krankenhaus und durch die Behandlung erworbene) Komplikationen (vgl. InEK 2011). Auch die Zahlung von Revisionsoperationen bspw. bei der Hüft-Endoprothetik setzt keinen Anreiz zur Vermeidung solcher (medizinisch unerwünschten) Ereignisse. Gleichzeitig jedoch lassen sich Elemente finden, die eine Qualitätsverbesserung fördern; u. a. im deutschen System die Fallzusammenführung als eine „Nicht-Bezahlung“ von Wiedereinweisungen oder verschiedene Strukturvorgaben in abrechnungsrelevanten OPS-Ziffern (insbesondere bei Komplexleistungen) (vgl. Mansky/Seeger 2011, S. 195) oder in den USA die Nicht-Erstattung von ausgewählten, im Krankenhaus erworbenen Komplikationen von Medicare und Medicaid (vgl. Valuck 2008).

Generell wird davon ausgegangen, dass DRG-Systeme einen wünschenswerten, positiven Anreizeffekt auf Effizienz und Qualität haben (vgl. Cots et al. 2011, S. 75). Dies begründet unter anderem die große Beliebtheit der DRG-Systeme in der internationalen Entwicklung. Gleichzeitig besteht das Risiko, dass für den Zweck der Kostenreduktion pro Fall Krankenhäuser die Qualität nicht in den Vordergrund stellen und dadurch die Qualität verringern können – gerade da Ergebnisparameter keinen Einfluss auf die Vergütung haben (vgl. Or/Häkkinen 2011, S. 115). In den USA zeigt sich Evidenz, dass DRG-Systeme die Organisationseffizienz und Versorgungsqualität bestimmter Bereiche und Patientengruppen verbessert (vgl. Or/Häkkinen 2011). Andere Studien dagegen deuten auf eine Verschlechterung von Sterblichkeit und Wiederaufnahmen aufgrund von Kostendruck hin, allerdings nur als kurzfristige Ergebnisse (vgl. Culter 1995, S. 29ff.; Shen 2003, S. 243ff.). Während einige Studien zeigen, dass die Qualität der Versorgung nach der Einführung des DRG-Systems zumindest nicht gesunken ist (vgl. Sens et al. 2010, S. 25), gibt es kaum ausreichende Evidenz in Europa zu Qualitätsveränderungen nach Einführung der DRG-Systeme (vgl. Or/Häkkinen 2011, S. 118ff.).

Insgesamt bleibt zu vermerken, dass – auch wenn weder eindeutig positive noch eindeutig negative Effekte nachweisbar sind – DRG-Systeme gewisse Risiken zu Qualitätsbeeinträchtigungen mit sich ziehen. Außerdem wird die Möglichkeit, das Vergütungssystem mit positiven Qualitätsanreizen zu versehen, nicht hinreichend genutzt. Speziell die hohe Datendichte und die stabile Qualität der Abrechnungsdaten der DRG-Vergütung konnten neue Qualitätsmessungen auf Basis von Routinedaten einläuten (und können in einer qualitätsorientierten Vergütung Anwendung finden).

4.2.1.2. Schwierigkeiten der Informationsasymmetrie

Neben der unzureichenden Qualitätsorientierung der derzeitigen Entgeltsysteme im stationären Sektor ist die Informationsasymmetrie eine der Schwierigkeiten, die Qualitätsdefizite erklären und durch Elemente der qualitätsorientierten Vergütung behoben werden könnten.

Vor allem die Informationsdefizite von Patienten sind aufgrund des hohen Komplexitätsgrads der Medizin verständlich und nur bedingt zu beheben. Doch auch zwischen Krankenkassen und den behandelnden Ärzten gibt es Informationsunterschiede. Diese Tatsache kann ebenfalls zu Qualitätsdefiziten führen. Ein „Wertewettbewerb“ (z. B. „value-based competition“ nach Porter/Teisberg 2006) kann aber nur dann stattfinden, wenn die „Käufer“, d. h. Patienten oder Kassen, den Wert der Leistung auch beurteilen können. Eine höhere Transparenz und ein besseres Verständnis sind daher notwendig. So fordert auch das IOM mehr Transparenz (vgl. Institute of Medicine 2001, S. 182).

Dass eine fehlende Transparenz zu Qualitätsdefiziten führt, ist vor allem daran festzumachen, dass bei erhöhter Transparenz die Qualität der Dienstleistungen steigen kann. So kann Datenveröffentlichung einen positiven Effekt auf Versorgungsqualität von Leistungserbringern zeigen, z. B. durch den Versuch der Vermeidung von veröffentlichten, schlechten Qualitätsergebnisse (vgl. unter anderem Chassin 2002).

Insgesamt zeigt sich jedoch, dass die Auswirkungen von Public Reporting als Erhöhung der Transparenz nur in ersten Ansätzen untersucht sind (vgl. Emmert 2008, S. 326ff.). Derzeit wird vor allem das Public Reporting entweder im Zusammenhang mit anderen, nicht-finanziellen Anreizen (vgl. Bufalino et al. 2007) oder aber als ein Teilaspekt der finanziellen Anreizsysteme im Rahmen von P4P gesehen (vgl. Lindenauer et al. 2007). Auch der Sachverständigenrat versteht P4P als eine konzeptionelle Weiterführung des „Public Disclosures“ (vgl. SVR 2007).

Die Erhöhung der Transparenz durch ein qualitätsorientiertes Anreizsystem ist nicht zwingend notwendig – theoretisch könnte eine rein finanzielle Vergütung mit variablem Qualitätsanteil auch ohne die Veröffentlichung der Qualitätsergebnisse gestaltet werden. Faktisch jedoch werden in einem Großteil der existierenden Systeme die Daten auch der Öffentlichkeit zur Verfügung gestellt. Dies ist sinnvoll, da die Daten bereits erhoben sind und somit der Mehraufwand einer Veröffentlichung vergleichsweise gering ist. Gleichzeitig ist mit einem zusätzlichen positiven Effekt auf die Qualitätsergebnisse zu rechnen. Daher soll auch in der vorliegenden Arbeit ein P4P-System stets mit dem Element des Public Reportings verstanden werden. Trotzdem liegt der Fokus der Arbeit auf den finanziellen Anreizen eines solchen Systems.

Insgesamt zeigt sich, dass verschiedene Strukturaspekte des Gesundheitssystems Qualitätsdefizite bedingen und fördern können. Defizite des Vergütungssystems und die Informationsasymmetrie können dabei zumindest zum Teil durch die Anwendung eines qualitätsorientierten Vergütungssystems verbessert werden.

4.2.2. Definition und Begrifflichkeit von Pay-for-Performance

Erste Ansätze von P4P-Projekten wurden Ende der 90er Jahre in den USA entwickelt. Sowohl in den USA als auch in Großbritannien gibt es eine Vielzahl von großangelegten Programmen, die qualitätsorientiert vergüten (vgl. Emmert 2008, S. 77ff.).

Eine einheitlich anerkannte Definition von Pay-for-Performance existiert kaum, vor allem da die meisten sich damit befassenden Arbeiten auf eine definitorische Beschreibung verzichten (vgl. SVR 2007, S. 171). Eine Liste von Beschreibungen sind unter anderem bei Emmert, Scheppach oder im BQS-Gutachten zu finden (vgl. Emmert 2008, S. 73f.; Scheppach et al. 2011, S. Anhang; Veit et al. 2012b, S. 7f.). Die Definition von Millenson beschreibt P4P als eine Vergütung, die Qualitäts- und Sicherheitsergebnisse mit Vergütung verknüpft (vgl. Millenson 2004, S. 323). Als „working definition“ wird für diese Arbeit folgende Beschreibung benutzt:

- Pay-for-Performance oder qualitätsorientierte Vergütung beschreibt Systeme im stationären Sektor, die abhängig von Qualitätsergebnissen des Leistungserbringers variable finanzielle[1] Anreize nutzen, um Behandlungsqualität nachhaltig zu verbessern.

Neben den verschiedenen Definitionen von P4P hat sich außerdem eine Vielzahl von Begrifflichkeiten für die qualitätsorientierte Vergütung etabliert. Neben „Pay-for-Performance (P4P)“ als bekanntester Begriff sind dies „value based purchasing“ (vgl. Porter/Teisberg 2006; Lindenauer et al. 2007), „Pay-for-Quality (P4Q)“ (vgl. Bufalino 2006; Seeger 2011), „performance-based“ oder „quality-based contracting“ (vgl. Shen 2003; Cannon 2007) und die deutschen Begriffe „Qualitätsorientierte Vergütung” (vgl. u. a. Amelung et al. 2009; Köhler 2010) und „Qualitätsbasierte oder Zielorientierte Finanzierung“ (vgl. SVR 2007, S. 172). Auch spezifische Varianten wie „value-based insurance design“ oder „Pay-for-Performance-for-Patients (P4P4P)” (vgl. Volpp et al. 2009) finden Verwendung.

Diese verschiedenen Definitionen und Begrifflichkeiten zeigen sowohl die schnelle Entwicklung der P4P-Forschung und -Umsetzung als auch die Fülle von unterschiedlichen Vorstellungen. Bei den erwarteten Zielen und möglichen Gestaltungsdimensionen ist dies ebenfalls sichtbar.

4.2.3. Ziele von P4P-Systemen

Das Ziel von qualitätsorientierten Vergütungssystemen ist die nachhaltige Verbesserung von Behandlungs- oder noch spezifischer Ergebnisqualität (vgl. Institute of Medicine 2007, S. 9). Dies bedeutet auch eine „Gleichschaltung“ der Interessen des Patienten und Leistungserbringers und der Kasse (vgl. Amelung/Zahn 2009, S. 16). Im Detail jedoch kann man verschiedene Zielsetzungen der P4P-Systeme unterscheiden, so z. B. zwischen verschiedenen Qualitätsvorstellungen (Einbeziehung Patientenzufriedenheit oder Fokus auf medizinische Qualität), weiteren Dimensionen der „Leistung“ (Einbeziehung von Kosten- oder Effizienzparametern) oder die Arten des Qualitätsvergleiches (Relativ vs. absolute Vergleiche) (vgl. Braun et al. 2012, S. 16).

[1] Nicht-finanzielle Anreize (wie z. B. Public Reporting) werden auch betrachtet, stehen aber nicht im Mittelpunkt.

Tabelle 6: Verschiedene Zielsetzungen von P4P-Programmen

	USA	UK	Estland	Israel	D
Qualitätsunterschiede nivellieren	x				
Qualitätsniveau erreichen bzw. halten (absolut)	x	x			
Versorgungsqualität verbessern (relativ)	x				
Chronikerversorgung verbessern	x		x		X
Prävention verbessern	x	x	x		
Praxismanagement verbessern (Service, Wartezeit)		x			
Effizienz steigern	x	x		x	X
Kosten senken	x				X
Koordination zwischen Leistungserbringern fördern	x	x		x	
Behandlungsverläufe per IT dokumentieren	x	x			
Evaluations- und Feedbacksysteme	x			x	
Patientenorientierung/ Patientenzufriedenheit steigern	x	x			X

Quelle: Auszug aus Braun et al. 2012, S. 16

Insgesamt zeigt sich, dass die Zielvorstellungen von P4P-Systemen ebenso vielfältig sind wie deren Ausgestaltung. Doch einige übergreifende Zielsetzungen können gesehen werden, so unter anderem in (vgl. auch SVR 2007, S. 171)[1]:

- Qualitätsverbesserungen
- Verbesserungen der Patientensicherheit
- Verbesserung der sektorübergreifenden Koordination
- Erhöhung der Effizienz

[1] Mehr Details zu existierenden P4P-Systemen in Abschnitt 5.1. und 5.2.

4.2.4. Funktionsweise von P4P-Systemen[1]

Die Funktionsweise von P4P-Systemen ist mit den Konzepten der Organisationstheorie und Verhaltenswissenschaften eng verknüpft. Gleichzeitig handelt es sich im Rahmen des stationären Sektors um ein System mit sehr hoher Komplexität (vgl. SVR 2007, S. 171; Veit et al. 2012e, S. 6f.). Trotz dieser Komplexität, die verschiedene Elemente und Faktoren wie bspw. Qualitätsmessung, Anreiz-Rahmen, Anreize, Feedback, Motivation, Ziele, Evaluation beinhalten können (vgl. Veit et al. 2012e, S. 6), ist der eigentliche Ablauf des Motivationsmechanismus mit den theoretischen Grundlagen von Anreizsystemen kompatibel – daher wird im Folgenden auf diese fokussiert.

Außerdem können verschiedene Ansätze der evidenz-basierten Medizin Einzug erhalten. Diese können im Rahmen von Leitlinien dem Arzt als operative Hilfestellung dienen. Insgesamt wird bei P4P-Systemen versucht, durch Verhaltensänderungen beim Leistungserbringer und besonders beim einzelnen Arzt das Ziel der Qualitätsverbesserung zu erreichen (vgl. hierzu zum einen die Ideen von „Stimulus und Response“ im Abschnitt 4.1.2.1 und SVR 2007, S. 171).

Finanzielle Anreize können nur dann wirksam sein, wenn es zu Veränderungen in den Ablauf- und Aufbauorganisationen des Leistungserbringers sowie zu Verhaltensänderungen des medizinischen Personals kommt. Diese Umsetzungen können bspw. im Rahmen eines Qualitätsmanagements geschehen. Hierfür bieten sich die verschiedenen Konzepte wie der PDCA-Zyklus und andere Aspekte von „Change Management“ und Managementsystemen an (vgl. u. a. Bleicher 2002, S. 1ff.; Funck 2002, S. 25ff.; Imai 1993). Ein qualitätsorientiertes Vergütungssystem sollte diese Implementierung nicht behindern sondern sogar durch die Anwendbarkeit im Managementkontext fördern.

Ein weiterer interessanter Aspekt von P4P können die relevanten Rahmenbedingungen oder „Prädiktoren“ für Projekte sein – so ist abhängig von bspw. der Konkurrenzsituation, den finanziellen Ressourcen, der Baseline-Performance oder der Grundhaltung des Unternehmens mit unterschiedlichen Effekten von P4P-Systemen zu rechnen (vgl. Veit et al. 2012c, S. 29f.).

4.2.5. Gestaltungsdimensionen von P4P-Systemen

Gestaltungsdimensionen von P4P sollen in diesem Kontext verstanden werden als diejenigen Parameter, anhand derer ein System beschrieben werden kann. Es handelt sich dabei nicht um Anforderungen, d. h. nicht um Bewertungskriterien, sondern um neutrale Charakteristika der Systeme. Gängige Begriffe sind unter anderem „design choices“ (van Herck et al. 2010) oder „Umsetzungsmodalitäten“ (Schrappe/Gültekin 2010). Die derzeitige Diskussion über P4P-Systeme in Deutschland konzentriert sich vorwiegend auf die Frage der Gestaltungsdimensionen (vgl. u. a. SVR 2007; Veit et al. 2012a;

[1] Die Evidenz von P4P-Projekten wird von einer Vielzahl von Studien betrachtet – dies wird in Abschnitt 5.1.2. näher diskutiert.

Schrappe/Gültekin 2010; Scheppach et al. 2011; van Herck et al. 2010; Amelung/Zahn 2009). Anforderungen werden seltener betrachtet (u. a. bei Mansky 2010b oder erste Mindeststandards bei Veit et al. 2012e, S. 73).

Beispiele für die Darstellung von Gestaltungsdimensionen von P4P sind:

Tabelle 7: Beispielhafte Gestaltungsdimensionen für P4P-Programme (Auswahl)

nach Scheppach	**nach Schrappe**
Empfängerebene	Adressaten
Anreizhöhe	Anreizhöhe
Bemessungsgrundlage	Monetäre Bewertung
Dokumentation des Leistungsdaten	Qualitätsmessung
Bonus oder Malus	Indikatorenauswahl
Dauer des Anreizes	Spezifizierung der Indikatoren
Zeitpunkt des Anreizes	Zahlungshäufigkeit
Versorgungsart	Spezifizierung der Anreize
Anreizfinanzierung	
Nicht-finanzielle Anreize	
Relevante Umweltbedingungen	

Quelle: Scheppach et al. 2011, S. 101ff.; Schrappe/Gültekin 2010, S. 110ff.

Dies ist nur eine Auswahl der möglichen Gestaltungsdimensionen, die bei der Implementierung von P4P-Systemen relevant sind. Bei einem Großteil dieser Parameter kann bislang keine Empfehlung gegeben werden, welche Ausprägung vorteilhaft für das P4P-System ist und eine positive Korrelation mit der Wirksamkeit hat. So ist bei der Zahlungshäufigkeit derzeit kein eindeutiger Zusammenhang zu Qualitätsverbesserungen in den durchgeführten Reviews erkennbar, während gleichzeitig zu wenig Evidenz bezüglich der sinnvollen Adressatenebene – d. h. individuelle Ärzte oder Krankenhäuser besteht (vgl. Schrappe/Gültekin 2010, S. 110; Scheppach et al. 2011, S. 110). Bei anderen Parametern dagegen ergeben sich die Empfehlungen auch sachlogisch, so bspw. dass die Spezifizierungen der Indikatoren detailliert beschrieben und definiert werden sollten (vgl. Schrappe/Gültekin 2010, S. 111).

Bei einigen Gestaltungsparametern jedoch gibt es bereits erste Ergebnisse und Aussagen bezüglich einer sinnvollen Ausprägung, so z. B. bei:

- *Bemessungsgrundlage für die Zahlungen*: Verschiedene Reviews so u. a. van Herck und Rosenthal konnten zeigen, dass die Anwendung mehrerer Bemessungsgrundlagen (d. h. der Vergleich der Leistung zu einem Schwellenwert, dem Vorjahres-Wert und innerhalb eines Rankings) die meisten

Vorteile bietet (vgl. Scheppach et al. 2011, S. 107f.; van Herck et al. 2010; Rosenthal 2008; Amelung/Zahn 2009, S. 28)

- *Auswahl von Indikatoren*: Für die Qualitätsindikatoren gibt es zum einen einige Stimmen, die sich für die stärkere Verwendung von Patientenzufriedenheit als Indikator aussprechen (so z. B. Amelung/Zahn 2009, S. 36). Zum anderen gibt es vor allem Äußerungen bezüglich der Qualitätsdimensionen nach Donabedian – d. h. der Struktur-, Prozess- und Ergebnisqualität. Derzeit werden vor allem Prozessindikatoren in P4P-Systemen angewandt (vgl. Scheppach et al. 2011, S. 118). Eine Kombination von allen drei Dimensionen ist laut Aussage verschiedener Studien sinnvoll (vgl. u. a. van Herck et al. 2010), einige Arbeiten weisen auch auf den Vorrang von Ergebnisqualität hin – allerdings mit dem Zusatz, dass bspw. Risikoadjustierungen hierfür erfolgen müssen (vgl. Veit et al. 2012e, S. 10ff.). Gerade für die Indikatorenauswahl werden innerhalb der Gestaltungsparameter auch einige Anforderungen diskutiert: Laut Scheppach sollten Indikatoren „klinisch relevant, standardisiert, verständlich, leicht messbar und valide sein“ (Scheppach et al. 2011, S. 120).

- *Nutzung nicht-finanzieller Anreize (v. a. Public Reporting*): Verschiedene nicht-finanzielle Elemente werden im Zusammenhang mit P4P-Systemen diskutiert, allen voran Public Reporting. Die Kombination finanzieller Anreize mit der Veröffentlichung der Qualitätsergebnisse wird nicht nur allgemein empfohlen (so unter anderem bei SVR 2007, S. 175; Institute of Medicine 2007). Gleichzeitig zeigen bereits eine Vielzahl von Studien positive Effekte des Public Reportings auf die Qualitätsverbesserung des P4P-Systems (so u. a. Lindenauer et al. 2007; Mehrotra et al. 2010a, S. 128). Zusätzlich ist zu erwähnen, dass allein schon aus Pragmatismusgründen Public Reporting als eine sinnvolle Zusatzmaßnahme erscheint: Wenn für die Grundlage der variablen, qualitätsorientierten Vergütung bereits akzeptierte detaillierte Qualitätsmessungen stattgefunden haben, wäre eine Veröffentlichung der Ergebnisse mit geringen Grenzkosten möglich.

- *Vergütungshöhe:* Aussagen zum relevanten Vergütungsanteil sind schwierig, da die Studienlage derzeit verschiedene Erkenntnisse zulässt (vgl. Scheppach et al. 2011, S. 110f.; Rosenthal/Frank 2006; Christianson et al. 2008). Erste Aussagen zur Vergütungshöhe unterteilen vor allem anhand einer 5 %-Grenze – für größere Einrichtungen sind ca. 3 % der Gesamtvergütung, für kleine Einheiten ca. 10 % als wirksam zu erachten (vgl. Veit et al. 2012e, S. 32; Amelung dagegen geht sogar von 10-20 % aus, vgl. Amelung/Zahn 2009, S. 58). Gleichzeitig müssen die Anreize mindestens die entstandenen Kosten der verbesserten Behandlung und der Systemimplementierung decken (vgl. Christianson et al. 2008, S. 12ff.). In einer Überblicksarbeit über 27 „Early Adopter“ von P4P in den USA wurden durchschnittlich 2,3 % der Gesamtvergütung als variable, qualitätsorientierte Vergütung genutzt (vgl. Rosenthal et al. 2008, S. 1678).

4.3. Beschreibung der Entwicklung des Anreizteils des Kriterienkatalogs

Während für Qualitätsindikatoren im stationären Sektor bereits für Teilaspekte Anforderungslisten existieren, gibt es für die vorliegenden Fragestellungen nach bestem Wissen noch keine konkreten Aufzählungen möglicher Kriterien. Wie bereits angesprochen befasst sich die Diskussion derzeit vor allem mit der Ausgestaltung von P4P-Systemen. Insgesamt gibt es zwei relevante Quellen für die Entwicklung des Katalogs:

- Bewertungsschemata für P4P-Systeme in – systematischen – Reviews (so z. B. in Deutschland SVR 2007; Scheppach et al. 2011; Veit et al. 2012a[1], Schrappe/Gültekin 2010; Mehrotra et al. 2010a und Amelung/Zahn 2009 oder international Rosenthal et al. 2005; van Herck et al. 2010; Bufalino et al. 2007; Institute of Medicine 2007 und Kuhmerker/Hartman 2007)
- Anforderungskataloge an Anreizsysteme aus der betriebswissenschaftlichen Literatur (so bspw. Becker 1990; Bleicher 1992; Seidel 1991; Bullinger 1996; Kerr 1997; Ortlieb 1993; Grewe 2000; Lehmann 2006; Laux 2006; Wünsch 2012)

4.3.1. Systemelemente des Anreizteils des Kriterienkatalogs

Die Systemelemente des Anreizteils des Kriterienkatalogs sind:

5. **Ziele:** Nachdem die Ergebnisse der Qualitätsmessung vorliegen, müssen anschließend Ziele für das Anreizsystem festgesetzt werden. Nur bei einem klaren Verständnis der Zielsetzung kann eine Anreizwirkung erfolgen.
6. **Vergütungsfunktion:** Der nächste Schritt zum Anreizsystem ist die Definition der sogenannten Vergütungsfunktion. Damit wird die Korrelation verstanden, die zwischen der Zielerreichung und der (zusätzlichen) variablen Vergütung besteht.
7. **Messergebnis:** Nachdem die Messung samt aller relevanten Elemente (Versorgungsbereich, Indikator) bereits im vorherigen Kapitel diskutiert wurden, sollen an dieser Stelle die Anforderungen an die Definition des Messergebnisses und dessen Verwendung besprochen werden. Da in der praktischen Anwendung von P4P-Systemen die Messergebnisse zur Umsetzung von Qualitätsverbesserungsmaßnahmen und daher meist nach der Vergütungsfunktion genutzt werden, sind die Anforderungen an das Messergebnis hier als letztes Systemelement dargestellt.

[1] Hier u. a. auch ein „Algorithmus“ zur Überprüfung der Kompatibilität und Eignung von Indikatoren für verschiedene P4P-Systeme, deren einzelne Anforderungen jeweils bei den Kriterien erwähnt werden (vgl. Veit et al. 2012e, S. 42ff. sowie Veit et al. 2012f, S. 69ff.)

4.3.2. Kriterienkategorien des Anreizteils des Kriterienkatalogs

Kriterienkategorie 6: Wirksamkeit

Sowohl die Ziele als auch die Vergütungsfunktion müssen wirksam und zielgerichtet sein. Vor allem die Anforderungen der „Anreizkompatibilität“ und „Anreizvalenz“ bei Lehmann sind hierfür relevant (vgl. Lehmann 2006, S. 16ff.). Themen wie Effizienz, Effektivität (für Vergütungsfunktion) und Kongruenz (für das Zielsystem) fallen unter diese Kriterienkategorie.

Kriterienkategorie 7: Operationalisierbarkeit

Die Operationalisierbarkeit gilt für alle Systemelemente und trägt eine besonders hohe Relevanz für diesen Teil des Anforderungskatalogs. Die Nutzung von Qualitätsmessungen innerhalb eines Anreizsystems kann bereits als Operationalisierung des Qualitätsbegriffes an sich verstanden werden. Dies bedeutet, dass alle drei Systemelemente so angewandt werden können, dass das eigentliche Ziel des Systems erfüllt werden kann, also die Qualität in den Strukturen des Leistungserbringers verbessert werden kann und gleichzeitig ein nachhaltiges System entsteht.

Kriterienkategorie 8: Vergleichbarkeit

Die letzte Kriterienkategorie ist für Ziele und Messergebnisse relevant. Gerade bei den Messergebnissen sind hier verschiedene Varianten der Vergleiche zwischen Leistungserbringern oder Fachabteilungen gemeint. Dazu muss eine subjektive und objektive Gerechtigkeit erreicht werden, um die Anreizwirkung zu ermöglichen. Diese Kriterienkategorie ergibt sich zum einen aus dem Kriterium der „Gerechtigkeit“ bei verschiedenen betriebswirtschaftlichen Quellen (vgl. Lehmann 2006; Grewe 2000; Bleicher 1992) und zum anderen bspw. aus verschiedenen Bemessungsgrundlagen und „Ranking-Arten“ aus P4P-spezifischen Gestaltungsdimensionen (vgl. Scheppach et al. 2011; SVR 2007).

4.4. Der 2. Teil des Kriterienkatalogs: Beschreibung und Diskussion der Kriterien des Anreizteils

Gerade für die standardisierte Bewertung der Einzelkriterien des zweiten Katalogteils ist langfristig eine Skala- und Grenzwertentwicklung möglich und wünschenswert. Eine Vielzahl dieser Kriterien (bspw. Aktualität oder Mindestanteil der Vergütung) sind bereits sehr quantitativ. Auch hier wäre langfristig ein „gewichteter Gesamtscore“ erstrebenswert (vgl. u. a. auch Mansky 2010a). Solange dies nicht entwickelt wurde, werden auch hier eine Delphibefragung im Rahmen eines Expertenpanels sowie eine Ordinalskala vorgeschlagen (vgl. Häder 2002; Mattke et al. 2006, S. 2 und Reiter et al. 2007).

Kriterium 7a: Zielkonsistenz

Systemelement: Ziele

Kriterienkategorie: Wirksamkeit

Beschreibung: Unter Zielkonsistenz ist zu verstehen, dass die Ziele, die für die handelnden Akteure gesetzt werden, widerspruchsfrei zu den eigenen internen Zielen sind. Die Ziele des P4P-Systems dürfen nicht den internen Zielen der Akteure widersprechen, indem bspw. Anreize für ein Ziel geschaffen werden, das mit dem Berufsethos im Konflikt steht.

Begründung/Bezug zu existierenden Systemen: Um eine ausreichende Anreizwirkung zu erreichen, ist es notwendig, ein Zielsystem für den Leistungserbringer und die dort agierenden Personen zu entwickeln, das konfliktfrei angestrebt werden kann. Dieses Kriterium kann aus der betriebswissenschaftlichen Lehre hergeleitet werden – Lehmann beschreibt hierfür die Anreizvalenz (vgl. Lehmann 2006, S. 17; siehe auch Horvath 2008). Zusätzlich ist die Diskussion bezüglich des „Crowding-Out" oder Verdrängungseffekts von intrinsischer Motivation durch externe Anreize hierfür relevant. Dieses Phänomen (wie bereits dargestellt) wird sowohl in der allgemeinen Motivationslehre als auch speziell für die Motivationseffekte bei der Ärzteschaft kontrovers diskutiert (vgl. allgemein Frey 2005; Matiaske/Weller 2006 und spezifisch für Ärzte SVR 2007, S. 184; Bufalino 2006; Custers et al. 2008). Um das Risiko zu minimieren, dass die Anreizwirkung von variabler Vergütung die intrinsische Motivation und den hohen Berufsethos von Ärzten verdrängt (auch wenn dies laut Steel und Willem nicht bestätigt werden kann (vgl. Scheppach et al. 2011, S. 126)), muss dafür Sorge getragen werden, dass die Ziele ohne Inkonsistenzen gesetzt werden.

Anwendungsdefinition und -beispiele: Aufgrund der derzeitigen Vergütungssystematik im stationären Sektor kann es bereits zu Fehlanreizen für den Leistungserbringer kommen. Gleichzeitig gibt es in vielen Krankenhäusern bislang keine Zielsetzungen für Führungskräfte oder weitere Mitarbeiter, in denen auch Qualitätsziele vergütungsvariabel festgelegt sind. Insgesamt sollte bei der Entwicklung von Qualitätszielen auf folgende Aspekte geachtet werden:

- Qualitätsziele des P4P-Systems werden zunächst auf Unternehmensebene[1] festgelegt, diese Ziele sollten anschließend für Führungskräfte und weitere Mitarbeiter heruntergebrochen werden. Dabei sollten die …
  - … Unternehmensqualitätsziele konsistent mit den weiteren Zielen des Unternehmens sein (beispielsweise die wirtschaftlichen Ziele mit betrachten)

[1] An dieser Stelle sind zunächst die Leistungserbringer und nicht die Krankenversicherungen gemeint. Die Zusammenwirkung von Kasse und Krankenhaus im Sinne einer Principal-Agent-Beziehung werden bei späteren Kriterien relevant.

- … Mitarbeiterqualitätsziele konsistent zu den persönlichen Zielen der Führungskräfte und Mitarbeiter sein (d. h. vor allem den medizinischen Berufsethos beachten)

- Um diese Anforderung zu erreichen, müssen zunächst sowohl die Ziele des P4P-Systems als auch die individuellen und Unternehmensziele transparent dargestellt werden. Außerdem ist eine Einbeziehung der beteiligten Mitarbeiter notwendig, um Konsistenz anzustreben.

- Jedoch muss erwähnt werden, dass bei einer zu einseitigen Anpassung der Qualitätsziele an bereits vorhandene Individual- oder Unternehmensziele, ggf. die eigentliche angestrebte Wirkung der Qualitätsverbesserung konterkariert wird. Daher muss ein sinnvolles Gleichgewicht zwischen den „externen“ Qualitätszielen des P4P-Systems und der „internen“ Ziele gefunden werden.

Informationsquellen: Um dieses Kriterium bewerten zu können, müssen zunächst die Ziele des P4P-Systems klar definiert sein. Die hauptsächliche Bewertungsquelle stellt daher die Definitionen der verschiedenen Zielsysteme dar. Für ein P4P-System sollte eine detaillierte Zielbeschreibung auf jeden Fall vorliegen, sonst sollte bereits eine negative Beurteilung stattfinden. Liegen allerdings keine Beschreibungen der zu vergleichenden Individual- und Unternehmensziele vor, sollten diese im Rahmen von Umfragen oder Experteninterviews etabliert werden.

Kriterium 7b: Zielkongruenz

Systemelement: Ziele

Kriterienkategorie: Wirksamkeit

Beschreibung: Die Zielkongruenz bezeichnet die Gleichgerichtetheit der Zielsetzungen innerhalb des Zielsystems – bspw. der Unternehmens- und der Führungs- und Mitarbeiterebene oder auch bei mehreren, gleichzeitig existierenden Zielen. Die Ziele dürfen sich nicht widersprechen und müssen die gleiche Zielrichtung tragen.

Begründung/Bezug zu existierenden Systemen: Zunächst ist bei Vorhandensein von mehreren Zielen in einem Zielsystem der qualitätsorientierten Vergütung generell deren Gleichgerichtetheit zu beachten – nur wenn alle Ziele „gleichzeitig“ erreicht werden können und die Ziele bspw. verschiedener Fachrichtungen widerspruchsfrei sind, kann eine Anreizwirkung entstehen. Aufgrund der hier vorliegenden Principal-Agent- und Anreiz-Beziehungen, die sowohl die „oberste“ Organisationsebene der Leistungserbringer aber auch die Führungs- und Mitarbeiterebene des medizinischen Personals als qualitätsbeeinflussende Akteure umfassen, muss zudem eine einheitliche Zielrichtung der verschiedenen Ebenen hergestellt werden. Dies wird zum einen in verschiedenen Kriterien der Anreiztheorien klar (so bei Becker 1990, S. 25; Bleicher 1992, S. 24; Lehmann 2006, S. 17; siehe auch Horvath 2008). Zum anderen wird das Thema im Rahmen der P4P-Gestaltungsdimensionen insoweit besprochen, dass die Adressatenebene (derzeit noch ergebnisoffen) diskutiert wird (vgl. u. a. Scheppach et al. 2011, S. 102f.;

vgl. Weiner et al. 2006). Dass auch bei einer Auszahlung an die „Gruppe“ bzw. Organisation die jeweiligen Ziele zueinander ausgerichtet sein müssen, wird nicht angesprochen[1]. Daher kann dieses Kriterium als recht neuartig angesehen werden. Es unterscheidet sich zur Zielkonsistenz, da hier nicht über „externe vs. interne“ Aspekte sondern über das Zusammenspiel zwischen Zielen der Unternehmen und Einzelpersonen nachgedacht wird.

Exkurs – gesellschaftliche Zielkonflikte: Ein Beispiel für divergierende Ziele ist im gesellschaftlichen Kontext zu sehen: Aufgrund der hohen gesellschaftlichen Bedeutung der Gesundheitsversorgung sind Vertreter der Gesundheitsbranche, der Politik und der Wirtschaft sehr an Ergebnissen der medizinischen Qualität interessiert. Doch durch unterschiedliche Vorstellungen entstehen Forderungen, die im Widerspruch zueinander stehen können, so z. B.:

- Beste Strukturqualität
- Wirtschaftlichkeit und effiziente Ressourcenverteilung
- Umfassende Facharztversorgung
- Wohnortnahe Versorgung

Diese Ansprüche können nicht gleichzeitig erfüllt werden. Daher müssen Entscheidungen getroffen werden, die die notwendigen Gewichtungen möglichst gerecht und effizient für die Bevölkerung einordnen. Ein Beispiel für diesen Zielkonflikt kann in der Behandlung von Herzinfarkten und Schlaganfällen gesehen werden. Für die optimale Versorgung wird heute ein Linksherzkatheter bei Herzinfarkt oder eine Stroke unit bei Schlaganfall als medizinisch sinnvoll angesehen. Sowohl Linksherzkathetermessplätze als auch Stroke units erfordern Mindestfallzahlen, um wirtschaftlich sinnvoll und ärztlich qualifiziert besetzbar zu sein. Die Versorgungsrealität zeigt jedoch, dass 25 % der Krankenhäuser in Deutschland „nur“ bis zu 36 Herzinfarkte und 15 Resektionen der Schilddrüse im Jahr 2010 behandelt haben (vgl. Nimptsch/Mansky 2012a, S. 1452). Wirtschaftlichkeit und hohe Versorgungsqualität können mit diesen Fallzahlen und dadurch bedingter geringer Erfahrung kaum erreicht werden (vgl. Nimptsch/Mansky 2012a, S. 1455).

Anwendungsdefinition und -beispiele: Für dieses Kriterium sind die unterschiedlichen Ziele im Zielsystem, der verschiedenen Adressatenebenen und in den unterschiedlichen Principal-Agent-Beziehungen relevant. Zunächst müssen die verschiedenen Ziele (beispielsweise von verschiedenen Versorgungsbereichen) eines Zielsystems widerspruchsfrei nebeneinander erreichbar sein.

[1] Bei Amelung wird kurz auf die „Widerspruchsfreiheit“ der Indikatoren eingegangen, ohne detaillierte Anforderungen zu entwickeln (vgl. Amelung/Zahn 2009, S. 20).

Auch die Ziele der verschiedenen Adressaten beim Leistungserbringer (d. h. vor allem Gesamtorganisation vs. Ärzte) können divergieren. Um dies zu verhindern, wird zunächst davon ausgegangen – wie zuvor bei der Zielkonsistenz –, dass nach der Definition der Qualitätsziele für die Unternehmensebene, diese dann für die Führungskräfte und Mitarbeiter heruntergebrochen werden (vgl. auch Beschreibung der Zielsetzung bei Schär 2002).

- Während dieses Prozesses muss sichergestellt werden, dass …
  - …die Summe aller Einzelziele (bspw. auf Fachabteilungsebene) das gesamte Zielsystem widerspiegeln und widerspruchsfrei gleichzeitig erreichbar sind
  - …die Einzelziele denen der Gesamtorganisation nicht widersprechen.
- Zusätzlich müssen auch die Ziele aller einzelnen Organisationen und Akteure mit denen des Gesundheitssystems kongruent sein (siehe auch Thematik des Exkurses).
- Vor allem durch die Einbeziehung aller Beteiligten und Interessierter in den Entwicklungs- und Evaluationsprozess kann hoffentlich ein für alle tragbarer Kompromiss gefunden werden.

Informationsquellen: Für dieses Kriterium müssen die definierten Zielsetzungen analysiert werden, so dass die Informationsquelle rein innersystemisch ist. Wenn keine klaren Ziele vorliegen, müssen die „indirekten“ Ziele anhand von qualitativen Interviews erfragt werden.

Kriterium 7c: Reichweite

Systemelement: Ziele

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

Beschreibung: Unter der „Reichweite“ des Zielsystems ist zu verstehen, dass die Gesamtheit der Ziele die Gesamtheit der Leistungen der Akteure (und vor allem des Leistungserbringers) möglichst vollständig abdeckt. Zumindest sollte ein ausreichender Anteil der Leistungen abgedeckt sein, um überhaupt eine ausreichende Anreizwirkung zu erlangen.

Begründung/Bezug zu existierenden Systemen: Wenn die Qualität „als Ganzes“ verbessert werden soll, muss auch das Zielsystem möglichst vollständig alle Bereiche der Leistungen umfassen. Zunächst ist hierfür die Betrachtung der Messung und Indikatoren (siehe hierfür auch Kriterium 1a), am Anschluss aber vor allem die Definition der Qualitätsziele relevant.

Bereits bei der Messung von Qualität ist es sinnvoll und notwendig Vollständigkeit zu beachten. Doch gerade bei Anreizsystemen ist dies von großer Bedeutung: Wenn nur einige Leistungsbereiche mit (finanziellen) Anreizen versehen sind, muss davon ausgegangen werden, dass das medizinische Personal gerade diese Bereiche beachtet – andere Leistungen können vernachlässigt werden, ggf. werden sogar

bestimmte Patienten bevorzugt (vgl. auch das „Cherry Picking“ u. a. bei Scheppach et al. 2011, S. 123). Je weiter der von den Qualitätszielen betroffene Leistungsbereich gefasst ist, desto unwahrscheinlicher wird ein solches Verhalten.

Während auch die Qualitätsmessung bereits relevant und „stellvertretend für einen Leistungsbereich“ sein sollte, so müssen gerade und vor allem die daraus erarbeiteten Ziele so gestaltet sein, dass ein möglichst großer Bereich abgedeckt ist. Dies ist nicht automatisch der Fall – die Ziele können, müssen aber nicht alle Messergebnisse der Qualitätsindikatoren beinhalten. Werden nur einzelne Leistungsbereiche betrachtet, kann es zu Fehlanreizen kommen.

Außerdem kann eine zu geringe Fallzahl zu statistischen Streuungen führen, so dass selbst größere Abweichungen nicht signifikant sind. Kliniken mit kleiner Fallzahl können qualitativ nie auffällig werden, da keine signifikanten Abweichungen festgestellt werden können. Gleichzeitig zeigen verschiedene Studien, dass gerade bei Kliniken mit kleinen Fallzahlen mit Qualitätsmängeln gerechnet werden muss und daher eine höhere Aufmerksamkeit geboten wäre (vgl. Geraedts/de Cruppe 2006). Ein Sonderproblem stellen Sentinelereignisse dar (mit Häufigkeit weniger als 1:100, z. B. Todesfälle bei risikoarmen Operationen). Diese Ereignisse sind Spezialfälle der Qualitätsindikatoren, werden aber oft auch Patient Safety Ereignisse verstanden. Im einzelnen Krankenhaus sind diese Ereignisse nie signifikant. Es kann aber auf Bundesebene klar erkennbare Verbesserungsmöglichkeiten geben (Beispiel: Todesfälle bei Cholezystektomie).

Diese Thematik wird nur indirekt von den diskutierten Quellen angesprochen, so z. B. bei der Spezifizierung der Indikatoren oder bei der Abdeckung eines Indikatorsystems von Todesfällen und Krankenhausfällen (vgl. Mansky et al. 2011, S. A-11). Die Bedeutung der Reichweite der Qualitätsziele wird aber nach bestem Wissen nicht näher betrachtet.

Anwendungsdefinition und -beispiele: Zunächst sollte das Zielsystem einen möglichst hohen Abdeckungsgrad haben, indem die Ziele anhand eines möglichst umfassenden Indikatorensystems gesetzt werden (siehe auch Kriterium 1a). Falls ein Problem von kleinen Fallzahlen bestehen bleibt, sind verschiedene Lösungsansätze denkbar (vgl. Mansky 2010a, S. 8f.; Mansky/Seeger 2011, S. 211[1]):

1. Zusammenfassung der Daten über mehrere Jahre für eine Klinik
2. Zusammenführung der Kassendaten für eine Klinik
3. Bildung integrierter Komplikationsscores aus einem Fachbereich (siehe kombinierte Komplikationsindizes in QSR)

[1] Siehe auch erste Arbeiten zur Korrelation von Fallzahlen bzw. Krankenhausgröße und Qualitätsergebnissen bei P4P (vgl. Davidson et al. 2007).

4. Punktebewertung einzelner Qualitätsindikatoren mit anschließender Aufsummierung über mehrere Fachbereiche
5. Betrachtung von Anbietergruppen (Krankenhaus-Konzerne, Netze etc.)
6. Fokussierung auf regionale Versorgungsqualität (sektorübergreifend)
7. Fokussierung auf Überprüfung von Einzelfällen, unabhängig von der Signifikanz (vor allem bei Sentinelereignissen)

Exkurs Bepunktungssystem: Noch ein Wort zu Bepunktungssystemen (vgl. Lösungsansatz 4): Hierbei werden Einzelergebnisse bepunktet und gewichtet (unabhängig von der Signifikanz), anschließend wird die Vergütungsmodifikation (z. B. +/- 3 %) an die Gesamtpunktzahl geknüpft (vgl. Mansky 2010a, S. 9; aber auch das Beispiel der CMS-Initiative, HHS - CMS 08.05.2011). Dies kann das Fallzahlproblem lösen und effektive Anreize setzen. Allerdings entsteht gleichzeitig eine abteilungsübergreifende Gesamtverantwortung (da mehrere Fachabteilungen in das Punktesystem integriert werden und die Bewertung nicht fachabteilungs- sondern versorgungsbereichsspezifisch stattfindet). Es ergibt sich der deutliche Vorteil einer engeren Zusammenarbeit der Fachabteilungen – gerade diese Interdisziplinarität kann die Gesamtqualität der stationären Behandlung deutlich verbessern.

Informationsquellen: Neben der Definition der Qualitätsziele des Anreizsystems und einer Auflistungen der erbrachten Leistungen, Vergleiche mit bspw. bundesweiten Fallzahldaten (vgl. Spindler 2010, S. 349ff.) kann dieses Kriterium auch anhand von Interviews mit Beteiligten bewertet werden..

Kriterium 8a: Erreichbarkeit

Systemelement: Ziele

Kriterienkategorie: Wirksamkeit

Beschreibung: Die Qualitätsziele des Anreizsystems müssen vom beteiligten Personenkreis langfristig erreichbar sein.

Begründung/Bezug zu existierenden Systemen: Die qualitätsorientierten Anreizziele richtig zu setzen, ist eine der Kernherausforderungen eines P4P-Systems (siehe auch Horvath 2008). Wie im sprichwörtlichen Bildnis der „Karotte vor der Nase“ müssen Ziele gefunden werden, die zwar herausfordernd aber trotzdem erreichbar sind. In den betrieblichen Anreizsystemen wird dies als „motivationsförderlich“ beschrieben (vgl. u. a. Bullinger 1996). In der derzeitigen P4P-Diskussion wird meist indirekt über diese Fragestellung gesprochen. Unter dem Vorzeichen der „Bemessungsgrundlage“ werden vor allem die verschiedenen Vergleichsformen und Vergütungselemente betrachtet (vgl. SVR 2007). Mit der Fragestellung dieses Kriteriums ist vor allem das „Erreichen eines festgelegten Schwellenwertes“ (Scheppach et al. 2011, S. 107) beschäftigt. Hier kann nur für solche Leistungserbringer eine

Verbesserungsmotivation entstehen, die derzeit unterhalb des Schwellenwertes operieren. Sobald der Schwellenwert überschritten ist, findet kein Anreiz mehr statt (vgl. Scheppach et al. 2011, S. 107f.).

Doch auch eine „zu ambitionierte“ Zielsetzung kann die Anreizwirkung reduzieren: So ist beispielsweise eine hundertprozentige Vermeidung von Revisionsoperationen nach Hüft-Endoprothesen nicht realistisch, eine 0 %-Rate von „Never-Events“ und Ereignissen der Patientensicherheit wie eine „wrong-side-surgery“ dagegen schon. Obschon Revisionsoperationen sicherlich unerwünscht sind, so sind sie im Einzelfall nicht immer vermeidbar – ein Qualitätsziel einer 0 %-Revisionsrate kann also ggf. zu Fehlanreizen führen – bspw. der Nicht-Dokumentation eines solchen Falles oder zum „cherry picking“.

Das Kriterium der „Erreichbarkeit“ soll die Fragestellung einer „richtigen“ Schwellenwertsetzung betrachtet. Ob ein solches Qualitätsziel nun absolut oder relativ gesetzt wird, ist weniger von Interesse als die Frage, für welche Leistungserbringer diese Ziele erreichbar aber doch herausfordernd sind, so dass eine möglichst große Anreizwirkung entsteht.

Anwendungsdefinition und -beispiele: Um dieses Kriterium bewerten zu können, müssen unter anderem folgende Fragen beantwortet werden:

- Ist es im Bereich des Möglichen, für den Leistungserbringer die gesetzten Ziele zu erreichen? (Zielsetzungen nicht „zu hoch“)
- Wurden die Ziele bereits erreicht oder können „quasi automatisch“ erreicht werden? (Zielsetzungen nicht „zu niedrig“)
- Wie viele und welche Leistungserbringer erreichen bereits den festgelegten Schwellenwert?
- Wie wird die weitere Entwicklung der Leistungserbringer für dieses Ziel aus medizinischer Sicht eingeschätzt? (ggf. durchschnittliche „Verbesserung“ der Ergebnisse z. B. auf Bundesebene zu erwarten)
- Sind die notwendigen Voraussetzungen (z. B. im Rahmen von Prozess- oder Strukturqualität) vorhanden, um das Ziel zu erreichen?
- Wann kann realistischerweise das angesetzte Ziel erreicht werden?

Informationsquellen: Zunächst muss die Definition der Qualitätsziele des P4P-Systems betrachtet werden. Um anschließend die Erreichbarkeit zu bewerten, können verschiedene Informationsquellen zum Vergleich herangezogen werden:

- Ziele von vergleichbaren P4P-Programmen und deren Zielerreichung
- Statistiken bezüglich der derzeitigen Zielerreichung anderer Leistungserbringer (bspw. Bundesstatistik, AQUA-Qualitätsbericht, Nimptsch/Mansky 2012a)

- Historische Zielerreichung der beteiligten Leistungserbringer (eigene Qualitätsmessungen)
- Qualitative Interviews mit Leistungserbringer und Expertenmeinungen

Kriterium 8b: Transparenz

Systemelement: Ziele

Kriterienkategorie: Operationalisierbarkeit

Beschreibung: Bei diesem Kriterium wird bewertet, ob die Qualitätsziele des Anreizsystems für die beteiligten Personenkreise (u. a. Leistungserbringer und deren medizinisches Personal, Krankenkassen, Patienten, interessierte Öffentlichkeit) transparent dargestellt sind und kommuniziert werden.

Begründung/Bezug zu existierenden Systemen: Das Thema der Transparenz wird bei einer Vielzahl von Anforderungskatalogen von Anreizsystemen diskutiert (so u. a. bei Becker 1990; Bleicher 1992; Grewe 2000). Damit das „S-O-R"-Prinzip tatsächlich funktionieren kann und so eine ausreichende Anreizwirkung für die qualitätsbeteiligten Personenkreise entsteht, müssen gerade diese Zielgruppen involviert sein und alle Elemente des Anreizsystems verstehen und nachvollziehen können. Im Rahmen der hier vorliegenden Principal-Agent-Beziehungen und der bereits angesprochenen Informationsasymmetrie ist eine weitergehende, transparente Kommunikation auch für sonstige Zielgruppen (wie bspw. Krankenversicherungen, Patienten oder interessierte Öffentlichkeit) sinnvoll.

Bereits im vorderen Teil des Anforderungskatalogs wurden über einige Aspekte der Transparenz gesprochen, bspw. über die Verständlichkeit der Indikatoren. Insgesamt zeigt sich jedoch, dass zwar die Kommunikation der Indikatoren an die Öffentlichkeit (u. a. innerhalb der Diskussion des „Public Reportings") von Interesse ist, die Transparenz der Zielsysteme aber nur bei Veit diskutiert wird (vgl. Veit et al. 2012e, S. 53). Im Rahmen von „Gestaltungsdimensionen" der Einführung der P4P-Projekte kann Van Herck aufzeigen, dass die Kommunikation(sstrategie) und die Bekanntheit der Projekteelemente bei den Teilnehmern einen Effekt auf den Projekterfolg haben kann (vgl. van Herck et al. 2010, S. 6).

Anwendungsdefinition und -beispiele: Um eine möglichst hohe Transparenz zu erreichen, sollten u. a. folgende Aspekte Beachtung finden:

- Beschreibung der Zielsetzungen des Anreizsystems gibt detailliert die Ziele wieder und ist verständlich für das medizinische Personal als erste Zielgruppe.
- Dabei wird deutlich, wie die Ziele zustande kommen, wer diese Ziele erreichen soll (vor allem welche Adressatenebene), wann die Zielerreichung überprüft wird und die Messergebnisse vorliegen. Inwieweit auch auf individueller Ebene die jeweiligen Ziele an den gesamten Leistungserbringer und seine Mitarbeiter kommuniziert werden, sollte im Einzelfall analysiert werden.

- Die Konsequenzen der Zielerreichung (d. h. vor allem die Vergütungsfunktion) sollten für den Leistungserbringer und individuell den betroffenen Mitarbeitergruppen transparent sein. Auch hier sollte aus Datenschutzgründen vermutlich nur die Leistungserbringerebene veröffentlicht werden.
- Potenzielle Zielanpassung (an medizinische Entwicklungen, Umweltbedingungen oder bereits erreichten Zielen) wird auch transparent dargestellt.
- Für weitere Zielgruppen wird ebenfalls eine transparente Kommunikation der Qualitätsziele sichergestellt. Diese ist in der Verständlichkeit aber auch im Umfang an die jeweilige Zielgruppe angepasst.

Informationsquellen: Neben der Detailbeschreibung der Qualitätsziele (soweit vorhanden) müssen für dieses Kriterium die Kommunikationsmittel an verschiedene Zielgruppen analysiert werden. Ob diese Darstellungen ein ausreichendes Maß an Transparenz beinhaltet, kann durch den Vergleich mit anderen P4P-Projekten, vor allem aber durch eine direkte Einschätzung der Beteiligten geschehen. Hierfür sind qualitative Interviews und ggf. Umfragen sinnvoll.

Kriterium 8c: Akzeptanz

Systemelement: Ziele

Kriterienkategorie: Operationalisierbarkeit

Beschreibung: Dieses Kriterium betrachtet die Akzeptanz, die die Qualitätsziele bei den verschiedenen Zielgruppen haben. Auch hier geht es um die Operationalisierbarkeit, d. h. die Möglichkeit einer effektiven Umsetzung der Qualitätsziele bei Leistungserbringern aber auch ggf. anderen Zielgruppen.

Begründung/Bezug zu existierenden Systemen: In den Quellen der betriebswissenschaftlichen Literatur wird nur einmal direkt auf die Akzeptanz als Anforderung an Anreizsysteme verwiesen (vgl. Lehmann 2006, S. 16), gleichzeitig jedoch wird in verschiedenen anderen Kriterien das Thema der Akzeptanz als Grundlage für die Anforderung angesprochen. So wird argumentiert, dass nur bei einer ausreichenden Flexibilität und subjektiv empfunden Gerechtigkeit eine Akzeptanz bei den Beteiligten erreicht werden kann und somit eine Anreizwirkung zustande kommt (vgl. u. a. Grewe 2000; Lehmann 2006; Wälchli 1995). Das gesamte Anreizsystem muss daher von den Personen, die einen Anreiz im Sinne einer „S-O-R"-Konstruktion erhalten sollen, akzeptiert werden. Daher wurde bereits im ersten Teil des Anforderungskatalogs auf die „Akzeptanz des Qualitätsindikators" und verschiedene „wahrgenommene Relevanzen" hingewiesen. Doch auch und gerade das Qualitätsziel, auf Grundlagen dessen die variable Vergütung stattfindet, muss akzeptiert werden. Bei verschiedenen Arbeiten wird auf dieses Thema eingegangen – bspw. in verschiedenen Umfragen und Meinungsäußerungen von Verwendern der Programme, die mögliche Aspekte der Akzeptanzerreichung beinhalten (vgl. u. a. Locke/Srinivasan 2008, S. 21; Beckman 2006, S. 192; Meterko et al. 2006, S. 1959 in Veit et al. 2012e, S. 61ff.).

Anwendungsdefinition und -beispiele: Akzeptanz von Qualitätszielen kann nur dann erreicht werden, wenn dieses Systemelement subjektiv auch tatsächlich akzeptiert wird. Hierunter fällt eine Vielzahl von weiteren Charakteristika, die erfüllt werden sollten, so z. B.:

- Individuelles Qualitätsziel wird als gerecht (unter anderem anforderungs- und qualifikationsgerecht (vgl. Bullinger 1996)), leistungsorientiert, erreichbar und somit motivierend empfunden. Gerade die persönliche Einschätzung der betroffenen Mitarbeiter sind relevant – auch im Vergleich zu anderen Leistungserbringern oder einzelnen Fachabteilungen.
- Weitere Zielgruppen können vor allem an der angestrebten Qualitätsverbesserung und „Höhe" des Qualitätsziels interessiert sein und diese akzeptieren (bspw. kann die Öffentlichkeit das Qualitätsziel einer „Null-Fehler-Rate" bei „Never-Events" akzeptieren, das Qualitätsziel einer Reduktion der Revisionsrate bei Hüft-TEPs auf 0 % aber als zu niedrig ablehnen).
- Insgesamt kann eine solche Akzeptanz vor allem dann erreicht werden, wenn die beteiligten Personenkreise in den Entwicklungsprozess des Anreizsystems und der Qualitätsziele im Speziellen integriert werden. Auch die Nutzung von etablierten, evaluierten Indikatoren- und Zielsystemen kann hilfreich sein (vgl. Endsley et al. 2004, S. 49f.).

Informationsquellen: Zunächst muss als Informationsquelle wiederum die Beschreibung der Ziele dienen. Die Einbeziehung verschiedenen Zielgruppen sollte außerdem aus den Details der Entwicklungsprozesse hervorgehen. Sofern eine solche Integration stattgefunden hat, können die ersten Einschätzungen der Beteiligten bspw. aus Gesprächsprotokollen während der Entwicklung gezogen werden. Ansonsten sollten Umfragen und qualitative Interviews für eine Bewertung genutzt werden.

Kriterium 8d: Langfristige Anreizwirkung der Ziele

Systemelement: Ziele

Kriterienkategorie: Operationalisierbarkeit

Beschreibung: Dieses Kriterium beschreibt inwieweit die Zielsetzungen auch langfristig und strategisch zu Anreizwirkung führen können.

Begründung/Bezug zu existierenden Systemen: In der betriebswissenschaftlichen Literatur wird zwar verschiedentlich das Thema der Motivationswirkung angesprochen (so u. a. bei Bullinger 1996; Becker 1990), diese muss auch langfristig begründet sein. Dazu ist unter anderem eine langfristige Anpassbarkeit der Ziele notwendig, um diese an Unternehmensumfeld oder Arbeitsbedingungen angleichen zu können (vgl. z. B. die Anforderung „Flexibilität" bei Becker 1990; Grewe 2000). Gleichzeitig müssen die Ziele auch an bspw. medizinische Veränderungen angepasst werden können.

Bei den Gestaltungsdimensionen von P4P-Systemen ist ein Aspekt mit zeitlicher Dimension zu finden, nämlich die sinnvolle, zeitliche Dauer der Anreize. Hierunter wird bspw. verstanden, dass erst eine „angemessene Projektlaufzeit“ zu einer langfristigen Qualitätsverbesserung ohne „Rückfall“ führen kann (vgl. Scheppach et al. 2011, S. 112f.). Die notwendigen Anpassungen innerhalb dieses Zeitraumes werden nicht erwähnt. Van Herck empfiehlt allerdings eine laufende „Re-Fokussierung“ des P4P-Systems auf veränderte Bedingungen, vor allem auf die erfolgte Zielerreichung (vgl. van Herck et al. 2010, S. 9; siehe auch Amelung/Zahn 2009, S. 17). Vor dem Hintergrund immer längerer Projektbeispiele und P4P-Programme ohne feste Laufzeiten ist eine Thematisierung einer langfristig etablierten Anreizwirkung relevant.

Anwendungsdefinition und -beispiele: Für die Bewertung dieses Kriteriums sollten folgende Fragen beantwortet werden können:

- Werden die Qualitätsziele des Vergütungssystems regelmäßig auf Anreizwirkung überprüft?
- Werden Qualitätsziele auf veränderte Bedingungen (z. B. medizinischen Fortschritt) angepasst, um auch langfristig die Qualität zu verbessern?
- Gibt es feststehende Mechanismen, um auch bei Leistungserbringern, die die Qualitätsziele erreichen, weiterhin Anreize zu schaffen? (Z. B. durch Nutzung von relativen Zielen im Vergleich zu anderen Leistungserbringern)
- Wird die durchschnittliche Qualitätsverbesserung (bspw. die kontinuierliche Reduktion der Sterblichkeit nach Herzinfarkt, Wegfall einen Versorgungsdefizits) in die Qualitätszielsetzung integriert?

Informationsquellen: Als Quelle für die Bewertung können vor allem Analysen bezüglich einer veränderten oder angepassten Anreizwirkung dienen. Auch detaillierte Darstellungen einer langfristigen Zielentwicklung sind relevant.

Kriterium 9a: Gerechtigkeit

Systemelement: Ziele

Kriterienkategorie: Vergleichbarkeit

Beschreibung: Das letzte Kriterium für die Qualitätsziele ist die Gerechtigkeit – d. h. sowohl die Frage, ob die Zielsetzungen gerecht sind oder angepasst werden können als auch ob sie von den beteiligten Personenkreisen als gerecht empfunden werden.

Begründung/Bezug zu existierenden Systemen: Die Anforderung der „Gerechtigkeit“ wird in mehreren Aufzählungen der betriebswissenschaftlichen Literatur behandelt, so unter anderem bei Grewe und

Lehmann (vgl. Grewe 2000; Lehmann 2006; Wälchli 1995; Wünsch 2012, S. 29). In diesen Arbeiten wird vor allem auf die Unterscheidung zwischen interner und externer Gerechtigkeit eingegangen – d. h. die Vergleiche der „Leistungs-Belohnungs-Korrelation“ innerhalb und außerhalb der Organisation. Auf das Systemelement der Qualitätsziele wird nicht gesondert eingegangen. Nur wenn die Qualitätsziele sowohl von der Organisation als auch ihren Führungskräften und Mitarbeitern als „gerecht“ empfunden werden, kann wiederum eine Akzeptanz entstehen, die zu einer Anreizwirkung führen kann. Somit kann das Kriterium der „Gerechtigkeit“ auch als Vorstufe der Akzeptanz verstanden werden.

Diese meist recht subjektiv zu bewertende Anforderung taucht in den Analysen zu Gestaltungsdimensionen von P4P-Systemen bislang nach bestem Wissen nicht auf. Generell gibt es subjektive Elemente nur bei einigen Anforderungen der Qualitätsindikatoren (vgl. hierfür auch den vorderen Teil dieses Anforderungskatalogs, u. a. „Wahrgenommene Relevanz“).

Anwendungsdefinition und -beispiele: Die Bewertung dieses Kriteriums ist aufgrund des subjektiven Charakters schwierig. Trotzdem soll versucht werden, möglichst objektive Aspekte zur Bewertung zu nutzen.

Qualitätsziele werden dann als „gerecht“ empfunden, wenn…

- …zur Zielerreichung generell die gleichen „Anstrengungen“ für unterschiedliche Krankenhäuser und Fachabteilungen notwendig sind - vor allem Größe, Trägerschaft und Region sollten keine Vorbelastungen darstellen (siehe auch „Leistungsgerechtigkeit“ nach Becker 1990).
- …besondere Gegebenheiten von Leistungserbringern und Fachabteilungen berücksichtigt werden, so dass verschiedene Akteure "gerechte" Möglichkeiten haben, ihr Ziel zu erreichen (u. a. durch Risikoadjustierungen).
- …Qualitätsziele vor allem für unterschiedliche Mitarbeitergruppe relativ zur Verantwortung der Leistung gesetzt werden (siehe auch „Anforderungsgerechtigkeit“ nach Becker 1990).

Informationsquellen: Für die Bewertung dieses Kriteriums können zunächst die Beschreibung und der Entwicklungsprozess der Qualitätsziele herangezogen werden, um die Anpassung der Ziele an verschiedene Gegebenheiten zu beurteilen. Zusätzlich sind qualitative Interviews mit den verschiedenen Personenkreisen sinnvoll.

Kriterium 10a: Effizienz

Systemelement: Vergütungsfunktion

Kriterienkategorie: Wirksamkeit

Beschreibung: Die nächsten Kriterien beziehen sich wiederum auf das Kriterienkategorie der „Wirksamkeit“, allerdings jetzt in Bezug auf das Systemelement der Vergütungsfunktion. Für eine hohe Effizienz muss der Nutzen der Vergütungsfunktion und somit der Erfolge des Anreizsystems im Verhältnis zum Aufwand für die beteiligten Personen und Organisationen stehen.

Begründung/Bezug zu existierenden Systemen: Wirtschaftlichkeit und Effizienz sind für die betriebswissenschaftliche Literatur höchst relevanten Themen. Dies ergibt sich unter anderem aus der Gewinnorientierung einer Vielzahl von Anreizsystemen (vgl. aber auch u. a. Becker 1990, S. 24; Bleicher 1992, S. 20; Wünsch 2012, S. 29), doch auch für qualitätsorientierte Anreizsysteme wird das Kriterium der „Wirtschaftlichkeit“ gestellt (vgl. Ortlieb 1993). Hier wird stets das Gesamtsystem unter der Prämisse und Anforderung der (meist Kosten-) Effizienz betrachtet.

Im Rahmen von P4P-Systemen wird derzeit vor allem über die Anreizhöhe diskutiert. Verschiedene Studien weisen darauf hin, dass die zusätzliche Vergütung für den Leistungserbringer zumindest die Kosten sowie Opportunitätskosten der Maßnahmen zur Qualitätsverbesserung sowie der Investitionen abdecken sollten (vgl. SVR 2007, S. 173). Über die Kosten-Nutzen-Verhältnisse und die Kosteneffektivität von P4P wird besonders im Ausland gesprochen – hierin wird deutlich, dass bei Qualitätsverbesserungen auch Kostensenkungen zu erwarten sind (vgl. u. a. Veit et al. 2012c, S. 39f.; Nahra et al. 2006, S. 49ff.; Golden 2006). Bei Amelung werden auch Effizienzindikatoren als Zielgrößen diskutiert (vgl. Amelung/Zahn 2009, S. 37). Eine Arbeit betrachtet einen „Gesamtkostenansatz“ für die Bewertung von P4P-Projekten (vgl. Dolan/Yanagihara 2011).

Anwendungsdefinition und -beispiele: Für die Bewertung des Kriteriums ist zunächst die Definition sowohl des „Nutzens“ als auch der „Kosten“ entscheidend:

- Als „Kosten“ stehen hier zum einen die *laufenden Implementierungskosten* an, so bspw.:
  - Kosten der Qualitätsverbesserungsmaßnahmen bei Leistungserbringern
  - Ggf. Opportunitätskosten auf Grundlage von „entgangenen Leistungen“ (z. B. durch Reduktion von „bezahlten“ Revisionsoperationen) beim Leistungserbringer
  - Implementierungskosten von Qualitätsmesssystemen beim Kostenträger
  - Zusätzlich können die Aufwendungen, die durch die Zahlung der monetären Anreize entstehen, beim Kostenträger als Kosten verstanden werden (die natürlich beim Leistungserbringer wiederum als Zusatzerlöse erscheinen)
- Zusätzlich sind einmalige *Investitionskosten* bei der Erstimplementierung solcher Projekte zu erwarten, z. B.:
  - Personalkosten für die Programmierung von EDV-Systemen zur Qualitätsmessung (z. B. beim Kostenträger)

  - Projektkosten zur Entwicklung des Anreizsystems (bspw. bei allen Beteiligten oder sogar übergeordneten Stellen)

- Als besonders schwierig stellt sich die Definition des Nutzens dar. Generell soll ein qualitätsorientiertes Anreizsystem den Zweck und Nutzen von Qualitätsverbesserungen erfüllen. Doch für den Vergleich mit den Kosten ist eine monetäre Nutzenbewertung notwendig. Wie bereits im vorherigen Kapitel diskutiert ist dies nur in einigen Qualitätsbereichen möglich. Für die unterschiedlichen Systemebenen bedeutet dies:

  - Für das Gesamtsystem ist die monetäre Nutzenbewertung der Qualitätsverbesserung aus verschiedenen Gründen kaum zielführend: Zum einen können nicht alle Qualitätsverbesserungen sinnvoll mit monetären Werten belegt werden. Zum anderen kann es gesellschaftlich erwünscht sein, dass Verbesserungen von Behandlungsqualität (nur) durch Mehrkosten erreicht werden.

  - Für die beteiligten Organisationen dagegen kann ein direkter Zusatznutzen dargestellt werden, so z. B. die Zahlung der Anreize für den Leistungserbringer oder die reduzierte Erstattung von Leistungen bspw. für Revisionsoperationen oder Komplikationen durch den Kostenträger nach Qualitätsverbesserungen.

Außerdem muss definiert werden, welche Aspekte der Effizienz für die verschiedenen Systemebenen relevant sind:

- *Für den Leistungserbringer*: Kosten/Aufwand, der für die Qualitätsverbesserung eingesetzt, wird ist geringer als die zusätzliche Vergütung (zumindest langfristig)

- *Für Kassen*: Mehrkosten des Systems (sowohl für Bonus als auch Implementierungs-/Wartungskosten) sind geringer als Ersparnis durch Qualitätsverbesserungen

- *Für das System*: Gesamtkosten für das System (generiert bei Krankenhäusern, Kassen und Politik) sind geringer als monetärer und gesellschaftlicher Mehrwert der Qualitätsverbesserung

Inwieweit jedoch müssen die Kosten eines qualitätsorientierten Vergütungssystems notwendig geringer sein als der monetär bewertete Nutzen? Langfristig muss diese Frage vor dem Hintergrund einer gesellschaftlichen Bewertung stattfinden und führt hier zu weit. Auch Aspekte eines „Pareto-Optimums“ – also ob das Anreizsystem die „effizienteste“ Alternative für eine Qualitätsverbesserung“ darstellt – sind hierbei zu beachten.

Für die Leistungserbringer und Kostenträger, die auch im wirtschaftlichen Wettbewerb bestehen müssen, stellt sich jedoch die Effizienzfrage sowohl langfristig als auch kurzfristig sehr konkret. Nur wenn eine ausreichende Wirtschaftlichkeit für alle Seiten sichergestellt wird, kann eine Anreizwirkung und Akzeptanz für das Anreizsystem erreicht werden. Erste Modellrechnungen eines solchen Kosten-Nutzen-

Verhältnisses wurden bei Mansky/Seeger durchgeführt und Implikationen daraus gezogen (vgl. Mansky/Seeger 2011).

Informationsquellen: Für die Bewertung dieses Kriteriums sind aufwendige Analysen zu Kosten- und Nutzenpotenzialen notwendig. Erste Ansätze dazu können zu Vergleichszwecken herangezogen werden (vgl. u. a. Nahra et al. 2006, S. 49ff.)

Kriterium 10b: Effektivität

Systemelement: Vergütungsfunktion

Kriterienkategorie: Wirksamkeit

Beschreibung: Neben einer hohen Effizienz muss die Vergütungsfunktion effektiv sein, also zielgerichtet zu den erhofften Veränderungen und zu Qualitätsverbesserungen führen.

Begründung/Bezug zu existierenden Systemen: Die Anforderung der Zielgerichtetheit bzw. Effektivität an das Anreizsystem oder im speziellen an die Vergütungsfunktion ist nur in einigen der hier diskutierten Arbeiten zu finden. Nur bei Becker und Bullinger wird die „Motivationswirkung“ als ähnliches Konstrukt erwähnt (vgl. Becker 1990; Bullinger 1996). Gleichzeitig jedoch kann davon ausgegangen werden, dass dies als „Hauptzweck“ des Anreizsystems angesehen wird und daher der Erfolg des Systems bewertet wird.

Bei den Qualitätsindikatoren wurde darauf verwiesen, dass überprüft werden muss, ob und inwieweit die Indikatoren auch tatsächlich zielgerichtet die erwarteten Ergebnisse messen. Bei der derzeitigen Bewertung von P4P-Systemen wird ebenfalls auf die Frage der Effektivität das Hauptaugenmerk gelegt – so wird in den Reviews analysiert, welche Auswirkung die P4P-Systeme auf die Versorgungsqualität haben und welche Qualitätsverbesserungen entstehen (vgl. z. B. SVR 2007, S. 175f.; siehe auch Abschnitt 5.1.2.). Auch hier zeigt sich, dass diese Anforderung zwar nicht explizit als Kriterium genannt wird, aber eindeutig einen sehr hohen Stellenwert hat.

Anwendungsdefinition und -beispiele: Da die Effektivität und damit die Wirksamkeit des Anreizsystems quasi als übergeordnetes Zielkriterium angesehen werden kann, ist die Definition der Bewertung schwierig. Viele verschiedene Aspekte sind zu beachten, so z. B.:

- Verbesserungen von Qualitätsergebnissen finden tatsächlich statt und korrelieren mit der Durchführung des Systems.
- Die Vergütungsfunktion ist so konzipiert, dass eine Verbesserung der Behandlungsqualität auch zu höherer Vergütung und somit Anreizen beim Leistungserbringer führt, aktiv Qualitätsverbesserungsmaßnahmen durchzuführen und zu entwickeln.

- Die Vergütungsfunktion wird nicht durch andere Effekte „überschattet“ (so müssen z. B. „entgangene Vergütung“ bei Komplikationsreduktion durch Vergütung der Qualitätsverbesserungen gedeckt werden, um nicht einen höheren Anreiz zu haben, keine Verbesserung durchzuführen).
- Mögliche Fehlanreize werden analysiert und weitestgehend vermieden.

Informationsquellen: Dieses übergeordnete Kriterium muss anhand von detaillierten Analysen bewertet werden. Erste Studien zu Reviews von P4P-Systemen stehen hierfür bereits zur Verfügung. Insgesamt sollten vor allem longitudinale Studien mit hoher wissenschaftlicher Qualität herangezogen werden. Ist dies nicht möglich, muss zumindest anhand von ersten Vergleichsdaten eine Aussage getroffen werden.

Kriterium 10c: Flexibilität

Systemelement: Vergütungsfunktion

Kriterienkategorie: Wirksamkeit

Beschreibung: Die Vergütungsfunktion muss so gestaltet werden, dass sie bei Bedarf flexibel an veränderte Situationen angepasst werden kann.

Begründung/Bezug zu existierenden Systemen: Flexibilität wird in einer Vielzahl von betriebswissenschaftlichen Anforderungskatalogen für Anreizsysteme diskutiert (so in Becker 1990; Bullinger 1996; Bleicher 1992; Grewe 2000; Lehmann 2006; Wünsch 2012, S. 30f.). Damit wird vor allem die Anpassung an über die Zeit stattfindende Veränderungen verstanden, aber auch die Flexibilität auf unterschiedliche Gegebenheiten der verschiedenen Leistungserbringer eingehen zu können. Diese Anpassung kann dann zu Akzeptanz, Motivation und Anreizwirkung führen.

Bei der derzeitigen P4P-Thematik wird Flexibilität im Rahmen von Gestaltungsdimensionen diskutiert. Bei Scheppach werden die zu berücksichtigen Umweltbedingungen diskutiert, zu denen bspw. demographische, organisatorische Unterschiede beim Leistungserbringer oder Verschiedenheiten im Gesundheitssystem gehören (vgl. Scheppach et al. 2011, S. 135). Auch van Herck empfiehlt eine laufende „Re-Fokussierung“ des P4P-Systems auf veränderte Bedingungen, vor allem auf die erfolgte Zielerreichung (vgl. van Herck et al. 2010, S. 9; siehe auch Amelung/Zahn 2009, S. 17). Gleichzeitig wird hier darauf hingewiesen, dass die P4P-Systeme für die verschiedenen Kostenträger einheitlich gestaltet werden müssen, um nicht „aufgeweicht“ zu werden (vgl. van Herck et al. 2010, S. 8).

Im Gegensatz zum vorherigen Kriterium der Qualitätsziele eine „langfristige Anreizwirkung“ zu erlangen soll hier spezifisch auf die Vergütungsfunktion eingegangen werden. Nur wenn neben einer kontinuierlichen Anpassung des Zielsystems auch die monetären Anreize flexibel sind, können Motivation und Akzeptanz erfolgen.

Anwendungsdefinition und -beispiele: Um dieses Kriterium bewerten zu können, sollten folgende Aspekte betrachtet werden:

- Der „Preis“ der Qualität wird flexibel (d. h. beispielsweise jährlich) an veränderte Bedingungen angepasst (bspw. an Inflation, technische Entwicklungen, gesetzliche Änderungen).
- Die Vergütungsfunktion ist so an die verschiedenen Ausgangsbedingungen der Leistungserbringer oder Kostenträger angepasst, dass Anreizwirkung sowie Effizienz und Effektivität auch für verschiedene Adressaten erreicht werden können. Bspw. dürfen kleinere Leistungserbringer keine höheren oder niedrigeren Anreize erhalten (aufgrund von absoluten Zahlungen).
- Flexibilität kann bspw. entstehen durch eine relative Vergütungsfunktion (d. h. anteilig am Gesamtumsatz des Leistungserbringers) oder die Anlehnung an die „Preisfunktion“ der stationären Vergütung, nämlich dem Basisfallwert[1].

Insgesamt kann festgehalten werden, dass wie bereits bei den Qualitätszielen eine gute Balance zwischen „einheitlicher“ und „gerechter“ Behandlung aller Leistungserbringer und Kostenträger gefunden werden muss. Während bei der Zielsetzung hierbei die Erreichbarkeit im Vordergrund steht, muss die Vergütungsfunktion flexibel auf die wirtschaftlichen Gegebenheiten reagieren können, ohne dabei alle Leistungserbringer „gleichzuschalten“. Eine relative Vergütungsfunktion bzw. die Nutzung eines qualitätsorientierten Basisfallwertes scheint hierfür eine gangbare Lösung.

Informationsquellen: Zur Bewertung dieses Kriteriums dient vor allem die Beschreibung der Vergütungsfunktion, deren Berechnung sowie die geplanten Veränderungen und Anpassungen. Außerdem sollte analysiert werden, welche variable Vergütung aufgrund der Vergütungsfunktion tatsächlich beim Leistungserbringer ankommt.

Kriterium 11a: Anteil der Vergütung

Systemelement: Vergütungsfunktion

Kriterienkategorie: *Operationalisierbarkeit*

Beschreibung: Für die Vergütungsfunktion des Anreizsystems sind weiterhin verschiedene Unteraspekte der Operationalisierbarkeit entscheidend, nur wenn auch die Vergütungsfunktion im operativen Alltag Anwendung findet, können die Systeme erfolgreich sein. Hier soll im speziellen bewertet werden, ob der Vergütungsanteil eine angemessene Höhe erreicht.

[1] Nähere Erläuterung im übernächsten Kapitel

Begründung/Bezug zu existierenden Systemen: Wie bereits diskutiert ist die Anreizhöhe und damit der Vergütungsanteil (als relative Anreizhöhe zur Gesamtvergütung bspw. eines Leistungserbringers) ein bei P4P-Systemen vielfach diskutierter Aspekt (vgl. u. a. SVR 2007; Schrappe/Gültekin 2010; Scheppach et al. 2011; Rosenthal/Frank 2006; Christianson et al. 2008)). Obschon bislang keine klare Korrelation von Anreizhöhe und Wirksamkeit des Systems gefunden wurde, müssen zumindest gewisse Schwellenwerte erreicht werden. Einige Arbeiten sprechen von einem wirksamen Vergütungsanteil bei 5 % - für größere Leistungserbringer 3 %, bei kleineren 10 % (vgl. SVR 2007, S. 173; Veit et al. 2012e, S. 32). Gleichzeitig sollten mindestens die Kosten der verbesserten Behandlung und der Systemimplementierung amortisierbar sein (vgl. Christianson et al. 2008, S. 12). In einer Überblicksarbeit über 27 „Early Adopter" von P4P in den USA wurden durchschnittlich 2,3% der Gesamtvergütung als variable, qualitätsorientierte Vergütung genutzt (vgl. Rosenthal et al. 2008, S. 1678). Auch erste Vorarbeiten und Modellrechnungen zur hier vorliegenden Arbeit thematisieren diese Anforderung (Mansky/Seeger 2011; Seeger 2011).

In der betriebswissenschaftlichen Literatur wird nur indirekt über die Höhe der variablen Vergütungsaspekte als Anforderung diskutiert, da es sich hierbei um einen Unteraspekt der Anreizwirkung und Wirtschaftlichkeit handelt (vgl. z. B. Bullinger 1996; Becker 1990 – bei Becker noch zusätzlich in der Anforderung „belohnend" wieder zu finden).

Anwendungsdefinition und -beispiele: Ein konkreter Schwellenwert für die Anreizhöhe kann hier nicht festgelegt werden, da dieser von verschiedenen Faktoren des Anreizsystems abhängt. Einige Vergleichs- und Annäherungsaspekte können aber bereits diskutiert werden:

- Die Vergütungshöhe muss auf den einzelnen Systemebenen einen Anteil der Gesamtvergütung erlangen, der als „motivierend" angesehen wird, bspw.:
  - Höhe der qualitätsorientierten Anreize ist vergleichbar mit anderen (bspw. wirtschaftlichen) Anreizen.
  - Höhe der Anreize wird von den beteiligten Personen im angemessenen Verhältnis zur (gerade auch individuellen) Anstrengung für die Zielerreichung gesehen.
  - Qualitätsorientierte Anreize sind (zumindest durchschnittlich und langfristig) höher als Vergütungsteile, die eher „zufällig" entstehen und nicht auf Grund einer individuellen oder Organisationsanstrengung – bspw. „wind-fall-profits" durch Gesetzesänderungen oder Tarifsteigerungen.
- Anreizhöhe ist angepasst an die Adressatenebene (d. h. Organisation vs. Führungskraft vs. Mitarbeiter).
  - Die Vergütungsfunktionen werden an die verschiedenen Gegebenheiten der Adressaten angepasst, so dass der angestrebte Anteil vom Gesamtumsatz erreicht wird (bspw. ist der

Gesamtanteil der qualitätsorientierten Vergütung sowohl für Fachkliniken als auch für Maximalversorger relevant – siehe auch Modellrechnung in Abbildung 16)

Abbildung 16: Umsatzanteile eines Modell-IV-Vertrages mit Gewährleistungsklausel

MODELLRECHNUNG

	1) Maximalversorger[1]	2) Fachklinik[2]
Fälle gesamt	40.000	4.000
Hüft-TEPs	100	400
Fallzahlanteil	0,25%	10%
Umsatzanteil	0,4%	10%
RevisionsOPs[3]	4,7	18,8
Fallzahlanteil	0,01%	0,47%
Umsatzanteil	**0,02%**	**0,53%**

1 Basisfallwert 2.963,82 Euro, CMI für Fälle gesamt 1,35, Hüft-Erst-OP 6.672 Euro, Revisions-Ops 7.611 Euro,
2 Relativgewicht von Fällen gesamt und Hüft-TEPs identisch 3 5-Jahres-Überlebensrate der Hüft-TEPs bei 95,3% → Revisionsanteil daher bei 4,7%

Quelle: Berechnungen aus Mansky/Seeger 2011 und Seeger 2011

- Beim Vergleich des Anteils der variablen Vergütung werden die Grenzkosten, die für die Zielerreichung notwendig sind, einbezogen.
- Verschiedene Stufen der Verteilungsfunktion (d. h. bspw. für maximale oder nur durchschnittliche Zielerreichung) werden unterschieden.

Obschon ein gewisser Mindestanteil erreicht werden sollte, damit eine Anreizwirkung entsteht, muss auch hier selbstverständlich ein angemessenes Maß gefunden werden. Durch einen zu hohen variablen Vergütungsanteil kann, sofern die Zielerreichung mit großen Anstrengungen verbunden ist, eine hohe Unsicherheit für die beteiligten Kreise entstehen und so die Anreizwirkung nivelliert werden.

Informationsquellen: Um dieses Kriterium zu bewerten, müssen die Vergütungsfunktion und die dadurch erreichten Vergütungsanteile analysiert werden. Hierfür muss die tatsächliche Ausschüttung der variablen Vergütung dokumentiert und ins Verhältnis zum sonstigen Umsatz bzw. individuellen Einkommen gesetzt werden. Verschiedene Reviews stehen hierfür bereits zur Verfügung, und auch die Ergebnisse dieser Arbeit können angewendet werden.

Kriterium 11b: Planbarkeit

Systemelement: Vergütungsfunktion

Kriterienkategorie: Operationalisierbarkeit

Beschreibung: Das nächste Kriterium der Operationalisierbarkeit der Vergütungsfunktion ist die Planbarkeit. Darunter ist zu verstehen, ob und inwieweit der Leistungserbringer und die Kostenträger mit den jeweiligen monetären Auswirkungen des qualitätsorientierten Anreizsystems planen können.

Begründung/Bezug zu existierenden Systemen: In der betriebswissenschaftlichen Literatur wird unter anderem von den Anforderungen der „Steuerbarkeit“ und „Organisierbarkeit“ gesprochen (vgl. Ortlieb 1993, S. 64f.). Um steuerbar zu sein, muss ein qualitätsorientiertes Anreizsystem in das etablierte Management integriert werden können. Dies betrifft vor allem die Messergebnisse, doch auch die Vergütungsfunktion muss in den betrieblichen Alltag implementiert werden. Speziell die Frage der „Planbarkeit“ ist relevant, da Managementtätigkeiten auch immer auf Entscheidungen basieren, die für die Zukunft relevant sind. Je „unplanbarer“ und „unsicherer“ die Zukunft ist, desto schwieriger sind Entscheidungen – auch qualitätsorientierte Managemententscheidungen.

Nach bestem Wissen wird dieses Kriterium bislang nicht bei Anforderungen und Gestaltungsdimensionen von P4P-Systemen diskutiert. Im Review von Scheppach wird zwar auf eine möglichst zeitnahe Auszahlung der qualitätsorientierten Vergütung eingegangen (vgl. Scheppach et al. 2011, S. 113) und Amelung erwähnt für die Anforderungen an Indikatoren die „Vorhersagbarkeit“ (vgl. Amelung/Zahn 2009, S. 40) – eine angemessen planbare Vergütung, um die Zukunftsunsicherheit zu reduzieren wird allerdings nicht betrachtet.

Anwendungsdefinition und -beispiele: Um dieses Kriterium bewerten zu können, sollten folgende Aspekte Beachtung finden:

- Vergütungsfunktion und deren Berechnung sind so transparent gestaltet, dass Organisation und Individuen (sowohl Leistungserbringer als auch Kasse) selbstständig die zukünftigen Vergütungsanteile berechnen und somit planen können (bspw. keine komplexen Berechnungen, die nur wenigen Experten bekannt sind).
- Vergütungsfunktion ist angemessen stabil und kontinuierlich, so dass mit einem gewissen „Mindestanteil“ und Trend für die nächste Planungsperiode (z. B. nächste 2-5 Jahre) gerechnet werden kann.
- Vergütungsfunktion ist „vorausschauend“, so dass der Leistungserbringer bei Einschätzung der eigenen Messergebnisse und kontinuierlichen Qualitätsverbesserungen auch langfristige Planbarkeit erlangt.
- Qualitätsorientierte Vergütungen werden somit nicht zu unplanbaren „Wind-Fall-Profits“, sondern können aktiv in die Wirtschaftsplanung und die individuellen Zielvereinbarungen der Führungskräfte und Mitarbeiter mit aufgenommen werden. Das zusätzliche Einkommen beim Leistungserbringer kann also auf verschiedenen Ebenen „ausgeschüttet“ werden.

Auf den ersten Blick besteht hier ein Gegensatz zwischen den Anforderungen der „Flexibilität“ und der „Planbarkeit“ – und auch das noch zu besprechende Kriterium der „Aktualität“ scheint zunächst der Forderung der Planbarkeit zu widersprechen. Tatsächlich müssen diese Kriterien geschickt in Einklang gebracht werden. So können bspw. flexible, aber planbare Vergütungsfunktion entwickelt werden, wenn Flexibilität durch relative Funktionen des Basisfallwertes entstehen, die für den Leistungserbringer vorhersehbar sind, sofern gewisse Qualitätsziele erreicht werden. Auch die Mechanismen für eine Anpassung an systembedingte Veränderungen können vorher festgelegt und transparent kommuniziert werden (z. B. klare Vorgaben für die Anpassung an Lohnsteigerungen).

Informationsquellen: Auch für dieses Kriterium sind die detaillierten Beschreibungen der Vergütungsfunktion und deren tatsächlichen berechneten Auswirkungen auf die Vergütung der Leistungserbringer entscheidend. Zusätzlich kann die Planbarkeit auch daran festgemacht werden, ob und wie dieser Vergütungsaspekt in den Wirtschaftsplänen der Leistungserbringer und Krankenkassen integriert wird bzw. warum dies nicht der Fall ist.

Kriterium 12a: Nutzung zur Steuerung

Systemelement: Messergebnis

Kriterienkategorie: Operationalisierbarkeit

Beschreibung: Die nächsten und damit letzten Kriterien des Kataloges beschreiben Anforderungen, die an das Messergebnis des qualitätsorientierten Vergütungssystems gestellt werden. Im Gegensatz zum Qualitätsziel oder der Qualitätsmessung geht es hierbei um die tatsächlich gemessenen Qualitätsergebnisse des Leistungserbringers und wie diese gestaltet und dargelegt werden sollten. Unter anderem müssen die Messergebnisse ebenso operationalisierbar sein, also in den beteiligten Organisationen umgesetzt werden und im Management nutzbar sein. Hierunter ist eine Nutzbarkeit in der Organisationssteuerung, bspw. durch Feedback-Anwendung zu verstehen. Insgesamt kann diese Anforderung als ein übergreifendes Kriterium verstanden werden, da verschiedene Aspekte der folgenden Kriterien relevant sind – so z. B. Aktualität oder die bereits angesprochene Planbarkeit der Vergütungsfunktion.

Begründung/Bezug zu existierenden Systemen: Wie bereits im vorherigen Kriterium erwähnt, wird die „Steuerbarkeit“ in den Anforderungen an qualitätsorientierte Anreizsysteme von Ortlieb erwähnt (vgl. Ortlieb 1993, S. 64f.). Hierunter kann z. B. die mögliche Einbindung der Messergebnisse in Managementsysteme verstanden werden.

Zwar werden diese Aspekte in der derzeitigen P4P-Diskussion nur selten explizit erwähnt, jedoch machen andere Diskussionen die hohe Relevanz dieser Anforderung deutlich, so bspw. die verbreiteten Arbeiten über Qualitätsmanagementsysteme im stationären Sektor (vgl. z. B. Haeske-Seeberg 2001; Busse et al.

2009; Mansky/Nimptsch 2010a). Gerade einige Weiterentwicklungen von Qualitätsmanagement- und -messsystemen haben gezeigt, wie relevant die Einbindung in bspw. einen PDCA-Zyklus sein kann. Als Anforderung an ein P4P-System wurde dies bislang allerdings nicht formuliert, so dass dieses Kriterium als recht neuartig betrachtet werden kann.

Anwendungsdefinition und -beispiele: Die Bewertung dieses Kriteriums kann bspw. entlang folgender Aspekte durchgeführt werden:

- Die Messergebnisse sind zur Steuerung der Organisation und der Führungskräfte bzw. Mitarbeiter nutzbar und werden bereits in etablierten Managementsystemen eingesetzt.
- Nutzbarkeit in der unternehmerischen Steuerung setzt sich unter anderem zusammen aus (vgl. hierzu die jeweiligen Kriterien):
  - Planbarkeit der Vergütungsfunktion
  - Aktualität der Messergebnisse
  - Kontrollierbarkeit
  - Verursachungsgerechte Zuordnung der Messergebnisse
  - Beeinflussbarkeit der Messergebnisse vom Leistungserbringer
- Außerdem wird die Einbindung in Managementsysteme vom qualitätsorientierten Anreizsystem gefördert und gefordert (bspw. durch detaillierte Handbücher, die die Umsetzung in Steuerungssystemen, Zielvereinbarungen, PDCA-Zyklen oder ähnlichem beschreiben)
- Messergebnisse werden für Feedback und/oder Zielvereinbarungsgespräche mit Führungskräften und Mitarbeitern genutzt.

Informationsquellen: Zur Beurteilung dieses Kriteriums müssen zunächst vor allem die Detailbeschreibungen und ggf. Umsetzungshandbücher und -empfehlungen des qualitätsorientierten Anreizsystems analysiert werden, um zu prüfen, ob die Messergebnisse für die Steuerung nutzbar sind. Ob diese dann auch tatsächlich genutzt werden, ist nur anhand von Untersuchungen im Rahmen von Fallstudien der umgesetzten Anreizsysteme analysierbar. Daher sollten im regelmäßigen Abstand die Umsetzung bspw. durch Interviews und Begehungen untersucht werden.

Kriterium 12b: Aktualität

Systemelement: Messergebnis

Kriterienkategorie: *Operationalisierbarkeit*

Beschreibung: Eines der konkreten Kriterien der Operationalisierbarkeit ist in der Aktualität und zeitnahen Überlieferung der Messergebnisse zu sehen. Diese können dann zur Steuerung der Organisation genutzt werden.

Begründung/Bezug zu existierenden Systemen: Die Anforderung aus der betriebswissenschaftlichen Literatur an die Steuerbarkeit und Organisierbarkeit wurde bereits beim vorherigen Kriterium erwähnt (vgl. Ortlieb 1993; siehe auch Horvath 2008). Hieraus lassen sich auch Aussagen zur Aktualität ableiten.

Noch direkter verweist allerdings die Diskussion zu P4P-Systemen auf dieses Kriterium, nämlich bei Scheppach bzw. Petersen (vgl. Scheppach et al. 2011, S. 113; Petersen et al. 2006). Hier wird festgestellt, dass eine Vergütung, die noch vor Jahresende ausgeschüttet wird, aufgrund von zwischenzeitlichem Feedback deutlich mehr Wirkung zeigen kann. Bei Veit wird auf die Wichtigkeit von „Intermediate Outcomes“ hingewiesen, neben den zeitnahen Ergebnissen zur unmittelbaren Steuerung auch solche, die ein längerfristiges Ergebnis besser vorhersagen können (vgl. Veit et al. 2012e, S. 12). Auch bei den Qualitätsindikatoren wird in einer Arbeit auf Aktualität und planmäßige Überprüfung der Messgrößen eingegangen (vgl. Geraedts et al. 2002, S. 94).

Anwendungsdefinition und -beispiele: Für die Bewertung dieser Anforderung sollten v.a. folgende Aspekte betrachtet werden:

- Indikatoren und Qualitätsmessung sind so gestaltet (und vor allem automatisiert darstellbar), dass die Auswertung und Messergebnisse möglichst zeitnah vorliegen.
  - Kurzfristergebnisse sollten auch unterjährig zur Einbindung ins Managementsystem zur Verfügung stehen – somit kann sogar kurzfristig auf bspw. Monatsergebnisse reagiert werden.
  - Langzeitergebnisse der Auswertung für den Leistungserbringer sind innerhalb von maximal zwei Jahren verfügbar. Somit ist eine mittelfristige Reaktion auf die Ergebnisse möglich.
- Sollte auf Systemebene eine solch hohe Aktualität nicht möglich sein, so stehen Mechanismen für den Leistungserbringer und den Kostenträger zur Verfügung, erste Auswertungen eigenständig durchzuführen und anhand der offiziellen Ergebnisse zu verifizieren.

Informationsquellen: Auch für dieses Kriterium stehen vorwiegend die Beschreibungen des P4P-Systems zur Verfügung. Hierbei ist neben der angegeben Aktualität der Ergebnisse vor allem zu überprüfen, ob die Überlieferung der Messergebnisse auch tatsächlich zum geplanten Zeitpunkt realisiert werden kann und konnte. Außerdem können stattfindende „Parallel-Messungen“ beim Leistungserbringer oder Kostenträger mit in die Bewertung einfließen – hier ist allerdings zu überprüfen, ob die Ergebnisse kohärent zu den offiziellen Messergebnissen des Systems sind oder Abweichungen vorliegen.

Kriterium 12c: Kontrollierbarkeit

Systemelement: Messergebnis

Kriterienkategorie: *Operationalisierbarkeit*

Beschreibung: Der Begriff der Kontrollierbarkeit soll hier im „engeren Sinne“, nämlich als „Nachprüfbarkeit“ verstanden werden (vgl. Gabler 1988; „Controllability/Kontrollierbarkeit“)[1]. Hierunter werden verschiedene Aspekte subsummiert, die es sowohl den beteiligten Organisation und Mitarbeiter als auch übergeordneten Instanzen erlauben, die Messergebnisse zu „kontrollieren“.

Begründung/Bezug zu existierenden Systemen: Bereits bei Ortlieb wird die „Kontrollierbarkeit“ als eine der Anforderungen an qualitätsorientierte Anreizsysteme beschrieben (vgl. Ortlieb 1993, S. 64f.; siehe auch Horvath 2008). Bei P4P-Systemen wird auf dieses Kriterium bislang nicht explizit eingegangen. Diese detaillierte Beschäftigung mit Anforderungen an die organisatorische Umsetzung der Messergebnisse ist daher nach bestem Wissen neu.

Anwendungsdefinition und -beispiele: Messergebnisse eines qualitätsorientierten Anreizsystems können dann als „nachprüfbar“ angesehen werden, wenn…

- … die Messergebnisse beim Kostenträger und Leistungserbringer vorliegen und von diesen auswertbar sind (der Leistungserbringer sollte wenn möglich die Ergebnisse auf Fallebene auswerten können).
- … die Messergebnisse z. B. aktuell, nachvollziehbar, verständlich und transparent sind.
- … die Messung den bereits betrachteten Kriterien der „Wissenschaftlichkeit“ entspricht, z. B. einer hohen Validität.

Insgesamt kann durch eine hohe Kontrollierbarkeit der Messergebnisse von verschiedenen Beteiligten die Akzeptanz und dadurch die Anreizwirkung des Systems deutlich erhöht werden.

Informationsquellen: Als Informationsquelle für diese Kriterienbewertung dienen vor allem die Beschreibung der Messergebnisse sowie Analysen bezüglich der Messergebnisse selbst.

Kriterium 12d: Praktische Umsetzung im Controlling

Systemelement: Messergebnis

Kriterienkategorie: *Operationalisierbarkeit*

[1] Aspekte der „Controllability“ im weiteren Sinne werden im Rahmen der nächsten zwei Kriterien diskutiert.

Beschreibung: Zusätzlich zur Kontrollierbarkeit soll das Kriterium 12d überprüfen, ob ein standardisiertes Reporting im Controlling des Leistungserbringer oder des Kostenträgers auch tatsächlich implementiert ist. Während das Kriterium 12c die potenzielle Umsetzbarkeit im Controlling betrachtet, soll hier die praktische, bereits bestehende Umsetzung bewertet werden.

Begründung/Bezug zu existierenden Systemen: Wie vor allem in den Fallstudien deutlich wird, ist derzeit eines der größten Schwierigkeiten von P4P-Systemen in der deutschen Krankenhauslandschaft die tatsächliche Umsetzung im Controlling – aufgrund von fehlender Priorität werden heutzutage die P4P-Aspekte sowohl beim Kostenträger als auch beim Leistungserbringer häufig nicht standardisiert kontrolliert oder umgesetzt. Nach bestem Wissen wird dieses Kriterium bislang nicht betrachtet.

Anwendungsdefinition und -beispiele: Zunächst kann dieses Kriterium durch die tatsächliche, praktische Umsetzung von Messergebnissen im Controlling bewertet werden. Messergebnisse eines qualitätsorientierten Anreizsystems werden dann in ihren Controlling-Mechanismen umgesetzt, wenn…

- … die Ergebnisse der Qualitätsmessung des qualitätsorientierten Vergütungssystems automatisiert und standardisiert in Controlling-Berichten enthalten sind (bspw. standardisierte, algorithmische Definitionen der Indikatoren, vgl. Kriterium 6d).
- … die Messergebnisse in Auswertungsgesprächen sowohl im Management als auch der operativen Unternehmensebene (beim Leistungserbringer und den Kostenträgern) diskutiert werden.
- … die Messergebnisse einen möglichst hohen Stellenwert auch im Vergleich zu wirtschaftlichen Kennzahlen haben.

Informationsquellen: Dieses Kriterium kann anhand einer Detailanalyse des jeweiligen P4P-Systems bewertet werden – ob und wie gut das System im Controlling etabliert ist, lässt sich vor allem anhand von Interviews bei Controlling-Mitarbeitern sowie die Analyse von Controlling-Berichten und -Mechanismen bewerten. Die folgenden Fallstudien können als Auswertungen dieser Art bereits erste Erkenntnisse liefern. Langfristig wären auch Umfrageergebnisse erstrebenswert.

Kriterium 12e: Verursachungsgerechte Zuordnung und Beeinflussbarkeit

Systemelement: Messergebnis

Kriterienkategorie: Operationalisierbarkeit

Beschreibung: Um die Messergebnisse für Steuerung und „Controlling-Zwecke“ nutzen zu können, müssen sie im Unternehmen „verursachungsgerecht“ zugeordnet werden können. Die Ergebnisse der Qualitätsmessung müssen demnach auf den verschiedenen Ebenen der Principal-Agent-Beziehungen – d. h. Organisationsebene des Leistungserbringer und der Kostenträger, Führungskräfte, Mitarbeiterebene

– jeweils denen zugeordnet werden können, die das Messergebnis „verursacht“ haben und hierfür (mit)verantwortlich sind.

Begründung/Bezug zu existierenden Systemen: Wenn auch nur untergeordnet in den Anforderungen der Anreizsysteme, so ist „Controllability“ für die betriebswissenschaftliche Literatur allgemein ein wichtiges Thema (vgl. u. a. Wünsch 2012, S. 30). Allerdings ergibt sich die Frage, wie weit es diese „Beherrschbarkeit/Beeinflussbarkeit“ der Ergebnisse geben kann und muss. Eng verknüpft wird die Anforderung der „Controllability“ nämlich zumeist mit der „Disponibilität“ oder „Beeinflussbarkeit“ der Messergebnisse (vgl. Gabler 1988; vgl. auch Horvath 2008). Neuerdings wird jedoch auch ein weiterer Begriff der „Controllability“ diskutiert, der auch indirekte Aussagen über die Leistungen eine wertvolle Information liefern kann (vgl. Schäffer/Pelster 2007). Dabei wird im Rahmen einer Principal-Agent-Beziehung argumentiert, dass zur Überwindung der Informationsasymmetrie auch solche „Variablen zur Leistungsbeurteilung herangezogen werden, die indirekt Auskunft über den Arbeitseinsatz des Agenten geben und von diesem nicht beeinflusst werden können“ (Schäffer/Pelster 2007, S. 424). Neben dem Postulat der „Beeinflussbarkeit“ ist es also möglich, eine verursachungsgerechte Zuordnung durchzuführen, wenn nur indirekte Beziehungen zwischen Leistung des Einzelnen und dem Messergebnis vorliegen.

Bei der Zuordenbarkeit werden bislang zum einen bei den Gestaltungsdimensionen der „Adressatenebene“ die Frage der „Verantwortlichkeit“ bereits angesprochen, aber nicht näher diskutiert (vgl. SVR 2007; Petersen et al. 2006). Bei den Anforderungen an Qualitätsindikatoren wird aber bspw. bei QUALIFY von „Beeinflussbarkeit“ gesprochen (vgl. Reiter et al. 2007; vgl. auch Collopy et al. 2000, S. 38; Geraedts et al. 2002, S. 94). Auch bei Veit wird die „zeitliche und persönliche Zuordenbarkeit von Performance und Qualität“ (Veit et al. 2012b, S. 49) – wie diese Dimension bewertbar ist, wird nicht diskutiert.

Anwendungsdefinition und -beispiele: Um dieses Kriterium zu bewerten sind folgende Überlegungen relevant:

- Messergebnisse werden pro Leistungserbringer gemessen, direkt von den beteiligten Leistungserbringern beeinflusst und diesen zugeordnet. Diese Zuordnung ist möglich, wenn Messergebnisse klar mit den Leistungen des Leistungserbringer in Verbindung gebracht werden können (d. h. fachlich zuordenbaren Indikationsergebnissen) und verschiedene Risikoadjustierungen stattfinden, damit kontextuell Unterschiede betrachtet werden (siehe hierfür auch Kriterium 5c).

- Messergebnisse können außerdem den einzelnen Fachabteilungen zugeordnet werden. Auch dies ist aufgrund des Wesens der Qualitätsmessung zumeist gut möglich – viele der existierenden Indikatoren beziehen sich großteils auf eine Fachabteilung. Hier können allerdings erste Schwierigkeiten auftreten, da einige Fachabteilungen der „Sekundärmedizin“ (z. B. Radiologie) nur sehr begrenzt in solchen Messungen abgebildet werden, obwohl es sicherlich eine (indirekte)

Beteiligung an den Ergebnissen gibt. Daher kann es sinnvoller sein, die Ausrichtung der Messergebnisse für einzelne Krankheitsbilder oder Eingriffe zu gestalten und die Zuordnung innerhalb des Leistungserbringers diesem zu überlassen.

- Auf individueller Ebene sollten die Zuordnungen stets durch den Leistungserbringer selbst vorgenommen werden. Von außen kann nur sehr begrenzt eine verursachungsgerechte, beeinflussungsgerechte Verteilung durchgeführt werden. Daher kann als eine der notwendigen Maßnahmen für den Leistungserbringer bereits das „Herunterbrechen“ der Messergebnisse und der dazugehörigen Vergütungsfunktion auf Führungskräfte- und vielleicht sogar Mitarbeiterebene gesehen werden. Im Rahmen von Zielvereinbarungsgesprächen wären somit ein Feedback und eine Zuordnung der Ergebnisse möglich.
- Die Beeinflussbarkeit kann unter anderem erreicht werden, in dem der Fokus auf solchen Indikatorenergebnissen gelegt wird, die nachweislich medizinisch veränderbar sind. So sollten bspw. infauste, präterminale Krankheitsbilder vermieden werden.

Informationsquellen: Auch für die Bewertung dieses Kriteriums dienen vor allem die Detailbeschreibungen und Umsetzungshandbücher als Quellen. Zusätzlich müssen auch Dokumentationen der tatsächlichen Umsetzung beim Leistungserbringer und ggf. Zielvereinbarungen erstellt und betrachtet werden. Gerade die subjektive, individuelle Einschätzung kann im Rahmen von Umfragen und qualitativen Interviews hinterfragt werden.

Kriterium 13a: Zeitverlauf

Systemelement: Messergebnis

Kriterienkategorie: *Vergleichbarkeit*

Beschreibung: Die letzte Kriterienkategorie für das Systemelement der Messergebnisse ist die Vergleichbarkeit – diese soll neben der Operationalisierbarkeit ausdrücken, inwieweit die Ergebnisse der Qualitätsmessung verglichen und daraus sinnvolle Aussagen für bspw. Prozessveränderungen abgeleitet werden können. In dem Kriterium des „Zeitverlauf“ geht es um die Möglichkeit eines Längsschnittvergleichs, der es dem Leistungserbringer ermöglicht, ggf. auch unterjährig auf Entwicklungen einzugehen.

Begründung/Bezug zu existierenden Systemen: Die longitudinale Vergleichbarkeit von Messergebnissen wird nach bestem Wissen derzeit weder in der betriebswissenschaftlichen Literatur noch in der P4P-Diskussion betrachtet. Zwar werden in den Gestaltungsdimensionen von qualitätsorientierten Vergütungssystemen ein Zeitpunkt und Dauer von Anreizsystemen erwähnt (vgl. z. B. Scheppach et al. 2011), eine spezifischen Bezug zu zeitlichen Entwicklungen und deren Darstellung für die beteiligten

Organisationen wird nicht betrachtet. Die Leitlinien der American College of Cardiology erwähnt die Wichtigkeit, transparente Rating-Methoden zu entwickeln (vgl. Brush et al. 2006, S. 2605).

Gleichzeitig werden sowohl in den derzeitigen Berichten des „Public-Reportings“ (vgl. bspw. AQUA-Institut GmbH 2011a) als auch bei der Bewertung existierender P4P-Systeme die meisten Ergebnisse tatsächlich im Längsschnitt betrachtet (vgl. z. B. historische Kohortenstudien mit longitudinalem Design bei SVR 2007, S. 181). Als Gestaltungsdimension oder gar Anforderung an die Darstellung der Messergebnisse – auch um bspw. die zeitliche Entwicklung für den Leistungserbringer verständlich zu machen – wird es nur bei Amelung gesehen (vgl. Amelung/Zahn 2009, S. 40).

Anwendungsdefinition und -beispiele: Um dieses Kriterium zu bewerten, sind folgende Aspekte relevant:

- Messergebnisse werden longitudinal dargestellt, um die zeitliche Entwicklung bewerten zu können.
- Möglichen Änderungen und Anpassungen der Qualitätsmessungen werden dabei Rechnung getragen, so dass der Längsschnitt zu vergleichbaren Aussagen führt. (Strukturelle Änderungen an Messergebnissen werden insgesamt vermieden)
- Für Leistungserbringer werden diese Ergebnisdarstellungen so gewählt, dass Aussagen zu Managemententscheidungen möglich sind – bspw. auch unterjährig, bereinigt um Messänderungen oder Fallzahlentwicklungen. So können auch eigene historische Maßnahmen begutachtet werden.

Insgesamt ist festzuhalten, dass es sich hierbei um eine sehr offensichtliche und einfach umzusetzende Anforderung handelt. Da die zeitliche (um Messänderungen bereinigte) Entwicklung für operative Änderungen beim Leistungserbringer von großem Interesse ist, ist das Kriterium aber trotzdem von hoher Relevanz.

Informationsquellen: Zur Information muss hier die Darstellung der Messergebnisse in den unterschiedlichen Quellen (d. h. für verschiedene Adressaten oder Veröffentlichungen) herangezogen werden.

Kriterium 13b: Querschnittsvergleich/Ranking

Systemelement: Messergebnis

Kriterienkategorie: *Vergleichbarkeit*

Beschreibung: Neben dem vorher diskutierten Längsschnittvergleich ist auch eine Darstellung der Messergebnisse im Querschnitt von Bedeutung. Gerade für den Leistungserbringer aber auch für die Kostenträger können so die Messergebnis ins Verhältnis zu anderen Leistungserbringern gesetzt werden und ermöglichen sachdienliche Aussagen.

Begründung/Bezug zu existierenden Systemen: Wie bereits beim vorherigen Kriterium gibt es für diese Anforderung bislang nur eine indirekte Herleitung aus der bisherigen Literatur. Während diese Thematik bei den Anforderungen an Anreizsysteme in der betriebswissenschaftlichen Literatur nicht betrachtet wird, findet sich bei den Gestaltungsdimensionen der P4P-Systeme die Frage der verschiedenen „Ranking-Arten“ (vgl. SVR 2007; Scheppach et al. 2011). In diesem Kriterium jedoch soll nicht entschieden werden, welche der „Bemessungsgrundlagen“ zu empfehlen ist – ausreichende Studienergebnisse liegen hierfür nicht vor, außerdem handelt es sich dabei um eine Gestaltungsdimension nicht um eine Anforderung. Als Anforderung ist hier zunächst zu sehen, dass die Messergebnisse des Leistungserbringers überhaupt im Vergleich zu anderen Leistungserbringern oder Fachabteilungen dargestellt werden.

Bislang wird in vielen Qualitätsberichten neben den individuellen Messergebnissen zumindest ein Referenz- oder Durchschnittswert angegeben (vgl. u. a. AQUA-Institut GmbH 2011a oder Helios Kliniken Gruppe 2005 – es sei allerdings zu beachten, dass einige „Qualitätsberichte“ selbst auf diesen ersten Vergleich verzichten, z. B. bei KTQ).

Anwendungsdefinition und -beispiele: Für die Bewertung dieses Kriteriums sind folgende Aspekte relevant:

- Messergebnisse werden im Querschnittsvergleich dargestellt, so dass sie zu den Ergebnissen anderer Leistungserbringer (oder Fachabteilungen) ins Verhältnis gebracht werden können.
- Hierfür ist entscheidend, dass die Messergebnisse an unterschiedlichen Voraussetzungen angepasst sind, aber diese Anpassungen auch transparent und verständlich dargestellt werden.
- Jeder Leistungserbringer oder jede Fachabteilung wird mindestens mit einem Referenzwert (bspw. aus der medizinischen Fachliteratur) oder Durchschnitt (z. B. auf Bundesebene) verglichen, so dass eine Einordnung der Messergebnisse möglich ist.
- Weiter differenzierte Querschnittsvergleiche (bspw. Rankings, Perzentile) sind erstrebenswert. Wenn die Vergütungsfunktion von einem detaillierten Ranking abhängen, sind diese Vergleiche sogar notwendig.
- Einzelfallvergleiche werden ggf. sogar im Rahmen von Peer-Reviews durchgeführt und anhand dieser Verfahren Prozessveränderungen unterstützt.

Auch diese Anforderung ist zunächst recht offensichtlich – je detaillierter der Querschnittsvergleich für den einzelnen Leistungserbringer, die Fachabteilung oder gar den Chefarzt jedoch ausfällt, desto besser kann eine Einschätzung und anschließend eine Verhaltensänderung durchgeführt werden. Daher ist eine Beschäftigung mit diesem Kriterium von hoher Relevanz.

Informationsquellen: Zur Information muss hier die Darstellung der Messergebnisse in den unterschiedlichen Quellen (d. h. für verschiedene Adressaten oder Veröffentlichungen) herangezogen werden.

Kriterium 13c: Justiziabilität

Systemelement: Messergebnis

Kriterienkategorie: *Vergleichbarkeit*

Beschreibung: Die letzte Anforderung an das Messergebnis unter der Kriterienkategorie der Vergleichbarkeit kann in der Justiziabilität der Ergebnisse gesehen werden. Hierunter wird verstanden, dass die Ergebnisse der Qualitätsmessung auch für andere Zwecke als die qualitätsorientierte Vergütung, nämlich in juristischen Prozessen, verwendet werden kann.

Begründung/Bezug zu existierenden Systemen: Ein juristisches Thema wird bislang nur in einer der hier diskutierten Quellen diskutiert – Lehmann beschreibt die Anforderung der „Rechtssicherheit" von Anreizsystemen (vgl. Lehmann 2006, S. 18). Hierunter ist jedoch vor allem die Einhaltung gesetzlicher Vorschriften zu verstehen. Dies wird zunächst einmal als gegeben vorausgesetzt – dass das Anreizsystem legal gestaltet sein muss, ist eine Anforderung, die keiner Diskussion bedarf.

Hier allerdings soll ein weiteres Thema betrachtet werden – nämlich die Möglichkeit einer Verwendung der Messergebnisse aus dem Anreizsystem für juristische Verfahren und somit ein weiterer Nutzen der vorhandenen Qualitätsmessung. Diese Anforderung wird bislang nur bei Veit diskutiert (vgl. Veit et al. 2012e, S. 40) – hier werden mehrere schon diskutierte Kriterien (Aussagekraft, Gerechtigkeit, u. a.) erwähnt. Dieses Kriterium hat einen hohen Mehrwert für die beteiligten Leistungserbringer und Kostenträger und soll hier näher detailliert werden.

Anwendungsdefinition und -beispiele: Für dieses Kriterium muss zwischen Aspekten des öffentlichen Rechts und Privatrechts unterschieden werden:

- Bei Fragen des öffentlichen Rechts (bspw. gesundheitsrechtliche Aspekte, Subventionsfragen u. ä.) ist für eine hohe Justiziabilität entscheidend, dass …
  - …Messergebnisse von unabhängigen, dritten Stellen (hoher Grad an „Kontrollierbarkeit" notwendig) auf diese Kriterien überprüft werden, so z. B. auf Manipulationssicherheit, Validität, Reliabilität aber auch Beeinflussbarkeit durch Leistungserbringer.
  - …diese Institutionen von allen Beteiligten inklusive der Patientenvertretungen und der Politik als Experteninstitutionen anerkannt sind (bspw. MDK).
  - …eine besonders hohe Kriterienerfüllung vorliegt.

- …es klare gesetzliche Vorgaben zum Umgang mit geringen Mengen gibt.

➢ Für privatrechtliche Streitfälle (bspw. Arzthaftungsfälle) müssen die Messergebnisse für eine hohe Justiziabilität andere Aspekte erfüllen, so z. B.

- Hoher Detailgrad der Dokumentation von Prozessen und Strukturen und Anlehnung an Leitlinien
- Hohe Evidenz von Leitlinien
- Erste Anwendung der Messergebnisse in vergleichbaren Gerichtsfällen
- Vorliegen bspw. von Sachverständigengutachten, die diese Ergebnisse verwenden

Insgesamt zeigt sich, dass die Justiziabilität einen höheren Grad an Erfüllung verschiedener Kriterien (Validität, Reliabilität aber auch Kontrollierbarkeit und Beeinflussbarkeit) verlangt. Inwieweit trotzdem weitere Analysen, Überprüfungen und Gutachten notwendig sind, muss im Einzelfall geprüft werden.

Informationsquellen: Da für die Bewertung dieses Kriteriums unter anderem die Erfüllung weiterer Kriterien relevant ist, steht als eine Informationsquelle die Anwendung dieses Kriterienkatalogs und die dazu benötigten Quellen zur Verfügung. Gleichzeitig sollte anhand der Beschreibung des P4P-Systems eine Aussage zur Beteiligung von Experteninstitutionen möglich sein – deren Bewertungen sollten zusätzlich in die Kriteriumsanalyse einbezogen werden.

4.5. Zwischenfazit: Der neu entwickelte Kriterienkatalog beschreibt und bewertet P4P-Systeme umfassend und beinhaltet neuartige Elemente sowie eine praxisorientierte Strukturierung

An dieser Stelle ist die Entwicklung des Kriterienkatalogs beendet. Dadurch ist eine Vielzahl neuer Erkenntnisse entstanden, die hier kurz zusammengefasst werden sollen:

➢ Zunächst wurden die theoretischen Grundlagen aus der volkswirtschaftlichen, vor allem aber der betriebswirtschaftlichen Sicht dargestellt. Hierbei wurde deutlich, dass eine Vielzahl der Theorien aus der BWL auch für qualitätsorientierte Anreizsysteme in der stationären Versorgung angewandt werden können.

➢ Gerade die Anwendung des Prinzipal-Agent-Theorems konnte die Zusammenhänge eines Anreizsystems für den stationären Sektor konkretisieren. Die Beziehung von Patienten, Ärzten, Krankenhaus und Krankenkassen konnte auf diese Weise in eine sinnvolle Struktur gefasst werden.

➢ Auch verhaltenstheoretische Ansätze können erste Erkenntnisse bringen, so bspw. eine Erklärung des „Crowding-Out-Effekts“, der die Gefahr einer Verdrängung der intrinsischen Motivation durch externe Anreize beschreibt, aber von verschiedenen Studien widerlegt wird.

➢ Die Analyse der Grundlagen von P4P konnte ebenso Erkenntnisse für den Anforderungskatalog liefern. So zeigten die historische Entwicklung und vor allem die Ziele und Funktionsweisen von P4P,

dass diese Ansätze zwar bereits angewandt und evaluiert werden, ein einheitlicher Anforderungs- oder Bewertungskatalog aber nicht besteht.

- Dagegen werden in der derzeitigen Diskussion und den existierenden Reviews vor allem erste Gestaltungsdimensionen von P4P-Systemen thematisiert. Unter anderem werden Vorschläge zu Bemessungsgrundlagen, Vergütungsanteilen oder Adressatenebenen gemacht. Gleichzeitig können für mehrere dieser Dimensionen bislang keine klare Aussagen oder Empfehlungen ausgesprochen werden, da die Studienlage unterschiedliche Ergebnisse aufweist. Diese Gestaltungsdimensionen entsprechen allerdings nicht Anforderungen im Sinne der vorliegenden Arbeit.
- Bei der Entwicklung des zweiten Teils des Anforderungskatalogs wurde wie bereits zuvor deutlich, dass die Strukturierung nach Systemelemente klare Vorteile und einen neuen Fokus mit sich bringt.
- Da ein Kriterienkatalog für den Anreizteil von P4P-Systemen derzeit nicht besteht, sind viele Einzelkriterien neuartig, so bspw. die Fragen der Praktischen Umsetzung im Controlling oder der Nutzung zur Steuerung. Zwar kann sowohl auf die Anforderungen an generelle Anreizsysteme als auch auf erste Aussagen bezüglich P4P zurückgegriffen werden, die Anwendungsdefinitionen wurden aber neu entwickelt und konkretisiert.
- Dabei ist durch die Beschreibung der Kriterien mittels Herleitung aber auch Anwendungsdefinitionen und -beispielen sowie möglichen Informationsquellen ein praktisch anwendbarer Katalog entstanden, der direkt existierende Systeme bewerten kann und Hilfestellungen zu Entwicklung eines neuen qualitätsorientierten Vergütungssystems liefert.
- Insgesamt ist ein in großen Teilen neuartiger, umfassender Katalog entstanden, dessen Anwendbarkeit und Nutzen durch die Anwendung an deutschen Beispielen analysiert und evaluiert werden soll.

5. Anwendung des entwickelten Katalogs bei P4P-Systemen

5.1. Umsetzungsbeispiele von P4P-Projekten

5.1.1. Entwicklung von P4P-Systemen – Schwerpunkt in den USA

Ein Großteil der konzeptionellen Entwicklung von P4P hat in den USA stattgefunden (vgl. u. a. Scheppach et al. 2011, S. 14). Vor allem das Institute of Medicine (IOM) war durch seine Empfehlungen maßgeblich daran beteiligt, die P4P-Diskussion voran zu bringen (vgl. Institute of Medicine 2001; Kohn et al. 1999). Auch in der Entwicklung und Anzahl von konkreten P4P-Projekten ist der angelsächsische Raum Vorreiter (vgl. Veit et al. 2012b, S. 31), vorwiegend die USA. Hier wurde 1993 eines der ersten P4P-Projekte etabliert, das vor allem bei kommerziellen Versicherten Anwendung fand (vgl. Christianson et al. 2008, S. 6; Scheppach et al. 2011, S. 14; SVR 2007, S. 169; Cannon 2007). Die Anzahl von P4P-Projekten stieg rasch an, allerdings können nur Zahlen aus Umfragen verwendet werden, da es keine Dokumentationspflicht oder Meldestelle gibt – 2005 waren 157 Projekte in den USA (vgl. Sorbero et al. 2006, S. xiii; SVR 2007, S. 169).

Insgesamt ist der größte Anteil der P4P-Projekte derzeit im ambulanten Sektor zu finden, 2005 sogar 90 % der P4P-Projekte privater Versicherer (vgl. SVR 2007, S. 169; Rosenthal/Frank 2006). In Großbritannien ist vor allem das Projekt der „Quality and Outcomes Framework (QOF)“ erwähnenswert, das seit 2004 etabliert ist – dieses P4P-System in der hausärztlichen Versorgung wird durch Steuergelder finanziert und ermöglicht einen bis zu 25 % Bonus in Höhe von 27.000 pro Hausarzt oder insgesamt 1,8 Mrd. Pfund (vgl. Roland 2004, S. 1448; Amelung/Zahn 2009, S. 52f.; McDonald et al. 2009, S. 751ff.). Die Messung findet anhand von 146 Qualitätsindikatoren statt, die zur Hälfte Ergebnisindikatoren sind (vgl. SVR 2007, S. 170). Eine aktuelle Studie zeigt eine regionale, signifikante Verbesserung der 30-Tages-Mortalität um 6 % (1,3 Prozentpunkte) in 18 Monaten – dies entspricht 890 Todesfällen (vgl. Sutton et al. 2012, S. 1824).

Weniger, dafür größere Projekte sind derzeit im stationären Sektor zu finden, insgesamt wird davon ausgegangen, dass es in 2009 ca. 40 Projekte in amerikanischen Krankenhäusern gab – insgesamt ca. 38% aller Krankenhäuser in den USA; Überblicksartikel über alle stationären Projekte existieren bisher nicht (vgl. Veit et al. 2012b, S. 31; Veit et al. 2012c, S. 21; Mehrotra et al. 2010a, S. 118; Amelung/Zahn 2009, S. 31). Die bekanntesten und am längsten laufenden Projekte sind „Premier Hospital Quality Incentive Demonstration Project (PHQID)“ der CMS seit 2003, „Hospital Quality Service and Recognition“ der Hawaii Medical Service Association seit 2001 sowie „Participating Hospital Agreement Incentive Program“ der Blue Cross Blue Shield of Michigan Versicherungsgesellschaft seit 2000 (vgl. Mehrotra et al. 2010a, S. 119ff.; Veit et al. 2012c, S. B 22ff.). Tabelle 9 zeigt eine kurze Übersicht der Projekte und deren Funktionsweise bzw. Anreizmodelle.

Tabelle 8: Beschreibung einer Auswahl stationärer P4P-Projekte in den USA

Projekt	Krankenhäuser	Qualitätsmessung	Vergütung
PHQID	Über 250 in 38 US-Staaten	➢ *5 Krankheiten*: akuter Myokardinfarkt, ambulant erworbene Pneumonie, Herzinsuffizienz, koronare Bypass-Operation, Hüft/Knie-TEP ➢ *33 Qualitätsindikatoren*, meist Prozessqualität, aber auch Mortalität, Wiederaufnahme- und Komplikationsrate	➢ Oberstes Zentil 2 %, zweites Zentil 1 % Zuschlag auf Medicare-Entgelt ➢ Nach 3 Jahren, Abschlag von 1-2 % für unterste zwei Zentile ➢ Durchschnittliche Vergütungshöhe 71.960 Dollar pro Jahr und Krankenhaus
Blue Cross Blue Shield of Michigan	86 Kranken-häuser in Michigan	➢ Fokus auf Herzinfarktversorgung ➢ Indikatoren der Patienten-sicherheit, Zugangs- und Prozessqualität ➢ Mindestens durchschnittliche Leistung	➢ Zuschlag von bis zu 4 % des DRG-Einkommens ➢ In 2004 pro Krankenhaus zwischen 30.000 und 4 Millionen Dollar
Hawaii Medical Service Association	17 (jedes Kranken-haus Hawaiis)	➢ Erfüllungsgrad von Prozess-maßnahmen bei akutem Myokardinfarkt ➢ Komplikationsrate und durch-schnittliche Verweildauer nach Geburt und 18 häufigen Operationen ➢ Patienten- und Arztzufriedenheit mit stationärer und Notfallbehandlung ➢ Umsetzung internes Qualitätsmanagementsystem	➢ Auszahlungen pro Jahr von durchschnittlich 500.000 Dollar pro Krankenhaus, insgesamt etwa 9 Millionen Dollar

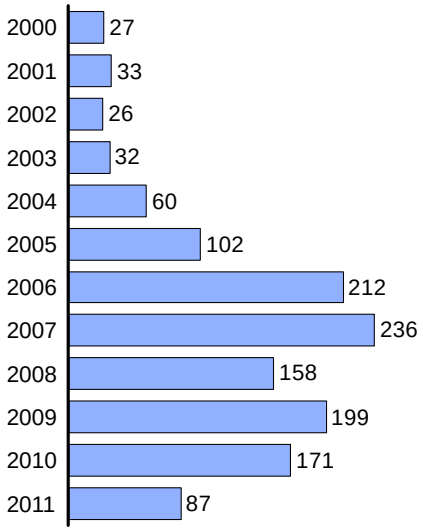
Quellen: SVR 2007, S. 171; Mehrotra et al. 2010a, S. 119ff.; Veit et al. 2012c, S. B 22ff.; Lindenauer et al. 2007; Berthiaume et al. 2004, S. 501f.; Berthiaume et al. 2006, S. 36f.; Sautter et al. 2007, S. 95

5.1.2. Derzeitige Bewertungen von P4P-Systemen in der Literatur

Eine weitere Entwicklung von P4P kann in der Vielzahl von Bewertungen und Evaluationen gesehen werden. Der erste, bekannte Review wurde bereits 2005 von Meredith Rosenthal veröffentlicht (Rosenthal et al. 2005). Weitere (auch systemische) Reviews folgten (vgl. unter anderem Rosenthal/Frank 2006; Rosenthal 2008; SVR 2007; van Herck et al. 2010; Amelung/Zahn 2009). Besonders die amerikanischen Projekte wurden mehrfach und sehr detailliert evaluiert, teils mit sehr speziellen Fragestellungen z. B. zur Effizienz (vgl. Robinson et al. 2009). Auch vergleichende Darstellungen von Projekten werden erstellt (vgl. z. B. Kuhmerker/Hartman 2007). Die Aussagen der Reviews beziehen sich derzeit auf die Ausgestaltungen der Projekte. Die meisten der Reviews jedoch analysieren die Qualitätsverbesserung, die aufgrund des Einsatzes der qualitätsabhängigen Vergütung stattgefunden hat. Hierfür werden zumeist randomisierte Studien genutzt.

Erste Überblicksarbeiten wurden auch in Deutschland durchgeführt. Das meist beachtete ist sicherlich das Gutachten des Sachverständigenrats, das 2007 den Stand der Entwicklung von P4P bewertete (vgl. SVR 2007). Ein systematisches Review wird auch bei Scheppach vorgenommen: Hier wurden anhand von 58 Studien nach einer Datenbankrecherche verschiedene Erkenntnisse unter anderem zu Anreizgestaltung, Bemessungsgrundlagen, Anreizhöhe, Anreizfinanzierung, Anreizzeitraum, Qualitätsmessung, Dokumentation sowie Qualitätsverbesserung diskutiert (vgl. Scheppach et al. 2011, S. 102ff.). Besondere Beachtung verdient außerdem das Gutachten der BQS, dass Ende des Jahres 2012 erschienen ist und vom Bundesministerium für Gesundheit in Auftrag gegeben wurde, um die Stand von P4P zu analysieren (vgl. Veit et al. 2012a). Hierin wurde sowohl ein ausführliches Review anhand von Sekundärliteratur als auch eine Internetumfrage, Experteninterviews und einen Workshop für die Analysen der deutschen P4P-Situation durchgeführt (vgl. Veit et al. 2012b, S. 20 ff.). Die Anzahl und Verteilung von eingeschlossenen Studien zeigt bereits das hohe wissenschaftliche Interesse für die P4P-Thematik, die allerdings 2007 ihren derzeitigen Höhepunkt erfahren hat (vgl. Abbildung 17).

Abbildung 17: Anzahl eingeschlossener Publikationen zu P4P im BQS-Gutachten



Quelle: Veit et al. 2012a, S. 22

Insgesamt zeigt sich in der derzeitigen Literatur ein Fokus auf der Evidenzanalyse, ob es zu erfolgreicher Qualitätsverbesserung durch P4P-Systeme kam. Einige Übersichtsarbeiten wie bspw. Van Herck et al. zeigen, dass es zu einer durchschnittlichen Verbesserung von Qualitätsergebnisse von bis zu 5 % durch finanzielle Anreize kommen kann (vgl. van Herck et al. 2010). Auch im SVR-Gutachten konnte ein systematisches Review 21 von 28 Studien mit positiven Ergebnissen identifizieren (vgl. SVR 2007, S. 176). Andere Reviews zeigen ein gemischtes Bild (vgl. u. a. Frølich et al. 2007; Institute of Medicine 2007; Sorbero et al. 2006; Lester et al. 2010). Auch einige Einzelstudien zeigen Verbesserungen eher in Prozess- als Ergebnisqualität bzw. können keine statisch signifikanten Verbesserungen nachweisen (vgl. Chien et al. 2007, S. 283f.).

Die wissenschaftliche Evidenz von P4P-Projekten im stationären Sektor zu evaluieren, ist derzeit noch schwieriger: So existieren bislang keine hochwertigen Studien speziell für die Evidenz von P4P in Krankenhäusern (vgl. Veit et al. 2012c, S. 22). Beispielhaft sollen hier kurz die Studienergebnisse für die zuvor vorgestellten stationären Projekte in den USA aufgezeigt werden.

PHQID: Während zwei Studien eine Verbesserung der teilnehmenden Krankenhäuser auch im Vergleich zu einer Kontrollgruppe von zwischen 2,6 % und 4,3 % zeigen konnten, waren die Ergebnisse einer dritten Studie statistisch nicht signifikant (vgl. Mehrotra et al. 2010a, S. 124ff.; Lindenauer et al. 2007, S. 490ff.). Spätere Analysen jedoch konnten keine Qualitätsverbesserung im Vergleich zur Kontrollgruppe mehr zeigen (vgl. Veit et al. 2012c, S. 22f.; Werner et al. 2011, S. 691ff.; Jha et al. 2012).

Blue Cross Blue Shield of Michigan: Für dieses Projekte konnten ebenso positive wie negative Ergebnisse nachgewiesen werden. Eine Befragung der Krankenhäuser ergab eine Struktur- oder Prozessveränderungen aufgrund des P4P-Projekts bei Dreiviertel der Teilnehmerorganisationen (vgl. Reiter et al. 2006, S. 124f.; Veit et al. 2012c, S. 24). Auch eine Analyse der Wirtschaftlichkeit zeigte positive Ergebnisse (vgl. Nahra et al. 2006, S. 49f.; Mehrotra et al. 2010a, S. 122) - doch auch das Ausbleiben von Verbesserungen kann nachgewiesen werden (vgl. Veit et al. 2012c, S. 24).

Hawaii Medical Service Association: Hier kann eine Verbesserung, d. h. Reduktion der Komplikationsrate nach Geburt und häufigen operativen Eingriffen, von 2 Prozentpunkten nachgewiesen werden, sowie eine Verweildauerreduktion von 1,2 - 0,4 Tage – allerdings ohne den Vergleich mit einer Kontrollgruppe (vgl. Berthiaume et al. 2006, S. 38f.). Auch hier gibt es andere Studien, die keine Verbesserung zeigen können (vgl. Mehrotra et al. 2010a, S. 121f.; Lindenauer et al. 2007, S. 494f.).

Insgesamt ist die empirische Evidenz weiterhin schwer zu beurteilen – weder eine Verbesserung noch ein Ausbleiben einer Verbesserung durch finanzielle Anreize ist statistisch nachweisbar. Dies kann u. a. anhand der folgenden Effekte erklärt werden (vgl. Veit et al. 2012c, S. 43–44):

- Vorwiegende Selektion von Teilnehmern mit bereits höherer Anfangsqualität bei freiwilliger Teilnahme
- Verbesserung der Ergebnisse von Versorgern mit guter und schlechter Anfangsqualität meist in marginaler Gesamtverbesserung nicht ausreichend differenzierbar
- Vermischung der Verbesserung durch das P4P-Programm mit kontinuierlichen Qualitätsverbesserungen, die auch ohne finanzielle Anreize auftreten
- Qualitätsverbesserung häufig durch P4P- aber auch weitere Initiativen erklärbar

Gleichzeitig zeigen die alltäglichen sowie die Expertenerfahrungen eine deutlich positive Einschätzung von P4P-Projekten (vgl. Veit et al. 2012c, S. 43f. sowie die Ergebnisse der folgenden Fallstudien dieser Arbeit).

Auch außerhalb von Reviews wurde das Thema in einer Vielzahl von Artikeln diskutiert – sowohl von wissenschaftlichen Autoren und Instituten (vgl. z. B. Amelung et al. 2009; Malzahn et al. 2010; Schrappe/Gültekin 2010; Mansky 2010a) als auch von klinischen Praktikern – vor allem in fachspezifischen Zeitschriften (vgl. z. B. Bauer 2010; Kostuj/Smektala 2010) oder in politischen, gesellschaftsrelevanten Kommentaren (vgl. z. B. Jaeckel 2009, S. 12). Neben der bereits erwähnten weit verbreiteten Diskussion von Gestaltungsdimensionen der P4P-Projekte, gibt es vor allem Beschreibungen der existierenden angelsächsischen Projekte und deren Erfolge, doch auch die generelle Idee von qualitätsvariabler Vergütung, finanziellen Anreizen und Wettbewerb werden diskutiert – teilweise mit sehr deutlichen Formulierungen bezüglich der Funktionsweise wie bspw. „Bei P4P reagieren die Ärztinnen und Ärzte auf die perversen Anreize damit, dass sie immer mehr unnötige Untersuchungen durchführen und vorschnell Medikamente verschreiben, weil das zu mehr gemessener «Qualität» führt.“ (Sax 2010, S. 2). Insgesamt jedoch findet eine sehr konstruktive Diskussion statt, die auch Vorschläge für die P4P-Umsetzung in Deutschland konkretisiert (vgl. Hertle/Veit 2012; Mansky 2010a).

5.1.3. Exkurs: Neueste Entwicklung von P4P in den USA

Besonders beachtenswert ist die neue Entwicklung eines bundesweiten P4P-Projektes in den USA. Erste verpflichtende Ansätze gibt es seit 2004: Der staatlichen Hauptversicherer Medicare bzw. dessen „Center for Medicare & Medicaid Services (CMS)“ verpflichtet seitdem ihre Krankenhäuser dazu, am „Inpatient Quality Reporting (IQR)“ und somit an zwei Qualitätsinitiativen teilzunehmen: 1) Ein 2 %iger, variabler Anteil der jährlichen Zahlung auf Grundlage der Dokumentationsvollständigkeit von Qualitätsergebnissen; 2) die verpflichtende Veröffentlichung von Qualitätsergebnissen im Internet (unter www.hospitalcompare.com). Die Qualitätsmessung umfasst neben Umfragen zur Patientenzufriedenheit vor allem Indikatoren aus Routinedaten, so bspw. AHRQ-Indikatoren, Wiederaufnahmen und verschiedene Mortalitätsraten (vgl. Gesetz „Medicare Prescription Drug, Improvement, and Modernization Act (MMA)“, Section 501(b) 2003, sowie Gesetz „Deficit Reduction Act“ 2005).

Doch vor allem die Entwicklungen der letzten Jahre in Richtung einer vermehrt qualitätsorientierten Vergütung sind fortschrittlich und bemerkenswert: Mit dem „Affordable Care Act“ aus dem Jahr 2010 verabschiedete das U.S. Department of Health and Human Services (HHS) das Programm des „Hospital Value-Based Purchasing (HVBP)“. Dies ist die erste Einführung eines bundesweiten P4P-Konzepts in den USA und weltweit. Ab dem Finanzjahr 2013 ist 1 % des Medicare-Einkommens abhängig von 12 klinischen Qualitätsindikatoren sowie Ergebnissen der Patientenzufriedenheit – mit einer Gewichtung von 70:30. Die meisten Indikatoren betreffen Prozessqualität, zusätzlich wird die „Patientenerfahrung“ betrachtet (vgl. HHS - CMS 08.05.2011, S. 26510; Ryan/Blustein 2012):

- **Klinische Indikatoren:**
  - **AMI:** (1) Fibrinolytic therapy received within 30 minutes of hospital arrival, (2) Primary PCI received within 90 minutes of hospital arrival
  - **Heart failure:** (3) Discharge instructions
  - **Pneumonia:** (4) Blood cultures performed in the emergency department prior to initial antibiotic received in hospital, (5) Initial antibiotic selection for CAP in immunocompetent patient
  - **Healthcare-associated infections:** (6) Prophylactic antibiotic received within 1 hour prior to surgical incision, (7) Prophylactic antibiotic selection for surgical patients, (8) Prophylactic antibiotics discontinued within 24 hours after surgery end time, (9) Cardiac surgery patients with controlled 6AM postoperative serum glucose
  - **Surgeries:** (10) Surgery patients on a beta blocker prior to arrival that received a beta blocker during the perioperative period, (11) Surgery patients with recommended venous thromboembolism prophylaxis ordered, (12) Surgery patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery
- **Indikatoren der Patientenerfahrung:** (1) Communication with nurses, (2) Communication with doctors, (3) Responsiveness of hospital staff, (4) Pain management, (5) Communication about medicines, (6) Discharge information, (7) Cleanliness of hospital environment, (8) Quietness of hospital environment, (9) Overall rating of hospital

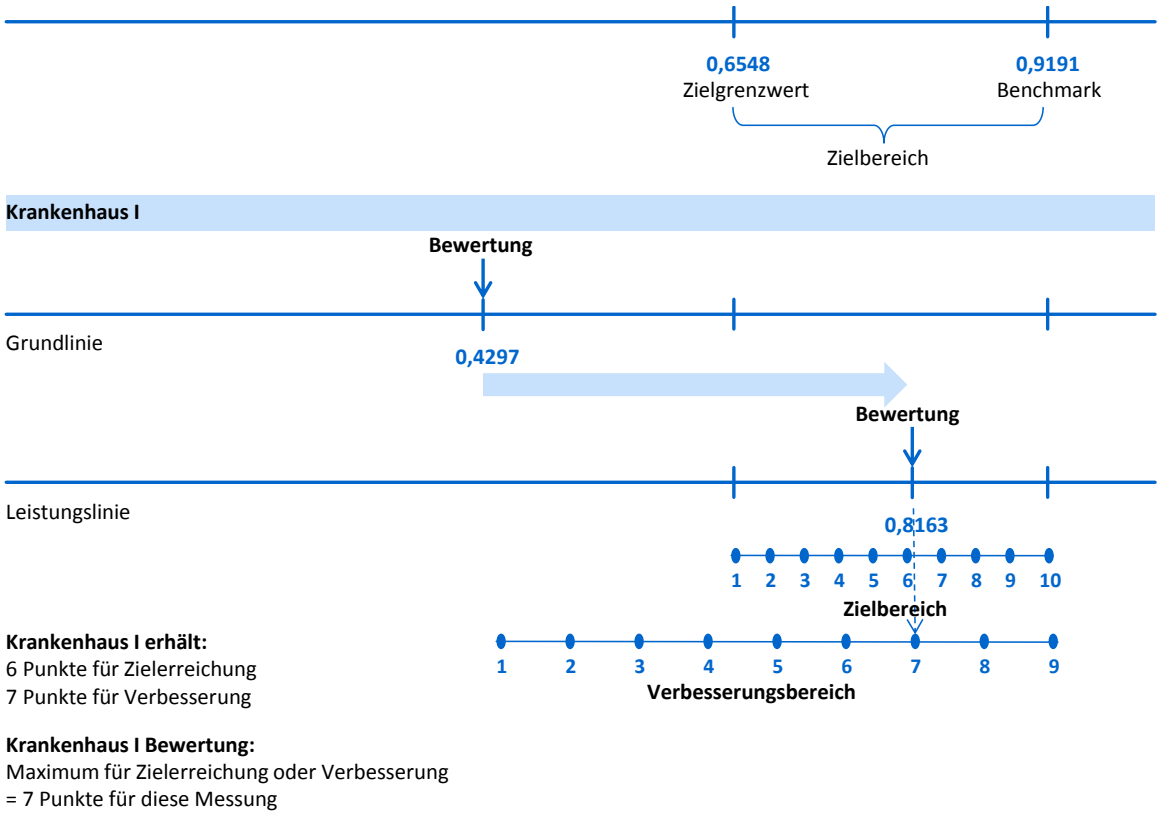
Ab 2014 kommen weitere Indikatoren der Ergebnisqualität dazu (vgl. HHS - CMS 08.05.2011, S. 26511):

- **Sterblichkeitsindikatoren:**
  - Akuter Myokardinfarkt: 30-Tages Sterblichkeitsrate
  - Herzinsuffizienz: 30-Tages Sterblichkeitsrate
  - Pneumonie: 30-Tages Sterblichkeitsrate
- **AHRQ – Patientensicherheits- und Inpatient Quality Indikatoren**
  - Zusammengesetzter Indikator zu Komplikationen und Patientensicherheit
  - Zusammengesetzter Indikator zur Sterblichkeit für ausgewählte Krankheiten
- **Indikatoren der „Hospital acquired conditions“-Initiative** – im Krankenhaus erworbene Krankheiten oder Zustände
  - Zurückgelassene Fremdkörper nach Operationen
  - Luftembolie
  - Blutunverträglichkeit
  - Dekubitus der Stadien III und IV
  - Stürze und Traumata
  - Infektionen durch Herzkatheter
  - Infektionen durch Harnkatheter

– Schlechte Blutzuckereinstellung

Die Ergebnisse der Qualitätsmessung werden in einen Gesamtscore zusammengefasst, anhand dessen die jährliche Vergütung berechnet wird. Sowohl die die Verbesserung des Krankenhauses als auch die Rankingposition gehen in die Bewertung mit ein. Insgesamt entsteht so eine sehr detaillierte aber auch komplizierte Berechnung der variablen Vergütung, wie bereits ein Beispiel der Bewertung des Prozessindikators „Fibrinolytic Therapy“ bei akutem Myokardinfarkt zeigt (siehe Abbildung 18):

Abbildung 18: Berechnungsbeispiels für den Indikator „AMI-7a - Fibrinolytic Therapy“



Quelle: (HHS - CMS 08.05.2011, S. 26521); siehe auch Amelung/Zahn 2009, S. 29; eigene Übersetzung

Bereits im ersten Projektjahr werden durch diese Vergütung 850 Mio. Dollar an 3.500 Krankenhäuser in den USA umverteilt, ab 2017 werden sogar 1,9 Mrd. Dollar Vergütung erwartet – die Initiative ist budgetneutral konzipiert (vgl. HHS - CMS 08.05.2011, S. 26543; Ryan/Blustein 2012, S. 1558). Dieses Programm wird in den USA sehr detailliert diskutiert, eine Vielzahl von Informationsmaterial und Diskussionsforen stehen den Krankenhäusern zu Verfügung – gleichzeitig ist deutlich, dass derzeit weniger die Güte und Umsetzbarkeit der Initiative sondern vielmehr Optionen zur Qualitätsverbesserung bspw. auf Kongressen thematisiert werden (bspw. PfP-Summit, http://www.pfpsummit.com/past7/index.html).

Drei weitere Initiativen wurden in den letzten Jahren vom CMS eingeführt, die ebenso eine Qualitätsorientierung aufweisen (vgl. Buhlman/Matthes 2012, S. 4ff.):

- Im „*Readmissions*“ Programm bezieht sich der qualitätsabhängige Vergütungsanteil auf Wiederaufnahmen nach Herzversagen, AMI und Pneumonie. Sofern ein Krankenhaus die risikoadjustierte, erwartete Wiederaufnahmerate überschreitet, werden die DRG-Zahlungen gekürzt und zwar ab 2013 um 1 %, ab 2014 um 2 % und ab 2015 sogar um 3 % der Gesamtzahlungen.

- Die Initiative der „*Hospital-acquired conditions (HAC)*“ umfasst 8 Zustände, die im Krankenhaus erworben werden und deren zusätzlichen Kosten bereits derzeit nicht vergütet werden. Da dies aber nur recht insignifikante Beträge betrifft, werden ab 2015 1 % der DRG-Zahlungen reduziert, wenn das Krankenhaus eine hohe Rate von HACs aufweist (oberes Quartil).

- Neben diesen Initiativen gibt es die bereits seit 2009 etablierte Regelung des „*Meaningful uses*“. Krankenhäuser werden dazu aufgefordert, zertifizierte Datenverarbeitungsprogramme zu nutzen, die anhand von elektronischen Patientendaten klinische Qualität darstellen und zeitnah messen können. So wird beispielsweise die Nutzung von Systemen gefördert, die Prozessindikatoren direkt am Krankenbett nicht nur messen, sondern aktiv auf Noch-nicht-Erfüllung von Leitlinien hinweisen. So kann die leitliniengerechte Behandlung unterstützt und gleichzeitig dokumentiert und gemessen werden. Für diese Initiative kann das Krankenhaus bis zu 1 % der DRG-Zahlungen erhalten.

Insgesamt sind ab 2015 demnach 7 % der Medicare-Zahlungen von der Krankenhausqualität abhängig. Diese 7 % beziehen sich nicht auf die Gesamteinnahmen von Krankenhäusern. Da allerdings Medicare der größte, staatliche Versicherer ist, betreffen diese Qualitätsinitiativen trotzdem einen hohen Anteil des Einkommens. Außerdem ist zu beobachten, dass auch weitere Krankenversicherungen die gleichen Qualitätsindikatoren und Vergütungssysteme anwenden (band-wagon-effect), so dass in einigen Regionen bereits bis zu 7 % des Gesamteinkommens qualitätsvariabel sind (vgl. Buhlman/Matthes 2012, S. 3).

Die Initiativen der CMS werden alle wissenschaftlich begleitet, sodass detaillierte Evidenzstudien erwartet werden. Sobald diese vorliegen, können auch Erkenntnisse für die deutsche Krankenhauslandschaft konkretisiert werden. Es zeigt sich, dass im angelsächsischen Raum und vor allem in den USA, das Thema der qualitätsorientierten Vergütung sowohl in der wissenschaftlichen Evaluation vor allem aber in der praktischen Umsetzung weit entwickelt ist.

5.2. P4P-Ansätze in Deutschland: Gesetzliche Rahmenbedingungen und IV-Verträge mit Qualitätsgarantien

In Deutschland wird spätestens seit dem SVR-Gutachten 2007 über P4P diskutiert (vgl. SVR 2007). Allerdings gibt es bislang nur erste implementierte Pilot-Ansätze; bundesweite Projekte sind kaum möglich, gerade im stationären Sektor. Die derzeitige Gesetzeslage sieht variabel, qualitätsorientierte Vergütung in der kollektiven Vergütung nicht vor. Es gibt selektivvertragliche Projekte, deren Funktionsweise kurz dargestellt wird.

5.2.1. Rahmenbedingungen der stationären Vergütung

Seit 2004 werden in akut-somatischen Krankenhäusern Behandlungsleistungen nach dem DRG-System abgerechnet und nicht mehr nach Tagessätzen (vgl. Gesetz zur Einführung des diagnose-orientierten Fallpauschalensystems für Krankenhäuser – Fallpauschalengesetz, 23. April 2002). Hierfür werden die Behandlungsfälle in verschiedene „Diagnosis-related groups“ kategorisiert und mit einem Zahlungsschlüssel (Basisfallwert) vergütet, der Basisfallwert wird seit 2010 vom InEK berechnet (vgl. u. a. Geissler et al. 2011b, S. 250).

Während eines der Ziele der DRG-Einführung in Deutschland die Förderung der Effizienz, Transparenz und Qualität im Krankenhaussektor war (vgl. Rau 2002, S. 498), werden im deutschen DRG-System keine expliziten Qualitätsanpassungen durchgeführt (vgl. Geissler et al. 2011b, S. 262). Einige Aspekte des DRG-Systems haben aber trotzdem einen Effekt auf die stationäre Qualität, so bspw. die Fallzusammenführung (vgl. Hertle/Veit 2012, S. 40). Welche Qualitätskonsequenzen die Einführung des DRG-Systems hatte, wurde bereits sehr früh und seitdem wiederholt untersucht, so unter anderem im Projekt „Wandel von Medizin und Pflege im DRG-System (WAMP)“ vom „Wissenschaftszentrum Berlin für Sozialforschung (WZB)“ (vgl. Buhr/Klinke 2006, S. 3). Insgesamt zeigt sich auch in späteren Studien, dass die Qualität der Versorgung nach der Einführung des DRG-Systems zumindest nicht gesunken ist (vgl. Sens et al. 2010, S. 25), auch wenn es kaum ausreichende Evidenz in Europa zu Qualitätsveränderungen nach Einführung der DRG-Systeme zu geben scheint (vgl. Or/Häkkinen 2011, S. 118ff.).

Das deutsche Gesundheitssystem weist einige Defizite, darunter Wettbewerbsdefizite, auf (vgl. Hartweg 2007, S. 38ff.) Diese können beispielsweise entstehen, da die Akteure im Gesundheitswesen meist verbandschaftlich organisiert sind und diese „kooperative Koordination“ gewisse Inflexibilität nach sich zieht, unter anderem die Schwierigkeit der Selbstverwaltung, Kompromissentscheidungen zu treffen (vgl. SVR 2005, S. 32ff.). Daher empfiehlt der Sachverständigenrat, durch selektive Vertragsverhandlungen den Wettbewerb im Gesundheitswesen zu dezentralisieren (vgl. SVR 2005, S. 32ff.). Zusätzlich erlauben die selektivvertraglichen Systeme mehr Gestaltungsfreiheit und daher auch eine konkrete Qualitätsvariabilität.

5.2.2. Verträge der Integrierten Versorgung - Beschreibung und Bestandsaufnahme

Verträge der Integrierten Versorgung nach Paragraphen §§ 140a-d SGB V wurden in Deutschland im Jahr 2004 durch das GKV-Gesundheitsmodernisierungsgesetz (GMG) deutlich gestärkt (vgl. BMG 14.11.2003). Ziel der Gesetzgebung war eine bessere Zusammenarbeit zwischen den verschiedenen Leistungsbereichen (vgl. Begründung zum GKV-Modernisierungsgesetz 2003). Diese Selektivverträge ermöglichen die Umsetzung eigener, innovativer Vorstellungen der Vertragspartner bezüglich der definierten Leistungen und der Vergütung – „privatrechtliche“ Einzelverträge zwischen Krankenkassen und Leistungserbringern wurden somit eingeführt (vgl. Henke et al. 2004, S. 21). Außerdem sind die

Vertragsteilnehmer zu einer „qualitätsgesicherten, wirksamen, ausreichenden, zweckmäßigen und wirtschaftlichen Versorgung der Versicherten“ verpflichtet (GKV-Modernisierungsgesetz 2003, BMG 14.11.2003, § 140). Es ergeben sich damit weitreichendere Umsetzungsoptionen, die auch eine qualitätsorientierte Vergütung ermöglichen. Diese Umsetzungsoptionen machen IV-Verträge zum möglichen Untersuchungsobjekt für diese Arbeit. Von besonderem Interesse sind IV-Verträge vor allem, weil die Vertragspartner zu „Einkäufern“ werden und eine größere Vertragsfreiheit besteht (vgl. Amelung/Zahn 2009, S. 10).

Allerdings muss zunächst festgehalten werden, dass IV-Verträge als Zielsetzung nicht primär die Verbesserung der Behandlungsqualität haben. Entwicklungsgrund für den Gesetzgeber war zunächst die verbesserte Zusammenarbeit der verschiedenen Sektoren, doch auch Wirtschaftlichkeits- und Qualitätsaspekte sind inkludiert: „Regelung [müssen] dem Sinn und der Eigenart der integrierten Versorgung [entsprechen], die Qualität, die Wirksamkeit und die Wirtschaftlichkeit der integrierten Versorgung [verbessern]“ (§140b (4) SGB V). Um dieses neue Gestaltungselement möglichst schnell umzusetzen, wurde zudem eine Anschubfinanzierung vereinbart: Ein Budget von 1% der Gesamtvergütung wurde nach Kürzung durch die Krankenhäuser und Kassenärztlichen Vereinigungen der Finanzierung von IV-Verträgen von 2004 bis 2008 zur Verfügung gestellt (vgl. Grothaus 2009, S. 101; § 140d SGB V).

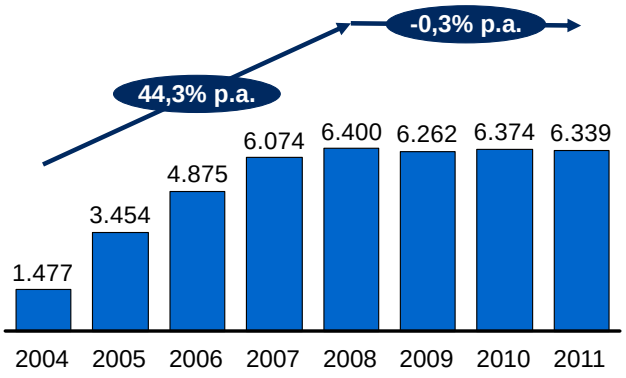
Aus Sicht der beteiligten Systempartner ist die Zielsetzung der Vertragsabschließung vor allem in folgenden Punkten zu sehen (vgl. vor allem die Ergebnisse der Fallstudien):

1. Nutzung Anschubfinanzierung
2. Rabattoptionen für die Krankenkasse
3. Fallzahlerhöhung für den Leistungserbringer
4. Vermarktungs-/Marketingoptionen bei Patienten und Versicherten

Ob unter diesen Zielsetzungen auch ein inhärentes Interesse an einer Qualitätsverbesserung zu finden ist, kann als eine Kernfrage dieser Arbeit gesehen werden. Nur wenn die Qualitätsverbesserung auch als wichtiges Ziel anerkannt wird, dass anhand des Instruments des IV-Vertrages umgesetzt werden kann, kann es aufgrund dieses Vertrages zu Verhaltensveränderungen und somit zu Verbesserungen kommen.

Während die Anzahl der IV-Verträge von 2004 bis 2008 deutlich anstieg, ist sie seit dem leicht rückläufig (vgl. Grothaus 2009, S. 15; SVR 2012, S. 345; siehe auch Abbildung 19). Dies ist vor allem mit dem Auslaufen der Anschubfinanzierung zu erklären. Rund 30 % der Krankenkassen haben 2011 weniger IV-Verträge als im Jahr 2008, nur bei den Betriebskrankenkassen haben sich die Anzahl der Verträge vervielfacht (vgl. SVR 2012, S. 345).

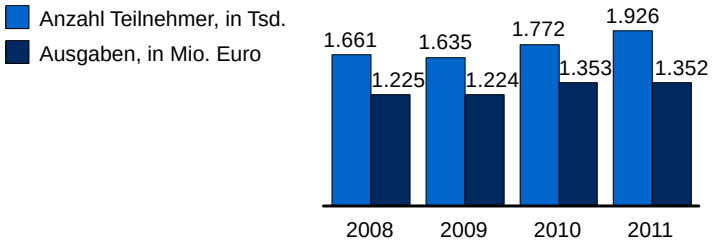
Abbildung 19: Anzahl IV-Verträge in Deutschland, 2004-2011



Quelle: Eigene Darstellung, Daten vgl. Grothaus 2009, S. 15; SVR 2012, S. 345

Die Anzahl der Teilnehmer und Ausgabenhöhe dagegen blieb von 2008 bis 2011 relativ konstant, nahm sogar leicht zu (vgl. SVR 2012, S. 345; siehe Abbildung 20). Die mittlere Teilnehmerzahl pro Vertrag liegt bei 249 Versicherten, dies variiert aber zwischen der Allgemeinen Ortskrankenkasse mit 2.141 und den BKK mit 63 Versicherten pro Vertrag deutlich (vgl. SVR 2012, S. 348). Insgesamt werden von den Ausgaben für die IV-Verträge nur 45 % für die Krankenhausbehandlung ausgeben, dies entspricht aber dem höchsten Ausgabenblock (vgl. SVR 2012, S. 350f.).

Abbildung 20: Teilnehmer von und Ausgaben für IV-Verträge, Deutschland, 2008-2011



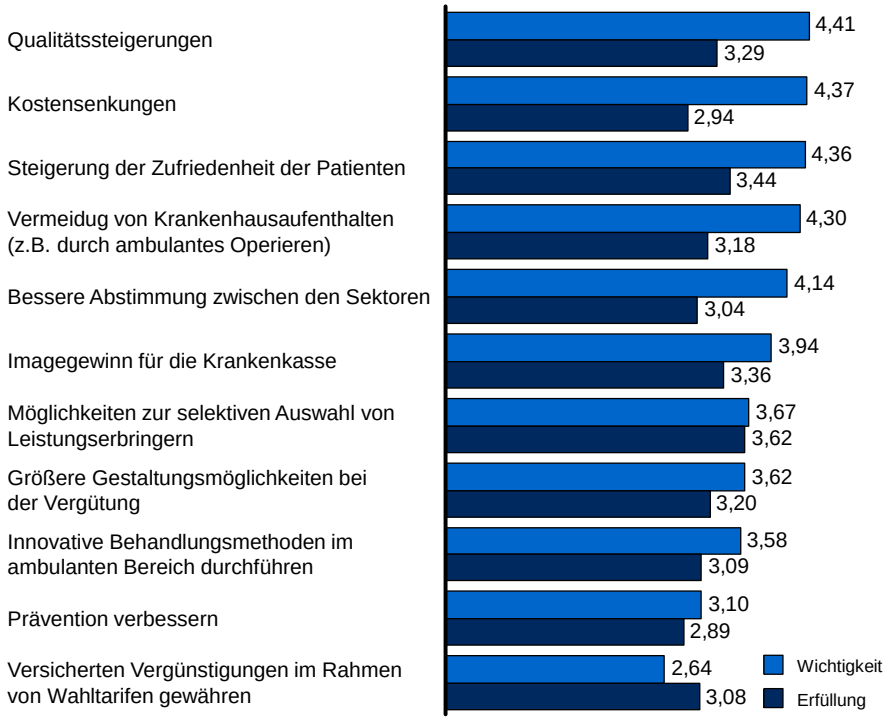
Quelle: Eigene Darstellung, Daten vgl. SVR 2012, S. 345

Die derzeitige Evaluation der IV-Verträge wird auch in Umfragen schon als sehr gering beschrieben: Nur 4,9 % der Krankenkassen evaluieren die Verträge immer, 17,1 % nie. Diese Evaluationen werden auch dann nur von 10,1 % der Krankenkassen zumindest meistens veröffentlicht. Dementsprechend sind die Möglichkeiten zu wissenschaftlichen Aussagen der IV-Verträge weiterhin sehr gering (vgl. SVR 2012, S. 352). Dies erklärt zusätzlich die Notwendigkeit zu einem qualitativen Forschungsansatz dieser Arbeit – ausreichend quantitative Daten liegen nicht vor.

Die wichtigsten Gründe für die Beendigung von IV-Verträgen sind wirtschaftlicher Natur und erklären sich vorwiegend durch zu hohe Kosten, zu geringe Teilnehmerzahl, Mengenausweitung durch den Leistungserbringer und zu lange Amortisationsdauer (vgl. SVR 2012, S. 355). Gleichzeitig konnte die

Erwartungen an die Verträge – die hier auch Qualitätssteigerungen beinhaltet – nur unzureichend erfüllt werden (siehe Abbildung 21).

Abbildung 21: Wichtigkeit und Erfüllung einzelnen Aspekte für die Teilnahme an IV-Verträgen, Umfrageergebnisse von 84 Krankenkassen



Quelle: SVR 2012, S. 359

Insgesamt zeigt sich, dass 71,6 % der Krankenkassen derzeit die Qualität für die Versicherten in der IV-Versorgung als höher einschätzen als in der Regelversorgung (vgl. SVR 2012, S. 361). Quantitative Daten hierfür liegen bislang kaum vor. Im empirischen Teil dieser Arbeit soll unter anderem diese Aussage überprüft werden, mit einem Fokus auf IV-Verträgen mit qualitätsorientierter Vergütung – d. h. vor allem mit Gewährleistungsklauseln.

5.2.3. Besonderheit von qualitätsorientierter Vergütung in IV-Verträgen - Beschreibung und Gestaltungsoptionen

Aufgrund der Möglichkeit eine vertragsspezifische Vergütung zu vereinbaren, sind auch qualitätsorientierte Vergütungen in IV-Verträgen eine Option. Während für die meisten Aspekte der gesetzlichen Qualitätssicherung Anforderungen der Zugangsqualität definiert sind, gehen einige Optionen zur qualitätsorientierten Vergütung auch der Ergebnisqualität darüber hinaus: nämlich die gesetzlichen Möglichkeiten nach §§73a-c SGB V (Strukturverträge §73a, hausarztzentrierten Versorgung §73b und Selektivverträge §73c) sowie die integrierte Versorgung nach §§140ff SGB (vgl. Meyer-Hofmann et al.

2012, S. 17). Gerade die Verträge der integrierten Versorgung ermöglichen einen besonders weiten Spielraum für Qualitätsorientierung, sowohl für variable Vergütung als auch Ergebnis- und Prozessqualität (vgl. Meyer-Hofmann et al. 2012, S. 16).

Es existieren im deutschen Gesundheitswesen derzeit verschiedene Gestaltungsvarianten für eine qualitätsorientierten Vergütung in Verträgen der integrierten Versorgung. Eine Auflistung aller IV-Verträge mit qualitätsorientierter Vergütung existiert nicht. Die ersten Ansätze von P4P in Deutschland fanden im ambulanten Sektor statt – bspw. von der AOK anhand der QISA-Qualitätsindikatoren oder von der BKK für Fertilisationszentren (vgl. Scheppach et al. 2011, S. 2ff.). Im letzten seit 2006 existierenden Projekt werden die Ärzte zusätzlich zu einer rabattierten Leistungsvergütung für den Eingriff der künstlichen Befruchtung bei einer gesunden Schwangerschaft mit einem Bonus belohnt.

Einige recht innovative Ausgestaltungen, die sich auf die stationäre, akut-somatische Versorgung beziehen, sind in Tabelle 9 beschrieben und wurden im BQS-Gutachten diskutiert[1]. Für das Gutachten wurde eine umfangreiche webbasierte Umfrage zur Meldung von P4P-Projekten durchgeführt, so dass eine recht ausführliche Liste entstanden ist. Dass diese Arbeit noch weitere Verträge mit qualitätsorientierten Vergütungsanteile aufzeigen kann, zeugt von der großen Variabilität dieses Themas.

Tabelle 9: Auswahl P4P-Projekte in Deutschland

Kasse	Projektname	Beschreibung
TK	IV-Vertrag Kopfschmerz	➢ Vertrag mit Schmerzklinik Kiel, Universitätsklinikum sowie niedergelassenen Ärzten ➢ Bonus-Malus-Regelung anhand von Komplexpauschalen: ➢ Bonus bei Arbeitsfähigkeit nach 8 Wochen sowie weniger als 7 Tagen Arbeitsunfähigkeit in 6 Monaten, Malus bei keiner Arbeitsfähigkeit nach 8 Wochen ➢ Ergebnis: Bonuszahlung bei 81,6 % der Fälle
TK	IV-Vertrag Rücken-schmerz	➢ Teilnahme von über 30 Standorten in Deutschland ➢ Bonuszahlung von 10 % bei Schmerzfreiheit innerhalb von 4 Wochen sowie Maluszahlung bei Nichterfolg von 5 % des Honorars ➢ Ergebnis: Teilnehmer kehren im Schnitt 72 Tage früher an den Arbeitsplatz zurück
AOK Hessen	IV-Verträge Schlaganfall, Endopro-thetik	➢ Rein stationäre Ausrichtung ➢ Wichtigste Qualitätsindikatoren 1-Jahres-Wiederaufnahme aber auch 30-Tages-Sterblichkeit ➢ Zuschläge durch qualitätsabhängige Reduktion der vereinbarten Rabatte

Quelle: Veit et al. 2012b, S. 40ff.

[1] Mit beiden hier genannten Krankenkassen bzw. Krankenkassenverbänden wurde im Rahmen der empirischen Untersuchung Kontakt aufgenommen und erste Diskussionen geführt. Leider konnte aus Zeit- und Datenschutzgründen seitens der Kostenträger keine weitere Zusammenarbeit stattfinden, so dass diese Arbeit andere, teilweise vergleichbare Projekte analysiert. Der Exkurs der Fallstudie 5 beschreibt eine Weiterentwicklung eines der hier beschriebenen Verträge. Fallstudie 3 beschreibt ebenso einen im Gutachten diskutierten Vertrag, allerdings nur am Rande, da er keine stationären Fälle beinhaltet.

Einige dieser Pilot-P4P-Projekte wurden auch bereits von wissenschaftlichen Studien evaluiert – so zeigen die zwei Schmerzverträge der TK Verbesserungen vor allem im Bereich der Kostensenkung auf. Diese sind sowohl mit den P4P-Ansätzen als auch mit dem Public Reporting zu erklären (vgl. Klusen et al. 2011, S. 111ff.).

Auch Gewährleistungsklauseln könne als eine Art der qualitätsvariablen Vergütung angesehen werden (vgl. Malzahn 2010, S. 4), da es hier qualitätsabhängig – nämlich bei Eintreten eines Gewährleistungsfalles, bspw. einer Revisionsoperationen – zu einer Nicht-Erstattung der Leistung und somit einem Malus kommt. Dieser Art der Qualitätsgarantien wird in einer Vielzahl von IV-Verträgen umgesetzt, gerade in der Anfangszeit der IV-Verträge hatten fast alle Verträge bestimmter Arten (vor allem Endoprothetik und Kardiochirurgie) Gewährleistungsklauseln (vgl. Ergebnisse der Fallstudien).

Trotz der großen Verbreitung dieser Umsetzungsform werden teilweise Zweifel an der Rechtmäßigkeit von Gewährleistungsklauseln ausgedrückt (bspw. bei Silber 2006, S. 37 und Schramm 2006, S. 33). Hier wird vor allem darauf hingewiesen, dass das ärztliche Handeln nicht als Werk- sondern als Dienstvertrag zu verstehen ist und daher eine verschuldensunabhängige Haftung für das Eintreten nicht vollständig beeinflussbarerer Ereignisse – wie in der Gewährleistung üblich – nicht durchgeführt werden sollte. Zumindest sollten Gewährleistungsklauseln in ihrem Umfang und Zeitfrist begrenzt werden (vgl. Schramm 2006, S. 33). Eine rechtliche Prüfung dieser Zweifel und des Sachverhaltes kann hier nicht abschließend durchgeführt werden, es wird aber auf den Abschnitt 2.2.3.1 verwiesen – in juristischen Abhandlungen wird die Nutzung von Gewährleistungsklauseln in IV-Verträgen nicht ausgeschlossen, Leistungserbringer werden aber davor gewarnt, dass sie hierdurch gefährdet sind (vgl. Möller et al. 2008, S. 549f.). Da gleichzeitig eine Vielzahl von Gewährleistungsklauseln existieren und weiterhin neu abgeschlossen werden und diese Verträge auch von den juristischen Abteilungen aller Vertragspartner geprüft wurden, soll hier zunächst von der Rechtmäßigkeit dieser Klauseln ausgegangen werden.

Im Gegensatz zu den innovativeren Vertragsformen von qualitätsorientierter Vergütung, sind IV-Verträge mit Gewährleistungsaspekten schon länger und verbreiteter eingeführt. Dies ermöglicht eine detaillierte Analyse auch über den Projektbeginn hinaus. Daher wurden diese Verträge ausgewählt, um daran im Folgenden im Rahmen der Fallstudien den Kriterienkatalog anzuwenden. Sofern andere Arten von qualitätsorientierter Vergütung (bspw. Bonusverträge oder interne Anreizmechanismen) vorlagen, wurden diese analysiert.

5.3. Methodische Ansätze für die Bewertung der Umsetzungsbeispiele

5.3.1. Methodische Argumentation für den Fallstudienansatz

Fallstudien sind ein in der Sozialwissenschaft häufig verwendeter Ansatz der qualitativen Forschung. Im Gegensatz zu quantitativen Ansätzen stehen hierbei die Wahrnehmungen, Interpretationen und

Bedeutungen von möglichen Fragestellungen im Vordergrund (vgl. Wrona 2005, S. 4). Dieser Ansatz ist zur Anwendung des Kriterienkatalogs an deutschen Umsetzungsbeispielen besonders angebracht und kann die fünf Grundsätze der qualitativen Sozialforschung (vgl. Mayring 2002) wie folgt umsetzen:

- *Subjektbezogenheit der Forschung:* Ausgangspunkt der zu analysierenden Umsetzungsbeispiele sind neben der Organisation vor allem die an den IV-Verträgen beteiligten Mitarbeiter der Krankenkassen und Krankenhäuser. Neben der quantitativen Analyse der Verträge sind vor allem die Einschätzung der Wirksamkeit der Anreizeffekte dieser Akteure bedeutsam für diese Arbeit
- *Deskription*: Zu Beginn jeder Fallstudie werden die IV-Verträge, die beteiligten Akteure sowie die Umsetzung der Anreizsysteme in den Organisationen (d. h. Kassen und Krankenhäusern) genau beschrieben um anschließend die Zusammenhänge analysieren zu können.
- *Interpretatives Paradigma*: Schon die Einführung von IV-Verträgen mit Qualitätsgarantien kann aus verschiedensten Beweggründen bei den unterschiedlichen Akteuren zustande kommen. Der Untersuchungsgegenstand dieser Entscheidungen ist daher weitgehend offen (im Gegensatz zur quantitativen Bewertung z. B. der Qualitätsverbesserungen)
- *Alltägliche Umgebung*: Die Entscheidungen der Akteure werden entweder durch Erzählungen in den Interviews oder direkte Beobachtungen bspw. der Controlling-Mechanismen und verwendeten Software gefällt.
- *Verallgemeinerung*: Die Generalisierung der Ergebnisse kann nicht automatisch vollzogen werden.

Die Nutzung eines Fallansatzes ist für diesen Teil der Arbeit aus verschiedenen Gründen besonders geeignet:

- Durch die Untersuchung der Fallstudien kann sehr detailliert auf die Fragestellungen der verschiedenen Organisationen eingegangen werden.
- Gerade die hohe Komplexität des Untersuchungsgegenstandes (der IV-Verträge, den beteiligten Organisationen und deren Zusammenspiel) machen eine qualitative Fallstudie sinnvoll und notwendig.
- Neben dem hohen Detailgrad ist durch den Fallstudienansatz eine umfassende Analyse aller relevanten Aspekte möglich.
- Es werden verschiedene Datenerhebungsstrategien genutzt: Eine quantitative Analyse sowie vor allem Interviews und Beobachtungen

Um die Aussagekraft der Ergebnisse zu erhöhen wird ein multipler Fallstudienansatz gewählt (vgl. Yin 2003). Drei Krankenkassen und vier Krankenhäuser mit jeweils bis zu drei unterschiedlichen IV-Verträgen

werden analysiert. Dadurch ist eine Triangulation der Ergebnisse möglich und anschließend eine Annäherung an allgemeingültige Aussagen.

Der gewählte Fallstudienansatz erlaubt allerdings keine Generalisierung. Durch die Analyse verschiedener Untersuchungsobjekte können breitere Schlüsse gezogen werden, Verallgemeinerungen sind jedoch nicht möglich. Der derzeitige Datenbestand des Untersuchungsobjektes „IV-Verträge mit Qualitätsgarantien“ ermöglicht keinen gangbaren Weg, alle Verträge statistisch auszuwerten. Daher kann der Fallstudienansatz bereits eine Vielzahl neuer Erkenntnisse hervorbringen. Sobald die Datensituation eine quantitativere Auswertung erlaubt, sollte diese auch vorgenommen werden.

5.3.2. Methodische Entscheidung für leitfadengestütztes Interview

Für die qualitative Sozialforschung gibt es eine Reihe von möglichen Interviewarten, die für verschiedene Fragestellungen angewandt werden (vgl. King/Horrocks 2010 oder Wrona 2005). Eines der Differenzierungsmerkmale ist das Vorhandensein eines theoretischen Konzeptes und von Hypothesen für die Befragung.

In dieser Arbeit soll der Kriterienkatalog angewandt und die Nutzbarkeit überprüft werden. Daher gibt es bereits deduktiv erarbeitete Konzepte und Hypothesen vor den Interviews. Gleichzeitig jedoch sollen die Ergebnisse der Interviews noch induktive Schlussfolgerungen ermöglichen. Um dieses Zusammenspiel aus induktiven und deduktiven Aspekten zu erreichen, ist vor allem ein problemzentriertes oder leitfadengestütztes Interview (PZI) geeignet (vgl. Witzel 2000).

Auch der Aspekt der Standardisierungsmöglichkeit spricht für ein problemzentriertes Interview. Während diese Interviewart zielorientiert fragt und z. B. anhand eines Interview-Leitfadens strukturiert sein kann, verbleibt bei der Durchführung eine gewisse Flexibilität (vgl. Lamnek 1995, S. 91).

Tabelle 10: Formen und Merkmale qualitativer Interviews

Methodologische Prämissen	narratives Interview	problem-zentriertes Interview	fokussiertes Interview	Tiefeninterview	rezeptives Interview
Offenheit	völlig	weitgehend	nur bedingt	Kaum	völlig
Kommunikation	erzählend	zielorientiert fragend	Leitfaden	fragend/ erzählend	erzählend/ beobachtend
Prozesshaftigkeit	gegeben	gegeben	nur bedingt	gegeben	gegeben
Flexibilität	hoch	Relativ hoch	relativ gering	relativ hoch	bedingt
Explikation	ja	ja	ja	ja	bedingt
theoretische Voraussetzungen	relativ ohne	Konzept vorhanden	weitgehendes Konzept	Konzept vorhanden	relativ ohne; nur Vorverständnis
Hypothesen	Generierung	Generierung; Prüfung	eher Prüfung; auch Generierung	eher Prüfung; auch Generierung	Generierung; Prüfung
Perspektive der Befragten	gegeben	gegeben	bedingt	bedingt	absolut

Quelle: Lamnek 1995, S. 91

Für die problemzentrierten Interviews werden drei Instrumente in dieser Arbeit benutzt (vgl. Witzel 2000): Leitfaden, Tonaufzeichnung des Gesprächs sowie Postskripte als Ergänzung zur Tonbandaufzeichnung.

5.3.2.1. Beschreibung des Interview-Leitfadens

Der Interview-Leitfaden bietet eine „Hintergrundfolie“ als Orientierungsmöglichkeit für den Interviewer (vgl. Wrona 2005, S. 26) – im Gegensatz zum strukturierten, standardisierten Fragebogen, der konkrete Fragen oder sogar Antwortmöglichkeiten bietet. Leitfäden enthalten vor allem Fragekategorien und unterschiedliche Fragearten (so z. B. Leitfragen, Aufrechterhaltungsfragen und konkrete Nachfragen, vgl. Helfferich 2005).

Auch für den hier vorliegenden Interviewleitfaden (zu finden im Anhang) wurden Leitfragen und Fragekategorien gebildet. Hierfür wurde der „SPSS-Methode“ nach Helfferich genutzt (Helfferich 2005):

Tabelle 11: „SPSS-Methode“ der Leitfaden-Erstellung

S	**Sammlung** von möglichen Fragen anhand eines möglichst offenen Brainstormings
P	**Prüfung** der Geeignetheit der Fragen und möglicherweise Streichung von ungeeigneten Fragen
S	**Sortierung** der Fragen anhand von Fragearten (Leitfaden, Aufrechterhaltungsfragen, konkrete Nachfragen) und Inhalten/Themen
S	**Subsumierung** der Fragen in Kategorien und somit Erstellung eines strukturierten Leitfadens

Quelle: Helfferich 2005, S. 9

Der Leitfaden soll die Anwendung des Kriterienkatalogs auf bestehende IV-Verträge unterstützen. Allerdings ist bereits im Vorfeld erkennbar, dass ein beträchtlicher Teil des Katalogs nicht oder nur wenig anwendbar sein wird – vor allem aufgrund der anderen Zielsetzung und geringeren Reichweite der Verträge im Vergleich zu anderen qualitätsorientierten Vergütungssystemen. Daher sind die Konzepte des Kriterienkatalogs sicherlich ein erster Ansatzpunkt für den Leitfaden – die konkreten Fragestellungen müssen jedoch angepasst werden.

Gleichzeitig sollen weitere qualitative Aussagen bezüglich der IV-Verträge möglich bleiben und die Fragestellungen somit über den Kriterienkatalog hinausgehen. Daher sind einige Fragen, die sich aus der derzeitigen Umsetzung der IV-Verträge ergeben, relevant und können indirekt Aufschluss bezüglich einiger Kriterien geben – ohne direkt aus dem Katalog herleitbar zu sein.

Für die Sammlung möglicher Leitfaden-Fragen sind somit zwei Richtungen zu erkennen: Der Kriterienkatalog und die realen Gegebenheit der IV-Verträge. Ein wichtiger Bestandteil des Leitfadens ist neben den offenen Fragen auch eine Überprüfung der vorher durchgeführten Datenanfrage. Für die vorliegende Untersuchung der IV-Verträge anhand des Kriterienkatalogs werden in einer Fallstudie quantitative Daten erhoben. Diese Erhebung wurde in der Organisation der Krankenkassen bei unterschiedlichen Personenkreisen angefragt – vor allem beim Vertragscontrolling aber auch bei den jeweiligen Sachbearbeitern des IV-Vertrages. Die beteiligten Personen, die diese Datenanfrage bearbeiten, waren auch jeweils Interview-Partner. Daher werden in den Interviews und anhand des Leitfadens die Datenerhebungen mit den Interviewten diskutiert – nicht allerdings die Ergebnisse der Datenanalyse. Wenn dies im Laufe des Dissertationsvorhabens notwendig war, wurden weitere (teilweise auch telefonische) Interviews durchgeführt.

Die Interviewleitfäden sind unterteilt nach den jeweiligen Interviewpartnern – sowohl bei Krankenkassen als auch bei Krankenhäusern. In diesen Organisationen sind jeweils unterschiedliche Mitarbeiter für die Umsetzung der IV-Verträge zuständig. Für jede dieser relevanten Positionen im Unternehmen wird ein Interview-Leitfaden entwickelt. Das ermöglicht sowohl eine detaillierte Untersuchung als auch einen Vergleich zwischen den Unternehmen.

5.3.3. Auswertungsverfahren nach Mayring

Im Anschluss an die Interviews wurden diese zunächst möglichst detailgenau transkribiert, um im Anschluss die wörtliche Textfassung des Interviews auswerten zu können. Dabei wird allerdings bereits vermehrt auf den Inhalt – nicht etwa auf das Verhalten des Interviewten geachtet (vgl. Langer 2010). Für die Auswertung wird die qualitative Inhaltsanalyse nach Mayring gewählt, deren Ziel es ist „… das Material so zu reduzieren, dass die wesentlichen Inhalte erhalten bleiben, durch Abstraktion einen überschaubaren Corpus zu schaffen, der immer noch Abbild des Grundmaterials ist“ (Mayring 2010a, S. 65). Bei der qualitativen Inhaltsanalyse finden drei Grundtechniken statt (vgl. Mayring 2010a):

1. Zusammenfassung
2. Explikation
3. Strukturierung

Für die Zusammenfassung muss hier zunächst ein Kategoriensystem entweder induktiv oder deduktiv erarbeitet werden (vgl. Mayring 2010b und Lamnek 2005). Da für diese Arbeit ja bereits ein Kriterienkatalog mit verschiedenen Skalen erstellt wurde und anhand der Interviews „abgefragt“ werden soll, ist das deduktive Kategoriensystem für die Auswertung bereits vorhanden. Die Auswertung findet demnach „entlang“ des Kriterienkataloges statt. Ggfs. auftretende zusätzliche Themen werden dann gesondert analysiert. Dadurch ist auch die Strukturierung inhaltlicher Natur und wird an die Kriterien angepasst (vgl. Mayring 2002).

Anhand der Auswertung der Interviews sollen am Ende folgende Ziele erreicht werden:

- Der Kriterienkatalog soll auf die diskutierten IV-Verträge und somit Beispiele von qualitätsorientierten Vergütungssystemen angewandt werden können (soweit möglich und vorhanden).
- Die Umsetzung von solchen Vergütungssystemen soll detailliert beschrieben und bezüglich derzeitiger Hindernisse diskutiert werden.
- Erste Ideen für neue Entwicklungen der jeweiligen Beteiligten an möglichen Systemen soll ergebnisoffen besprochen werden.

5.3.3.1. Methodische Beschreibung der Projektauswahl

Die Projekte für diesen empirischen Teil der Dissertationsarbeit wurden nach „maximaler Kontrastierung“ ausgewählt. Im Rahmen eines theoretischen Samplings wurden die beteiligten Krankenkassen und Krankenhäuser ausgewählt (vgl. Glaser/Strauss 1998; Wrona 2005, S. 23). Es wurde versucht, möglichst viele unterschiedliche Arten und Größen von Krankenkassen und Krankenhäusern zu erreichen.

Durch die Kontrastierung der Kassengröße wurden neben der Versichertenzahl auch unterschiedliche Professionalisierungsgrade betrachtet – inwieweit sich dies auch im Grad der Operationalisierung der IV-Verträge und qualitätsorientierten Aspekte widerspiegelt, wird im weiteren Verlauf diskutiert. Insgesamt wurden drei Krankenkassen in die Analyse mit aufgenommen. Zwei Ersatzkassen und eine Betriebskasse. Aus Gründen des Datenschutzes werden die Namen der Krankenkassen nicht genannt.

Für die Krankenhäuser wurde ebenso eine maximale Kontrastierung durchgeführt und zwar auf Grundlage der Trägerart einerseits und der Krankenhausgröße andererseits. In Deutschland stehen drei Trägerarten (öffentlich, privat, freigemeinnützig) zur Verfügung. Alle drei sind in den Fallstudien vertreten. Gleichzeitig wurden sowohl größere Klinikverbünde als auch einzelne, kleine Krankenhäuser mit aufgenommen. Ein weiteres Auswahlkriterium war das Vorhandensein eines IV-Vertrages. Dies wurde in einem kurzen, telefonischen Vorgespräch geklärt.

Nach der Auswahl der beteiligten Organisationen wurden die zu bewertenden IV-Verträgen im Rahmen der Interviews ausgewählt und diskutiert. Auch hierbei wurde versucht, möglichst unterschiedliche Größen und Charakteristika (wie bspw. Indikation und beteiligte Fachabteilungen) der IV-Verträge abzudecken. Da allerdings gerade in der stationären Versorgung selbst große Maximalversorger häufig nur ein oder zwei IV-Verträge haben, musste diese maximale Kontrastierung teilweise gering ausfallen. In diesen Fällen wird vor allem die mögliche Umsetzung von qualitätsorientierten Vergütungsaspekten oder etablierte Qualitätsmanagementsystem anhand der Interviews diskutiert.

5.4. Beschreibung der Fallstudien: Derzeitige Umsetzung von P4P-Systemen in Deutschland[1]

5.4.1. Fallstudie 1: Krankenkasse A

5.4.1.1. Beschreibung der Krankenkasse und der IV-Verträge

Seit Einführung der IV-Verträge wurden ca. 100 Verträge von dieser Krankenkasse abgeschlossen. Die Verträge betreffen vor allem die Bereiche des ambulanten Operierens, der Hüft-Prothetik, Migräne und

[1] Aus Gründen des Datenschutzes, der Wahrung der Anonymität und auf Wunsch der Befragten sind nur die Interviewleitfäden, nicht aber die Volltranskripte der Interviews im Anhang zu finden. Bei Interesse kann mit der Autorin Kontakt aufgenommen werden, um nach Rücksprache mit den Befragten Auszüge aus den Volltranskripten zu übermitteln. Auch die Beschreibungen der Krankenkassen, Krankenhäuser und IV-Verträgen werden aus Gründen des Datenschutzes nur soweit detailliert, wie es für die Fallstudienergebnisse notwendig ist und ansonsten anonymisiert.

Die ersten, beschreibenden Teile der Fallstudien beinhalten naturgemäß auch Expertenmeinungen der Befragten, vor allem in den Abschnitten „Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen“. Diese Meinungen werden aus Gründen der Lesbarkeit meist im Indikativ wiedergegeben. Eine Beurteilung der Fallstudien und Analyse der Expertenmeinungen finden vor allem anhand des Kriterienkataloges statt und ist dort jeweils als solche gekennzeichnet.

Schwindel. Eine Vielzahl der Verträge konzentriert sich auf die ambulante Versorgung, bei Verträgen mit dem stationären Sektor sind stets Rehabilitationseinrichtungen integriert.

Zwei Arten von Qualitätsaspekten sind in die Verträge integriert: Zum Ersten Vertragsklauseln zur Qualitätssicherung, zum Zweiten Gewährleistungsklauseln mit Qualitätsaspekten.

Für die Qualitätssicherung wurde zunächst vereinbart, dass die Vertragskrankenhäuser an der externen Qualitätssicherung nach § 137 SGB V teilnehmen. Zusätzlich wurden Qualitätsziele anhand einiger Indikatoren der AQUA-Auswertungen definiert: Die Vertragskrankenhäuser sollen dabei unterhalb des Mittelwerts, Medians bzw. des Referenzwertes liegen. Bei Verfehlen dieser Ziele werden zunächst Gespräche aufgenommen – außerdem kann dann ein Grund für eine außerordentliche Kündigung vorliegen.

Die Gewährleistungsklauseln betreffen vor allem Verträge der Orthopädie. Hier wird vereinbart, dass nach der Erstimplantation die Revisionsoperationen nicht erstattet werden. Die Gewährleistungszeiträume hierfür liegen vertragsabhängig zwischen 8 Wochen und 10 Jahren. Einige (aber nicht alle) Verträge beinhalten Garantieeinschränkungen, beispielsweise abhängig vom Patientenverhalten oder Erkrankungen. Einige Gewährleistungsklauseln verpflichten dazu, dass eventuell notwendige Revisionseingriffe in der teilnehmenden Klinik stattfinden müssen. Außerdem soll den Patienten von der Klinik ein „Garantiezertifikat“ ausgehändigt werden.

5.4.1.2. Analyse der Interviews entlang der Systemelemente

Die Interviewpartner der ersten Fallstudie kommen aus den Abteilungen Entwicklung, Controlling und Abrechnung. Alle drei Mitarbeiter beschäftigen sich großteils mit Verträgen der Integrierten Versorgung sowie anderen Selektivverträgen.

Die Kernaussagen werden in sechs Themenblöcken zusammengefasst:

1. Entwicklungsprozesse der IV-Verträge mit qualitätsabhängiger Vergütung
2. Gründe für die Vertragsschließung
3. Evaluation und Monitoring der Verträge
4. Beschreibung der Systemelemente (Versorgungsbereich, Qualitätsindikator, Qualitätsmessung, Messergebnis, Qualitätsziel und Vergütungsfunktion/Konsequenzenmanagement)
5. Praktische Umsetzung
6. Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen

<u>Entwicklungsprozesse der IV-Verträge mit qualitätsabhängiger Vergütung</u>

Für den Entwicklungsprozess von IV-Verträgen bzw. Musterverträgen gibt es in dieser Fallstudie keinen klar standardisierten Prozess sondern er ist recht individualisiert. Ideen für die Verträge (sowohl der Versorgungsbereiche als auch der Ausgestaltung) kommen entweder durch die Analyse von Versorgungslücken anhand von kasseneigenen Daten und wissenschaftlichen Trends oder aufgrund von Vorschlägen vom Leistungserbringer (meist Chefärzten) zustande. Der Prozess von der Idee bis zum fertigen Vertrag kann bis zu 2 Jahre in Anspruch nehmen und wird von einer Vielzahl von Personen beim Leistungserbringer und der Kasse unterstützt. Auch Abrechnungsaspekte werden durch Integration der Abrechnungsabteilung betrachtet.

Gründe für die Vertragsschließung

Ein Grund für die Entwicklung von IV-Verträgen war früher zunächst die Möglichkeit der Umsetzung innovativerer Vertragsformen. Heutzutage stehen Kosteneinsparungen häufig im Vordergrund, sowohl durch die Verhandlung von Rabatten als auch besseres Arbeitsunfähigkeits- oder Komplikationsmanagement. Auch zu Marketingzwecken können Verträge genutzt werden, allerdings werden dadurch eher (vor allem chronisch) kranke, nicht aber gesunde Versicherte adressiert. In seltenen Fällen wurden auch Verträge umgesetzt, die sich aus vorhandenen Konzepten anderer konkurrierenden Krankenkassen ergaben.

Wenn ein Qualitätsmangel hohe Kosten verursacht, steht Qualitätsverbesserung als Mangelbeseitigung im Fokus der Entwicklung. Auch die gezielte Positiv-Auswahl von Leistungserbringern anhand von Qualitätsaspekten wird für die Nutzung eines Selektivvertrages als wichtiges Kriterium gesehen. So kann von vorneherein mit einer hohen Behandlungsqualität gerechnet werden

Evaluation und Monitoring der Verträge

Es finden detaillierte Vorabevaluationen für die IV-Verträge und neue Musterverträge statt. Hierbei wird vor allem die Wirtschaftlichkeit anhand Szenarienanalysen mit unterschiedlichen Annahmen (z. B. bezüglich der erreichbaren Qualitätsverbesserung) evaluiert. Weitere Elemente der Analyse sind bspw. zu erwartende Anzahl der Behandlungen, regionale Verteilung, Auswahl der Krankenhäuser, derzeitige Qualität der Krankenhäuser, Arbeitsunfähigkeit oder Alter der Patienten. Qualitätsverbesserungen werden meist anhand von monetären Aspekten oder indirekt monetären Aspekten (bspw. Reduktion stationärer Aufenthalte) evaluiert, reine Qualitätsaspekte wie Verbesserungen von QALYs, DALYs spielen kaum eine Rolle. Wenn die Qualitätsverbesserung keine monetäre Auswirkung hat, wird sie eher untergeordnet betrachtet. Dokumentations-, Monitorings- oder Verhandlungskosten werden in den Vorabevaluationen meist nicht miteinberechnet.

Da die Verträge sehr unterschiedlich gestaltet sind, existieren kaum Standards zur Vertragsbewertung und -entwicklung. Für das Monitoring nach Vertragsbeginn werden ähnliche Aspekte evaluiert, allerdings

können die Bewertungsdefinitionen der Vorabevaluation nicht immer für das Monitoring übernommen werden, so dass ein zusätzlicher Aufwand für die Analysen entsteht. Es werden Qualitätsberichte der Verträge erstellt, aber nur für den internen Gebrauch genutzt.

Beschreibung der Systemelemente

- Versorgungsbereich

Die Ideen für die zu berücksichtigenden Krankheitsbilder oder Behandlungsarten der Verträge kommen oft vom Leistungserbringer. Die Analyse dieses Versorgungsbereiches bei der Krankenkasse findet dann sehr datenbasiert statt – wenn die vorgeschlagenen potenziellen Qualitätsverbesserungen auch in den eigenen Kassendaten (oder im Ausnahmefall in wissenschaftlicher Literatur) nachvollziehbar sind, werden die Potenziale wirtschaftlich bewertet und ggfs. zu Verträgen weiterentwickelt.

- Qualitätsindikator

Für die Auswahl der verwendeten Qualitätsindikatoren werden keine festen Kriterien angewandt. Bislang werden stets existierende Indikatoren genutzt, die nicht sehr komplex sind. Die wichtigsten verwendeten Qualitätsindikatoren sind: Anzahl der stationären Aufenthalte, Arbeitsunfähigkeitstage, Krankengeldtage und Kosten für Arzneimittel. Auch AQUA-Indikatoren werden in den (stationären) Verträgen genannt, führen aber nicht zu vergütungsrelevanten Konsequenzen.

- Qualitätsmessung

Der Aufwand für die Qualitätsdokumentation und -messung ist eines der wichtigsten Themen für den Leistungserbringer. Während die derzeitigen Messungen für das Qualitätsmonitoring der einzelnen Verträge recht aufwendig sind, u. a. da kein Standard möglich ist, wären die Auswertung der Abrechnungsdaten für die Krankenkasse recht aufwandsarm und manipulationsresistent möglich. Auch komplexe, aufwendige Datenauswertungen können so durchgeführt werden, solange die Datenqualität hoch ist (bspw. im stationären Sektor). Einige Datenpunkte jedoch (wie beispielsweise die Dokumentation von Gewährleistungen) müssen derzeit händisch erfasst werden und verursachen daher einen großen Aufwand – aus diesem Grund gibt es bislang keine Auswertungen zur Anzahl aller Gewährleistungsfälle.

- Qualitätsziel

Qualitätsziele, die mit einem Konsequenzenmanagement verknüpft sind, werden nur im Rahmen von Kostenzielen gesetzt, da andere Qualitätsziele mit den derzeit eingesetzten Methoden in den Abrechnungsdaten nicht einfach auffindbar sind. Dadurch können anhand der Daten eindeutig erkennbare Reduktionen von Komplikationen oder Revisionsoperationen als Ziele gesetzt werden. Für eine Businessplanrechnung werden Hypothesen und Szenarien entwickelt, die die möglichen Reduktionen und Zielerreichungen monetär bewerten. Es werden zunächst Grenzwerte für die

Qualitätsziele definiert, die anschließend in der Verhandlung mit den Vertragspartnern diskutiert und abgestimmt werden. Für die Definition von Grenzwerten werden vor allem Erfahrungswerte aus anderen Verhandlungen herangezogen – die eigentliche Zielsetzung dagegen ist Verhandlungssache.

- Vergütungsfunktion/Konsequenzenmanagement

Derzeit findet im stationären Sektor kaum qualitätsorientierte Vergütung oder eine andere Art des Konsequenzenmanagements statt, während es im ambulanten Sektor erste Verträge mit Bonus-Malussystemen gibt. Derzeit werden Leistungserbringer, die die Qualitätsziele nicht erreichen, „zur Brust genommen“. Da die Qualitätsziele Verhandlungssache sind, werden Konsequenzen bei Nicht-Erreichung als „weiche Faktoren“ gesehen.

Praktische Umsetzung der IV-Verträge und Gewährleistungen

Aufgrund historischer Entwicklungen müssen derzeit für die Bearbeitung von IV-Fällen drei unterschiedliche Software-Benutzeroberflächen und ein zusätzliches Teilnehmerverzeichnis bedient werden. Zusätzlich existiert eine Excel-Datenbank zu den wichtigsten Inhalten der IV-Verträge.

Die Zuordnung der Fälle zu IV-Verträgen findet im Krankenhaus anhand von festgelegten Entgeltschlüsseln und in der Aufnahmedokumentation statt. Nur wenn sowohl der richtige Aufnahme- als auch der richtige Entgeltschlüssel genutzt wird, wird der Fall als IV-Fall erkannt. Ansonsten wird der Fall in der Regelversorgung abgerechnet, selbst wenn er von den Voraussetzungen her ein IV-Fall wäre. Nur wenn der Abrechner der Regelversorgung solche möglichen Fehler erkennt, kann eine Korrektur der Zuordnung nachträglich stattfinden.

Automatische Warnsignale bei Eintreten eines potenziellen Gewährleistungsfalles existieren auch bei der richtigen Verschlüsselung von IV-Fällen nicht. Der Abrechner kann nur im Rahmen der angezeigten Patientenhistorie „händisch“ überprüfen, ob ggfs. ein Gewährleistungsfall in einem bestimmten Zeitrahmen stattgefunden hat. Eine automatisierte oder teilautomatisierte Erkennung findet nicht statt. Sobald die Patientenhistorie zu lang ist und die auslösenden Anfangsfälle nicht mehr dargestellt werden bzw. zu lange zurück liegen, werden die Gewährleistungsfälle kaum identifiziert (5 Jahreszeiträume bspw. eignen sich hierfür bereits nicht mehr). Auch Revisionsoperationen, die in anderen als den IV-Vertragskrankenhäusern stattfinden und somit dort als Regelversorgungsfälle dokumentiert werden, können nicht identifiziert werden. Sichere Identifikationen solcher Fälle wären nur über ein spezielles Monitoring oder Controlling denkbar.

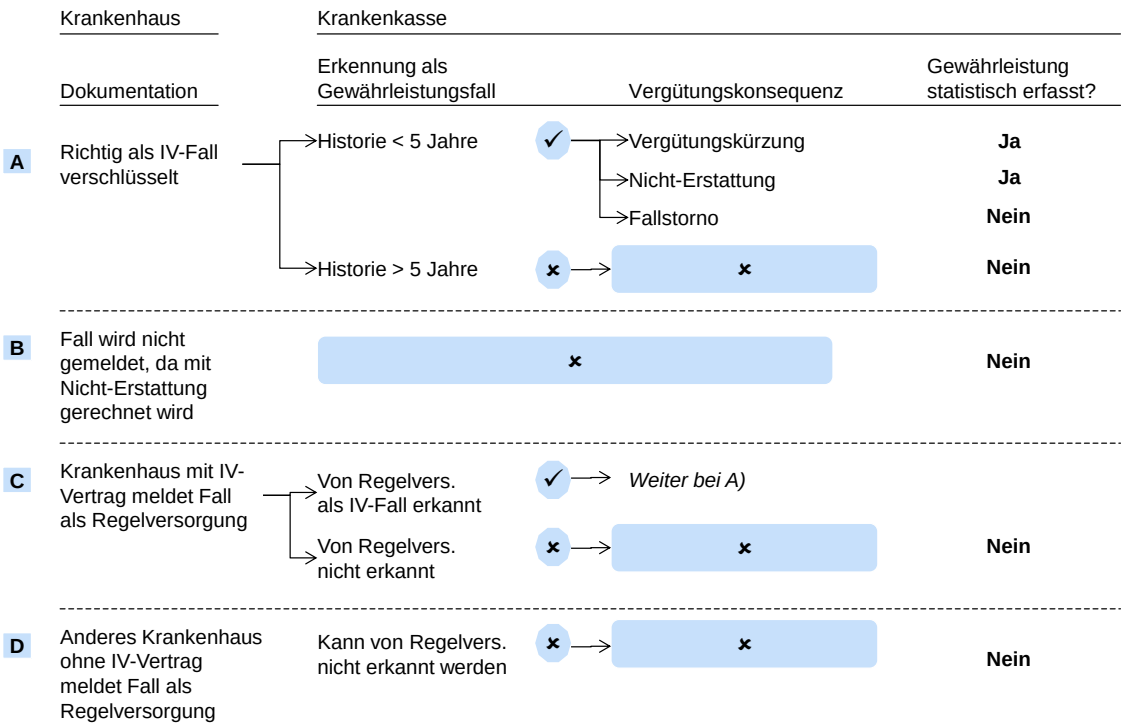
Wenn ein Fall als möglicher Gewährleistungsfall manuell vom Abrechner identifiziert wurde, muss der Abrechner zunächst im Vertrag nachschauen, welche Vergütungsform hier angewandt werden muss. Anschließend wird dem Krankenhaus die veränderte Vergütung (bspw. Nicht-Erstattung) mitgeteilt. Wenn es tatsächlich zu einer Nicht-Erstattung kommt, kann ein Fallstorno eingeleitet werden, und der

Fall ist somit in der Statistik aller Gewährleistungsfälle oder sogar aller Revisionsoperation nicht mehr auffindbar.

Während ein einfacher IV-Fall 2-3 Minuten Bearbeitungszeit in Anspruch nimmt, dauern schwierige IV-Fälle (bspw. solche mit veränderter Vergütung) in der Bearbeitung dagegen bis zu 15 Minuten. Der Befragte ist mit der derzeitigen Unterstützung der EDV in der Abrechnung der IV-Fälle nicht zufrieden.

Der Vorgang der Identifikation möglicher Gewährleistungsfälle bzw. der Hindernisse wird zur besseren Veranschaulichung grafisch dargestellt:

Abbildung 22: Möglichkeiten der Erkennung, Umsetzung und Messung von Gewährleistungsfällen in Fallstudie 1



Einschätzung der Interviewten bezüglich P4P und weiterer Entwicklungen

Es wird davon ausgegangen, dass P4P eine wichtige zukünftige Entwicklung ist und sowohl im selektivvertraglichen als auch im kollektivvertraglichen Bereich genutzt werden wird. Für die Zukunft wird die verbleibende und steigende Flexibilität in der Gestaltung von Selektivverträgen, bessere Datenverfügbarkeit und Datenzusammenführung sowie mehr Durchlässigkeit bei Stichproben der Kassenärztlichen Vereinigung gewünscht.

Da IV-Fälle immer recht flexibel und komplex sind, ist eine weitere variable Vergütung auch für die Abrechnung prinzipiell kein Problem. Eine bessere EDV-Unterstützung wäre wünschenswert. Wichtiger

als ein besseres Abrechnungsprogramm wäre allerdings ein besserer Datenaustausch, gerade bei ambulanten Verträgen mit viel Papierübertragung von Rechnungen.

5.4.1.3. Kurzanalyse des Datenmaterials

Für die Bewertung der IV-Verträge wurde die Kasse um eine Analyse gebeten. Aus datenschutzrechtlichen Gründen war es nicht möglich, die Analysen selbst durchzuführen. Daher wurde vorab die Analyse besprochen, und anschließend wurden die Ergebnisse diskutiert.

Ziel der Analyse war es, erste grobe Aussagen zu Anzahlen und Anteilen der Revisionsoperationen bei Hüft- und Knie-TEP-Operationen zu treffen, die in IV-Verträgen mit Gewährleistungsklauseln vorkommen. Zunächst wurden die Startfälle anhand der OPS-Codes ausgewählt und die möglichen Revisionsoperationen anhand der Komplikationsdiagnose identifiziert. Für die Analyse wurden die Daten von IV-Teilnehmern von vier Verträgen betrachtet, im Zeitraum 2005 bis Ende 2011. Als Startfall wurden die folgenden OPS-Codes berücksichtigt:

- 5-820 Implantation einer Endoprothese am Hüftgelenk
- 5-821 Revision, Wechsel und Entfernung einer Endoprothese am Hüftgelenk
- 5-822 Implantation einer Endoprothese am Kniegelenk
- 5-823 Revision, Wechsel und Entfernung einer Endoprothese am Kniegelenk

Es gab ca. 250 relevante Fälle im Zeitraum der vier IV-Verträge[1].

Anschließend wurden potenzielle Gewährleistungsfälle anhand von behandelten, diagnostizierten Komplikationen identifiziert, nämlich:

- T84 Komplikationen durch orthopädische Endoprothesen, Implantate oder Transplantate

Die Anzahl von „umgesetzten“, d. h. nicht-erstatteten Gewährleistungsfällen kann aufgrund der derzeitigen Datenlage nicht vollzählig ermittelt werden bzw. würde sehr viel manueller Analysearbeit bedürfen, die hier nicht möglich war. Daher kann an dieser Stelle nur die Anzahl und der Anteil der T84 Komplikationen als Revisionsoperationen evaluiert werden.

Ergebnisse: Von den rund 250 identifizierten Startfällen hatten 15 im Zeitraum eine Kodierung mit T84 im stationären Sektor. Dies entspricht ca. 6 %. Von diesen 15 Fällen hatten 9 vorher eine ambulante Behandlung nach der Kodierung T84.

[1] Die exakte Anzahl wird aus datenschutzrechtlichen Gründen nicht genannt und hat auf Analyseergebnisse keine Auswirkung.

Diskussion: Diese 6 % Komplikationen nach TEP-Operationen müssen nun ins Verhältnis zu vergleichbaren Studien gesetzt werden. Nach Recherche konnten keine Studien identifiziert werden, die alleinig die Komplikationsrate nach T84 aufzeigen[1]. Hier soll zunächst der Revisionsanteil von Hüft- und Knie-TEP-Operationen erwähnt werden. In Deutschland stehen vor allem die Daten der AOK für die Fünf-Jahres-„Überlebensrate" der Implantate von 95,3 % zur Verfügung. Daraus ergibt sich eine Fünf-Jahres-Revisionsrate von 4,7 % (vgl. Heller 2010). Das Ergebnis von 6 % allerdings betraf einen 7-Jahres-Zeitraum. Eine 10-Jahres-Revisionsrate aus Schweden ist bekannt. Sie lag zunächst in der Kohorte 1979 bei 9 % und konnte auf 2,5 % im Jahr 1991 gesenkt werden (vgl. Malchau et al. 2002). Nach erstem Ermessen sind die Ergebnisse der Fallstudie gemessen am deutschen Durchschnitt damit in keiner Richtung auffällig, sondern liegen prinzipiell im erwarteten Bereich von Revisionsoperationen. Die Vergleichbarkeit dieser Auswertungen ist zwar aufgrund methodisch unterschiedlicher Herangehensweisen gering, es stehen aber kaum bessere Daten zur Verfügung. Derzeit kann die Frage, wie viele der erkannten Revisionseingriffe tatsächlich unter die Gewährleistungsklausel gefallen sind, nicht beantwortet werden.

Theoretisch könnte davon ausgegangen werden, dass die Revisionsrate (hier als Approximation der Behandlungsqualität) bei einem IV-Vertrag mit Gewährleistungsklauseln aus zwei Gründen besser, d. h. niedriger sein sollte:

1. *Auswahl der IV-Leistungserbringer*: Wie bereits bei den Interviewergebnissen erwähnt, werden einige der Leistungserbringer bereits anhand von Qualitätsindikatoren für die IV-Verträge ausgewählt. Ist dies der Fall, ist davon auszugehen, dass auch bspw. die Revisionsrate besser ausfallen sollte. Ob gerade für diese vier IV-Verträge eine spezifische Auswahl durchgeführt wurde ist unbekannt. Gleichzeitig ist nicht davon auszugehen, dass das Auswahlkriterium die Revisionsrate war, da diese nicht standardisiert analysiert wurde.
2. *Qualitätsverbesserung durch Malus-Anreiz*: Durch die Nicht-Erstattung von Gewährleistungsfällen ist von einem Malus und somit einem Anreiz zur Qualitätsverbesserung auszugehen. Da es in dieser Fallstudie aber bislang nur sehr wenige, nicht statistisch auswertbare Gewährleistungen gab, ist eine große Anreizwirkung unwahrscheinlich. Natürlich könnte bereits das Wissen um einen Malus zu einem Anreizeffekt führen, doch auch dies ist nach Einschätzung der Interviewpartner eher nicht der Fall.

[1] Zwar gibt es Studien zum Anteil von T84-Komplikationen, allerdings entweder als festgestellte Behandlungsfehler der Schlichtungsstellen als Anteile an stationären Aufnahmen gesamtheitlich (vgl. Wirth 2009, S. 14; Datenquelle: Schlichtungsstelle und des Statistischen Bundesamtes 2006 – nur 1,8 von 10.000 stationären Aufnahmen) oder als ein Element von verschiedenen Arten von Komplikationsraten nach Hüft-TEPs (vgl. Zacher/Rohn 2010, S. 96).

Insgesamt wird deutlich, dass die bisherigen IV-Verträge vermutlich keinen messbaren, positiven Effekt auf die Behandlungsqualität haben. Wenn es bessere Qualitätsergebnisse gibt, so ist dies vermutlich mit der Auswahl der Leistungserbringer (also dem eigentlichen Selektivgedanken) und weniger mit dem Anreizsystem der Gewährleistungen zu erklären.

Verbesserungsmöglichkeiten der Analyse: Um eindeutigere Analyseergebnisse zu erlangen, wären für zukünftige Analysen folgende Verbesserungsvorschläge denkbar.

- Es sind klare, reproduzierbare Definitionen der Ausgangsfälle und der potenziell vergütungsrelevanten Folgekomplikationen erforderlich. Dazu wäre beispielsweise wünschenswert, wenn aufeinander bezogene OPS-Codes ins Verhältnis zueinander gesetzt werden: z. B. OPS 5-821 (Revision, Wechsel und Entfernung einer Endoprothese am Hüftgelenk) patientenbezogen als Folge von OPS 5-820 (Implantation einer Endoprothese am Hüftgelenk) im Gewährleistungszeitraum. Der Bezug auf den Diagnosecode T84 ist zumindest problematisch, da dieser in der Kodierung auch ausgelassen werden kann.
- Die Ergebnisstatistik für die IV-Vertragsteilnehmer sollte mit Versicherten der Regelversorgung verglichen werden.
- Es sollte möglichst ein standardisierter Nachverfolgungszeitraum beispielsweise von 5 Jahren gewählt werden, für den Vergleichswerte aus wissenschaftlichen Studien verfügbar sein sollten.

Insgesamt zeigt sich, dass zum heutigen Zeitpunkt anhand der bestehenden Daten gute Analysen möglich sind. Allerdings werden diese noch nicht standardisiert durchgeführt oder zu Monitoring-Zwecken genutzt. Gleichzeitig könnte mit einer besseren Datenbasis in Bezug auf die Gewährleistungen die eigentliche Anreizwirkung besser dargestellt werden.

5.4.1.4. Auswertung anhand des Kriterienkatalogs

Im nächsten Schritt soll die Fallstudie anhand des Kriterienkataloges bewertet werden. Für die Einschätzung der Erfüllung der Kriterien wurden hier als Primärquellen die Aussagen der Interviews und die Ergebnisse der Datenanalyse sowie der IV-Verträge genutzt. Hierbei wird ein Fokus auf Verträge von Hüft- und Knie-TEPs mit Gewährleistungsklauseln gelegt. Da nicht alle Kriterien abschließend beantwortet werden können, wurden zusätzlich wissenschaftliche Literatur und Expertenmeinungen analysiert und ggfs. eigene Einschätzungen genutzt.

Wie bereits in der Beschreibung der Einzelkriterien erwähnt, bestehen noch keine einheitlichen Skalen oder Grenzwerte. Daher wird hier eine 3-stufige, qualitative Skala für den Erfüllungsgrad angewandt. Die Stufen sind „Hoch“, „Mittel“ und „Niedrig“ bzw. „Nicht bewertbar“. Aufgrund der derzeit noch recht qualitativen Bewertung, ist ein quantitativer Gesamtvergleich bspw. der Anzahl von als „hoch“ bewerteten Kriterien nicht möglich.

Bewertung der Systemelemente, Kriterienkategorien und Einzelkriterien

➢ Gesamtsystem/Versorgungsbereich

Tabelle 12: Bewertungen Fallstudie 1, Systemelemente Gesamtsystem und Versorgungsbereich

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
1a	Abdeckung des Gesamtsystems	Wenig, Vertrag nicht als Gesamtsystem entwickelt	**Niedrig**	Vertrag betrifft nur einen Versorgungsbereich	Mehr Versorgungsbereiche hinzuziehen	Im selektivvertraglichen Rahmen kaum umsetzbar	Vertrag
2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet	Ja, indirekt durch Auswahl des Versorgungs-bereiches	**Hoch**	Hüft-TEPs sind "Volkskrankheit", häufige und medizinisch wichtige Behandlung für Fachgebiet	Überlegungen aktiv angestellt, Entwicklungsprozess gut: nach Lücken wird in den eigenen Daten gesucht	Weitere Daten und Studien hinzuziehen, auch nach "nicht-Volkskrankheiten" suchen	Interview, Vertrag, wissen-schaftliche Literatur
2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich	Ja, indirekt durch Auswahl des Versorgungs-bereiches	**Hoch**	Hüft-TEPs sind häufige DRG, auch ökonomisch relevant, hohe Gesamtkosten für Gesundheitssystem	Überlegungen aktiv angestellt, Entwicklungsprozess gut: nach Lücken wird in den eigenen Daten gesucht	Weitere Daten und Studien hinzuziehen, auch nach "nicht-Volkskrankheiten" suchen	Interview, Vertrag, wissen-schaftliche Literatur

Die Auswahl des Versorgungsbereiches findet zwar nicht systematisch statt. Doch die Kriterien der Aussagekraft des Versorgungsbereiches werden sehr gut erfüllt: Hüft- und Knie-TEPs sind bedeutsame Bereiche, sowohl für die medizinische als auch die ökonomische Relevanz (siehe u. a. Mansky/Seeger 2011; Seeger 2011). Der Abdeckungsgrad des Gesamtsystems ist aufgrund des indikationsspezifischen Selektivvertrages dagegen gering – die Abdeckung ist allerdings ein systemisches Problem, das von einem Kostenträger allein nicht lösbar ist.

➢ Qualitätsindikator

Tabelle 13: Bewertungen Fallstudie 1, Systemelement Qualitätsindikator

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten	Ja, wenn Revisions-operation als Indikator angesehen wird	**Hoch**	Revisionsoperation ist Ergebnisqualität, wird von Fachgesellschaften genutzt	Indikator nicht vollständig systematisch ausgewählt	Verbessertes Verständnis für Indikator, standardisierte Definition des Indikators	Interview, wissenschaft-liche Literatur
3b	Relevanz des Indikators für ein bestimmtes Fachgebiet	Ja, wenn Revisions-operation als Indikator angesehen wird	**Hoch**	Revisionsrate wichtiges Qualitätskriterium und relevantes Ziel auch aus Sicht der Fachgesellschaften	Zwar keine sehr systematische Auswahl, aber enge Einbindung des Leistungserbringers und der Ärzte	Verbessertes Verständnis für Indikator, standardisierte Nachverfolgung, Reporting erforderlich	Vertrag, wissenschaft-liche Literatur
3c	Sektorbezug vs. sektorübergreifende Nutzung *(Informationskriterium)*	Teilweise, Vertragspartner ja, Intention unklar	**Hoch**	IV automatisch mehrere Sektoren, Indikator misst mehrere, aber nur mit einem Vertrag	Nicht relevant für den Entwicklungsprozess	Einbeziehen in Überlegungen	Vertrag

Kriterienkategorie: Akzeptanz

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
4a	Medizinische Bedeutung in der medizinischen Fachliteratur	Ja, wenn Revisions-operation als Indikator gesehen wird	**Hoch**	Revisionsoperationen werden in einer Vielzahl von Literatur als Indikator angewandt	Sehr gut bewertbar, allerdings in Überlegungen nur indirekt enthalten	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	Vertrag, Interview, medizinische Fachliteratur
4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften	Ja, wenn Revisions-operation als Indikator gesehen wird	**Hoch**	Revisionsoperationen werden auch von den Fachgesellschaften als Qualitätskriterium genutzt	Derzeit keine systematische Auswahl, Indikator erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	Interview
4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern	Ja, wenn Revisions-operation als Indikator gesehen wird	**Hoch**	Wird von Chefärzten als Indikator vorgeschlagen	Derzeit keine systematische Auswahl, Indikator erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	Interview
4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit	Ja, wenn Revisions-operation als Indikator gesehen wird	**Hoch**	Laienverständlichkeit recht hoch	Derzeit keine systematische Auswahl, Indikator erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	Experten-meinung
4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit	Ja, wenn Revisions-operation als Indikator gesehen wird	**Hoch**	Wird von verschiedenen Stellen häufig genutzt	Derzeit keine systematische Auswahl, Indikator erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	Interview

Für die Anwendung des Kataloges können zwei Elemente als Qualitätsindikator bezeichnet werden:

- Revisionsoperationen mit Gewährleistungsklauseln, die eine Nicht-Erstattung nach sich ziehen
- Indikatoren der BQS/LQS-Auswertung nach §137 SGB V, die als Qualitätsziele vereinbart sind, die derzeit aber nicht qualitätsorientierte vergütet werden.

Da jedoch die Qualitätsmessung der BQS-Indikatoren in den Verträgen keine direkte Vergütungskonsequenz beinhaltet, werden hier nur die Gewährleistungsklauseln betrachtet.

Die Aussagekraft des Indikators Revisionsoperationen ist nicht weiter fraglich: Dass die Anzahl oder der Anteil von Revisionsoperationen eine sinnvolle Aussage bezüglich der Behandlungsqualität der Erstimplantation ermöglicht (vgl. u. a. Heller 2010; Engesæter et al. 2012, S. 436; Schuh et al. 2012, S. 1349), wird von Ärzten, der Öffentlichkeit und auch Fachgesellschaften wahrgenommen und kann auch

objektiv so betrachtet werden. Insgesamt werden die Kriterien des Qualitätsindikators also mit hoch bewertet.

- Qualitätsmessung

Tabelle 14: Bewertungen Fallstudie 1, Systemelement Qualitätsmessung

Kriterienkategorie: Wissenschaftlichkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
5a	Reliabilität	Kaum, Messung findet nicht automatisiert statt	**Niedrig**	Messung nur adhoc, daher Definitionen nicht überprüfbar	Messung findet nicht standardisiert oder automatisiert statt	Messung standardisieren, klare Indikatorendefinition, Kriterium beachten	Interview, Daten
5b	Validität	Kaum, Messung findet nicht automatisiert statt	**Niedrig**	Messung nur adhoc, daher Definitionen nicht überprüfbar	Messung findet nicht standardisiert oder automatisiert statt	Messung standardisieren, klare Indikatorendefinition, Kriterium beachten	Interview, Daten
5c	Risikoadjustierung und Risikoadjustierbarkeit	Trifft auf einzelfall-bezogenen Indikator nicht zu, da über Ein- und Ausschluss-kriterien geregelt	*nicht anwend-bar*	*nicht anwendbar*	*nicht anwendbar*	*nicht anwendbar*	*nicht anwendbar*

Kriterienkategorie: Messaufwand / -sicherheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
6a	Objektiver Erhebungsaufwand und -komplexität	Ja, theoretischer Aufwand bewertbar / praktisch findet Erhebung nicht standardisiert statt	**Niedrig (pot. Mittel)**	Erhebung findet nicht automatisiert statt, Analyse nur für Dissertation erstellt und daher derzeit erheblicher Aufwand	Derzeit keine Auswertung, langfristig geplant und (nach IT-Investition) aufwandsarm möglich	Wirtschaftlichkeit der IT-Investition prüfen und bei positiver Bewertung durchführen	Interview, Beobachtung
6b	Manipulationssicherheit / Objektivität der Messung	Ja, theoretisch bewertbar	**Pot. Hoch**	Erhebung nicht automatisiert, Daten würden aber auf Datensatz des §301 basieren	Datenbasis recht manipulationsresistent, da basierend auf Routinedaten - Nicht-Dokumentation möglich aber im Sinne der Vergütungsfunktion unschädlich, allerdings Verfälschung der Statistik	Für Vergütung gut, für statistische Auswertungen Mechanismen für eine garantierte Dokumentation einbauen	Interview, Beobachtung
6c	Grad der Integration in automatische Prozesse / IT-Systeme	Ja, anhand Interviews und Beobachtung der Prozesse	**Niedrig (pot. technisch hoch, wirtschaft-lich mittel)**	Weder Erhebung noch Erkennung noch Controlling der Gewährleistungsfälle automatisiert, derzeit IT-Investitionen geplant, dann Verbesserung	Automatisierung prinzipiell möglich, aber Investition wird nicht als prioritär gesehen	Wirtschaftlichkeit der IT-Investition prüfen, parallel Wirtschaftlich-keit durch Standard-isierung der Verträge und Messverfahren verbessern	Interview, Beobachtung

Die Qualitätsmessung der Gewährleistungs- und Revisionsfälle kann in dieser Fallstudie nicht so positiv bewertet werden. Es gibt keine standardisierten Definitionen der Qualitätsindikatoren[1] und daher auch keine automatisierten Messungen. Dadurch ist auch nicht sichergestellt, dass die Auswertungen der Vorabevaluation und des nachfolgenden Controllings sich jeweils auf die gleichen Sachverhalte beziehen. Somit kann der für diese Arbeit stattgefundene Messvorgang weder als vollständig reliabel noch valide eingestuft werden. Auch eine Risikoadjustierung findet nicht statt. Da die Gewährleistungsklauseln aber nicht als Form von P4P gesehen werden und nicht davon ausgegangen wird, dass eine verstärkte Umsetzung von Gewährleistungen die Qualität nachhaltig verbessern kann, wird der Standardisierung und Automatisierung dieses Sachverhaltes im kleinen Gebiet der Selektivverträge keine hohe Priorität beigemessen.

Vor allem der Messaufwand muss hinterfragt werden. Die vorher diskutierte Datenanalyse fand zu Zwecken dieser Dissertationsarbeit statt und wird nicht standardisiert oder automatisiert durchgeführt. Da zum einen die Datenbasis aufgrund von nicht-dokumentierten Gewährleistungsfällen nicht vollständig ist, zum anderen die Dokumentation und Auswertung derzeit mit viel manueller Arbeit verbunden sind, ist die Messung aufwendig und nicht in den IT-Workflow integriert. Potenziell könnten die Messungen relativ einfach anhand bei der Kasse schon vorhandener Routinedaten ausgewertet werden – nur die Gewährleistungen müssten weiterhin separat erfasst werden. Auch diese könnten aber bei Änderung der EDV als nicht erstattungsfähig im System registriert werden. Bei der untersuchten Kasse müssten jedoch hierfür Änderungen in der bestehenden Software durchgeführt werden, die derzeit nicht prioritär bearbeitet werden. Diese nachrangige Priorisierung ist auf eine zu geringe Größe des IV-Vertragsbereiches zurückzuführen.

Die Messsicherheit des Indikators im Sinne einer Manipulationsresistenz wäre prinzipiell bei richtiger Umsetzung sehr hoch. Die Auswertungen können anhand der Routinedaten stattfinden und besitzen eine hohe Manipulationsresistenz, da der Leistungserbringer einen Anreiz zur Meldung der Revisionsoperation hat. Aufgrund der Vertragsgestaltung und der Nichterfassung der nicht abrechenbaren Garantiefälle können allerdings die Statistiken bezogen auf diese Fälle unvollständig sein, da nicht erstattete Revisionsoperationen im System nicht erfasst werden (für den Leistungserbringer besteht im 100 %igen Gewährleistungsfall kein Anreiz, diese Fälle zu melden). In diesen Fällen ist zwar die Garantieregelung wirtschaftlich korrekt umgesetzt, aufgrund der fehlenden Erfassung dieser Fälle bleibt aber sowohl die Qualitätsdokumentation als auch die Kalkulation des wirtschaftlichen Vorteils der Garantieregelung für die Kasse unvollständig.

Während die Erfüllung dieser Kriterien in der derzeitigen Umsetzungsvariante der Messung mit „niedrig“ eingestuft werden muss, könnte sie potenziell einen hohen Erfüllungsgrad erreichen, wenn die Messung

[1] Also Definition der Gewährleistungsfälle für alle Verträge und in algorithmischer Form

standardisiert und automatisiert stattfände. Technisch wäre dies ohne weiteres realisierbar. Allerdings ist wegen des geringen wirtschaftlichen Umfangs und der Unterschiedlichkeit der Vertragsinhalte die technische Umsetzung aufwendig und teuer. Die technisch möglichen Lösungen werden daher aus wirtschaftlichen Gründen hier nicht umgesetzt.

- Qualitätsziele

Tabelle 15: Bewertungen Fallstudie 1, Systemelement Qualitätsziele

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
7a	Zielkonsistenz	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen	**Pot. Hoch**	Klares Ziel der Reduktion von Revisionsoperationen (aber eher als Kostenziel)	Hohe Zielkonsistenz bei Leistungserbringer und Kasse möglich, da Reduktion von Revisionsoperationen bei allen erwünscht ist	Klarere Definition der Ziele	Interview
7b	Zielkongruenz	Nein, da derzeit nur ein Ziel vorhanden, kein Zielsystem	*Nicht bewertbar*	*Nicht bewertbar*	*Nicht bewertbar*	*Nicht bewertbar*	*Nicht bewertbar*
7c	Reichweite	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen	**Niedrig**	Zwar ist durch den Versorgungsbereich ein recht hoher Anteil des Fachbereichs abgedeckt, der Gesamtanteil des Ziels für Gesamt-unternehmen ist aber gering	Unterscheidung zwischen Fachabteilung und Gesamtunter-nehmen notwendig, da kein Zielsystem sondern nur Vertrag, eher gering, außerdem statistische Fallzahlproblematik	Zielsystem erweitern und Punktesystem (Scoring), ggf. Zusammenfassung mehrerer Häuser (Regionen) oder Kassen	Vertrag, Interview

Kriterienkategorie: Operationalisierbarkeit und Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
8a	Erreichbarkeit	Ja, wenn die Reduktion von Revisions-operationen als Ziel gesehen	**Mittel**	Keine Revisionsoperation zu erreichen ist unrealistisch, Reduktion der Rate aber in Deutschland erreichbar	Muss mit Leistungserbringer abgesprochen werden, Ziel sollte daher nicht als gänzliche Vermeidung empfunden werden	Klare Definition der Ziele (dadurch kein "Empfinden" der gänzlichen Vermeidung)	Vertrag, Interview
8b	Transparenz	Schwierig, nicht bekannt, welche Ziele für wen transparent sind	**Hoch**	Ziele sind klar definiert und verständlich	Hier nicht bekannt, wer beim Leistungserbringer die Ziele kennt	Klare Definition der Ziele und transparente Darstellung	Vertrag, Interview
8c	Akzeptanz	Ja, wenn die Reduktion Revisions-operationen als Ziel gesehen	**Mittel**	Generell hohe Akzeptanz (siehe auch Indikator), aber bei gänzlicher Vermeidung ist Akzeptanz gering, da nicht erreichbar	Leistungserbringer muss vermittelt werden, dass Ziel nicht 0% Revisionsrate ist, dann Akzeptanz vermutlich hoch	Klare Definition der Ziele und transparente Darstellung, Einbeziehung aller Beteiligten	Vertrag, Interview
8d	Langfristige Anreizwirkung der Ziele	Schwierig, da kein Entwicklungs-system	**Mittel**	Kein Zielsystem mit ggf. Anpassungen der Ziele zu finden, aber Reduktion der Revisionsrate und Beibehaltung einer niedrigeren Rate kann langfristig inzentivieren	Generell Ziele eher zu hoch als zu niedrig, keine Entwicklung integriert	Lieber niedrigere Ziele mit Entwicklung, Rankings, etc.	Vertrag, Interview
9a	Gerechtigkeit	Ja, wenn die Reduktion Revisions-operationen als Ziel gesehen	**Mittel**	Nichtadjustierung der Gewährleistung wird von den Leistungserbringern teilweise als ungerecht empfunden (Aus Patientensicht wird die Umsetzung der Garantien häufig als gerecht angesehen)	Revisionsoperationen nach Hüft-TEPs nicht risiko-adjustiert da es sich um Einzelfallindikatoren handelt, Aus- und Einschlusskriterium mit den Leistungserbringern ggfs. zu besprechen	Anpassungen mit integrieren	Vertrag, Interview

Zunächst ist festzuhalten, dass für die Qualitätssicherung (ohne Gewährleistung) klar definierte Qualitätsziele festgehalten werden, deren nicht-Erreichung aber keine Vergütungsrelevanz hat. Die Qualitätssicherungsaspekte können lediglich in den Vertragsverhandlungen diskutiert werden, haben aber auch hier keine im Voraus festgelegten Auswirkungen. Da in der hier zugrunde liegenden Beurteilungssystematik nur Qualitätsaspekte mit definierter Vergütungskonsequenz untersucht werden sollen, verbleiben in diesem Fall nur die Gewährleistungsklauseln.

Die Kriterien der „Wirksamkeit/Zielgerichtetheit“ sind kaum bewertbar. Da nur ein vergütungsrelevantes Ziel verbleibt und dadurch kein Zielsystem vorliegt, kann die Zielkongruenz hier nicht bewertet werden. Die Zielkonsistenz ist hoch, da davon auszugehen ist, dass die Reduktion von Revisionsoperationen sowohl Ziel der Kasse als auch der Leistungserbringer und der beteiligten Ärzte ist. Die Reichweite der Ziele ist je nach Klinikstruktur unterschiedlich hoch, da die Endoprothesen bei Spezialversorgen einen hohen Anteil der Fälle, bei Maximalversorger aber auch nur einen kleinen Teil der Fallzahlen ausmachen können (vgl. Seeger 2011).

Auch die Kriterien der Operationalisierbarkeit der Qualitätsziele werden nur bedingt erfüllt:

- Die Transparenz ist gegeben, da der Sachverhalt der Vermeidung von Revisionsoperationen klar ist.
- Hinsichtlich der Erreichbarkeit und Akzeptanz könnte es bei der Umsetzung zu Problemen kommen. Die Nichtzahlung von Revisionen impliziert quasi eine ideale Revisionsrate von Null, die in der Praxis nicht erreichbar ist. Tatsächlich erreichbar ist aber eine Reduktion der Revisionsrate. Hinsichtlich der Akzeptanz gilt es in der Praxis, den Leistungserbringern zu vermitteln, dass es auch bei den in der

Praxis verbleibenden, nicht vermeidbaren Revisionen aus Gründen der Abwicklungseffizienz zu Malusregelungen kommt.

- Die anfängliche Senkung der Revisionsraten und anschließende Beibehaltung möglichst niedriger Revisionsraten ist per se ein langfristiges Anliegen. Bei Erreichung optimaler Raten müsste aber im System eine Weiterentwicklung der Ziele erfolgen. Dies ist aber im untersuchten Vertrag nicht vorgesehen.

Insgesamt zeigt sich, dass die Kriterien der Qualitätsziele hinsichtlich der Ziele der Qualitätssicherung gering, vom Qualitätsziel der Gewährleistung her dagegen mittel erfüllt werden.

- Vergütungsfunktion

Tabelle 16: Bewertungen Fallstudie 1, Systemelement Vergütungsfunktion

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
10a	Effizienz	Im konkreten Fall schwierig, Vergütungs-funktion an sich zwar klar, aber keine Daten zur Effizienz, daher nur Abschätzung	**Niedrig**	Daten für Kasse nicht vorhanden, aber anhand einer Abschätzung ist die Effizienz eines der Hauptprobleme	Zu klein mit zu hohen Investitions- und laufenden Kosten um wirklich effizient zu sein, wird außerdem von Kasse nicht analysiert	Analyse der Effizienz, Einbeziehen Gesamtkosten, Controlling der Effizienz	Vertrag, Interview
10b	Effektivität	Ja, da Daten vorhanden, aber Analyse mit Schwierigkeiten	**Niedrig**	Aus der Datenanalyse können keine eindeutigen Qualitätsverbesserungen ermittelt werden	Soweit anhand der vorhandenen Daten erkennbar, liegt die Revisionsrate auf einer zum Durchschnitt vergleichbaren Höhe	Standardisierte Daten-analyse, Konsequenzen-management bei Nicht-Erreichung der Ziele einführen	Daten
10c	Flexibilität	Ja, Vergütungs-mechanismus ist klar	**Mittel**	"Preis" der Vergütung richtet sich nach DRG-Katalog und Basisfallwert, daher absoluter Wert, der aber an die Hausgröße zumindest indirekt angepasst ist, aber keine weitere Anpassung vorgesehen	Nicht wirklich in Entwicklung integriert	Einbeziehen in Überlegungen, ggf. anpassen	Vertrag

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
11a	Anteil der Vergütung	Kaum, Vergütungs-funktion an sich zwar klar, aber keine Daten zur Gesamt-vergütung, nur Abschätzung möglich	**Niedrig**	Daten für Kasse nicht vorhanden, aber anhand Abschätzung ist der Vergütungsanteil eines der Hauptprobleme	Wird nicht analysiert, aber bereits ersichtlich, dass die qualitätsorientierte Vergütung keinen hohen Stellenwert innehat	Kollektivvertragliche Regelungen, mehrere Krankenhäuser, etc.	Interview, Vertrag, Daten
11b	Planbarkeit	Ja, Weiterentwick-lung der Vergütungs-funktion zwar nicht dargelegt, aber relativ klar	**Mittel bis hoch**	Anzahl von Revisionsoperationen beim Leistungserbringer ermittelbar und abgesehen von statistischen Streuungen gut planbar, Preisentwicklung einschätzbar	Planbarkeit im statistischen Mittel gut, angesichts nicht zu vermeidender statistischer Streuung bei niedrigen Fallzahlen sind aber die tatsächlichen Risiken im jeweiligen Berichtsjahr nur schwer vorhersehbar	Einbeziehen in Überlegungen und Entwicklung, ggf. Techniken zur Risikobegrenzung (z.B. Bildung von Rückstellungen, Begrenzung der Haftung, Rückversicherung)	Vertrag, Interview

Die Effektivität der Maßnahme wurde anhand der Daten bereits diskutiert: Eine Qualitätsverbesserung und sinkende Revisionsrate aufgrund der Anreize der Gewährleistungsklauseln kann nicht eindeutig nachgewiesen werden. Einerseits ist ein routinemäßiges Qualitätscontrolling der Gewährleistungen nicht implementiert, andererseits zeigen auch die für diese Auswertung erstellten Statistiken keinen adäquaten Effekt. Die zusätzlichen Aussagen in den Interviews lassen insgesamt darauf schließen, dass diese qualitätsorientierte Vergütung derzeit kaum Auswirkungen auf das Qualitätsverhalten der Leistungserbringer haben kann.

Die IV-Verträge werden zwar insgesamt auf Wirtschaftlichkeit überprüft und nur dann unterschrieben, wenn eine Effizienz nachzuweisen ist. Die reinen Qualitätsaspekte werden allerdings nicht monetär bewertet. Obschon hierzu keine Daten vorliegen, kann aufgrund einer ersten Abschätzung davon ausgegangen werden, dass die Verträge hinsichtlich der Qualitätswirkungen vermutlich nicht effizient sind (vgl. Mansky/Seeger 2011).

Die Operationalisierbarkeit der Vergütungsfunktion ergibt sich aus zwei Kriterien, die gering bis mittel bewertet werden. Während die Planbarkeit gut ist, da sie von der Anzahl der Revisionsoperationen abhängt und daher gut abzuschätzen ist, ist der Anteil der Gewährleistungsvergütung am Gesamtvolumen sehr gering. Nach einer beispielhaften Schätzung wären bei einem Maximalversorger durch einen solchen Vertrag nur etwa 0,02 % des Umsatzes betroffen (vgl. Seeger 2011). Studien gehen davon aus, dass eine Vergütungsvariabilität bei Krankenhäusern ab einer Höhe von 5 % des Gesamtumsatzes eine relevante Größenordnung erreicht (vgl. Veit et al. 2012e, S. 32).

- Messergebnis

Tabelle 17: Bewertungen Fallstudie 1, Systemelement Messergebnis

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
12a	Nutzung zur Steuerung	Ja, Nutzung diskutiert und Nutzbarkeit relativ klar	**Niedrig**	Wird weder in der Kasse noch beim Leistungserbringer zur Steuerung genutzt (nach Einschätzung der Kasse)	Nutzbarkeit auch niedrig, da statistisch zu klein und als Malus-System sehr "alles oder nichts" getrieben	Nur bei Anpassung tatsächlich nutzbar, muss dann auch aktiv in Steuerung integriert werden	Vertrag, Interview
12b	Aktualität	Mittel, potenzieller Messvorgang recht klar aber angesichts des praktischen Umsetzungs-grades nicht praktisch bewertbar	**Mittel**	Zwar finden die Rechnungskürzungen vermutlich fortlaufend und somit zeitnah statt, wenn es allerdings zu Unstimmigkeiten kommt, kann der Prozess sehr lang dauern	Bei Streitfällen kann es eine sehr langfristige Abwicklung geben	Durch klarere Definition sollten Streitfälle möglichst vermieden werden	Vertrag, Interview
12c	Kontrollierbarkeit	Mittel, potenzieller Messvorgang recht klar aber angesichts des praktischen Umsetzungs-grades nicht praktisch bewertbar	**Niedrig (pot. Hoch)**	Derzeit findet kein Controlling der Gewährleistungen statt, Berichte werden vor allem zu Fallzahlen, Versichertenzahlen und Zahlungen erstellt	Controlling könnte - wenn automatisierte Erhebungen stattfinden würden - sehr leicht integriert werden	Berichtswesen könnte ggf. sogar verpflichtend sein, bei höherer Relevanz würde es automatisch mehr Berichte geben	Interview, Vertrag, Daten
12d	Praktische Umsetzung im Controlling	Ja, derzeitige Evaluationen bekannt	**Niedrig bis mittel**	Derzeit kein standardisiertes Controlling der Gewährleistung der IV-Verträge	Analysen anhand der Daten gut möglich, erste Analyse für diese Arbeit erstellt	Aufbau eines standardisierten Qualitätsberichtswesens höher priorisieren	Interview, Vertrag, Daten
12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit	Ja, Zuordnung Messergebnis klar	**Hoch**	Sowohl auf Krankenhaus als auch auf Fachabteilungsebene ist die Zuordnung des Messergebnisses der Revisionsoperation klar	Auf Mitarbeiterebene wäre es schwieriger, das ist aber erst beim Leistungserbringer relevant	Nicht erforderlich	Vertrag

Kriterienkategorie: Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Quelle
13a	Zeitverlauf	Ja, zwar derzeit nicht vorhanden, Möglichkeiten aber klar	**Niedrig**	Generell könnten die Messergebnisse natürlich in zeitlichen Zusammenhang gesetzt werden, allerdings ist durch die geringe Fallzahl die Aussagekraft begrenzt	Derzeit findet keine standardisierte Messung und daher keine Zeitverlaufsanalyse statt	In die Betrachtung einbeziehen	Vertrag
13b	Querschnittsvergleich/Ranking	Ja, derzeit aber nicht vorhanden und aufgrund der Verträge nicht vorgesehen	**Niedrig**	Wegen individueller Gestaltung der Verträge ist ein Ranking derzeit praktisch kaum möglich	Krankenkassen könnten Analysen machen, allerdings fallen die Selektivverträge oft unterschiedlich aus, so dass ein Vergleich kaum möglich wäre	Nur in Kollektivvertrag oder bei zumindest kassenbezogener Standardisierung möglich	Vertrag
13c	Justiziabilität	Ja, derzeit aber nur selten umgesetzt	**Hoch**	Der Revisionsfall ist klar und eindeutig beschrieben und könnte daher (bei Erkennung) auch gerichtlich durchgesetzt werden	Regelung bezüglich Justizierbarkeit gut	Nicht erforderlich	Vertrag, Leitlinien

Die Operationalisierbarkeit der Messergebnisse fällt recht gemischt aus: Während die „verursachungsgerechte Zuordnung“ der Ergebnisse auf den einzelnen Leistungserbringer und die Fachabteilung sehr gut möglich sind und hier mit „hoch“ bewertet wird, ist die „Nutzung zur Steuerung“ sehr gering – die Messergebnisse bezüglich Revisionsoperationen werden nicht verfolgt und weder in der Krankenkasse noch beim Leistungserbringer für interne Steuerung genutzt. Auch die „Kontrollierbarkeit“ ist derzeit gering, könnte aber potenziell deutlich höher ausfallen – durch die prinzipiell gut standardisierbaren Messmöglichkeiten könnten Controllingberichte auch zu Qualitätsergebnissen erstellt werden, wie entsprechende Studien an anderer Stelle zeigen (vgl. u. a. Malzahn et al. 2010; Heller 2010).

Die Aktualität der Messergebnisse wird derzeit mit „mittel“ bewertet: Zwar könnten die Rechnungskürzungen fortlaufend und somit zeitnah stattfinden, wenn es allerdings zu Unstimmigkeiten kommt, kann der Prozess sehr lang dauern.

Die Vergleichbarkeit der Messergebnisse ist derzeit eher niedrig: Da bislang keine standardisierten, reproduzierten Messungen stattfinden, wird weder ein Zeitverlauf noch ein Vergleich über mehrere Leistungserbringer verfolgt. Das Ranking wird auch langfristig aufgrund der unterschiedlich gestalteten Verträge nicht stattfinden können.

5.4.1.5. Entwicklungsmöglichkeiten des untersuchten Systems

Allgemein ist festzuhalten, dass die Qualitätsgarantien nicht Hauptzweck beim Abschluss der IV-Verträge waren. Insbesondere war die systematische Entwicklung eines qualitätsorientierten Anreizsystems nicht bewusst Gegenstand dieses Prozesses, obwohl mit den Qualitätsgarantien unweigerlich ein solches System implizit eingeführt wurde. Dies erklärt auch, warum viele Aspekte nicht systematisch in die Entwicklungsüberlegungen einbezogen wurden.

Hinsichtlich der Qualitätsaspekte der Kriterienkatalogs wären zunächst nur wenige Verbesserungen notwendig, um die hier untersuchten IV-Verträge wirksamer zu gestalten. Zunächst sollten die Versorgungsbereiche, Qualitätsindikatoren und -messungen systematischer ausgewählt und bewertet werden – die hier formulierten Kriterien können hierfür als Hilfestellung dienen. Zusätzlich müsste vor allem das Kriterium des Erhebungsaufwandes und der Integration in automatische Prozesse überprüft werden, um auf dieser Basis auch ein praktisch wirksames Qualitätscontrolling der Gewährleistungen aufzubauen. Nach einer eingehenden Investitionsrechnung für mögliche Kosten der Softwareimplementierung könnte eine solche Integration die Messung und ebenso die mögliche Umsetzung deutlich verbessern.

Gerade die Kriterien des Anreizsystems sollten in einem erweiterten System besser erfüllt werden. Für die Messergebnisse können sich die Kriterienerfüllungen verbessern, sobald die Qualitätsergebnisse der Gewährleistungen systematisch, standardisiert und automatisiert gemessen und in Berichten dokumentiert werden. Für die Zielkriterien dagegen ist vor allem eine Erweiterung auf ein aufeinander abgestimmtes Zielsystem notwendig.

Die Vergütungsfunktion kann erst dann alle Kriterien erfüllen, wenn einige Elemente grundlegend verändert werden, so bspw. den Vergütungsanteil. Im Rahmen von Selektivverträgen ist dies nur schwer möglich. Insgesamt lässt sich daher diskutieren, ob einige der Kriterien vor allem die Umsetzung von P4P-Aspekten innerhalb des Kollektivvertrages nicht besser erfüllt werden könnten.

5.4.2. Fallstudie 2: Krankenkasse B

5.4.2.1. Beschreibung und erste Ergebnisse der IV-Verträgen

Die Krankenkasse B hat derzeit insgesamt mehr als 200 IV-Verträge. Davon sind ca. ein Drittel Verträge der Kardiologie und Endoprothetik. In den Endoprothetik-Verträgen sind ca. 4.000 Fälle enthalten. Darunter gibt es auch recht kleine Verträge mit nur 10 Fällen. Der größte Vertrag hat 480 Fälle pro Jahr.

Der Fokus der Analyse der zweiten Fallstudie liegt auf zwei Muster-IV-Verträgen, die den zu analysierenden Aspekten am besten entsprechen. Sie wurden für zwei Versorgungsbereiche erstellt, die im Fokus der IV-Vertragsgestaltung stehen. Diese Musterverträge werden für die gesamte Krankenkasse erstellt und anschließend in den einzelnen Regionen mit den jeweiligen Vertragspartnern abgeschlossen. Individuelle Variationen von den Musterverträgen sind möglich, müssen aber mit der Entwicklungsabteilung abgestimmt werden.

Der erste Mustervertrag ist im Bereich Kardiologie und gilt seit 2010. Die Versorgungsziele sind die Verlagerung von rein stationärer, invasiv-kardiologischer Leistungen in den quasi-ambulanten Bereich, Vermeidung von Doppeluntersuchungen und Reduktion der kardiovaskulären Morbiditäten. Im Bereich der Qualitätssicherung werden Ziele des Qualitätsmanagement definiert, so bspw. die Reduzierung des Anteils der Patienten mit Auftreten eines akuten Koronarsyndroms oder Vermeidung der Entwicklung einer Herzinsuffizienz NYHA Stadium III und IV. Zusätzlich werden jährlich interne Qualitätsberichte erstellt, die z. B. Anteil der Patienten mit leitliniengerechter Medikation oder Anteil PCI Eingriffe in gleicher Sitzung mit der Katheterdiagnostik darstellen. Zusätzlich werden die Vertragsteilnehmer dazu verpflichtet, an der externen Qualitätssicherung nach § 135a und § 137 SGB V teilzunehmen, anhand derer Qualitätsziele festgesetzt werden – das primäre Qualitätsziel ist hierbei, dass die Ergebnisse bei allen Qualitätsindikatoren besser als die Referenzwerte sind. Bei Nicht-Erreichung der Qualitätsziele liegt ein Grund für eine außerordentliche Kündigung des Vertrages vor. Gewährleistungsvereinbarungen können ergänzt werden, sind derzeit aber nicht detailliert dargestellt.

Der zweite Mustervertrag betrifft endoprothetische Eingriffe. Zwei Arten der Musterverträge liegen vor – einer wirksam von 2005-2009, der zweite Mustervertrag gilt seit 2009. Beide Verträge beinhalten Abschnitte bezüglich der Qualitätssicherung. Während allerdings der ältere Vertrag vereinbarte Qualitätsziele – bspw. Verringerung von Prothesenluxation als postoperative Komplikation – und Gewährleistungsklauseln beinhaltet, sind diese im neuen Vertrag nicht mehr enthalten. Die „alte“ Gewährleistung-Musterklausel beinhaltet die Option, Gewährleistungsfristen (frei) zu definieren, während derer eine erneute endoprothetische Versorgung desselben Gelenkes nicht vergütet wird. Allerdings können Krankenhäuser die Notwendigkeit der Revisionsoperation überprüfen lassen (bspw. vom MDK) – die Beweispflicht, dass die Ursache der Revisionsoperation außerhalb des von der Gewährleistung umfassenden Bereich liegt, ist beim Leistungserbringer.

5.4.2.2. Erste Ergebnisse der Interviews

Die Interviews der zweiten Fallstudie wurden ebenfalls in den Bereichen Abrechnung, Entwicklung und Controlling der IV-Verträge durchgeführt. Hier sollen vor allem die Unterschiede zur ersten Fallstudie dargestellt werden.

Entwicklungsprozesse der IV-Verträge mit qualitätsabhängiger Vergütung

Ein Standardprozess zur Vertragsentwicklung existiert. Zunächst werden Festlegungen bezüglich der Indikationsbereiche bundesweit getroffen. Für diese sogenannte „Versorgungsprogramme“ gibt es jeweils einen Programmleiter, der mithilfe einer Arbeitsgruppe die Entscheidung für bestimmte IV-Verträge oder andere Maßnahmen trifft.

Für IV-Verträge wurden Musterverträge erstellt. Diese beinhalten Erfahrungen aus vorangegangenen Pilotverträgen und -projekten. Anschließend werden verschiedene Abteilungen und regionale Spezialisten eingebunden. Musterverträge werden als „Leitplanken“ gesehen – Änderungen können integriert werden, wenn sie von der Hauptverwaltung genehmigt wurden. Bei großen, relevanten Änderungen werden diese teilweise in den Mustervertrag aufgenommen.

Die Entwicklung von Musterverträgen dauert ca. 6-12 Monate, unter anderem aufgrund der komplexen Organisationsstruktur und der hohen Anzahl beteiligter Abteilungen. Die Entwicklung und der Abschluss von Individualverträgen dauern ca. 6 Monate. An dem Prozess der Musterverträge sind 6-8 Personen beteiligt, bei Einzelverträgen 2-3. Bei den Musterverträgen sind die Leistungserbringer zunächst nicht integriert, erst bei den Einzelverträgen sind ca. 3-4 Personen vom Leistungserbringer beteiligt. Mindestens fünf Treffen sind mit dem Leistungserbringer nötig. Die Abrechnungsabteilung ist in den Entwicklungsprozess allerdings nicht eingebunden. Es wird der Wunsch geäußert, dass der Entwicklungsprozess zukünftig schlanker gestaltet werden sollte.

Evaluation und Monitoring der Verträge

Für die Entwicklung der Musterverträge werden Annahmen getroffen, welche Qualitätsverbesserungen durch die Auswahl der Vertragsteilnehmer entstehen können. Außerdem werden Einnahmeeffekte (z. B. aufgrund der Morbi-RSA-Zuweisung) und Vertriebseffekte analysiert – die Qualitätsveränderungen werden nicht monetär bewertet. Auch die Einzelverträge werden vorab evaluiert, anhand von Planung der Umsatz- und Fallzahlen – Qualitätsaspekte werden für individuelle Verträge nicht vorab analysiert.

Die Auswahl der Vertragsteilnehmer findet anhand von Qualitäts- und Beziehungsaspekten statt. Beispielsweise werden zunächst nur Leistungserbringer angesprochen, deren AQUA-Indikatorenergebnisse oberhalb des Referenzwertes liegen. Nachdem die Vertragsteilnehmer ausgewählt und die Verträge abgeschlossen sind, finden keine standardisierten Qualitätsevaluationen mehr statt. Zwar gibt es auf regionaler Ebene einige Qualitätsanalysen, diese werden aber nicht zentral gesammelt

oder ausgewertet. In den einzelnen Geschäftsstellen werden die Ergebnisse vor allem zu Verhandlungszwecken genutzt. Da bereits die Entgeltschlüssel, die für die IV-Vertragscodierung genutzt wird, in den Geschäftsregionen unterschiedlich sind, ist ein gemeinsames Konzept derzeit kaum möglich.

Ein übergreifendes Controlling-Konzept für die IV-Verträge besteht derzeit nicht. Auch hat Qualität im Controlling noch eine untergeordnete Rolle, weil vor allem rein wirtschaftliche Aspekte relevant sind. Derzeit wird eine Datenbank erstellt, die alle Entgeltschlüssel zusammenfasst. So sind dann schnellere und detaillierte Auswertungen möglich. Außerdem sollen mittelfristig Selektivverträge zukünftig im Softwareprogramm abgebildet werden. Aufgrund anderer Prioritäten jedoch, steht diese Umwandlung derzeit nicht im Fokus.

Beschreibung der Systemelemente

- Versorgungsbereich

Die Verträge und die Auswahl des Versorgungsbereiches entstanden früher meist auf Initiative der Krankenhäuser. Seit 2010 werden die Versorgungsbereiche anhand von Schwerpunktthemen ausgewählt und die Kasse geht selbst aktiv auf die Leistungserbringer zu. Diese Schwerpunktthemen sollen ca. 80 % des Vertragsvolumens ausmachen.

Anhand von drei Kriterien werden die Versorgungsbereiche ausgewählt:

1. Morbi-RSA – Auswahl von Indikationen mit hohen Umsätzen oder höheren Kosten als im GKV-Durchschnitt
2. Marketing-strategische Gründe – Fokus auf Familien und Kinder als Zielgruppe
3. Konkurrenzsituation – Auswahl von Versorgungsbereichen, die auch andere Kassen bereits anbieten

- Qualitätsindikator

Die AQUA-Indikatoren wurden für einige IV-Verträge ausgewählt, da sie derzeit einfach und schnell auswertbar sind. Zusätzliche Indikatoren gibt es bspw. in der Kardiologie, die gemeinsam mit den Leistungserbringern entwickelt wurden. Derzeit ist geplant, für die Kardiologie weitere Indikatoren mit Routinedaten einzusetzen, vor allem Sterblichkeit und Wiedereinweisung. Auch Patientenbefragungen sollen einbezogen werden.

Ein Entwicklungskriterium für die existierenden kardiologischen Indikatoren war die medizinische Bedeutung. Im Regelfall werden die Qualitätsindikator vor allem nach Praktikabilität, d. h. vor allem einfachen Mess- und Darstellbarkeit ausgewählt. Aber auch in der wissenschaftlichen Literatur und unter medizinischen Gesichtspunkten (vor allem von Medizinkollegen) wurden die Indikatoren nach Relevanzkriterien geprüft.

Die größte Herausforderung der Qualitätsindikatoren ist die Praktikabilität. Anhand der kasseneigenen Daten ist nicht alles abbildbar. Gleichzeitig ist bei den AQUA-Daten die Information nur so gut wie die Dokumentation, die dahinter steckt. Die größte Herausforderung ist die Identifikation von Indikatoren, die nicht manipulationsanfällig sind und die man ohne größeren Erfassungsaufwand abbilden kann. Hierfür werden verschiedene Indikatoren ausprobiert und mit Leistungserbringern zusammengearbeitet, die bereits Erfahrung haben – bspw. für die kardiologischen Indikatoren gemeinsam mit gesundheitsökonomischen Lehrstühlen. Die Entwicklung dieser Indikatoren dauert ca. 6-12 Monate.

- Qualitätsmessung

Plausibilitätschecks finden nur für einzelne Verträge und nicht übergreifend statt, auch nicht anhand der Routinedaten für die selbstdokumentierten Daten. Validität und Messaufwand sind wichtige Themen, ein Fokus der Qualitätsmessung liegt aber vor allem auf der Praktikabilität.

Qualitätsberichte sollen regelmäßig erstellt und zugesandt werden, dies wird aber nicht zentral überprüft. Auch regional werden Berichte nicht strukturiert angefordert. Nur bei großem Interesse der lokalen Kollegen werden die Qualitätsergebnisse analysiert.

- Messergebnis

Rankings von verschiedenen Leistungserbringern finden nur einzeln, auf regionaler Ebene statt – standardisierte Vergleiche gibt es nur für die Auswahl der Vertragsteilnehmer. Zeitliche Verläufe werden derzeit nicht dargestellt, sind aber geplant. Wissenschaftliche Evaluationen finden selten statt. Die stattgefundenen wissenschaftlichen Analysen beziehen sich auf einzelne Verträge und wurden bei Gelegenheit durchgeführt.

Aus den Daten ist derzeit die Anzahl der Gewährleistungsfälle nicht abbildbar. Gewährleistungen wurden wieder aus den Verträgen herausgenommen, nachdem festgestellt wurde, dass nicht viele Gewährleistungsfälle tatsächlich aufgegriffen und umgesetzt wurden. Es sollte dadurch getestet werden, ob der Abschluss eines IV-Vertrages ohne Gewährleistung beim Leistungserbringer besser akzeptiert wird. Mittlerweise wird die Wiederaufnahme der Garantieklauseln erörtert, um die möglichen Anreizeffekte zur Qualitätsverbesserung zu erzielen.

- Qualitätsziel

Qualitätsziele werden nicht erstellt. Nur für die Auswahl der Vertragsteilnehmer gibt es Grenzwerte. Zukünftig sollen bei neuen Indikatoren und Verträgen Zielwerte gesetzt werden, auch mit Grenzwertentwicklungen über mehrere Jahre mit kontinuierlichen Steigerungen.

- Vergütungsfunktion/Konsequenzenmanagement

Im regionalen Arbeitsbereich des Interviewten hat seit 2000 nur ein Fall von potenzieller Gewährleistung stattgefunden, der nach Zusammenarbeit mit der Landesgeschäftsstelle und Anforderung von Unterlagen nicht als Gewährleistungsfall anerkannt wurde. Weiter 3-4 Fälle sind im gesamten regionalen, stationären Abrechnungsbereich aufgetaucht, von dem einer ein Streitfall wurde. Andere Regionen haben einen höheren Anteil von Gewährleistungsfällen.

Konsequenzenmanagement wird im Rahmen der jährlichen Vertragsgespräche thematisiert. Hierfür gibt es allerdings keine standardisierte Herangehensweise. Außerordentliche Kündigungen aufgrund von Qualitätsmängeln gab es bislang nicht. Eine ordentliche Kündigung allerdings fand wegen vorhandener Qualitätsmängel aufgrund von Beschwerden der Versicherten zu Behandlungsqualität und Serviceaspekten statt. In neueren Verträgen soll das Konsequenzenmanagement vermehrt intergiert werden.

Praktische Umsetzung

Derzeit ist die Analyse der IV-Verträge nicht ausreichend automatisiert, da beispielsweise die relevanten Abrechnungsschlüssel der IV-Verträge in separaten Dateien gespeichert werden. Die Qualitätsdatenerhebung und -analyse erfolgt dezentral und wird auch nicht übergeordnet zusammengefasst. Die Qualitätsdaten der Leistungserbringer liegen nur den regionalen Verantwortlichen vor.

Im Rahmen der jährlichen Vertragsplanung wird überprüft, ob Revisionsoperationen beispielsweise als Regelfall kodiert wurden – vor allem, wenn der Leistungserbringer bereits auffällig war, wie z. B. zu wenig abgerechnete Fälle im IV-Vertrag. Besonders Gewährleistungen, bei denen die Revisionsoperationen nicht auf diesem Wege erkannt wurden, können nicht erkannt werden.

Die Abrechnung in dieser Krankenkasse ist organisatorisch zunächst regional, anschließend sektoral aufgeteilt – für den betrachteten regionalen stationären Bereich sind 3,5 Mitarbeiter zuständig. Diese sind sowohl für die Regelversorgung als auch die IV-Verträge zuständig, so dass alle Abrechner die Inhalte der Selektivverträge kennen müssen.

Der Aufwand der Abrechnung für IV-Verträge ist hoch – vor allem aufgrund verschiedener Arbeitsschritte die z. T manuelle Überprüfungen beinhalten bspw. der Teilnahmebestätigung. Es gibt eine automatische Erkennung möglicher IV-Fälle, die anhand der Entgeltschlüssel identifiziert werden – anschließend müssen diese Fälle manuell überprüft werden, um festzustellen, ob es sich tatsächlich um IV-Fälle handelt. Wenn der Leistungserbringer den IV-Fall nicht als solchen verschlüsselt, werden die Fälle nur erkannt, wenn die Teilnahmebedingung bereits im System hinterlegt sind, oder können nur im Nachhinein vom Controlling herausgefiltert werden. Sobald jedoch die Patientenhistorie den IV-Vertrag enthält, werden alle potenziellen Gewährleistungsfälle ausnahmslos erkannt – auch wenn bspw. die

Revisionsoperation in einem Nicht-Vertragskrankenhaus stattgefunden hat. Die Identifizierung der potenziellen Gewährleistungsfälle findet anhand der ICD/OPS-Codes statt.

Sobald ein potenzieller Gewährleistungsfall erkannt ist, beginnt das Verfahren mit der Kontaktaufnahme mit dem Leistungserbringer. Wenn das Krankenhaus die Nicht-Erstattung akzeptiert wird ein Rechnungsstorno eingeleitet – der Fall kann trotzdem nachverfolgt werden. Manchmal werden die Patienten über die Gewährleistungsfälle informiert. Gewährleistungsfälle sind auch schon vor dem Sozialgericht gelandet, vorher findet dann eine MDK-Überprüfung statt.

Während ein „normaler“ IV-Fall ca. 5 Minuten in Anspruch nimmt, können sich Gewährleistungsfälle über Wochen hinziehen. Zukünftig sollen einige Schritte automatisiert werden, sowohl die elektrische Erfassung von Teilnahmebedingungen als auch die Integration der Gewährleistungsfristen in die Abrechnungssoftware. Ab dem 3. Quartal 2012 soll es eine automatische Erkennung potenzieller Gewährleistungsfälle anhand der Patientenhistorie und den Gewährleistungsklauseln der IV-Verträge geben, derzeit ist dies nicht ausreichend standardisiert und nicht automatisiert in Controllingberichten inkludiert.

Einschätzung der Interviewten bezüglich P4P und weiteren Entwicklungen

Die Qualitätsaspekte der IV-Verträge werden für den Leistungserbringer vor allem als Marketinginstrument angesehen. Für eine gute Qualitätsdarstellung werden auch Qualitätsverbesserungen angestrebt.

Die Verbesserung der EDV-Unterstützung wird als wichtigste Entwicklung der nächsten Jahre gesehen. Die nächsten Schritte im Controlling sind die Überprüfung der Erhebung und Analyse der neuen Indikatoren. Anschließend werden diese Erfahrungen für mehr Verträge und standardisiert angefragt.

Neue Verträge werden derzeit mit neuen Indikatoren, wissenschaftlichen Ergebnissen und verbesserten Vergütungsfunktionen erstellt. Qualitätsorientierte Vergütungen nur im Rahmen einzelner Leistungen sind schwierig. Nur ein freiwilliges Bonus-System kann langfristig erfolgreich sein – vor allem aufgrund der notwendigen Akzeptanz des Leistungserbringers. Wenn es neue Entwicklungen und vernünftige Qualitätsindikatoren gibt, sollen auch Anreize anhand von Vergütungsanteilen gesetzt werden. Dafür muss allerdings die Qualitätsverbesserung auch einen positiven Effekt auf die Gesamtwirtschaftlichkeit haben. Auch die Leistungserbringer wollen die Qualität mehr abbilden, vor allem zu Transparenz, und weniger zu Vergütungszwecken.

Die Verständlichkeit der Qualitätsindikatoren für die Patienten und Versicherten ist zukünftig sehr wichtig, bspw. für die Darstellung in Internetportalen.

5.4.2.3. Auswertung anhand des Kriterienkatalogs

Da hier vor allem die Unterschiede zur ersten Fallstudie dargestellt werden, wird zunächst nur der Kriterienkatalog abgebildet und anschließend die Differenzen diskutiert. Einer der Kernunterschiede bei den Versorgungsbereichen und Indikatoren zur ersten Fallstudie besteht darin, dass hier neben den Revisionsoperationen von Hüft-TEPs auch kardiologische Indikatoren bestehen und entwickelt wurden. Zwar gibt es für diese Indikatoren, die im Rahmen der IV-Verträge gemessen werden, derzeit keine qualitätsorientierte Vergütung. Da dies jedoch in nähere Zukunft geplant ist, sollen diese hier auch analysiert werden. Zwar sind die Indikatoren noch nicht gänzlich entwickelt, werden aber zumindest auch prozessqualitative Aspekte enthalten – dies soll hier betrachtet werden.

Tabelle 18: Bewertungen Fallstudie 2, Systemelemente Gesamtsystem und Versorgungsbereich

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
1a	Abdeckung des Gesamtsystems	Wenig, Vertrag nicht als Gesamtsystem entwickelt	**Niedrig**	Betrachtete Verträge betreffen nur zwei Versorgungsbereiche	Mehr Versorgungs-bereiche mit hinzuziehen	Im selektiwertraglichen Rahmen kaum umsetzbar	*Zweiter Versorgungs-bereich, aber trotzdem kein Gesamtssystem*
2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet	Ja, indirekt durch Auswahl des Versorgungs-bereiches	**Hoch**	Kardiologie und Hüft-TEPs beide mit hoher Relevanz	Schwerpunktthemen anhand strategischer Prioritäten ausgewählt, Auswahl vor allem aufgrund wirtschaftlicher Aspekte	Weitere Daten und Studien hinzuziehen, auch nach "nicht-Volkskrankheiten" suchen	*Kardiologie und Hüft-TEPs, strategische Schwerpunkt-themen, Auswahl mehr durch Kasse*
2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich	Ja, indirekt durch Auswahl des Versorgungs-bereiches	**Hoch**	Kardiologie und Hüft-TEPs beides "Volkskrankheiten" mit hoher Relevanz	Schwerpunktthemen anhand strategischer Prioritäten ausgewählt, Auswahl vor allem aufgrund wirtschaftlicher Aspekte	Weitere Daten und Studien hinzuziehen, auch nach "nicht-Volkskrankheiten" suchen	*Kardiologie und Hüft-TEPs strategische Schwerpunkt-themen, Auswahl mehr durch Kasse*

Die Aussagekraft von Hüft-TEPs in dieser Fallstudie ist sehr hoch. Auch die kardiologische Krankheiten haben eine hohe ökonomische und medizinische Bedeutung (vgl. z. B. Spindler 2010, S. 350f.). Vor allem ist zu erwähnen, dass der Auswahlprozess die strategischen Schwerpunktthemen betrachtet und bereits einige Kriterien diskutiert werden.

Tabelle 19: Bewertungen Fallstudie 2, Systemelement Qualitätsindikator

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Revisionsoperation ist Ergebnisqualität, wird von Fachgesellschaften genutzt; kardiologische Indikatoren dagegen auch prozesslastig	Indikator nur zum Teil systematisch ausgewählt	Verbesserung Verständnis für Indikator, standardisierte Definition des Indikators	*Kardiologische Indikatoren, systematischere Auswahl*
3b	Relevanz des Indikators für ein bestimmtes Fachgebiet	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Revisionsoperation , wird von Fachgesellschaften genutzt; kardiologische Indikatoren eher prozesslastig	Gerade zukünftige Indikatoren gemeinsam mit gesundheits-ökonomischen Lehrstühlen ausgewählt	Verbesserung Verständnis für Indikator, standardisierte Nachverfolgung	*Kardiologische Indikatoren, systematischere Auswahl*
3c	Sektorbezug vs. sektorübergreifende Nutzung *(Informationskriterium)*	Teilweise, Vertragspartner ja, Intention unklar	**Hoch**	IV automatisch mehrere Sektoren, allerdings Kriterien speziell für stationären Sektor	Derzeit nicht sehr relevant für den Entwicklungsprozess	Einbeziehen in Überlegungen	*Keine*

Die Kriterienerfüllung für die orthopädischen Indikatoren (Revisionsoperationen) ist mit der ersten Fallstudie vergleichbar. Die kardiologischen Indikatoren dagegen sind vermutlich meist Prozessindikatoren. Diese werden sind zwar von den Fachpersonen anerkannt, Ergebnisindikatoren wie Mortalitätsraten könnten aber besonders für kardiologische Behandlungen eine höhere medizinische Relevanz haben.

Kriterienkategorie: Akzeptanz

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
4a	Medizinische Bedeutung in der medizinischen Fachliteratur	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Rev. werden in einer Vielzahl von Literatur als Indikator angewandt, kardiologische Indikatoren werden mit Lehrstuhl entwickelt	Sehr gut bewertbar, allerdings in Überlegungen nur indirekt enthalten	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	*Kardiologische Indikatoren, systematischere Auswahl u.a. mit Lehrstuhl*
4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Rev. werden auch von den Fachgesellschaften als Qualitätskriterium genutzt	Derzeit keine systematische Auswahl, Revision erfüllt Kriterium aber gut, wird vermutlich für Kardioindikatoren mit betrachtet	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium	*Kardiologische Indikatoren, systematischere Auswahl u.a. mit Lehrstuhl*
4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Rev. wird von Chefärzten als Indikator vorgeschlagen	Derzeit keine systematische Auswahl, Revision erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium, sollte für Kardioindikatoren betrachtet werden	*Kardiologische Indikatoren, systematischere Auswahl u.a. mit Lehrstuhl und Leistungserbringer*
4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Laienverständlichkeit der Rev. recht hoch, kardiologische Indikatoren nicht bekannt	Derzeit keine systematische Auswahl, Indikator erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium, sollte bei Kardioindikatoren betrachtet werden	*Kardiologische Indikatoren, systematischere Auswahl, Patienten-befragungen für Kasse interessant*
4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit	Ja, Revisions-operation und kardiologische Indikatoren	**Hoch bzw. mittel**	Wird von verschiedenen Stellen häufig genutzt, kardiologische Indikatoren nicht bekannt aber Patientenbefragungen integriert	Derzeit keine systematische Auswahl, Indikator erfüllt Kriterium aber gut	Standardisierte Auswahl der Indikatoren, Einbeziehung der Akzeptanz als Kriterium, sollte bei Kardioindikatoren betrachtet werden	*Kardiologische Indikatoren, systematischere Auswahl, Patienten-befragungen relevant*

Die Kriterienkategorie der Akzeptanz für die Gewährleistungen dieser Fallstudie unterscheidet sich auch nur gering für den Indikator der Revisionsoperation. Bei den kardiologischen Indikatoren dagegen ist die Akzeptanz schwerer zu beurteilen. Zwar sind die Auswahlprozesse standardisierter und systematischer – Akzeptanz und wahrgenommener Nutzern verschiedener Akteure wird jedoch nicht explizit als Auswahlkriterium genutzt. Gleichzeitig werden aber wissenschaftliche Akteure (Lehrstuhl) und

Leistungserbringer mit in die Entwicklung neuer Indikatoren einbezogen. Langfristig sollte die Akzeptanz dieser Akteure daher hoch sein – derzeit jedoch werden Prozessindikatoren in der Fachliteratur divers diskutiert, während die Akzeptanz von medizinischem Personal und Fachgesellschaften meist hoch ist (vgl. unter anderem Fonarow et al. 2007; Bradley et al. 2006; Werner/Bradlow 2006 und Mehta et al. 2008).

Tabelle 20: Bewertungen Fallstudie 2, Systemelement Qualitätsmessung

Kriterienkategorie: Wissenschaftlichkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
5a	Reliabilität	Kaum, Messung findet nicht automatisiert statt	**Niedrig**	Keine automatisierte Messung der Gewährleistungen, daher Definitionen nicht überprüfbar	Messung findet nicht standardisiert oder automatisiert statt	Messung standardisieren, klare Indikatorendefinition, Kriterium beachten	*Kaum Unterschiede zur Fallstudie 1, bei Kardioindikatoren nicht bekannt*
5b	Validität	Kaum, Messung findet nicht automatisiert statt	**Niedrig**	Keine automatisierte Messung der Gewährleistungen, daher Definitionen nicht überprüfbar	Messung findet nicht standardisiert oder automatisiert statt, keine Entwicklung relevant	Messung standardisieren, klare Indikatorendefinition, Kriterium beachten	*Kaum Unterschiede zur Fallstudie 1, bei Kardioindikatoren nicht bekannt*
5c	Risikoadjustierung und Risikoadjustierbarkeit	Trifft nur auf kardiologische Indikatoren zu	*nicht bekannt*	*nicht bekannt*	*nicht bekannt*	*nicht bekannt*	*Kaum Unterschiede zur Fallstudie 1, bei Kardioindikatoren notwendig*

Da auch die Evaluation und Analyse der Gewährleistung der zweiten Fallstudie nicht regelmäßig, standardisiert oder automatisiert durchgeführt wird und gleichzeitig die Entwicklung der Qualitätsmessung keinen wissenschaftlichen Kriterien unterliegt, ist auch hier die Validität und Reliabilität niedrig. Die Risikoadjustierung trifft auf die einzelfallbezogenen Revisionsoperationen nicht zu, die genauen Prozesse zur Risikoadjustierung der kardiologischen Indikatoren sind nicht bekannt. Da die genauen Messmethoden der kardiologischen Indikatoren nicht bekannt sind, können hier keine Aussagen zur Reliabilität und Validität getroffen werden.

Kriterienkategorie: Messaufwand / -sicherheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
6a	Objektiver Erhebungsaufwand und -komplexität	Ja, theoretischer Aufwand bewertbar / praktisch findet Erhebung nicht statt	**Niedrig (pot. Mittel)**	Erhebung findet nicht automatisiert statt, nur eine wissenschaftliche Erhebung mit hohem Aufwand / Prozessindikatoren der Kardiologie aufwendiger	Derzeit keine Auswertung, langfristig geplant und (nach IT-Investition) für Revisionsoperationen aufwandsarm möglich, Prozessindikatoren recht aufwendig	Wirtschaftlichkeit der IT-Investition prüfen und bei positiver Bewertung durchführen, für neue Indikatoren als Kriterium betrachten	*Kardiologische Indikatoren vermutlich höherer Aufwand, da prozesslastig - ist der Kasse als Kriterium sehr wichtig*
6b	Manipulationssicherheit / Objektivität der Messung	Ja, theoretisch bewertbar	**Niedrig, für Revisions-operation pot. Hoch, für Kardio-indikatoren pot. Mittel**	Erhebung nicht automatisiert, Daten für Revisionsoperationen würden aber auf §301 basieren, Prozessindikatoren müssten überprüft werden	Datenbasis recht manipulationsresistent, da basierend auf Routinedaten - Nicht-Dokumentation möglich aber im Sinne der Vergütungsfunktion unschädlich, allerdings Verfälschung der Statistik	Für Vergütung gut, für statistische Auswertungen Mechanismen für eine garantierte Dokumentation einbauen	*Kardiologische Indikatoren eventuell geringere Manipulations-sicherheit da prozesslastig - ist der Kasse als Kriterium wichtig*
6c	Grad der Integration in automatische Prozesse / IT-Systeme	Ja, anhand Interviews und Beobachtung der Prozesse	**Niedrig (pot. technisch hoch, wirtschaft-lich mittel)**	Erhebung und Erkennung im Controlling besser automatisiert als in Fallstudie 1 aber noch nicht ausreichen	Automatisierung prinzipiell möglich, für prozesslastige Indikatoren ist die Integration aufwendiger aber auch möglich, derzeit IT-Investitionen geplant, dann Verbesserung	Wirtschaftlichkeit der IT-Investition prüfen, parallel Wirtschaftlich-keit durch Standard-isierung der Verträge und Messverfahren verbessern	*Wie Fallstudie 1, da auch die Prozessindika-toren integriert werden können, dies aber zu höheren Umstellungs-kosten*

Die Kriterien des Messaufwands und der Messsicherheit sind für die Revisionsoperationen identisch zur ersten Fallstudie. Auch hier gibt es keine automatisierten Erhebungen – die existierende wissenschaftliche Studie ist mit hohem Aufwand und wissenschaftlicher Unterstützung entstanden. Prinzipiell ist für die Revisionsoperationen auch hier der mögliche Grad der Integration in EDV-Systeme hoch – die Umsetzung und Erkennung der Gewährleistungsfälle ist allerdings schon jetzt höher automatisiert als in Fallstudie 1, außerdem sind weitere Verbesserungen geplant. Bei den derzeit im Entwicklungsprozess stehenden kardiologischen Kriterien jedoch sind vermutlich Prozessindikatoren integriert die zu höherem Erhebungsaufwand und manipulationsunsicheren Messergebnissen führen können.

Tabelle 21: Bewertungen Fallstudie 2, Systemelement Qualitätsziele

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
7a	Zielkonsistenz	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen , kardiologische Indikatoren nicht bekannt	**Pot. Hoch**	Klares Ziel der Reduktion von Revisionsoperationen (aber eher als Kostenziel)	Hohe Zielkonsistenz beim Leistungserbringer und Kasse möglich, da Reduktion von Revisionsoperationen bei allen erwünscht ist, sobald Vergütung angepasst ist	Klarere Definition der Ziele	*Wie Fallstudie 1, kardiologische Indikatoren nicht bekannt, sollte in der Entwicklung beachtet werden*
7b	Zielkongruenz	Nein, da derzeit nur ein Ziel vorhanden, kein Zielsystem	**Nicht bewertbar**	*Nicht bewertbar*	*Nicht bewertbar*	*Nicht bewertbar*	*Wie Fallstudie 1*
7c	Reichweite	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen , kardiologische Indikatoren nicht bekannt	**Niedrig**	Zwar ist durch den Versorgungsbereich ein recht hoher Anteil des Fachbereichs abgedeckt, der Gesamtanteil des Ziels für Gesamt-unternehmen ist aber gering	Unterscheidung zwischen Fachabteilung und Gesamtunter-nehmen notwendig, da kein Zielsystem sondern nur Vertrag, eher gering, außerdem statistische Fallzahlproblematik	Zielsystem erweitern und Punktesystem (Scoring), ggf. Zusammenfassung mehrerer Häuser (Regionen) oder Kassen	*Wie Fallstudie 1, gilt vermutlich auch für kardiologische Indikatoren*

Auch für die Qualitätsziele ist diese Fallstudie der ersten sehr ähnlich. Die genaue Zielsetzung für die kardiologischen Indikatoren ist nicht bekannt – die Kriterien der Konsistenz, Kongruenz und Reichweite sollten aber im Entwicklungsprozess Beachtung finden.

Kriterienkategorie: Operationalisierbarkeit und Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
8a	Erreichbarkeit	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen , kardiologische Indikatoren nicht bekannt	**Mittel**	Keine Revisionsoperation zu erreichen ist unrealistisch, Reduktion der Rate aber in Deutschland erreichbar	Muss mit Leistungserbringer abgesprochen werden, Ziel sollte daher nicht als gänzliche Vermeidung empfunden werden	Klare Definition der Ziele (dadurch kein "Empfinden" der gänzlichen Vermeidung), kardiologische Indikatoren auf Erreichbarkeit zu prüfen	*Wie Fallstudie 1, kardiologische Indikatoren nicht bekannt, sollte in der Entwicklung beachtet werden*
8b	Transparenz	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen , kardiologische Indikatoren nicht bekannt	**Hoch**	Ziele sind klar definiert und verständlich	Hier nicht bekannt, wer beim Leistungserbringer die Ziele kennt	Klare Definition der Ziele und transparente Darstellung	*Wie Fallstudie 1, gilt vermutlich auch für kardiologische Indikatoren*
8c	Akzeptanz	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen , kardiologische Indikatoren nicht bekannt	**Mittel**	Generell hohe Akzeptanz (siehe auch Indikator), aber bei gänzlicher Vermeidung ist Akzeptanz gering, da nicht erreichbar, eher prozesslastige Indikatoren häufig gut akzeptiert	Leistungserbringer muss vermittelt werden, dass Ziel nicht 0% Revisionsrate ist	Klare Definition der Ziele und transparente Darstellung, Einbeziehung aller Beteiligter	*Wie Fallstudie 1, kardiologische Indikatoren zusätzlich*
8d	Langfristige Anreizwirkung der Ziele	Schwierig, da kein Entwicklungs-system	**Mittel**	Kein Zielsystem mit ggf. Anpassungen der Ziele zu finden, aber Reduktion der Revisionsrate und Beibehaltung einer niedrigeren Rate kann inzentivieren	Generell Ziele eher zu hoch als zu niedrig, keine Entwicklung integriert	Lieber niedrigere Ziele mit Entwicklung, Rankings, etc.	*Wie Fallstudie 1, gilt vermutlich auch für kardiologische Indikatoren*
9a	Gerechtigkeit	Ja, wenn die Vermeidung von (jeglichen) Revisions-operationen als Ziel gesehen , kardiologische Indikatoren nicht bekannt	**Mittel**	Die Nichtadjustierung der Gewährleistung wird von den Leistungserbringern teilweise als ungerecht empfunden (Aus Patientensicht wird die Umsetzung der Garantien häufig als gerecht angesehen), kardiologische Indikatoren müssen adjustiert werden	Revisionsoperationen nach Hüft-TEPs nicht risiko-adjustiert da es sich um Einzelfallindikatoren handelt, Aus- und Einschlusskriterium mit den Leistungserbringern ggfs. zu besprechen	Anpassungen mit integrieren	*Wie Fallstudie 1, kardiologische Indikatoren nicht bekannt, sollte in der Entwicklung beachtet werden*

Auch hier sind die Kriterienerfüllungen der Gewährleistungen mit der ersten Fallstudie vergleichbar. Für die kardiologischen Indikatoren sollten diese Aspekte, vor allem die Akzeptanz und Gerechtigkeit, durch valide und aussagekräftige Adjustierungen mitentwickelt werden.

Tabelle 22: Bewertungen Fallstudie 2, Systemelement Vergütungsfunktion

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
10a	Effizienz	Im konkreten Fall schwierig, Vergütungs-funktion an sich zwar klar, aber keine Daten zur Effizienz, daher nur Abschätzung	**Niedrig**	Daten für Kasse nicht vorhanden, aber anhand von Einschätzung in Interviews ist die Effizienz eines der Hauptprobleme	Zu klein mit zu hohen Investitions- und laufenden Kosten um wirklich effizient zu sein, wird außerdem von Kasse nicht analysiert	Analyse der Effizienz, Einbeziehen Gesamtkosten, Controlling der Effizienz	*Wie Fallstudie 1, keine Datenanalyse für Dissertation erstellbar*
10b	Effektivität	Ja, da Daten vorhanden, aber Analyse mit Schwierigkeiten	**Niedrig**	Keine Datenanalyse aber mit Qualitätsverbesserungen wird nicht gerechnet	Soweit anhand der vorhandenen Daten erkennbar, liegt die Revisionsrate auf einer zum Durchschnitt vergleichbaren Höhe	Standardisierte Daten-analyse, Konsequenzen-management bei Nicht-Erreichung der Ziele einführen	*Wie Fallstudie 1, keine Datenanalyse für Dissertation erstellbar*
10c	Flexibilität	Ja, Vergütungs-mechanismus ist klar	**Mittel**	"Preis" der Vergütung richtet sich nach DRG-Katalog und Basisfallwert, daher absoluter Wert, der aber an die Hausgröße zumindest indirekt angepasst ist, aber keine weitere Anpassung vorgesehen	Nicht wirklich in Entwicklung integriert	Einbeziehen in Überlegungen, ggf. anpassen	*Wie Fallstudie 1, keine Datenanalyse für Dissertation erstellbar*

Für die zweite Fallstudie konnte keine spezielle Datenanalyse stattfinden. Daher wird die Effizienz und Effektivität aufgrund von Interviewaussagen bewertet – die Kriterienerfüllung entspricht aber der der ersten Fallstudie.

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
11a	Anteil der Vergütung	Kaum, Vergütungs-funktion an sich zwar klar, aber keine Daten zur Gesamt-vergütung, nur Abschätzung möglich	**Niedrig**	Daten für Kasse nicht vorhanden, aber anhand Abschätzung ist der Vergütungsanteil eines der Hauptprobleme	Wird nicht analysiert, aber bereits ersichtlich, dass die qualitätsorientierte Vergütung keinen hohen Stellenwert innehat	Kollektivvertragliche Regelungen, mehrere Krankenhäuser, etc.	*Wie Fallstudie 1, für Kardioindikatoren nicht bekannt*
11b	Planbarkeit	Ja, Weiterentwick-lung der Vergütungs-funktion zwar nicht dargelegt, aber relativ klar	**Mittel bis hoch**	Anzahl von Revisionsoperationen beim Leistungserbringer ermittelbar und abgesehen von statistischen Streuungen gut planbar, Preisentwicklung einschätzbar	Planbarkeit im statistischen Mittel gut, angesichts nicht zu vermeidender statistischer Streuung bei niedrigen Fallzahlen sind aber die tatsächlichen Risiken im jeweiligen Berichtsjahr nur schwer vorhersehbar	Einbeziehen in Überlegungen und Entwicklung, ggf. Techniken zur Risikobegrenzung (z.B. Bildung von Rückstellungen, Begrenzung der Haftung, Rückversicherung)	*Wie Fallstudie 1, für Kardioindikatoren nicht bekannt*

Es gibt kaum Unterschiede zur ersten Fallstudie für die Vergütungsfunktion. Dies ist vor allem in der Tatsache begründet, dass bislang nur für die Indikatoren der orthopädischen IV-Verträge tatsächlich ein Konsequenzen- und Vergütungsmanagement besteht. Daher können auch nur hier die Gewährleistungsklauseln bei Revisionsoperationen analysiert werden. Für die neuen kardiologischen Indikatoren ist derzeit nicht klar, welche Art von Vergütungsfunktion integriert wird, daher können diese nicht analysiert werden. Die Kriterienerfüllung bleibt somit identisch zur ersten Fallstudie.

Tabelle 23: Bewertungen Fallstudie 2, Systemelement Messergebnis

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
12a	Nutzung zur Steuerung	Ja, Nutzung diskutiert und Nutzbarkeit relativ klar	**Niedrig bzw. Mittel**	Wird weder in der Kasse noch beim Leistungserbringer zur Steuerung genutzt (nach Einschätzung der Kasse)	Revisionsoperationen aufgrund des Einzelfallcharakters nur bei größerer Zahl für die Kasse zur Steuerung nutzbar, kardiologische Prozessindikatoren für Leistungserbringer gut nutzbar	Revisionsoperationen nur schwer nutzbar, kardiologische Indikatoren können besser genutzt werden wenn bspw. Managementsysteme integriert werden	*Kardiologische Indikatoren wären bei Prozesslastigkeit gut nutzbar*
12b	Aktualität	Mittel, potenzieller Messvorgang recht klar aber angesichts des praktischen Umsetzungsgrad es nicht praktisch bewertbar	**Mittel**	Zwar finden die Rechnungskürzungen vermutlich fortlaufend und somit zeitnah statt, wenn es allerdings zu Unstimmigkeiten kommt, kann der Prozess sehr lang dauern, Prozessindikatoren für Leistungserbringer sehr aktuell, für Kasse erst bei Übermittlung und ggf. nach Überprüfung	Bei Streitfällen kann es eine sehr langfristige Abwicklung geben, Prozessindikatoren bei Kardio wären auch vom Leistungserbringer evaluierbar	Durch klarere Definition sollten Streitfälle möglichst vermieden werden, Prozessindikatoren können schnell übermittelt werden	*Wie Fallstudie 1, kardiologische Indikatoren zusätzlich*
12c	Kontrollierbarkeit	Mittel, potenzieller Messvorgang recht klar aber angesichts des praktischen Umsetzungsgrad es nicht praktisch bewertbar	**Mittel (pot. Hoch)**	Derzeit findet kein standardisiertes Controlling der Gewährleistungen statt, Berichte werden vor allem zu Fallzahlen, Versichertenzahlen und Zahlungen erstellt, aber weiter entwickeltes Controllingsystem als in Fallstudie 1	Controlling könnte - wenn automatisierte Erhebungen stattfinden würden - sehr leicht integriert werden	Berichtswesen könnte ggf. sogar verpflichtend sein, bei höherer Relevanz würde es automatisch mehr Berichte geben	*Besser als in Fallstudie 1*
12d	Praktische Umsetzung im Controlling	Ja, derzeitige Evaluationen bekannt	**Mittel**	Bessere automatische Erkennung der Gewährleistungen, aber auch hier vermutlich nicht vollzählig	Softwareänderungen für bessere automatische Erkennung in Planung	Aufbau eines standardisierten Qualitätsberichtswesens höher priorisieren	*Höher priorisierte Software-anpassung*
12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit	Ja, Zuordnung Messergebnis klar	**Hoch**	Sowohl auf Krankenhaus als auch auf Fachabteilungsebene ist die Zuordnung des Messergebnisses der Revisionsoperation klar	Auf Mitarbeiterebene wäre es schwieriger, das ist aber erst beim Leistungserbringer relevant	Nicht erforderlich, bei kardiologischen Indikatoren zu beachten	*Wie Fallstudie 1, kardiologische Indikatoren nicht bekannt, sollte in der Entwicklung beachtet werden*

Die Messergebnisse der Revisionsoperationen unterscheiden sich nur wenig von der ersten Fallstudie, allerdings sind die derzeitige Umsetzung im Controlling sowie die Kontrollierbarkeit aufgrund von weiterentwickelten Erkennungsmechanismen der Gewährleistungsfälle höher. Auch für die kardiologischen Indikatoren könnten die Erfüllungsgrade der Kontrollierbarkeit, praktische Umsetzung des Controllings und Aktualität hoch sein, auch können Prozessindikatoren teilweise gut für die Steuerung genutzt werden – die Vergütung sollte allerdings weiterhin anhand von Ergebnisqualität stattfinden.

Kriterienkategorie: Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Bewertung	Verbesserungen	Unterschied zu Fallstudie 1
13a	Zeitverlauf	Ja, zwar derzeit nicht vorhanden, Möglichkeiten aber klar	**Niedrig**	Generell könnten die Messergebnisse natürlich in zeitlichen Zusammenhang gesetzt werden, allerdings ist durch die geringe Fallzahl die Aussagekraft gering	Derzeit findet keine Zeitverlaufsanalyse statt	In die Betrachtung einbeziehen, auch bei Kardioindikatoren	*Wie Fallstudie 1, kardiologische Indikatoren nicht bekannt, sollte in der Entwicklung beachtet werden*
13b	Querschnittsvergleich/Ranking	Ja, derzeit aber nicht vorhanden und aufgrund der Verträge nicht vorgesehen	**Niedrig (pot. Mittel)**	Wegen individueller Gestaltung der Verträge ist ein Ranking derzeit praktisch kaum möglich	Krankenkasse könnte Analysen machen, allerdings fallen die Selektivverträge oft unterschiedlich aus, so dass ein Vergleich kaum möglich wäre	Nur in Kollektivvertrag oder bei zumindest kassenbezogener Standardisierung möglich	*Für neu zu entwickelnde kardiologische Indikatoren nur nutzbar, wenn für mehrere Verträge genutzt*
13c	Justiziabilität	Ja, derzeit aber nur selten umgesetzt	**Hoch**	Der Revisionsfall ist klar und eindeutig beschrieben und könnte daher (bei Erkennung) auch gerichtlich durchgesetzt werden	Regelung bezüglich Justizierbarkeit gut	Nicht erforderlich	*Wie Fallstudie 1, kardiologische Indikatoren nicht bekannt, sollte in der Entwicklung beachtet werden*

Während der Zeitverlauf und die Justizierbarkeit mit der ersten Fallstudie vergleichbar sind, kann ein Querschnittsvergleich für die eigens entwickelten kardiologischen Indikatoren besser möglich sein, wenn diese für mehrere Verträge genutzt werden.

- <u>Insgesamt</u>

Die eigentliche Umsetzung ist zwar im Detail verschieden, die Bewertung der einzelnen Kriterien verändert sich dadurch aber nur teilweise. Dies kann vor allem dadurch erklärt werden, dass die Kriterien des zweiten Katalogteils vor allem die praktische Umsetzung des Systems bewertet – aufgrund der systemischen Grundbedingungen in Deutschland ist gerade diese Umsetzung derzeit noch nicht besonders fortgeschritten. Denn auch in dieser Fallstudie wird deutlich, dass zwar viele Verbesserungen (bspw. der EDV-Unterstützung) geplant sind, die Prioritäten aber bei Umsetzungsfragen des Kollektivvertrages liegen. Trotzdem ist darauf hinzuweisen, dass die Erkennung von Gewährleistungen in dieser Fallstudie besser funktioniert als in Fallstudie 1. Auch die Tatsache, dass es sich um eine größere Kasse handelt, hilft bei der systematischeren, strategischeren Auswahl von Versorgungsbereichen. Einer der Kerngründe hierfür ist die deutlich höhere wirtschaftliche Relevanz der kollektivvertraglichen Vergütung.

Gleichzeitig wird sichtbar, dass die vermuteten prozesslastigen Indikatoren der kardiologischen Verträge einige Kriterien der Indikatoren schlechter, einige Umsetzungsparameter zumindest gleich gut erfüllen können – hier ist aber darauf zu achten, dass die Prozessindikatoren eher als flankierende Unterstützung gesehen werden. Die Vergütung sollte weiterhin auf Ergebnisindikatoren aufbauen.

5.4.3. Fallstudie 3: Krankenkasse C

5.4.3.1. Beschreibung der IV-Verträge

Die dritte Fallstudie betrifft eine Krankenkasse mit verschiedenen regionalen Geschäftsstellen. Die Geschäftsstellen sind recht eigenständig organisiert und können freiwillig an gemeinsamen Verträgen teilnehmen. Sie sind häufig eher klein und haben nur eine geringe Anzahl von Versicherten[1].

In der Region gibt es 14 IV-Verträge mit insgesamt knapp 500.000 Versicherten darunter auch sehr große Verträge, vor allem der hausarztzentrierten Versorgung. Nur ein Vertrag fokussiert sich auf den stationären Bereich und betrifft endoprothetische Behandlungen mit Gewährleistungen und somit eine Vergütungsfunktion. Daher wird dieser Vertrag hier näher betrachtet. Der Vertrag beinhaltet mehrere Indikatoren der Ergebnisqualitätsziele, unter anderem: Reduktion der Anzahl stationärer Wiederaufnahmen, Verkürzung der Verweildauer, Reduktion der Rezidive, Verbesserung der Patientencompliance, Verbesserung der sektoralen Kommunikation, Verbesserung der sektoralen Kooperation. Auch Qualitätssicherungsklauseln sind inkludiert – hier sind bspw. gesetzliche Mindestvoraussetzungen sowie die Zertifizierung nach DIN ISO 9001 vorgegeben. In dem Vertrag sind ca. 20-30 Versicherte eingeschrieben. Die Gewährleistungsregelungen beinhalten eine 5-Jahresgarantie und verschiedene Ausschlusskriterien für Versicherte, so bspw. schwere Allgemeinerkrankungen (Diabetes, neurologische Begleiterkrankungen, etc.), Gerinnungsstörungen, akutes Trauma, Osteoporose, technisch aufwendige Implantationen, Sonderimplantate, erhebliche Zweifel an der Rehabilitationsfähigkeit. Auch Infektionen, Lockerungen oder Fehllagen der durchgeführten Revisionsoperationen werden ausgenommen. Bislang allerdings sind keine Gewährleistungsfälle bekannt.

Da in dieser Fallstudie nur ein recht kleiner Vertrag im stationären Bereich mit Vergütungsfunktion besteht, wird dieser zwar fokussiert betrachtet, andere Verträge mit Vergütungskonsequenzen aus anderen sektoralen Bereichen werden aber an dieser Stelle kurz beschrieben.

Ein Vertrag, der aufgrund des ambulanten Charakters hier nicht analysiert wird aber bereits recht weit in der Bonus-Thematik entwickelt wurde, ist ein Vertrag der Reproduktionsmedizin. In diesem Vertrag sind ca. 4.000 Versicherte eingeschrieben. Als Grundlage gelten die Behandlungsleitlinien sowie die Richtlinien des SGB V. Umfangreiche Qualitätsberichte werden von einem unabhängigen Institut erstellt, und eine Vergütung erfolgt nur bei erfolgreicher Schwangerschaft.

Der zweite Vertrag versucht, die Arbeitsfähigkeit nach Fällen von Rückenschmerzen zu erhöhen. Gerade die Thematik der Patientensteuerung ist hierfür relevant. Der Vertrag führt derzeit zu einem sehr positiven Ergebnis: 99 % der Patienten sind bei guter Compliance wieder arbeitsfähig. Aber gerade hier

[1] Aufgrund von Wünschen zum Datenschutz wurden für diese Fallstudie die Interviews nicht aufgenommen oder transkribiert – die Analyse findet anhand eines Gedächtnisprotokolls statt.

zeigt sich, dass verschiedene Umsetzungsschwierigkeiten bestehen, bspw. wenn der Patient den niedergelassenen Arzt wechselt oder die Teilnahme am Programm vergisst. Während die Auswahl des Leistungserbringers auf Grundlage von Qualitätsparametern gut funktioniert, ist die Auswahl und Steuerung der Patienten sehr aufwendig und langfristig kaum wirtschaftlich möglich.

Der letzte relevante Vertrag betrachtet die radiochirurgische Therapie zur Verhinderung von OPs. Ähnlich wie bei der Endoprothetik werden auch hier möglicherweise auftretende Rezidive nicht zusätzlich vergütet sondern mit der zuvor vereinbarten Pauschale dauerhaft abgegolten. Die Attraktivität für die Patienten entsteht allerdings vor allem durch die Neuartigkeit der Behandlungsmethode, die qualitätsorientierten Anreize sind hier kaum relevant. Insgesamt ist dieser Vertrag noch sehr klein, soll aber ausgeweitet werden.

5.4.3.2. Erste Ergebnisse der Interviews

Entwicklungsprozesse der IV-Verträge mit qualitätsabhängiger Vergütung

Alle Verträge werden von einer übergreifenden Arbeitsgemeinschaft entwickelt und abgeschlossen. Die einzelnen Kassen können diesen Verträgen freiwillig beitreten. Diese Arbeitsgemeinschaften sind regional und eigenständig organisiert und agieren als „Lobby“ für die Verträge gegenüber den Einzelkassen. Die Verträge der Arbeitsgemeinschaften werden oft vom Leistungserbringer angestoßen.

Zwar werden insgesamt nur wenige Verträge entwickelt, diese sind aber sehr detailliert analysiert und mit einer Vielzahl von Personen bei den Kostenträgern und bei Leistungserbringern diskutiert. Allerdings müssen die meisten Verträge im Rahmen einer Ausschreibung abgeschlossen werden, wodurch der zeitliche und administrative Aufwand deutlich steigt, die Wahrscheinlichkeit des Abschlusses jedoch sinkt.

Gründe für die Vertragsschließung

IV-Verträge werden von der Kasse nicht als besonders wirtschaftliches oder effektives Instrument eingestuft. Wenn dadurch die Leistungen extrabudgetär abgerechnet werden können, ist gerade für den Kassenverband der Gesamteffekt meist kaum wirtschaftlich. Für die einzelne Kasse dagegen kann sich der Vertrag aufgrund des Rabattes lohnen. Insgesamt führt allerdings gerade die geringe Anzahl Versicherter dazu, dass die Relevanz niedrig bleibt – gerade im Vergleich zur Regelversorgung. Nur die hausarztzentrierte Versorgung ist für die strategischen Entscheidungen der Kasse bedeutsam.

Gleichzeitig jedoch ermöglichen es Selektiv- und vor allem IV-Verträge, innovative, eigenständige Ideen umzusetzen. So können sich Kassen von anderen abgrenzen und dies als Wettbewerbsvorteil nutzen. Außerdem bietet dies die Möglichkeit, mit dem Leistungserbringer gemeinsam Projekte zu realisieren. Gerade dieser Aspekt ist dem Befragten sehr wichtig: Mit dem Leistungserbringer zusammen zu arbeiten und die Versorgung aktiv mitzugestalten, ist für die tägliche Arbeit entscheidend und wird sehr geschätzt.

Insgesamt ist Qualität für Selektivverträge ein wichtiges Thema, aber vor allem für die Kommunikation mit den Versicherten. Gleichzeitig sind Selektivverträge die einzige Möglichkeit für Krankenkassen, Patientensteuerung tatsächlich umzusetzen.

Evaluation und Monitoring der Verträge

Ein standardisiertes Monitoring der Verträge besteht nicht – vor allem, da die meisten Verträge eine zu geringe Fallzahl haben, um relevante Auswertungen durchführen zu können. Wenn Analysen durchgeführt werden, betreffen sie vor allem die wirtschaftlichen Auswirkungen der Rabatte und dem Vergleich mit Mehr- und Mindererlösausgleichen. Die analysierten Daten sind: Versichertenanzahl, Entgeltschlüssel, Entgelthöhe, stationäre und ambulante Daten, BQS-Daten. Gewährleistungsfälle würden auch abgebildet werden, bislang fanden aber keine statt. Diese Vertragsanalysen müssen mit großer Vorsicht ausgewertet werden, da ein „Schönrechnen“ von Verträgen durchaus möglich ist. Transaktionskosten der Verträge werden in der wirtschaftlichen Analyse nicht berücksichtigt, da dann vermutlich kein Vertrag mehr wirtschaftlich lohnend wäre.

Beschreibung der Systemelemente

- Versorgungsbereich

Wie bei den vorherigen Fallstudien ist die Auswahl der Versorgungsbereiche auch bei Fallstudie 3 meist den Ideen von Leistungserbringern zu zuordnen. Im Auswahlprozess sind medizinische Kollegen des Kostenträgers beteiligt und überprüfen die medizinische Relevanz. Kernauswahlkriterien jedoch sind auch hier die potenzielle Fallzahl, Kosten und Einsparungen – d. h. eine wirtschaftliche Analyse. Die ersten Verträge, die im Bereich der Endoprothetik geschlossen wurden, konnten gerade zu Anfang schnell und einfach abgeschlossen werden und ermöglichten stationäre Reha bei der sektorübergreifenden Versorgung.

- Qualitätsindikator

Bei der Auswahl der Qualitätsindikatoren wird unter anderem betrachtet, dass eine zu hohe Anzahl von Indikatoren eine große und somit nicht mehr kontrollierbare Komplexität verursacht. Außerdem müssen die Ergebnisse der Indikatoren in den eigenen Daten auffindbar sein, um eine hohe Anwendbarkeit zu erlangen. Daher ist beispielsweise „Wiederaufnahme“ ein in verschiedenen Verträgen genutzter Qualitätsindikator.

Da bereits Verträge aufgrund eines untragbar hohen Dokumentationsaufwandes beendet wurden, ist auch dieses Kriterium bei der Indikatorauswahl entscheidend. So wird beispielsweise versucht, Fragebögen sowohl für Ärzte als auch für Patienten zu vermeiden.

- Qualitätsmessung

Bis 2011 war für mehrere Verträge die Datenerhebung sehr komplex, da die meisten Abrechnungen der Einzelverträge in Papierform stattfanden. Vor allem der übergeordnete Verbund konnte so Qualitätsmessungen kaum selbst durchführen. Nach viel Aufwand ist heutzutage eine Datenanalyse möglich. Bisher war somit nicht der Messaufwand sondern die Messmöglichkeit vorrangiges Thema. Verträge der Endoprothetik sind deutlich einfacher zu messen, derzeit findet aber keine standardisierte Analyse von Gewährleistungsfällen statt.

- Messergebnis

Da es bislang keine Messergebnisse für die Gewährleistungen gibt und die anderen nicht-stationären Verträge nicht bis zum letzten Detail analysiert werden konnten, wird dies hier nicht näher ausgewertet.

- Qualitätsziel

In den Verträgen werden zwar Vertragsziele gesetzt, die eine Verbesserung verschiedener Ergebnisindikatoren beinhalten. Jedoch sind diese Ziele nicht mit Vergütungsfunktion hinterlegt, so dass die Vergütungsziele derzeit nur in der Malus-Regelung der Garantie zu finden sind. Diese sind identisch mit den vorherigen Fallstudien. Wie auch in den vorherigen Fallstudien, gibt es auch hier eine Auswahl der beteiligten Leistungserbringer anhand von Fallzahlen und Qualitätskriterien.

- Vergütungsfunktion / Konsequenzenmanagement

Für die Vergütungsfunktion der Gewährleistungsklauseln der Endoprothetik-Verträge sind die gleichen Aspekte relevant wie bei den vorherigen Fallstudien. Vor allem weil es hier noch keine Gewährleistungsfälle gab, sind die Einschränkungen des Konsequenzenmanagements sichtbar. Da gerade bei „Vertragsverletzung“, d. h. Nicht-Erreichen der oben genannten Qualitäts- und Vertragsziele, die Maßnahmen nur qualitativ sind, wird nicht mit großen Einspareffekten gerechnet. Für den Mangel an Effizienz und Effektivität werden drei Hauptgründe angegeben:

- Fehlendes Konsequenzenmanagement
- Effekte der IV-Verträge eher prozessuale als qualitative Verbesserungen
- Hohe Transaktionskosten für Vertragsentwicklung, Messung und Auswertung

Um zukünftig ein effektives Vergütungsmanagement erreichen zu können, wird empfohlen, vor allem darauf zu achten, dass ein solcher Vertrag budgetrelevant umsetzbar ist und sich für alle Akteure insgesamt lohnt.

Praktische Umsetzung

Die Datenerhebung und der Datenzugriff sind ein wichtiges Thema in dieser Fallstudie und werden vom übergeordneten Regionalverband versucht, so weitreichend wie möglich umzusetzen.

Bei IV-Verträgen ist die erste Herausforderung die korrekte Abrechnung der Fälle beim Leistungserbringer. Bei der Krankenkasse wird geschätzt, dass ca. 90 % der Fälle der IV-Verträge doppelt abgerechnet werden (d. h. sowohl für die Regel- als auch die Selektivversorgung). Hierfür sind Controllingmechanismen integriert, die in den einzelnen Kassen unterschiedlich umgesetzt werden. IT-Unterstützung gibt es für diese Fragestellung nicht, allerdings wird im internen Datawarehouse die Regel- und Selektivversorgung kontinuierlich miteinander verglichen und Überschneidungen werden eliminiert.

Die Erkennung von potenziellen und tatsächlichen Gewährleistungsfällen ist auch in dieser Fallstudie nicht automatisiert. Zwar müsste es eine automatisierte IT-Erkennung geben, die tatsächliche Erkennung hängt dann aber von der Aufmerksamkeit des Sachbearbeiters ab. Wie in der vorherigen Fallstudie ist auch hier die Erkennung von Gewährleistungsfällen, die in anderen als den Vertragskrankenhäusern stattgefunden haben, nicht möglich. Insgesamt zeigt sich, dass für den überregionalen Verband das Thema der Gewährleistungserkennung derzeit wenig thematisiert wird, vorwiegend weil bislang keine Gewährleistungsfälle bekannt geworden sind.

Auch bei dieser Fallstudie beschäftigen sich die Kollegen aus dem Controlling eher mit Fragestellungen aus der Regelversorgung, da die Selektivverträge vergleichsweise wenig Fallzahlen haben (bis auf Hausarztverträge). Aufgrund dieser „Mickey-Mouse-Verträge“ liegt die wirtschaftliche Priorität der Umsetzung stets bei der Regelversorgung.

Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen

Insgesamt wird die Entwicklung von P4P als sehr interessant und innovativ eingestuft. Gleichzeitig gibt es jedoch eine Vielzahl an Bedenken zur derzeitigen Umsetzung:

Die meisten Selektivverträge mit P4P-Aspekten sind heutzutage vor allem „Marketingverträge“. Verträge, die versorgungsrelevant sein könnten, sind für die Kasse häufig wirtschaftlich kaum interessant, da damit vor allem kranke Versicherte angezogen werden. Gleichzeitig ist der administrative und finanzielle Aufwand für solche Verträge unverhältnismäßig hoch. Insgesamt wird somit angezweifelt, ob man derzeit einen P4P-Vertrag budgetrelevant umsetzen kann.

Besonders schwierig ist nicht die Auswahl der Leistungserbringer, sondern die Auswahl der Versicherten, die sich durch einen solchen Vertrag steuern lassen. Derzeit konnte die Feststellung, dass Patienten nicht steuerbar sind, nicht widerlegt werden.

Gerade durch die begrenzten Möglichkeiten des Controllings gehen in dieser Thematik die Praxis und Theorie diametral auseinander. Dies wird durch die schwierige Integration der IV-Verträge in die Softwareprogramme verstärkt.

Im Endeffekt werden viele der P4P-Projekte daher nicht umgesetzt, da angenommen wird, dass dies negative Konsequenzen bei der Zusammenarbeit mit den (politisch und wirtschaftlich sehr relevanten) Leistungserbringern nach sich ziehen könnte.

Es wird davon ausgegangen, dass langfristig P4P-Aspekte in den Kollektivvertrag integriert werden. Hierfür ist eine Gesetzesänderung notwendig. Für diese Entwicklung muss darauf geachtet werden, dass zum einen die richtigen Anreize gesetzt werden und zum anderen die Balance der Interessen von Leistungserbringer und Krankenkasse gewahrt bleiben.

5.4.3.3. Auswertung anhand des Kriterienkatalogs

Die Fallstudie 3 ermöglicht aufgrund der Organisationsstruktur des betrachteten Kostenträgers keine Auswertung anhand des Kriterienkatalogs. Da es sich um einen Kassenverband handelt, kann die operationelle Umsetzung der Gewährleistungen und daher das mögliche P4P-System nicht evaluiert werden. Die Grundzüge der Definition der Gewährleistung dagegen sind identisch zu den vorherigen Fallstudien, so dass hier kein Erkenntnisgewinn zu erwarten ist. Die Ergebnisse ergeben sich daher aus der Auswertung anhand der Systemelemente.

5.4.4. Fallstudie 4: Krankenhaus A

5.4.4.1. Beschreibung des Krankenhauses und der Qualitätsorientierung

Der Leistungserbringer der Fallstudie 4 ist ein Krankenhaus der Maximalversorgung in Trägerschaft eines privaten Klinikkonzerns. Die Klinik verteilt sich auf zwei Standorte mit einer Vielzahl von Fachabteilungen (unter anderem einer renommierten Kardiochirurgie), fast 1.000 Betten und 2.500 Mitarbeitern.

Die Qualitätsorientierung des Leistungserbringers ist vor allem durch das Qualitätsmanagementsystem des privaten Mutterkonzerns geprägt. Der Konzern wird mittels Qualitätsindikatoren gesteuert, die wissenschaftlich entwickelt wurden und seit vielen Jahren angewandt und evaluiert werden. Die Indikatoren basieren auf Routinedaten und werden in Rahmen eines Qualitätsmanagementsystems mit PDCA-Zyklus und Peer-Reviews umgesetzt. Auch ein finanzieller Anreiz ist durch die Integration von Qualitätsaspekten in die variable Vergütung der Zielvereinbarungen für die obere Führungsebene gegeben.

5.4.4.2. Beschreibung der IV-Verträge

Im Rahmen der Fallstudie wurden drei IV-Verträge analysiert: Zwei Verträge der Kardiologie und Kardiochirurgie mit Gewährleistungsklauseln, und ein Vertrag, der die Nachsorge nach Infarkten regelt. Die Verträge wurden mit gesetzlichen Krankenkassen mit bundesweit hohen Marktanteilen geschlossen.

Die zwei größeren IV-Verträge wurden von zwei verschiedenen Krankenkassen für die Fachabteilungen der Kardiologie und Kardiochirurgie abgeschlossen. Außerdem sind niedergelassene Ärzte und Rehabilitationseinrichtungen beteiligt. Gegenstand der Verträge sind verschiedene herzchirurgische und kardiologische Erkrankungen. Auch in diesen Verträgen sind sowohl Qualitätssicherungsaspekte als auch Gewährleistungsklauseln inkludiert. Diese beiden IV-Verträge umfassen zusammen ca. 700 Fälle pro Jahr (bei knapp 50.000 stationären Fällen des Klinikums pro Jahr).

Die *Qualitätssicherungsklauseln* beinhalten zunächst eine Verpflichtung zur Nutzung eines internen Qualitätsmanagementsystems sowie zur Teilnahme am externen QS-Verfahren. Gleichzeitig wurden auf Basis einer Auswahl der AQUA-Qualitätsindikatoren verschiedene Qualitätsziele vereinbart. Zielwerte sind, dass die Ergebnisse des Krankenhauses grundsätzlich unterhalb des Mittelwertes, Medians und – wenn vorhanden – Referenzwertes liegen müssen. Wenn dies nicht erreicht wird, liegt ein Grund für eine außerordentliche Kündigung vor. Außerdem wurde vereinbart, dass die Patientenversorgung durch indikationsbezogene Behandlungspfade optimiert werden soll. Insgesamt sollen innovative Versorgungsformen ermöglicht werden. Auch für die anderen in den Vertrag einbezogenen Sektoren (z. B. Reha) wurden Qualitätsziele gesetzt.

In den *Gewährleistungsklauseln* für die Kardiochirurgie ist die Frist auf 2 Jahre gesetzt. Verschiedene Patienten sind von der Gewährleistung ausgenommen, so beispielsweise bei Kombinationseingriffen oder Endokarditis. Ansonsten werden Revisionseingriffe nicht vergütet. Es wird erwartet, dass Revisionseingriffe nur im Vertragskrankenhaus stattfinden. Dem Krankenhaus wird allerdings das Recht eingeräumt, die Notwendigkeit eines Revisionseingriffes bspw. vom MDK überprüfen zu lassen. In der Kardiologie wird die Restenoserate für Stents und/oder Dilatationen der Koronargefäße über 2 Jahre global auf 30 % festgelegt – dies hat allerdings keine Vergütungskonsequenzen.

Diese Gewährleistungsklauseln sind alles in allem mit denen anderer Fallstudien vergleichbar.

Der dritte IV-Vertrag soll die Nachsorge nach Infarkten optimieren. So koordiniert beispielsweise der Leistungserbringer die Termine bei niedergelassenen Kardiologen im Anschluss an den stationären Aufenthalt. Dieser Vertrag umfasst ca. 200 Patienten und ist in der Umsetzung recht aufwendig, da die Patientenauswahl und -einschreibung sowie die Terminkoordination sehr zeit- und personalaufwendig sind. Zwar handelt es sich hierbei nicht um einen P4P-Vertrag im strikten Sinne, da die Umsetzung beim Leistungserbringer aber sehr konsequent durchgeführt wird, soll dieser Aspekt im Folgenden auch diskutiert werden. Eine Qualitätsmessung findet hier allerdings nicht statt.

5.4.4.3. Ergebnisse der Interviews

Gründe für die Vertragsschließung

Wie bei den Krankenkassen sind die Hauptgründe für die Vertragsschließung bei den Leistungserbringern wirtschaftlicher Natur: Da die Klinik zum Zeitpunkt der Vertragsentwicklung (und bis zum heutigen Tag) stets im Bereich der Mehrleistung war, war und ist die Möglichkeit der extrabudgetären Vergütung ohne Mehrleistungsabschlag wirtschaftlich interessant. Die meisten anderen IV-Verträge ermöglichen extrabudgetäre Bezahlung nur mit Rabatt auf die Fallzahlvergütung. In diesem Fall jedoch werden alle Fälle mit vollem Relativgewicht und Basisfallwert bezahlt. Für den Leistungserbringer stellt dies eine attraktive Vergütung dar, die Motivation für die Krankenkasse liegt hier vermutlich eher bei Marketing- und Steuerungsaspekten.

Der Hauptgrund für den Leistungserbringer, den Vertrag mit kardiologischer Nachsorge abzuschließen – der wirtschaftlich nicht attraktiv ist und keine positiven Geldflüsse verursacht – war vor allem wettbewerblicher Natur: Die Krankenkasse kommunizierte deutlich, dass dieser Vertrag mit nur einer Fachabteilung der Region abgeschlossen werden würde und es große Steuerungsmöglichkeiten der Versicherten gäbe.

Evaluation, Monitoring und Controlling der Verträge

Die bestehenden IV-Verträge werden anhand von wirtschaftlichen Kennzahlen evaluiert und kontrolliert. Hierfür werden in den Monatsberichten die Fallzahlen, Bewertungsrelationen und Umsätze für alle IV-Verträge kumuliert ausgewertet. Gewährleistungsfälle oder Qualitätsergebnisse sind in diesen Berichten nicht integriert. Sobald es zu einem Erlös- oder Fallzahlrückgang kommt, können weitere Details der IV-Verträge analysiert werden, dies kam allerdings bislang nicht vor.

Qualitätsaspekte sind in das kaufmännische Controlling nicht integriert. Eine Qualitätsüberwachung findet primär in den Kliniken statt und wird vom Medizincontrolling geleitet. Zielvereinbarungen können Qualitätsaspekte beinhalten, wenn sich in diesem Bereich Auffälligkeiten ergeben. Andernfalls stehen wirtschaftliche Kennzahlen im Vordergrund. Sie werden im Personalmanagement verfasst und verhandelt. Risikomanagement ist für das kaufmännische Controlling nicht relevant. Da die Klinik keine Kostenträgerrechnung (d. h. Fallkostenrechnung) hat, lassen sich Qualitätsabweichungen derzeit wirtschaftlich nicht bewerten.

Beschreibung der Systemelemente

- Versorgungsbereich

Die Auswahl des Versorgungsbereiches der Kardiologie/Kardiochirurgie wurde für die IV-Verträge von den Krankenkassen durchgeführt. Der Leistungserbringer stand dem Vorschlag sehr positiv gegenüber, da

gerade die dortige Kardiologie und Kardiochirurgie für ihre sehr hohe Behandlungsqualität bekannt sind. Dies war auch ein Grund für die Kasse, gerade auf diesen Leistungserbringer zuzukommen. Nach Angaben des Krankenhauses wurden die Versorgungsbereich bei der Kasse aufgrund eines existierenden Rahmenvertrages ausgewählt – da die Verträge sich auch im Wortlaut sehr ähneln, kann vermutet werden, dass es sich eventuell um ein Vertragsmuster des Kassenverbandes handelt.

Die von der Krankenkasse an den Leistungserbringer kommunizierten Auswahlgründe waren vor allem die einfache Strukturierbarkeit und Definierbarkeit der zum Vertrag zugehörigen DRGs. Je klarer die Abgrenzung der für den Vertrag relevanten Fälle, desto besser geeignet ist der Versorgungsbereich – dies wird auch vom Leistungserbringer bestätigt.

- Qualitätsindikator

Die Qualitätsindikatoren der IV-Verträge sind im Grundsatz ähnlich wie in den vorherigen Fallstudien: Die Qualitätssicherungsklauseln basieren hier auf AQUA- bzw. BQS-Indikatoren, die aus dem gesamten Indikatorenset spezifisch für die Fachbereiche ausgewählt wurden. Die Ergebnisse dieser Kriterien werden jedoch nicht zu Bonus- oder Malus-Zwecken genutzt. Bei den ausgewählten AQUA-Indikatoren handelt es sich großteils um Prozess- und Strukturindikatoren. Auch die Auswahl des Leistungserbringers wird auf Grundlage von Strukturqualität (nämlich bspw. Anzahl Oberärzte, Geräte und Ausstattung) getroffen.

Die Gewährleistungsregelungen beziehen sich nicht auf AQUA-Vorgaben. Sie betreffen Revisionseingriffe bzw. Restenoseraten (und somit Ergebnisqualitätsindikatoren), die im Detail von den Vertragspartnern definiert sind. Der Leistungserbringer selbst bezeichnet dies als die Hauptqualitätsindikatoren der IV-Verträge.

- Qualitätsmessung

Die in der Klinik eingesetzte Qualitätsmessung im „normalen“ Klinikalltag unterscheidet sich deutlich von der Messung, die speziell für die IV-Verträge erforderlich wäre. Es existiert ein detailliertes, etabliertes Qualitätsmess- und -managementsystem, für das in regelmäßigen Abständen Berichte erstellt und diskutiert werden. Gleichzeitig werden Daten für die Qualitätsmessung der gesetzlichen Qualitätssicherung dokumentiert und erhoben. Diese Qualitätsindikatoren entsprechen denen der IV-Verträge, die Qualitätsergebnisse der Teilnehmer des IV-Vertrages sind aber nicht gesondert ausgewiesen.

Die Gewährleistungsfälle werden nicht systematisch und standardisiert gemessen oder dokumentiert, d. h. für die IV-Verträge finden keine speziellen Qualitätsmessungen statt.

- Messergebnis

Da hier keine Qualitätsmessung für die IV-Verträge stattfindet, muss auch das Messergebnis der IV-Verträge nicht näher diskutiert werden. Auf der anderen Seite werden die Ergebnisse der internen Qualitätsmessungen auf Grundlage von Routinedaten sehr detailliert analysiert und im Managementsystem eingebunden.

- Qualitätsziel

Qualitätsziele sind in den IV-Verträgen im Bereich der Qualitätssicherung und der Gewährleistung inkludiert. Die Ziele der Qualitätssicherung sind unter anderem mittels Mittelwerten, Medianen oder Referenzwerten der AQUA-Indikatoren vereinbart. Dies stellt recht konkrete Ziele dar, allerdings werden die Qualitätsergebnisse nicht für den IV-Vertrag ausgewiesen, so dass eine Nachverfolgung der Zielerreichung nicht möglich ist (sie hätten aber auch im Falle einer separaten Darstellung keine Konsequenzen, da der IV-Vertrag für diesen Bereich keine Vergütungsfunktion hat).

Für die Gewährleistungsklauseln gibt es zum einen eine festgelegte Restenoserate von 30 %. Auch dieses Ziel wird in der Klinik nicht nachverfolgt. Es gab aber auch seitens der Kassen bisher keinen Gewährleistungsfall.

Im klinikeigenen internen Qualitätsmanagementsystem (außerhalb des IV-Vertrages) sind Qualitätsziele definiert. Diese sind mit Grenzwerten versehen, so dass bei Qualitätsdefiziten mit Konsequenzen zu rechnen ist.

- Vergütungsfunktion/Konsequenzenmanagement

In dieser Fallstudie ist sehr spannend, dass die klinikinterne Vergütungsfunktion (Bonus/Malus) des klinikeigenen Qualitätsmanagementsystems weiter entwickelt ist als die Gewährleistungsklauseln der IV-Verträge. Die internen Qualitätsmessungen und deren Ergebnisse sind in die Zielvereinbarungen integriert und haben damit einen direkt Einfluss zumindest auf das Chefarztgehalt und das Gehalt der Klinikleitung. Gleichzeitig gibt es eine sehr konkrete operative Konsequenz bei Nichterreichung der Qualitätsziele – nämlich der Einsatz eines Peer-Review-Verfahrens bei dem Chefärzte von Fachabteilungen anderer Kliniken beteiligt werden. Für die jeweilige Abteilung beinhaltet dies klare Konsequenzen beispielsweise für die Strukturierung der Prozesse.

Die IV-Verträge und deren Qualitätsaspekte dagegen haben derzeit keinerlei Konsequenzenmanagement. Selbstverständlich sind die Behandlungsfälle der IV-Verträge in den Qualitätsmanagementaktivitäten aller Kassenpatienten inkludiert, so dass generell vom Wirksamwerden der allgemein im Klinikum hohen Qualität ausgegangen wird. Da jedoch weder die Gewährleistungsfälle noch die Indikatorenergebnisse der IV-Patienten in der Qualitätssicherung separat ausgewiesen werden, kann auf dieser Grundlage nicht gesteuert werden. Die Argumentation der Klinik verweist genau auf den Aspekt, dass bei einer hohen Allgemeinqualität diese auch den IV-Patienten zugutekommt. Für die Patienten sind die Vorteile des IV-

Vertrages daher weniger in der Qualität als vielmehr in zusätzlichen Leistungen zu sehen, so bspw. kürzere Wartezeiten oder Wahlleistungsessen. Dadurch entsteht auch ein Wettbewerbsvorteil, der zusätzlich zum Steuerungseffekt der Krankenkassen zur Annahme führt, dass durch die IV-Verträge mehr Fälle behandelt werden. Gleichzeitig wird aber nicht davon ausgegangen, dass es durch die IV-Verträge zu einer zusätzlichen, separaten Qualitätsverbesserung kommt.

Praktische Umsetzung im klinischen Alltag

Wie bereits erwähnt, werden die Gewährleistungsfälle im klinischen Alltag nicht standardisiert ausgewiesen. Wenn es zu solchen Fällen kommen sollte, wird davon ausgegangen, dass die Patientenabrechnung den Verantwortlichen aus Verwaltung und Klinik benachrichtigen würde. Bislang (obschon die Verträge bereits seit 2004 existieren) sind jedoch noch keine Gewährleistungsfälle oder potenziellen Garantiefälle von den Kassen eingefordert worden. Daher wird angezweifelt, ob es überhaupt zu solchen Fällen kam. Auch „Warnsysteme“ oder IT-Unterstützung für diesen speziellen Sachverhalt bei der Patientenabrechnung gibt es nicht. Aus diesem Grunde gibt es keine automatisierte Unterstützung für die Erkennung der Gewährleistungsfälle in der Routine.

Die praktische Berücksichtigung der IV-Verträge in der nicht-medizinischen Patientenversorgung findet jedoch statt. So werden für alle Verträge zunächst die Patienten, die sich für den Vertrag qualifizieren, ausgewählt. Dies kann abhängig vom Vertrag sehr zeitaufwendig sein, da nicht nur die Kassenzugehörigkeit und Hauptdiagnose sondern auch noch eine Vielzahl von Ausschlusskriterien überprüft werden müssen. Anschließend müssen sich die Patienten einschreiben, was den aufwendigsten Prozessschritt darstellt, weil den Patienten der Vorteil des Einschreibens sowie die Formulare erklärt werden müssen. Trotz der Leistungsvorteile der IV-Verträge (Wartezeit, Wahlleistungsessen) müssen viele Patienten noch persönlich überzeugt werden. Auch für die Krankenkassen sind die Einschreibungsquoten ein wichtiges Kriterium, das – anders als die Behandlungsqualität – regelmäßig überprüft und diskutiert wird.

Für die Behandlungsqualität werden als wichtiger Aspekt des klinischen Alltags die „Standard Operating Procedures (SOPs)“ genannt. Diese sind sowohl für die Ärzte als auch die Pflege richtungsweisend und werden klar strukturiert umgesetzt. Ein großes Verbesserungspotenzial wird hier nicht mehr vermutet.

Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen

Insgesamt wird von allen Befragten das Thema P4P als sehr wichtig und zukunftsrelevant eingeschätzt. Es wird davon ausgegangen, dass auch in Deutschland zukünftig vermehrt qualitätsorientierte Vergütung existieren wird. Hierfür sei die eigene Unternehmenskultur durch die integrierten Qualitätsmanagementsysteme bereits sehr gut vorbereitet, und es sei mit weniger Hindernissen als bei anderen Leistungserbringern zu rechnen.

Gerade im Controlling wird dieser Trend sehr positiv bewertet: P4P sei eine sehr gute Möglichkeit, um auch die monetären Aspekte der Qualität zu betrachten. Dies finde bislang nicht ausreichend statt. Eine höhere Finanzrelevanz würde dann auch unweigerlich zu einer größeren Aufmerksamkeit beim Leistungserbringer führen – sowohl in der Klinik als auch im Management. Insgesamt könnte ein solcher Vergütungsmechanismus dazu führen, dass die strategische Ausrichtung sich aufgrund der externen Systemvorgaben bzw. Anreizwirkungen nicht mehr vorwiegend auf Fallzahlen und Bewertungsrelationen fokussieren müsste. Vor allem die Frage, wie diese Qualität gemessen und auch kaufmännisch kontrolliert werden kann, sollte hierbei im Vordergrund stehen.

Auch im klinischen Alltag wird angenommen, dass die Einführung einer qualitätsorientierten Vergütung zu einer größeren Aufmerksamkeit des Qualitätsthemas führen könnte – große Prozess- oder Kulturveränderungen werden dagegen nicht erwartet. Aus Sicht der Klinikverwaltung könnte eine solche Vergütung, insbesondere bei Ausführung als episodenbezogene Gesamtvergütung, mit dazu beitragen, dass weniger Fälle stationär durchgeführt werden, da der Gesamtprozess und die Gesamtbehandlungsqualität stärker im Vordergrund stehen.

Insgesamt zeigt sich, dass ein Leistungserbringer, der bereits eine etabliert Qualitätskultur mit eigenen internen Managementsystemen und Verbesserungsmechanismen hat, einem P4P-System sehr offen gegenübersteht und darauf vorbereitet ist. Andererseits wird deutlich, dass spezifische Qualitätsregelungen von IV-Verträgen bei einem existierenden komplexen Qualitätsmanagement kaum zusätzliche Wirkung entfalten.

5.4.4.4. Auswertung anhand der Systemelemente

Da in dieser Fallstudie die IV-Verträge mit Gewährleistung keinen Charakter eines P4P-Systems gewinnen, soll hier vor allem anhand der Systemelemente, nicht aber durch Anwendung des Kriterienkatalogs, Erkenntnisse gewonnen werden. Der Katalog wird im Anschluss an das existierende Qualitätsmanagementsystem angewandt.

Systemelement: Versorgungsbereich, Kriterienkategorie: Aussagekraft

Sowohl die medizinische als auch die wirtschaftliche Relevanz von kardiologischen und kardiochirurgischen Eingriffen ist als hoch einzustufen. Interessant ist jedoch, dass anders als bei den Krankenkassen, in dieser Fallstudie berichtet wird, dass die Kassen und nicht die Leistungserbringer oder einzelne Chefärzte die Ideen für die IV-Verträge entwickelt und vorgeschlagen haben. Der Versorgungsbereich wurde auch hier aus wirtschaftlichen Gründen ausgewählt – gleichzeitig war für die Kassen die einfache Mess- und Definierbarkeit des Versorgungsbereiches entscheidend.

Systemelement: Qualitätsindikatoren, Kriterienkategorie: Aussagekraft und Akzeptanz

Da in diesem Fall nur die Gewährleistungsklauseln eine Vergütungskonsequenz nach sich ziehen könnten, sollen hier nur diese ausgewertet werden. Die Revisionseingriffe bei kardiochirurgischen Behandlungen sind zunächst ein sehr relevanter Indikator, der die Kriterien 2a, 2b sowie 3a-f gut erfüllt. Die Akzeptanz der Indikatoren beim medizinischen Personal ist sehr hoch, da die Fachabteilungen auch der internen Qualitätsmessung gegenüber sehr aufgeschlossen sind. Gleichzeitig sind jedoch die speziellen Qualitätsaspekte des IV-Vertrages für die Fachabteilung nicht besonders relevant, da deren Steuerung vorwiegend auf den klinikinternen Indikatoren beruht.

Systemelement: Qualitätsmessung, Kriterienkategorie: Wissenschaftlichkeit und Messaufwand/-sicherheit

Die Wissenschaftlichkeit der Revisionsraten und Restenoseraten ist prinzipiell gegeben, da dies auch in wissenschaftlichen Arbeiten relevante Indikatoren sind (vgl. u. a. Hehrlein et al. 1998, S. s166; Kovacs 2002; Kamke et al. 2002, S. 233; Gorenoi et al. 2005).

Eine Qualitätsmessung für die Indikatoren der Gewährleistung findet nicht statt, daher können die Kriterien dieses Systemelements nicht bewertet werden. Wenn Gewährleistungsfälle überhaupt bemerkt werden, würden sie von der Kasse aufgedeckt und an den Leistungserbringer weitergeleitet werden. Die Messung von im Klinikum versorgten Revisionen könnte prinzipiell auch beim Leistungserbringer durchgeführt werden, würde aber (einmalig) kostenintensive Sonderauswertungen in der SAP-Software bedingen. Eine übergreifende Messung (vor allem die Messung von Gewährleistungsfällen außerhalb des Vertragskrankenhauses) ist aber im Klinikum nicht möglich, so dass langfristig davon ausgegangen werden muss, dass die Messung bei der Krankenkasse stattfinden sollte.

Systemelement: Qualitätsziele, Kriterienkategorie: Wirksamkeit, Operationalisierbarkeit, Vergleichbarkeit

Die Qualitätsziele der Gewährleistung und deren Kriterienerfüllung sind zunächst vergleichbar mit denen der vorherigen Fallstudien. Zwar gibt es ein recht konkretes Ziel für die Restenoserate und die Grenzwerte der Qualitätssicherung, da diese aber nicht mit Konsequenzen hinterlegt sind, sind diese Zielwerte nicht relevant. Das Ziel einer Verhinderung von Revisionseingriffen ist in der besprochen Fachabteilung sicherlich immanent und wird im internen Qualitätsmanagement ohnehin aktiv verfolgt – eine spezifische Verbindung zu den IV-Verträgen besteht jedoch nicht. Es kann davon ausgegangen werden, dass die Ziele akzeptiert, konsistent, erreichbar, gerecht und langfristig inzentivierend wären, aber dies ist auch ohne IV-Verträge der Fall.

Systemelement: Vergütungsfunktion, Kriterienkategorie: Wirksamkeit, Operationalisierbarkeit

Die Vergütungsfunktion kann zunächst nicht bewertet werden, da keine Gewährleistungsfälle umgesetzt wurden und die anderen Qualitätsziele nicht mit Konsequenzenmanagement verbunden sind. Die tatsächlichen Auswirkungen der P4P-Anteile der IV-Verträge sind daher noch geringer als in den vorherigen Fallstudien.

Auch bei Auftreten von Gewährleistungsfällen wären die Effizienz und die Vergütungsanteile allerdings in solchen Selektivverträgen nur sehr gering, wie in Musterrechnungen bereits gezeigt wurde (vgl. Seeger 2011). Eine höhere Relevanz könnte nur entstehen, wenn alle Krankenkassen gemeinsam einen IV-Vertrag mit gleichen Indikatoren und Vergütungsfunktion umsetzen würden, da für die beteiligte Fachabteilung dann ein hoher Anteil von Fällen und Umsatz betroffen wäre. Dies käme allerdings einer kollektivvertraglichen Regelung nahe.

Der allgemeine Vergütungsaspekt des IV-Vertrages, nämlich die außerbudgetäre Vergütung aller Fälle, ist andererseits für die Fachabteilung und den Leistungserbringer wirtschaftlich relevant. Daher wird die Einschreibung von Patienten in die Verträge aktiv gefördert – Qualitätsaspekte stehen bei diesem Aspekt aber nicht im Vordergrund.

Systemelement: Messergebnis, Kriterienkategorie: Vergleichbarkeit, Operationalisierbarkeit

Gleichzeitig ergäbe sich aber bei beabsichtigten Rankings die Schwierigkeit, dass der Leistungserbringer selbst einerseits keine einrichtungsübergreifende Messung durchführen könnte und andererseits auch keine Vergleichsdaten aus anderen Projekten oder auf Bundesebene zur Verfügung stünden. Ein Ranking wäre daher aus Sicht des Leistungserbringers nicht möglich. Falls die Messungen bei der Kasse durchgeführt würden, ergäben sich die gleichen Kriterienerfüllungen wie in den vorherigen Fallstudien bei den Krankenkassen.

5.4.5. Exkurs: Auswertung des internen Qualitätsmanagementsystems anhand des Kriterienkatalogs

An dieser Stelle soll das interne Qualitätsmanagementsystem des Krankenhauses A anhand des Kriterienkataloges bewertet werden. Dieses steht nicht im Zusammenhang mit dem untersuchten IV-Vertrag. Da es sich hierbei aber um ein standardisiertes Qualitätsmesssystem mit Peer-Review-Verfahren und variabler Vergütung bei Chefärzten handelt, kann es im Sinne eines (trägerinternen) P4P-Systems gesehen und evaluiert werden.

Tabelle 24: Bewertungen Exkurs Fallstudie 4, Systemelemente Gesamtsystem und Versorgungsbereich

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
1a	Abdeckung des Gesamtsystems	Ja, Versorgungs-bereiche klar definiert	**Hoch**	Derzeit umfassendstes Indikatorensystem mit 38,7% aller Krankenhausfälle und 61,8% aller im Krankenhaus stattfinden Todesfälle	Langfristig ggf. noch mehr Versorgungs-bereiche hinzuziehen
2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet	Ja, Versorgungs-bereich klar definiert	**Hoch**	Versorgungsbereiche anhand von medizinischer Relevanz ausgewählt, relevantesten Bereiche inkludieren so z.B. Herzinfarkt, Schlaganfall, Pneumonie, große Darmchirurgie usw.	
2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich	Ja, Versorgungs-bereich klar definiert	**Hoch**	Auswahl der Versorgungsbereiche eher nach medizinischer als nach wirtschaftlicher Relevanz, aufgrund hoher Fallzahlabdeckung und hoher Kosten aber auch wirtschaftlich relevant	Wirtschaftliche Relevanz quantifizieren

Zur Messung wird das derzeit umfassendste Indikatorensystem eingesetzt, das 38,7 % aller Krankenhausfälle und 61,8 % aller im Krankenhaus stattfindenden Todesfälle abdeckt (vgl. Nimptsch/Mansky 2012a, S. 1454). Darüber werden für das interne Qualitätsmanagementsystem Versorgungsbereiche ausgewählt wie bspw. „häufige Standardbehandlungen (Beispiel: Herniotomie in der Viszeralchirurgie), komplexere, aber vergleichsweise verbreitete Verfahren (Beispiel: kolorektale Chirurgie), spezialisierte, im Fachgebiet bedeutende Verfahren (Beispiel: Ösophaguschirurgie)“ (Mansky et al. 2011, S. A-7). Die medizinische Bedeutung ist sehr hoch. Die wirtschaftliche Bedeutung wird zwar nicht als Auswahlkriterium beschrieben, allerdings ist schon aufgrund der Auswahl von Bereichen mit hoher Fallzahl und überwiegend hoher Bewertung im DRG-System die wirtschaftliche Bedeutung für den Leistungserbringer hoch. Gleichzeitig sind auch Versorgungsbereiche mit hohem Erlösvolumen inkludiert, so bspw. Hüft- und Knieimplantate (vgl. Spindler 2010, S. 435).

Tabelle 25: Bewertungen Exkurs Fallstudie 4, Systemelement Qualitätsindikator

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten	Ja	**Hoch**	Viele Indikatoren stellen wichtige Teile des Behandlungsergebnisses direkt dar, insbesondere die Sterblichkeit als wichtiger "Endpunkt" gerade komplexere Behandlungen sind eingeschlossen	Einbeziehung weiterer Endpunkte (z.B. über QSR)
3b	Relevanz des Indikators für ein bestimmtes Fachgebiet	Ja	**Hoch**	Viele Indikatoren stellen relevante Bereiche des jeweiligen Fachgebiets dar	Einbeziehung weiterer Fachgebiete
3c	Sektorbezug vs. sektorübergreifende Nutzung *(Informationskriterium)*	Ja	Niedrig	Klarer Krankenhausbezug, nur Krankenhausfälle und -sterblichkeit inkludiert, nur Krankenhausdaten verfügbar	Übergreifendes Indikatorenset mit aufnehmen, bspw. QSR

Kriterienkategorie: Akzeptanz

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
4a	Medizinische Bedeutung in der medizinischen Fachliteratur	Ja	**Hoch**	Auswahl der Indikatoren vor allem anhand von wissenschaftlicher Literatur, Indikatoren wurden u.a. als "Studienendpunkte" einer Vielzahl von wissenschaftlichen Studien ausgewählt	
4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften	Ja	**Hoch**	Hoher, zunehmender Verbreitungsgrad der Indikatoren auch international u.a. aufgrund der Akzeptanz bei den Fachgesellschaften	Stärkere Einbeziehung der Fachgesellschaften
4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern	Ja	**Hoch**	Akzeptanz beim medizinischen Personal wird als hoch eingeschätzt, allerdings einige Chefärzte auch nach Jahren noch zweifelnd	
4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit	Ja	**Hoch**	Sterblichkeit gut verständlich, Relevanz muss teilweise laienverständlich gemacht werden	Vermehrte Aufklärungs-arbeit
4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit	Ja	**Hoch**	Bedeutung der Sterblichkeit naturgemäß hoch, in anderen Fällen muss die Relevanz teilweise laienverständlich gemacht werden	Vermehrte Aufklärungs-arbeit

Auch die Qualitätsindikatoren wurden aufgrund von medizinischem Interesse bzw. Relevanz ausgewählt (vgl. Mansky et al. 2011, S. A-7). Sowohl Ergebnisqualitätsindikatoren als auch Beobachtungswerte, Prozesskennzahlen, Komplikationskennzahlen und Mengeninformationen sind abgebildet (vgl. Mansky et al. 2011, S. A-13ff.). Gerade Sterblichkeitsraten als „Endpunkte“ werden häufig in klinischen Studien angewandt (vgl. beispielhaft Wells 1864, S. 384f.; Werner/Bradlow 2006, S. 2694f.). Somit ist auch die wahrgenommene Bedeutung in der medizinischen Fachliteratur, bei Fachgesellschaften und beim medizinischen Personal hoch. Die Verständlichkeit und wahrgenommene Bedeutung für die Öffentlichkeit ist bei Sterblichkeitsraten sowie Mengenangaben naturgemäß hoch, die konkrete medizinische Relevanz und die genaue Interpretation der Messwerte muss aber ggfs. laienverständlich erklärt und hinsichtlich der relativen Einordnung erläutert werden. Für das interne Qualitätsmanagementsystem ist aber das Kriterium der Laienverständlichkeit nur von geringer Bedeutung: Zwar werden die Ergebnisse auch in Qualitätsberichten veröffentlicht, der Steuerungsmechanismus allerdings findet beim Leistungserbringer selbst statt.

Tabelle 26: Bewertungen Exkurs Fallstudie 4, Systemelement Messung

Kriterienkategorie: Wissenschaftlichkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
5a	Reliabilität	Ja	**Hoch**	Durch detaillierte, algorithmische Definition der Indikatoren und Messparameter kann eine hohe Reliabilität erreicht werden	
5b	Validität	Ja	**Hoch**	Messergebnisse vielfach in verschiedenen Krankenhäusern überprüft	
5c	Risikoadjustierung und Risikoadjustierbarkeit	Ja	**Hoch**	Durch Anwendung der SMRs werden die Ergebnisse risikoadjustiert	

Kriterienkategorie: Messaufwand / -sicherheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
6a	Objektiver Erhebungsaufwand und -komplexität	Ja	**Hoch**	Routinedaten sind zur Abrechnung ohnehin erhoben, Auswertung findet standardisiert, automatisiert und ohne zusätzlichen Messaufwand statt	
6b	Manipulationssicherheit / Objektivität der Messung	Ja	**Hoch**	Ein wichtiger Vorteil der Nutzung von Routinedaten ist die hohe Manipulationssicherheit, da die Daten zu Abrechnungszwecken erhoben und vielfältig überprüft werden	
6c	Grad der Integration in automatische Prozesse / IT-Systeme	Ja	**Hoch**	Automatisierte Messung für alle beteiligten Krankenhäuser ist Standard	

Die Qualitätsmessung findet anhand von Routinedaten statt. Dies ermöglicht einen geringen Erhebungs- und Messaufwand und eine hohe Manipulationsresistenz, da beispielsweise eine „Nichtmeldung“ von Komplikationsfällen praktisch auszuschließen ist oder ggf. mit Nicht-Bezahlung „bestraft“ würde. Da die Daten für Abrechnungszwecke erhoben werden, sind sie ferner u. a. durch die Kasse und ggf. den MDK mehrfach überprüft.

Tabelle 27: Bewertungen Exkurs Fallstudie 4, Systemelement Ziele

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
7a	Zielkonsistenz	Ja	**Hoch**	Ziele des QM-Systems entsprechen alle den akzeptieren Qualitätszielen und sind mit denen des medizinischen Personals konsistent	
7b	Zielkongruenz	Ja	**Hoch**	Gesamtes Zielsystem koordiniert entwickelt und aufeinander abgestimmt, Ziele gleichzeitig erreichbar und in sich widerspruchsfrei	
7c	Reichweite	Ja	**Hoch**	Mit 38,7% aller Fallzahlen derzeit umfassendstes Ziel- und Indikatorensystem	Kontinuierlich weitere Indikatoren/Ziele entwickeln

Kriterienkategorie: Operationalisierbarkeit/Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
8a	Erreichbarkeit	Ja	**Hoch**	Studien zeigen, dass Ziele von einer Vielzahl von Fachabteilungen und Krankenhäusern erreicht werden können	
8b	Transparenz	Ja	**Hoch**	Ziele sowie Messergebnisse sind bei allen Beteiligten bekannt und werden sogar extern publiziert	
8c	Akzeptanz	Ja	**Hoch**	Akzeptanz der Indikatoren und auch Qualitätsziele bei Management- und Fachebene hoch	
8d	Langfristige Anreizwirkung der Ziele	Ja	**Hoch**	Studien zeigen die Anreizwirkung nach Implementierung des Systems sowie nach mehreren Jahren, Anreizwirkungen bestehen insbesondere für "Schwarze Schafe", System eignet sich für die langfristige Überwachung der wichtigsten Leistungsbereiche	Evenutell Anpassung der Ziele
9a	Gerechtigkeit	Ja	**Hoch**	Durch Risikoadjustierung werden Unterschiede im Patientenkollektiv berücksichtigt, einige Beteiligte zweifeln Gerechtigkeit jedoch an	Langfristig stärkere Risikoadjustierung?

Die Qualitätsziele des Systems sind sehr detailliert beschrieben und für die in Frage kommenden Indikatoren hinterlegt. Alle Ziele wurden mit der Unterstützung vom medizinischen Personal entwickelt, so dass Zielkonsistenz, -kongruenz und Akzeptanz gewährleistet sind. Die Zielwerte orientieren sich am Bundesdurchschnitt, was die Akzeptanz verbessert und die Erreichbarkeit ermöglicht. Die Anwendung im Klinikkonzern ermöglicht beinhaltet betriebsintern eine hohe Reichweite. Die Ziele werden sehr

transparent für die beteiligten Fachabteilungen sowie den Gesamtkonzern dargelegt und haben nachweislich eine hohe Anreizwirkung (vgl. Busse et al. 2009, S. 300f.).

Tabelle 28: Bewertungen Exkurs Fallstudie 4, Systemelement Vergütungsfunktion

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
10a	Effizienz	Schwierig, da Details der Ziel-vereinbarung nicht bekannt	**Niedrig**	Vergütungsfunktion derzeit nur für Chefärzte relevant, Zielvereinbarungen laut Interviews vor allem basierend auf wirtschaftlichen Aspekten	Analyse der Effizienz, Ausweitung des Systems auf weitere Mitarbeiterebenen
10b	Effektivität	Schwierig, da Details der Ziel-vereinbarung nicht bekannt	**?**	Verbesserung der Qualitätsergebnisse nachweisbar, Korrelation zur Vergütungsfunktion aber nicht klar (auch andere Faktoren wie z.B. internes Ranking sind relevant)	Verknüpfung der Ergebnisanalysen mit Details der Vergütungsfunktion
10c	Flexibilität	Schwierig, da Details der Ziel-vereinbarung nicht bekannt	**Hoch**	Vergütungsfunktion in Zielvereinbarungsgesprächen verhandelt und daher flexible auf Einzelsituation angepasst	

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
11a	Anteil der Vergütung	Schwierig, da Details der Ziel-vereinbarung nicht bekannt	**Niedrig**	Laut Interviews im untersuchten Haus nur geringer Anteil der variablen Chefarztvergütung qualitätsorientiert	Ggfs. Vergütungsanteil erhöhen, auch andere Mitarbeiterebenen einbeziehen
11b	Planbarkeit	Schwierig, da Details der Ziel-vereinbarung nicht bekannt	**Hoch**	Durch vermutlich klare Verknüpfung von Qualitätsergebnissen (die abschätzbar sein sollten) und Vergütung recht hohe Planbarkeit für Chefärzte	

Da das Qualitätsmanagementsystem anhand von Anreizmechanismen des Public Reportings, internen Vergleichen sowie eines verpflichtenden Peer-Review-Verfahrens aufgebaut ist, spielen Vergütungsaspekte eine untergeordnete Rolle. Eine Vergütungsfunktion ist vor allem in der qualitätsvariablen Vergütung von Chefärzten zu sehen – die Höhe des variablen Anteils ist allerdings nicht bekannt, wird von den Befragten aber im untersuchten Haus, welches die Ziele überwiegend erreicht, als gering vermutet. Die konkrete Gestaltung der Vergütung ist aber verhandlungsbasiert, so dass von einer hohen Flexibilität ausgegangen werden kann und somit bei geringerer Zielerreichung (beispielsweise in anderen Kliniken) auch ein deutlich höherer qualitätsabhängiger Bonusanteil vereinbart werden könnte. Der Träger hat in diesem Fall keine einheitliche, hausübergreifende Vergütungsfunktion festgelegt. Bei vermutet höherer Relevanz von Fallzahlen und Wirtschaftskennzahlen in den Zielvereinbarungen des untersuchten Hauses muss die Effizienz hier zunächst mit niedrig bewertet werden. Da zwar die Qualitätsverbesserung nicht aber die Korrelation dieser zur variablen Vergütung bekannt sind, kann die Effektivität der Vergütungsfunktion nicht genau bewertet werden. Aufgrund des im vorliegenden Fall kleinen qualitätsvariablen Anteils sowie der Beschränkung der variablen vergüteten Mitarbeiterkreise auf

Chefärzte ist der Vergütungsanteil am Gesamtumsatz gering. Die Planbarkeit dürfte aber gleichzeitig wegen recht gut abschätzbarer Qualitätsergebnisse hoch sein.

Tabelle 29: Bewertungen Exkurs Fallstudie 4, Systemelement Messergebnis

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
12a	Nutzung zur Steuerung	Ja	**Mittel**	Wird bereits aktiv zur Steuerung der Kliniken eingesetzt, Relevanz für das lokale Management könnte durch qualitätsorientierte Vergütung erhöht werden	(Noch) stärkerer Fokus auf qualitätsorientierter Vergütung
12b	Aktualität	Ja	**Hoch**	Ergebnisse werden monatlich reported	
12c	Kontrollierbarkeit	Ja	**Hoch**	Aufgrund des Bezuges auf Abrechnungsdaten mit automatisierter Verarbeitung sehr hoch	
12d	Praktische Umsetzung im Controlling	Ja	**Niedrig**	Derzeit im medizinischen Controlling relevant, im untersuchten Haus keine Auswirkung auf lokales, kaufmännisches Controlling	Erhöhung der lokalen wirtschaftlichen Relevanz anzustreben
12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit	Ja	**Hoch**	Sowohl auf Krankenhaus- als auch auf Fachabteilungsebene ist die Zuordnung der Messergebnisse nachvollziehbar	

Kriterienkategorie: Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
13a	Zeitverlauf	Ja	**Hoch**	Ergebnisse werden über bis zu 12 Jahre rückwirkend im Zeitverlauf dargestellt	
13b	Querschnittsvergleich/Ranking	Ja	**Hoch**	Rankings auf Fachabteilungs- und Klinik-Ebene für gesamten Klinikkonzern dargestellt, außerdem ist der Vergleich zum Bundesdurchschnitt und zu weiteren externen Kliniken möglich	
13c	Justiziabilität	*Nicht bewertbar*	**?**	*nicht bewertbar*	

Die Messergebnisse des QM-Systems sind aktuell, kontrollierbar und werden verursachungsgerecht den Kliniken zugeordnet. Außerdem werden Rankings auf Konzernebene erstellt. Gleichzeitig jedoch findet das Controlling derzeit nur im Medizincontrolling statt – für das kaufmännische Controlling sind die Qualitätsergebnisse nur von geringer Relevanz, da im untersuchten Haus (bei ohnehin guter Zielerreichung) keine direkten wirtschaftlichen Konsequenzen entstehen.

Die Ergebnisdarstellung im langjährigen Zeitverlauf sowie im Vergleich mit namhaften externen Kliniken (darunter viele Universitätskliniken) schafft eine hohe Transparenz mit vielfältigen Vergleichsmöglichkeiten und damit auch eine hohe Akzeptanz bei den Fachbeteiligten.

5.4.6. Fallstudie 5: Krankenhaus B

5.4.6.1. Beschreibung des Krankenhauses und der IV-Verträge

Das Krankenhaus der fünften Fallstudie ist ein Maximalversorger in einer deutschen Großstadt. Die IV-Verträge werden zentral in der Verwaltungszentrale geschlossen. Der Leistungserbringer hat derzeit einen IV-Vertrag der 2004 für endoprothetische Eingriffe mit einem Kostenträger der Ersatzkassen geschlossen. Im Vertrag sind ca. 100-130 Fälle pro Jahr in mehreren Standorten des Verbundes.

Mehrere Verträge wurden diskutiert und verhandelt, aber nach Einführung des Gesundheitsfonds nicht weiter verfolgt. Ein weiterer Vertrag der Endoprothetik wurde 2011 verhandelt und sollte ein qualitätsbezogenes Bonussystem für bis zu 900 Fälle jährlich einschließen. Der Vertrag hätte einen qualitätsabhängigen Anteil beinhaltet, der auf Grundlage von drei Qualitätsindikatoren bestimmt würde. Anhand von drei Indikatoren wäre die Behandlungsqualität der Klinik bzw. eines Standortes mit Routinedaten gemessen und die Zielerreichung von vordefinierten Referenzwerten analysiert worden, so bspw. Komplikationsraten. Die Vergütung hätte mittels eines Qualitätszuschlages stattgefunden, der auf Grundlage von zwei Skalastufen unterschiedliche Zahlungshöhen erreicht. Die Grundvergütung wäre angepasst und dementsprechend niedriger gewesen. Allerdings hatte der Leistungserbringer bezüglich des Indikatoren- und Messsystems deutliche Bedenken. Zum einen gab es juristische Fragen, ob der Kostenträger die Daten, die durch Behandlungen im Anschluss an den Krankenhaus-Aufenthalt entstehen, zusammenfügen und für Qualitätsmesszwecke nutzen dürfe. Außerdem wurde gezweifelt, ob eine ausreichende Risikoadjustierung stattfindet. Die Nutzung von einer „rein statistischen“ Sterblichkeitsrate (mit „all-cause mortality“) und eine unzureichende Kausalitätsbetrachtung der Mortalitätsfälle werden ebenfalls mit Bedenken gesehen. Der Leistungserbringer hat daraufhin einen eigenen Vorschlag für ein qualitätsorientiertes System gemacht, nämlich die Verwendung einer Auswahl von AQUA-Indikatoren die besonders aussagekräftig und relativ manipulationssicher sind. Nachdem der Kostenträger eine erste Qualitätsmessung auf Grundlage der eigenen Qualitätsindikatoren durchgeführt hatte, wurden die Verhandlungen abgebrochen, da die Ergebnisse der Klinikstandorte für eine Vertragsschließung nicht ausreichten. Für den Leistungserbringer bedeutete dies einen bereits investierten Aufwand und langen Vertragsprozess ohne Abschluss.

5.4.6.2. Ergebnisse der Interviews

<u>Gründe für die Vertragsschließung</u>

Der 2004er IV-Vertrag wurde vor allem abgeschlossen, weil mit einer Steuerungsmöglichkeit der Kasse gerechnet wurde, die allerdings nicht eintrat. Die Möglichkeit einer außerbudgetären Vergütung ist auch ein Vorteil, durch den verringerten Mehrerlösausgleich ist dies jedoch kaum noch relevant. Gleichzeitig erhofft sich der Leistungserbringer eine bessere Kommunikation mit dem Kostenträger und einen optimierten Prozess zwischen den Sektoren.

Die Kostenträger konnten durch den IV-Vertrag eine Rabattierung verhandeln. Insgesamt geht der Leistungserbringer davon aus, dass durch die Rabatte und Anschubfinanzierung die Preise „verdorben“ wurden.

Evaluation, Monitoring und Controlling der Verträge

Vor der Vertragsschließung wurden besonders die zu erwartenden Mehrfälle evaluiert. Die Gewährleistungsklauseln wurden in der Berechnung nicht betrachtet, im Nachhinein hätten die nicht-erstatteten Revisionsoperationen mit eingepreist werden müssen, sie müssten demnach höher sein als nach der Rabattierung von ca. 8-10 %. Für die Vertragsverhandlung wurden vor allem Erfahrungswerte genutzt. Der bestehende Endoprothetik-Vertrag wird bezüglich der Fallzahlen und der Einschreibequote evaluiert. Im klinischen Alltag wird der IV-Vertrag von einem Fallmanager geführt und kontrolliert.

Für den qualitätsorientierteren Vertrag mit Bonus-Aspekt hätte ein spezielles Controlling aufgebaut werden müssen, dies wurde auch bereits vorbereitet.

Beschreibung der Systemelemente

- Versorgungsbereich

Wie in der vorherigen Fallstudie kam auch hier der Vorschlag des Vertrages und damit des Versorgungsbereiches von der Kasse.

- Qualitätsindikator

Auch die Idee für die Qualitätsindikatoren kam vom Kostenträger. Der Leistungserbringer steht Vorschlägen offen gegenüber auch wenn bspw. Gewährleistungen als „notwendiges Übel“ gesehen werden. Während die Gewährleistungsklauseln als Qualitätsindikator bislang kaum Auswirkungen haben und seit 2004 nur 2-3 Fälle aufgetaucht sind, waren sie zum Zeitpunkt des Vertragsabschlusses für den Kostenträger sehr wichtig.

Beim vorgeschlagenen „Bonus-Vertrag“ waren die Qualitätsindikatoren ein wichtiges Diskussionsthema und ein Grund für das Nicht-Zustandekommen des Vertrages. Während der Kostenträger das eigene Indikatorensystem vorschlug, empfahl der Leistungserbringer eine Auswahl von AQUA-Indikatoren. Diese Auswahl wurde aufgrund von Aussagekraft und Wissenschaftlichkeit getroffen.

- Qualitätsmessung

Die Messung der Qualität der IV-Verträge ist kaum standardisiert. Es existiert ein Softwareprogramm zum Monitoring der IV-Verträge, dies ist aber vor allem für die Prozesse und sektorübergreifenden Kommunikation und Aktivitäten hilfreich. Dieses Programm ermöglicht die Freigabe verschiedener Stammdaten zwischen den Vertragspartnern. Hierdurch kann der Prozess optimal gesteuert werden, bereits der einweisende Arzt kann die Fälle anmelden und den Prozess beginnen. Der Leistungserbringer hat in das Programm investiert, als der „Bonus-Vertrag“ mit zusätzlichen 800 jährlichen IV-Fällen erhofft wurde. Derzeit können die Investitionskosten nicht amortisiert werden, außerdem werden nicht alle Funktionen des Programms genutzt.

Wenn der Revisionseingriff eines potenziellen Gewährleistungsfalles recht zeitnah nach der Erstimplantation stattfindet, ist die Identifizierung einfach möglich und würde auch außerhalb des IV-Vertrages in der Patientenabrechnung stattfinden. Für den IV-Vertrag ist dann zusätzlich der IV-Fallmanager für solche Fragen zuständig. Dies ermöglicht eine schnellere Abwicklung der Gewährleistungsfälle. Die 2-3 Gewährleistungsfälle wurden ziemlich kooperativ mit dem Kostenträger geregelt. Ein klar standardisierter Prozess wurde hingegen nicht etabliert.

- Messergebnis

Das Messergebnis der Gewährleistungsfälle ist sehr vergleichbar mit den vorherigen Fallstudien. Die Ergebnisse sind allerdings im Controlling nicht standardisiert implementiert. Die Gewährleistungen werden nicht gemessen, weil die Daten außerhalb des Krankenhausaufenthaltes vom Leistungserbringer nicht gemessen werden können. Da allerdings die meisten Komplikationen sehr schnell und meist schon während des Krankenhausaufenthaltes stattfinden, sind die meisten Messergebnisse sehr aktuell

- Qualitätsziel

Das Qualitätsziel des IV-Vertrages ist hier das gleiche wie in den vorherigen Fallstudien. Bezüglich der Anreizwirkung geht der Befragte nicht davon aus, dass die Gewährleistungsklauseln eine qualitätsverbessernde Wirkung haben, da Ärzte und medizinisches Personal nicht absichtlich Fehler machen. Daher sind die Qualitätsgarantie eines IV-Vertrages keine Maßnahmen, die die Qualität langfristig erhöhen.

Die Qualitätsziele des Bonusvertrages wären dagegen anhand von Referenzwerten sehr konkret gesetzt worden. Bei einer Gewährleistungsklausel wären hier nur 2-3 Jahre vereinbart worden, da Kausalitäten danach nicht mehr klar sind. Dieser Trend – nämlich eine Verkürzung der Gewährleistungsfrist – zeigt sich auch in den Diskussionen mit anderen Kostenträgern.

- Vergütungsfunktion/Konsequenzenmanagement

Die Gewährleistung des IV-Vertrages wird nicht als Vergütungsfunktion gesehen. Da auch außerhalb des IV-Vertrages ein ähnliches Konsequenzenmanagement bei offensichtlichen Komplikationen und Revisionseingriffen stattfinden würde, scheint zwar durch Gewährleistungen generell eine gewisse Anreizwirkung zu entstehen, diese sind aber nicht speziell an die Klauseln und das Management der IV-Verträge zu knüpfen.

Für den „Bonus-Vertrag“ wurden zwar noch keine konkreten Vergütungsdetails diskutiert, ein variabler Anteil von bis zu 7-8 % wäre für die Befragten noch akzeptabel – auch wenn angenommen wird, dass es einen zusätzlichen Preisrabatt von bis zu 10 % geben würde. Ein höherer qualitätsvariabler Anteil (über ca. 10 %) wäre schwer akzeptierbar. Der Vergütungsanteil hängt jedoch davon ab, ob das Indikatorensystem akzeptiert wird.

Praktische Umsetzung im klinischen Alltag

Der Prozess zur Umsetzung der Gewährleistungen funktioniert wie folgt:

- Aufdeckung des potenziellen Gewährleistungsfalles durch die Patientenabrechnung (speziell ein Mitarbeiter in der Patientenabrechnung für IV-Verträge) oder den IV-Fallmanager (vor allem bei Revisionsoperation innerhalb eines Jahres)
- Anfrage bei der Klinik bezüglich der Notwendigkeit des Revisionseingriffes
- Einschaltung der Qualitätsabteilung
- Diskussion und Analyse der Fehlersuche
- Bei Materialfehler Einbeziehung der Lieferanten und Einigung der Zahlung
- Einigung mit der Kasse bezüglich der Erstattung – meist sehr kooperative Regelung

Der gesamte Prozess ist relativ aufwendig und nicht standardisiert, da es sich nur um sehr wenige Fälle handelt. Auch das medizinische Personal ist sehr kooperativ in der Fehlersuche, da es im eigenen Interesse ist, Revisionseingriffe zu verhindern.

Beim IV-Vertrag wird kein MDK bestellt und ein Streitfall weitestgehend verhindert, um die Vertragsbeziehung nicht zu gefährden – so kam es bislang auch zu keinem Streitfall. Dadurch ist der Prozess für den IV-Vertrag schneller als bei Kollektivvertragsfällen mit Haftungsdiskussionen. Die Sorge, dass ggf. die potenziellen Gewährleistungsfälle nicht gemeldet werden, ist nicht relevant, da dies eine offene, transparente Kommunikation mit dem Vertragspartner gefährdet.

Gleichzeitig jedoch kann ein Selektivvertrag die Prozessqualität von integrierter Versorgung deutlich verbessern, nämlich durch

- Auswahl der Vertragspartner anhand von Qualitätsaspekten
- Standardisiertes Feedback an Patienten
- Standardisiertes Feedback von medizinischem Personal
- Enge Kooperation mit Rehabilitationseinrichtungen
- Verbesserte Kommunikation zwischen den Sektoren
- Freigabe von Patientendaten zur besseren Prozessabstimmung

Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen

In dieser Fallstudie gibt es keine große Zustimmung zu P4P. Einerseits seien für Ärzte und medizinisches Personal finanzielle Anreize nicht entscheidend, da eine hohe Qualität der klinischen Arbeit immanent

sei. Die tägliche Arbeit richtet sich nach dem Können und Wissen der Beteiligten sowie der Überzeugung zur Patientenhilfe. Für die Verwaltung und Geschäftsführung dagegen könnte eine qualitätsorientierte Vergütung ermöglichen, dass man Investitionen zur Qualitätsverbesserung besser wirtschaftlich argumentieren kann. Hierfür muss jedoch der Aufwand der Implementierung eines P4P-Systems mitbetrachtet werden. Die Schwierigkeit einer objektiven Messung von Qualität wird immer bestehen bleiben, da es sich nicht um einen Produktionsbetrieb handelt und Patienten immer mit sehr unterschiedlichen Voraussetzungen zur Behandlung kommen.

5.4.7. Exkurs: Auswertung des „Bonus-Vertrages“ anhand des Kriterienkatalogs

Da die Kriterienerfüllung des bestehenden IV-Vertrages denen der bisherigen Fallstudien zu sehr ähnelt, wird hier der Kriterienkatalog an den nicht-abgeschlossenen „Bonus-Vertrag“ angewandt. Hier sind zwar einige Umsetzungsparameter kaum bewertbar, dafür gibt es neue Erkenntnisse für die Qualitätsteile des Katalogs.

Das Indikatorensystem des „Bonus-Vertrages“ gleicht in vielen Aspekten den Indikatoren des internen Qualitätsmanagementsystems der Fallstudie 4. Daher werden hier nur die Excel-Auswertungen gezeigt und die Unterschiede beschrieben.

Tabelle 30: Bewertungen Exkurs Fallstudie 5, Systemelement Gesamtsystem und Versorgungsbereich

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
1a	Abdeckung des Gesamtsystems	Ja, Gesamtsystem ist klar	**Niedrig**	Nur orthopädische Fachrichtung abgebildet, daher recht geringe Abdeckung
2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet	Ja, Versorgungs-bereich klar definiert	**Hoch**	Versorgungsbereich anhand von medizinischer Relevanz ausgewählt
2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich	Ja, Versorgungs-bereich klar definiert	**Hoch**	Auswahl der Versorgungsbereiche eher nach medizinischer als nach wirtschaftlicher Relevanz, aufgrund hoher Fallzahlabdeckung und hoher Kosten aber auch wirtschaftlich relevant

Beim Versorgungsbereich gibt es keine Unterschiede zum Exkurs der Fallstudie 4, auch hier werden die Versorgungsbereiche systematisch ausgewählt. Allerdings bezieht sich der Vertrag nur auf den orthopädischen Bereich, daher ist die Abdeckung des Gesamtsystems gering.

Tabelle 31: Bewertungen Exkurs Fallstudie 5, Systemelement Qualitätsindikator

Kriterienkategorie: Aussagekraft

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten	Ja, Indikatoren sind bekannt	**Hoch**	Indikatoren sind vor allem Ergebnisqualität, die das Behandlungsergebnis direkt darstellen
3b	Relevanz des Indikators für ein bestimmtes Fachgebiet	Ja, Indikatoren sind bekannt	**Hoch**	Die betrachteten Indikatoren werden von den Fachbereichen als relevant erachtet, trotzdem wird die Aussagekraft teilweise diskutiert
3c	Sektorbezug vs. sektorübergreifende Nutzung *(Informationskriterium)*	Ja, Indikatoren sind bekannt	**Hoch**	Kostenträger verknüpft auf Patientenebene die Daten verschiedener Sektoren und ermöglicht sektorübergreifende Auswertungen

Kriterienkategorie: Akzeptanz

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
4a	Medizinische Bedeutung in der medizinischen Fachliteratur	Ja, Indikatoren sind bekannt	**Hoch**	Auswahl der Indikatoren vor allem anhand von wissenschaftlicher Literatur, Indikatoren als "Studienendpunkt" einer Vielzahl von wissenschaftlichen Studien
4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften	Ja, Indikatoren sind bekannt	**Hoch**	Steigender Verbreitungsgrad der Indikatoren auch international u.a. aufgrund der Akzeptanz bei den Fachgesellschaften
4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern	Ja, Indikatoren sind bekannt	**Mittel**	Akzeptanz beim medizinischen Personal wird als mittel eingeschätzt, Krankenhaus B nur teilweise überzeugt
4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit	Ja, Indikatoren sind bekannt	**Mittel**	Ergebnisqualität gut verständlich, Relevanz bspw. von Sterblichkeit muss teilweise erklärt werden, Messung beim Leistungserbringer-Management nicht vollständig nachvollzogen
4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit	Ja, Indikatoren sind bekannt	**Hoch**	Sterblichkeit gut verständlich, Relevanz muss teilweise laienverständlich gemacht werden

Ein großer Unterschied dieses Systems ist die Möglichkeit, einen echten sektorübergreifenden Bezug herzustellen, da die Daten verschiedener Sektoren vom Kostenträger zusammengefügt werden. Gerade für einen IV-Vertrag ist dies ein großer Vorteil. Gleichzeitig erscheint jedoch das Indikatorensystem vom Leistungserbringer weniger akzeptiert zu sein, und es gab Schwierigkeiten bei der Verständlichkeit der Indikatorenbeschreibung.

Tabelle 32: Bewertungen Exkurs Fallstudie 5, Systemelement Messung

Kriterienkategorie: Wissenschaftlichkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
5a	Reliabilität	Ja, Messung ist bekannt	**Hoch**	Durch detaillierte, algorithmische Definition der Indikatoren und Messparameter kann eine hohe Reliabilität erreicht werden
5b	Validität	Ja, Messung ist bekannt	**Hoch**	Messergebnisse vielfach in verschiedenen Krankenhäusern überprüft
5c	Risikoadjustierung und Risikoadjustierbarkeit	Ja, Messung ist bekannt	**Hoch**	Durch Anwendung der SMRs werden die Ergebnisse risikoadjustiert

Kriterienkategorie: Messaufwand / -sicherheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
6a	Objektiver Erhebungsaufwand und -komplexität	Ja, Messung ist bekannt	**Niedrig**	Routinedaten sind bereits erhoben, Auswertung findet standardisiert, automatisiert und quasi ohne zusätzlichen Messaufwand statt
6b	Manipulationssicherheit / Objektivität der Messung	Ja, Messung ist bekannt	**Hoch**	Ein wichtiger Vorteil der Nutzung von Routinedaten ist die hohe Manipulationssicherheit, da die Daten zu Abrechnungszwecken erhoben und vielfältig überprüft werden
6c	Grad der Integration in automatische Prozesse / IT-Systeme	Ja, Messung ist bekannt	**Hoch**	Automatisierte Messung für das Krankenhaus sollte etabliert werden

Bei der Messung gibt es keine Unterschiede zum Exkurs der Fallstudie 4.

Tabelle 33: Bewertungen Exkurs Fallstudie 5, Systemelement Ziele

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
7a	Zielkonsistenz	Ja, Ziele sind bekannt	**Hoch**	Ziele des QM-Systems entsprechen alle den akzeptierten Qualitätszielen und sind mit denen des medizinischen Personals konsistent
7b	Zielkongruenz	Ja, Ziele sind bekannt	**Hoch**	Nur eine Fachrichtung inkludiert, Ziele koordiniert entwickelt und gleichzeitig, widerspruchsfrei erreichbar
7c	Reichweite	Ja, Ziele sind bekannt	**Mittel**	Hoher Anteil der Fallzahlen und daher recht umfassendes Ziel- und Indikatorensystem, hier aber nur eine Fachrichtung

Kriterienkategorie: Operationalisierbarkeit/Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
8a	Erreichbarkeit	Ja, Ziele sind bekannt	**Hoch**	Ziele werden von einigen Krankenhäusern bereits erreicht
8b	Transparenz	Ja, Ziele sind bekannt	**Hoch**	Ziele sowie Messergebnisse wären bei Beteiligten bekannt
8c	Akzeptanz	Ja, Ziele sind bekannt	**Mittel**	Beim Leistungserbringer wurden die Indikatoren sowie die Ziele aufgrund von fehlenden Erklärungen der Herleitung und Berechnung eher angezweifelt
8d	Langfristige Anreizwirkung der Ziele	Ja, Ziele sind bekannt	**Hoch**	Wirkung in einigen Studien gezeigt, Anreiz gerade für "Schwarze Schafe"
9a	Gerechtigkeit	Ja, Ziele sind bekannt	**Hoch**	Durch Risikoadjustierung werden Unterschiede im Patientenkollektiv berücksichtigt, einige Beteiligte zweifeln Gerechtigkeit jedoch an

Bei den Qualitätszielen gibt es kaum Unterschiede zum Exkurs der Fallstudie 4, allerdings werden sowohl die Akzeptanz als auch Gerechtigkeit vom Leistungserbringer hinterfragt – trotz durchgeführter Risikoadjustierung. Eine Aufgabe wäre hier vermutlich eher eine bessere Information aller Beteiligten. Auch die Reichweite ist aufgrund der Umsetzung innerhalb eines Selektivvertrag mit einer Fachabteilung geringer.

Tabelle 34: Bewertungen Exkurs Fallstudie 5, Systemelement Vergütungsfunktion

Kriterienkategorie: Wirksamkeit/Zielgerichtetheit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
10a	Effizienz	Schwierig, da Details des Vertrages nicht bekannt	**Wahr-scheinlich mittel**	Vorab-Investition für Software und Verhandlung relativ hoch, mit vermutetem variablen Anteil von ca. 7-8% variabler Anteil aber potenziell amortisierbar
10b	Effektivität	Schwierig, da Details des Vertrages nicht bekannt	**Potenziell hoch**	Bislang keine Studien zur Verbesserung mit qualitätsvariabler Vergütung bekannt
10c	Flexibilität	Schwierig, da Details des Vertrages nicht bekannt	**Wahr-scheinlich mittel**	Vermutlich im Voraus festgelegte Vergütungsfunktion, aber relative Vergütungsanteile, dadurch mittlere Flexibilität

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
11a	Anteil der Vergütung	Schwierig, da Details des Vertrages nicht bekannt	**Hoch**	Variabler Anteil zwischen 7-8% vermutet, damit höher als Vielzahl von P4P-Systemen und Empfehlungen
11b	Planbarkeit	Schwierig, da Details des Vertrages nicht bekannt	**Mittel**	Vergütungsfunktion gut planbar, allerdings Aktualität der Ergebnisse eher schlecht aufgrund der mehrjährigen Datenanalyse

Der Vergütungsanteil wird vom Leistungserbringer bei 7-10 % eingeschätzt. Dies würde einen recht hohen Anteil ergeben, da es bereits relativ hohe Investitionen für Softwareprogramme und einen aufwendigen Verhandlungsprozess gab, ist eine hohe Effizienz des IV-Vertrages zumindest für den Leistungserbringer vorerst nicht wahrscheinlich. Gleichzeitig wird davon ausgegangen, dass der Vergütungsanteil im Voraus aber durch relative Gestaltung recht flexibel festgelegt wird. Die Planbarkeit der Vergütungsfunktion ergibt sich aus der planbaren Vergütungsfunktion – die hier vermutlich gegeben ist – und der Aktualität der Messergebnisse, die aufgrund der Datenlage etwas schwieriger wäre.

Tabelle 35: Bewertungen Exkurs Fallstudie 5, Systemelement Messergebnis

Kriterienkategorie: Operationalisierbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
12a	Nutzung zur Steuerung	Ja	**Mittel**	Derzeit eher ein System zur Analyse der Daten von Seiten der Kostenträger, Nutzbarkeit zur Steuerung beim Leistungserbringer noch zu überprüfen
12b	Aktualität	Ja	**Niedrig**	Ergebnisse liegen aufgrund von Daten von mehreren Jahren (1-Jahres-Sterblichkeit) nicht zeitnah vor, nach Aussage des Leistungserbringers sind Daten mindestens 3-4 Jahre alt
12c	Kontrollierbarkeit	Ja	**Mittel**	Ergebnisse aufgrund der Verfügbarkeit gut kontrollierbar, Problem nur wegen der Aktualität und der notwendigen Übertragung vom Kostenträger
12d	Praktische Umsetzung im Controlling	*Nein, da keine Implementierung in Fallstudie*	*nicht bekannt*	*nicht bekannt*
12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit	Ja	**Hoch**	Sowohl auf Krankenhaus als auch auf Fachabteilungsebene ist die Zuordnung der Messergebnisse klar

Kriterienkategorie: Vergleichbarkeit

#	Einzelkriterium	Bewertbar	Erfüllungs-grad	Beschreibung
13a	Zeitverlauf	Ja	**Hoch**	Daten können im Jahresvergleich und -entwicklung gezeigt werden
13b	Querschnittsvergleich/Ranking	Ja	**Hoch**	Rankings auf Fachabteilungsebene im Vergleich zu anderen Kliniken und Bundesdurchschnitt möglich
13c	Justiziabilität	*nicht bekannt*	*nicht bekannt*	*nicht bekannt*

Die Messergebnisse sind nur wenig aktuell – der Leistungserbringer nimmt an, dass die Daten für die Qualitätsanalyse mindestens 3-4 Jahre alt sind. Dies führt außerdem zu einer schwierigeren Nutzbarkeit zu Steuerung und Kontrollierbarkeit. Gleichzeitig ist ein Zeitverlauf und Ranking integriert und sehr transparent möglich.

5.4.8. Fallstudie 6: Krankenhaus C

5.4.8.1. Beschreibung des Krankenhauses und der IV-Verträge

Das analysierte Krankenhaus C ist ein Haus der Maximalversorgung. Es hat mehrere Standorte in einer deutschen Metropolregion. Das Krankenhaus ist KTQ-zertifiziert und nimmt an einer regionalen Transparenzinitiative sowie an „4QD – qualitätskliniken.de“ teil.

Das Krankenhaus hat mehrere IV-Verträge abgeschlossen, keines davon mit qualitätsorientierter Vergütung oder Gewährleistungsklauseln. Die IV-Verträge betreffen unter anderem Herzoperationen, Endoprothetik-Behandlungen sowie Operationen, die sowohl ambulant als auch stationär durchgeführt werden können. Der einzige noch bestehende Vertrag betrifft kardiologische Operationen und beinhaltet

50-60 Fälle. Der Endoprothetikvertrag, der mit einer kleineren Kasse geschlossen wurde, enthielt nur sehr wenige Fälle – einige Qualitätsaspekte waren inkludiert, sowie u. a. clinical pathways.

Zum Thema Gewährleistung geben die beiden Befragten unterschiedliche Meinungen wieder: Während ein Befragter Gewährleistungen nicht akzeptiert und als rechtlich fragwürdig betrachtet, unterstützt die andere Befragte die Idee der Gewährleistung – wenn auch das derzeitige System deren Umsetzung nicht fördert. Die Zweifel basieren vor allem auf der Grundidee, dass eine Lastenverschiebung von Ereignissen, die nicht vom Krankenhaus beeinflussbar sind (wie Lebensdauer von Hüftimplantaten oder Infektionen), auf den Leistungserbringer nicht rechtens sei. Krankenhäuser, die eine solche Vereinbarung abschließen wollen, sollten dies zunächst rechtlich prüfen – eine wirtschaftliche Prüfung muss anschließend eruieren, ob mit einer deutlichen Fallzahlsteigerung zu rechnen ist. Eine anders strukturierte Gewährleistung – wie beispielsweise die Vereinbarung auf eine abgesenkte Revisionsrate – könnte schon eher akzeptiert werden[1].

Das Krankenhaus verfügt außerdem über ein System sehr detaillierter Zielvereinbarungen – sowohl mit allen Mitarbeitern als auch mit Chefärzten. Diese beinhalten unter anderem auch Qualitätsindikatoren, häufig Aspekte der Mitarbeiter- und Patientenzufriedenheit. Zusätzlich werden einige Indikatoren der externen Qualitätssicherung ausgewählt und in Auswertungsgesprächen entweder im Rahmen von Medical Boards (standortübergreifende Treffen von Fachabteilungen) oder in Einzelgesprächen zwischen ärztlicher Leitung, Chefärzten und Mitarbeitern der Qualitätsmanagementabteilung besprochen. Außerdem werden Zielwerte für das kommende Jahr gemeinsam festgelegt. Mittelfristig soll anhand dieser Qualitätsindikatoren auch die Zielvereinbarungen und somit die variable Vergütung der Chefärzte bestimmt werden. Daher wird diese Systematik der Auswahl, Bewertung und Zielsetzung von Qualitätsindikatoren im Folgenden bei den Systemelementen besprochen.

Auch ein Risikomanagementsystem besteht, das allerdings eher auf finanzielle Aspekte fokussiert. Risiken werden relevant, sobald sie Auswirkungen auf das Leistungsgeschehen haben – bspw. Wettbewerberverhalten oder aber auch Qualitätsrisiken wie Hygienerisiken.

Das Krankenhaus befindet sich derzeit in einer Sanierungsphase. Die Qualitätsdiskussion ist nicht mit der finanziellen Seite der Sanierung gleichrangig – es war bisher nicht möglich, den Zusammenhang zwischen guter Qualität und niedrigen Kosten im gesamten Krankenhaus zu verbreiten.

[1] Vergleiche hierzu die juristische Diskussion bei Abschnitt 5.2.3.

5.4.8.2. Ergebnisse der Interviews

Gründe für die Vertragsschließung

Die abgeschlossenen IV-Verträge waren weitestgehend interessant, da sie extra-budgetäre zusätzliche Fälle ermöglichen konnten. Gleichzeitig waren auch hier die Rabattvorstellungen des Kostenträgers relevant. Wenn allerdings – wie hier der Fall – keine Mehrfälle durch Patientensteuerung erreicht werden, ist der administrative Aufwand zu hoch und der Vertrag wird nicht weiter umgesetzt.

Derzeit werden vor allem IV-Verträge diskutiert, die dem Krankenhaus eine freie Entscheidung ermöglichen, ob eine Leistung ambulant, teil-stationär oder stationär durchgeführt wird, indem die Vergütung stets anhand des gleichen Mechanismus stattfindet. Solch ein innovativer Vertrag ist für das Krankenhaus von großem Interesse, konnte aber bislang keine ausreichende Fallzahl erreichen.

Beschreibung der Systemelemente

Im Folgenden werden nur die Systemelemente besprochen, die in den IV-Verträgen, vor allem aber den Zielvereinbarungs- und Auswertungsgesprächen relevant sind.

- Qualitätsindikator

Im Krankenhaus wurden zwar keine neuen Indikatoren entwickelt, allerdings wird sehr aktiv eine Auswahl der Indikatoren des AQUA-Berichtes getroffen, um damit im internen Qualitätsmanagementsystem die Kliniken steuern zu können. Diese Auswahl entsteht im kollegialen Austausch zwischen den Chefärzten, den ärztlichen Direktoren der Standorte sowie der Qualitätsmanagementabteilung. Für jede Fachrichtung werden jährlich drei Kennzahlen ausgewählt, anhand derer eine detaillierte Auswertung sowie eine Steuerung stattfindet.

Die Auswahlkriterien sind hierbei vor allem der Patientennutzen: Ein Indikator, dessen Ergebnisverbesserung zu einer aktiven Verbesserung der Behandlung des Patienten führt, wird stets bevorzugt. Beispiele sind eine sichere Behandlung, saubere Diagnostik und Diagnostik bei Entlassung. Zweite Priorität ist die Frage, welcher Indikator eine Prozessoptimierung unterstützen kann.

Einige dieser Indikatoren werden anschließend auch für Zielvereinbarungen genutzt – diese umfassen jedoch meist Ergebnisse von Patienten- oder Mitarbeiterbefragungen. Zukünftig sollen auch die ausgewählten Indikatoren der externen Qualitätssicherung (mit Fokus auf der Ergebnisqualität) in die Zielvereinbarungen mit aufgenommen werden.

- Qualitätsmessung

Als wichtigstes Kriterium der Qualitätsmessung werden Manipulationssicherheit, Validität, Evidenz und Relevanz für Patientennutzen beschrieben.

- Messergebnis

Ein wichtiger Aspekt der Messergebnisse für das Krankenhaus ist der richtige Vergleichskreis. Um eine detailliertere Auswertung der Indikatorenergebnisse zu ermöglichen und möglichst nachvollziehbare Vergleiche zu haben, werden beispielsweise die Ergebnisse der „4QD"-Initiative hinzugezogen. Auch die Messergebnisse von Universitätskliniken werden gesondert ausgewiesen und ermöglichen eine detailliertere Diskussion mit einigen Chefärzten (obschon das Krankenhaus selbst keine Universitätsklinik ist). Gerade ein möglichst spezifisches Ranking ist daher für dieses Krankenhaus relevant.

Gleichzeitig werden die Messergebnisse auch derzeit bereits zu Steuerungszwecken genutzt: Sie stellen einen Teil der Kennzahlen, mittels derer die einzelnen Kliniken im Planungs- und Controllingprozess gemessen werden.

- Qualitätsziel

Qualitätsziele sind bei diesem Unternehmen auf verschiedenen Ebenen etabliert. Zunächst werden generelle Qualitätsziele bereits strategisch für das Gesamtunternehmen gesetzt – diese strategische Dimension soll allerdings detaillierter und ernsthafter verfolgt werden.

Relevanter sind hier die Zielwerte, die für die zuvor ausgewählten Qualitätsindikatoren gesetzt werden. Diese Grenzwerte orientieren sich zunächst an den Referenzwerten bzw. dem Bundesdurchschnitt der Indikatorenergebnisse des AQUA-Berichtes. Die konkreten Zielwerte werden im kollegialen Austausch mit den Fachabteilungen bestimmt. Hierfür werden zunächst die Ergebnisse der Kliniken analysiert und anhand verschiedener Rankings verglichen. Außerdem werden bereits Ursachen für mögliche Ausreißer sowie die Einhaltung von internen Leitlinien überprüft.

Eine unternehmensweite Richtlinie ist zunächst, dass das Ziel immer mindestens der Bundesdurchschnitt sein muss und zusätzlich von Jahr zu Jahr ansteigt. Eine langfristige Anreizwirkung wird somit sichergestellt. Die Einigung auf Zielwerte im Dialog ist manchmal schwierig und zeitaufwendig – einige Chefärzte blocken die Diskussion recht schnell ab. Mittelfristig soll es daher Zielwerte geben, die direkt von der Geschäftsführung vorgegeben werden.

- Vergütungsfunktion/Konsequenzenmanagement

Die hier beschriebenen Zielwerte sind nicht mit Konsequenzenmanagement hinterlegt. Zwar findet die Diskussion regelmäßig mit der Klinikleitung und den ärztlichen Direktoren statt, eine Vergütungsfunktion gibt es jedoch nicht. Auch erste Versuche der Einführung von „Peer-Reviews" sind aufgrund fehlender Bereitschaft der Chefärzte wieder zurückgenommen worden.

Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen

Die Relevanz und Umsetzbarkeit von P4P wird wiederum von den Befragten sehr unterschiedlich eingeschätzt. Zwar wünschen sich beide einen offensiveren Umgang mit Qualitätsaspekten auch im Management, die Vorstellungen für eine sinnvolle Qualitätsausrichtung weichen aber voneinander ab:

Der erste Befragte erhofft sich zukünftig ein stärkeres Zusammenwachsen der verschiedenen Sektoren. Hierfür sei beispielsweise eine „Regionalbudgetierung“ möglich, die die Leistungen unabhängig vom Sektor vergüten. Hierfür könnten beispielsweise IV-Verträge genutzt werden. Eine qualitätsorientierte Vergütung im Sinne von P4P wird dagegen abgelehnt. Vor allem, da die Gefahr gesehen wird, dass eine qualitätsorientierte Vergütung vordergründig dafür genutzt wird, um das Budget für Krankenhäuser zu schmälern. Die Kostenträger wollen dadurch neben einer Kostensenkung unter anderem ein „Einkaufsmodell“ einführen und den Kontrahierungszwang umgehen. Außerdem bestehe das Risiko, dass das Qualitätsmesssystem Ungerechtigkeiten und Unterfinanzierung fördere.

Für eine akzeptierte Qualitätsmessung ist zunächst eine gute Dokumentation mit der Möglichkeit von unterjähriger Auswertung wichtig. Das Messsystem muss valide, evident, sehr transparent und gerecht sein. Aufgrund der hohen Komplexität der stationären Qualität sei eine breite Akzeptanz schwer zu erreichen. Für ein System der qualitätsorientierten Vergütung allerdings seien die derzeitigen Messinstrumente bei weiten nicht ausreichend valide, um Akzeptanz zu erlangen. Gerade die notwendige Risikoadjustierung sei kaum zu erreichen.

Eine sinnvollere Entwicklung wird in einer möglichen Versorgungssteuerung gesehen, d. h. die Konzentration von Behandlungen an einigen, qualitativ hochwertigen Standorten. Hierfür müssen Krankenhäuser vor allem Strukturvoraussetzungen erfüllen und ausreichend finanziert sein, um gute Qualität zu erreichen. Eine vernünftige, sektorübergreifende Versorgung mit Fokus auf Strukturqualität wird bevorzugt.

Die zweite Befragte befürwortet qualitätsorientierte Vergütung. Gerade für verschiedene Zusatzleistungen (sowohl Routineeingriffe als auch hochspezialisierte Leistungen) wäre P4P gut vorstellbar. Einzelne Indikatoren der externen Qualitätssicherungen wären hierfür bereits geeignet – allerdings müsste eine klare Auswahl wissenschaftlich erfolgen. Speziell das Thema der Manipulationssicherheit ist dann von höchster Bedeutung – gerade bei Routinedaten jedoch sei der Vorwurf von bewussten Manipulationen schon heute illusorisch. Außerdem müssen die Indikatoren „sinnvoll“ sein und dem Patientennutzen entsprechen. Auch eine finanzielle Relevanz müsste dann sichergestellt sein. Ein wichtiger Aspekt ist zum Beispiel die Feststellung der Indikationssicherheit, damit es nicht weiter zu Überversorgung kommen kann. Ein regionaler Ansatz von P4P würde zwar bevorzugt, Deutschland sei davon aber mindestens 15-20 Jahre entfernt.

Derzeit gibt es im Unternehmen noch keine ausreichende Qualitätskultur um ein System von qualitätsorientierter Vergütung einzuführen – zwar könne man sich auf „internes P4P“ im Rahmen von Zielvereinbarungen einigen, für einen gemeinsamen Ansatz mit den Kostenträgern fehlt aber intern noch die Akzeptanz.

Auch außerhalb des Krankenhauses sei die Qualitätskultur in der deutschen Krankenhauslandschaft noch nicht bereit für eine qualitätsorientierte Vergütung. Vor allem aufgrund der hohen Bedeutung der Lobbyvereinigungen sei mit einem Konsens zum Thema Qualitätsmessung nicht zu rechnen.

5.4.8.3. Auswertung anhand des Kriterienkatalogs

In dieser Fallstudie soll das System der Zielvereinbarungen und internen Auswertungsgespräche mit Qualitätsaspekt analysiert werden. Dabei werden nur die Kriterien diskutiert, bei denen eine Aussage besonders erkenntnisreich ist – diese beziehen sich ausschließlich auf die Systemelemente der Messergebnisse und Qualitätsziele.

Tabelle 36: Bewertungen Fallstudie 6, Systemelement Ziele

Kriterienkategorie: Operationalisierbarkeit

Sub-Kriterium		Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
8a	Erreichbarkeit	Ja	**Hoch**	Zielwerte werden in kollegialem Austausch bestimmt und Ziele werden teilweise bereits erreicht	Keine
8b	Transparenz	Ja	**Mittel**	Ziele sowie Messergebnisse werden den Fachabteilungen standortübergreifend präsentiert und diskutiert	Transparenzerhöhung durch Veröffentlichung der derzeit internen Ziele
8c	Akzeptanz	Ja	**Hoch**	Durch gemeinsame Entscheidung bezüglich der Ziele mit Fachabteilungen, ärztlichen Direktoren und Qualitätsabteilung Akzeptanz sehr hoch	Keine
8d	Langfristige Anreizwirkung der Ziele	Ja	**Hoch**	Ziele werden jährlich überprüft und ggf. angepasst, Hauptrichtlinie ist kontinuierliche Verbesserung	Keine

Aufgrund der Konkretisierung der Zielwerte im kollegialen Dialog und Verhandlungsmodus sind Akzeptanz und Erreichbarkeit sehr hoch. Auch die Transparenz innerhalb des Krankenhauses ist hoch, könnte aber für die Öffentlichkeit erhöht werden. Gerade die langfristige Anreizwirkung ist aufgrund der kontinuierlichen Anpassung der Ziele und Richtgröße einer ständigen Verbesserung sehr hoch.

Tabelle 37: Bewertungen Fallstudie 6, Systemelement Messergebnis

Kriterienkategorie: Operationalisierbarkeit/Vergleichbarkeit

Sub-Kriterium		Bewertbar	Erfüllungs-grad	Beschreibung	Verbesserungen
12a	Nutzung zur Steuerung	Ja	**Mittel**	Derzeit in Auswertungsgesprächen genutzt, Einbindung in Zielvereinbarungen geplant	Verknüpfung mit Zielvereinbarungen
12b	Aktualität	Ja	**Mittel**	Messergebnisse anhand der AQUA-Qualitätsberichte, meist 2-3 Jahre alt	Einige Ergebnisse zeitnaher auswerten
13b	Querschnitts-vergleich/Ranking	Ja	**Hoch**	Vergleiche mit Bundesdaten, Universitätskliniken und 4QD-Kliniken	Erstellung von Rankings

Die Messergebnisse sind nur wenig aktuell – die Daten der Qualitätsberichte der externen Qualitätssicherung sind ca. 2-3 Jahre alt sind. Trotzdem werden die Messergebnisse derzeit bereits in

„Auswertungsgesprächen“ genutzt und diskutiert – eine Einbindung in die Zielvereinbarungen der Chefärzte und damit höhere Steuerungsnutzung ist geplant. Vergleiche der Messergebnisse werden mit Bundesdurchschnitten, Universitätskliniken sowie den Teilnehmern der „4QD“-Initiative durchgeführt.

5.4.9. Fallstudie 7: Krankenhaus D

5.4.9.1. Beschreibung des Krankenhauses und der IV-Verträge

Das Krankenhaus D ist ein freigemeinnütziges Haus mit ca. 600 Betten in einer deutschen Großstadt. Es gibt einen IV-Vertrag, der gynäkologische Operationen umfasst. Da dieser jedoch keine Gewährleistung inkludiert, wird in dieser Fallstudie unter anderem auf die allgemeine Qualitätsorientierung des Leistungserbringers sowie Qualitätseinschätzungen eingegangen.

Der IV-Vertrag wurde 2010 geschlossen und zwischen 10 und 30 Fälle pro Jahr sind eingeschrieben. Dies ist relativ konstant, während die Anzahl von Regelversorgungsfällen in der Fachklinik und von dem Kostenträger in dem Zeitraum gestiegen ist. Die Einschreibungsquote ist dementsprechend rückläufig. Für die Fallvergütung ist ein Rabatt von 12,5 % vereinbart. Dadurch hat das Krankenhaus derzeit keine finanzielle Inzentivierung, möglichst viele Fälle in den Vertrag einzuschreiben. Der IV-Vertrag soll vor allem eine enge Beziehung zu einem der Kerneinweiser und niedergelassenen Gynäkologen mit großer Fallzahl ermöglichen. Der Vertrag ist auf Prozess- und Ablaufqualität fokussiert. Ein weiterer IV-Vertrag ist mit einem anderen Kostenträger vereinbart, jedoch gibt es keine Patienten in dem Vertrag, da der Niedergelassene nicht zur Klinik steuert. Zur Qualitätssicherung sind auch hier nur die gesetzlichen Regelungen wiederholt – so muss ein internes Qualitätsmanagementsystem vorgehalten sein, und auf Anfrage können die Ergebnisse eingefordert werden. Eine Gewährleistungsklausel wurde explizit abgelehnt, vor allem da es sich um sehr große, seltene, gynäkologische Operationen handelt, für die eine Haftung für den Operateur nicht in Frage kam.

Weitere IV-Verträge sind in Diskussion: Ein Vertrag der Endoprothetik, der eventuell eine Gewährleistungsklauseln beinhalten würde, sowie ein Vertrag mit einem MVZ aus der Region bezüglich Prozessqualität.

In den Zielvereinbarungen der Führungskräfte sind ca. 30 % qualitätsorientiert. Diese sind jedoch eher „weiche“ Ziele, die nicht anhand von Qualitätsindikatoren gemessen werden können. Diese Zielvereinbarungsgespräche werden mit Chef-, Ober- und Assistenzärzten durchgeführt, nur die Vergütung der oberen Führungsebene ist jedoch variabel. Auch in allen ärztlichen Dienstvereinbarungen werden als strategisches Unternehmensziel Wirtschaftlichkeit, Mitarbeiterzufriedenheit und Qualität festgelegt. Gleichzeitig sollen die Ziele stets messbar und nachvollziehbar sein – hierfür verpflichtet sich der Arbeitgeber dazu, die nötigen Ressourcen bereitzustellen.

Das Krankenhaus hat bereits an mehreren Qualitätsinitiativen teilgenommen oder nimmt daran teil. Zunächst hat der Leistungserbringer am QSR-Verfahren teilgenommen – ohne auffällige Qualitätsergebnisse. Aufgrund des hohen zeitlichen Aufwands der Peer-Review-Verfahren wurde die Teilnahme wieder beendet. Derzeit ist das Krankenhaus am Internetportal der „Qualitätskliniken“ beteiligt. Außerdem wurde der Prozess begonnen, das gesamte Krankenhaus (und nicht nur einzelne Kliniken) mit dem ISO-Verfahren zertifizieren zu lassen. Außerdem gibt es ein zertifiziertes Brustzentrum. In den nächsten 3-5 Jahren soll außerdem an IQ^{M} teilgenommen werden, da Peer-Review langfristig als sinnvollstes Verfahren angesehen wird.

5.4.9.2. Ergebnisse der Interviews

Gründe für die Vertragsschließung

Für diesen Leistungserbringer wurde der IV-Vertrag geschlossen, um eine enge Bindung zum beteiligten niedergelassenen Gynäkologen zu ermöglichen und die Patientenflüsse mindestens beizubehalten. Dieser ambulante Vertragspartner ist zunächst auf den beteiligten Kostenträger zugegangen, bevor im Anschluss der Chefarzt des Leistungserbringers angesprochen wurde. Die Auswahl des stationären Leistungserbringers fand auf Wunsch des niedergelassenen Gynäkologen statt. Grund für die Zustimmung des stationären Leistungserbringers – der kaum mit Mengeneffekte rechnen konnte und hohe Rabattabschläge akzeptieren musste – war vor allem, nicht in eine Verweigerungsstellung gegenüber einem wichtigem Einweiser zu gelangen. Qualitative – vor allem ergebnisqualitative – Gründe waren nicht relevant.

Eine Patientensteuerung hat bislang nicht stattgefunden, da zum einen der Leistungserbringer für die Patienten recht weit weg ist, und zum anderen die Krankenkasse kein großes Interesse an Steuerung zeigt. Der Kostenträger wollte vermutlich vor allem verhindern, dass Operationen und die dadurch entstehenden Kosten sowohl im ambulanten als auch stationären Sektor stattfinden und schloss den Vertrag zu Kontrollzwecken ab. Politische Gründe für die Vertragsschließung werden in einem sektorübergreifenden Zusammenwachsen sowie einer Verbesserung der Ablaufprozesse gesehen.

Evaluation, Monitoring und Controlling der Verträge

Der bestehende IV-Vertrag wird nicht evaluiert – es gibt aber ein Instrumentarium für das Risikomanagement bei Haftpflichtfällen und für die Evaluierung des Brustzentrums. Auch bei der wirtschaftlichen Vorabevaluation dieses Zentrums wurden Verhandlungsaufwand, Personalaufwand der Umsetzung und Qualitätsmessungen nicht mitbetrachtet. Wichtiger als diese wirtschaftlichen Aspekte waren Themen der Mitarbeiterakquise, Weiterbildung und Marketingaspekte gegenüber den Patienten.

Beschreibung der Systemelemente [1]

- Versorgungsbereich

Mit Ausnahme des derzeit existierenden IV-Vertrages wird der Versorgungsbereich von Selektivvertragsverhandlungen meist vom Leistungserbringer indiziert. Gerade die Chefärzte zeigen für die eigene Fachabteilung dann eine hohe Initiative.

- Qualitätsindikator

Die Befragten zeigen ein deutliches Interesse an Qualitätsindikatoren, die Prozessqualität messen – vor allem anhand von Patientenumfragen. Hierbei wird die Prozessqualität bspw. mittels Wartezeiten und Ablaufqualität bewertet. Auch die ISO Zertifizierung wird vor allem aufgrund des starken Fokus auf Prozesse geschätzt. Umfragen zur Wiederempfehlungsrate und Patientenzufriedenheit werden ebenfalls vermehrt durchgeführt.

Zwar werden gleichzeitig auch die Ergebnisse von Indikatoren wie bspw. der Komplikationsrate als sehr relevant betrachtet. Es wird jedoch davon ausgegangen, dass Ergebnisqualität vom medizinischen Personal nicht akzeptiert wird. Da gerade bei Gewährleistungsfällen der Haftungsaspekt wichtig ist, sollte in P4P nur solche Indikatoren mit aufgenommen werden, die vom Leistungserbringer und vom medizinischen Personal möglichst direkt gesteuert werden können.

- Qualitätsmessung

Bislang wurde vom Kostenträger keine Veröffentlichung der Ergebnisse für den IV-Vertrag speziell angefragt. Auch bei der Auswahl des Leistungserbringers wurde nicht mittels einer Qualitätsmessung entschieden.

Der wichtigste Aspekt bei der Qualitätsmessung wird in der Manipulationssicherheit gesehen – daher werden Routinedatenauswertungen bevorzugt. Eine Teilnahme im IQ^{M}-System wurde überlegt, aus Zeitgründen aber wieder verworfen worden. Wichtig ist außerdem, dass die Qualitätsmessung so aussagekräftig stattfindet, dass nicht reine Dokumentationsfehler zu Malus führen können.

- Qualitätsziel

Zielwerte sind beim Leistungserbringer nur in den Zielvereinbarungen enthalten. Diese beinhalten zwar 30 % Qualitätsaspekte, diese sind jedoch eher „weiche“ Grenzwerte. Auch der GB-A zeige derzeit gewisse Hemmnisse, Richtgrenzen konkret anzugeben.

[1] Da der vorhandene IV-Vertrag nur zu vereinzelten Systemelementen Erkenntnisse ermöglicht, wurden die Befragten vor allem nach Expertenmeinungen und eigenen Einschätzungen zu einigen Elementen befragt. Die Ergebnisse dieser Diskussion werden im Folgenden dargelegt.

- Vergütungsfunktion/Konsequenzenmanagement

Wichtig für die Höhe der Vergütung ist eine möglichst konkrete Vorstellung– so können beispielsweise die Beträge der variablen Vergütung für den Leistungserbringer dann gut verständlich gemacht werden, wenn sie beispielsweise die Einstellung einer weiteren Pflegekraft ermöglichen.

Zur Frage, ob besser ein Bonus- und/oder Malussystem etabliert werden sollte, sind die Befragten unterschiedlicher Meinung. Während der erste Befragte ein reines Bonussystem für sinnvoller hält da Strafen nicht erfolgsversprechend bei der Ärzteschaft sind, zeigt sich der zweite Befragte auch einem Bonus- und Malussystem gegenüber offen – vor allem, um dem Kostenträger entgegen zu kommen.

Praktische Umsetzung im klinischen Alltag

Die Umsetzung des IV-Vertrages findet anhand des SAP-Abrechnungsprogrammes statt – hier ist die Kodierung für die IV-Vertragsnummer anhand der Behandlungskategorie hinterlegt. Auch der Abschlag von 12,5 % auf jeden Kostenpunkt dieses Vertrages ist programmiert. Diese Programmierung war recht aufwendig. Auch die Einschreibung in den IV-Vertrag ist weiterhin mit Aufwand verbunden. Insgesamt hat der Leistungserbringer wenig Interesse daran, möglichst viele Patienten in den IV-Vertrag aufzunehmen, da zum einen ein höherer Aufwand, zum anderen geringe Einnahmen entstehen. Auch der Kostenträger ist nicht weiter daran interessiert. Eine Umsetzung von Gewährleistungen könnte auch in das SAP-Programm integriert werden, dies wird aber ebenfalls als aufwendig eingeschätzt.

Einschätzung der Befragten bezüglich P4P und weiteren Entwicklungen

Die Strategie des Leistungserbringers ist in drei Säulen unterteilt, die alle gleichgewichtig und miteinander verknüpft sind: Qualität, Mitarbeiter und Wirtschaftlichkeit. Die Qualität hängt auch bspw. von der Mitarbeiterzufriedenheit und -kompetenz ab. Die Behandlungsqualität ist aber für IV-Verträge, Regelversorgung und Privatpatienten gleich. Dies entspricht auch dem Anspruch eines Leistungserbringers der freigemeinnützigen Trägerschaft.

Zunächst wird bei diesem Leistungserbringer festgestellt, dass die Ärzteschaft intrinsisch an einer hohen Ergebnisqualität interessiert ist. Daher sei auch eine variable Vergütung für Ärzte immer schwierig durchzusetzen, da auch eine große Gefahr von Fehlanreizen besteht. Gleichzeitig fördert das DRG-System bereits Qualität, weil die rationellste Leistung immer die qualitativste ist. Komplikationen werden ebenso vermieden, weil sich die Erhöhung des Casemixes in diesen Fällen wirtschaftlich nicht lohnt. Insgesamt werden dadurch Prozesse gestrafft und verbessert – unbewusst wird also bereits an Qualitätsschrauben gedreht.

Die erste Schwierigkeit bei P4P ist zunächst die Frage, wer die Entscheidung zu Qualitätsmessung trifft. Da unter anderem die Selbstverwaltungsorgane häufig ein starkes Lobbyverhalten aufweisen, ist in

diesem System eine Einigung auf Qualitätsindikatoren schwierig. Hier wäre eine neutrale Stelle wie der TÜV sinnvoll.

Der Leistungserbringer geht nicht davon aus, dass die Kultur im eigenen Krankenhaus eine Einführung von P4P-Aspekten bereits ermöglicht. Gerade die Fehlerkultur – auch in anderen Häusern in Deutschland – ist noch nicht ausreichend ausgeprägt. Qualität muss erst diskutabel sein, bevor es variabel vergütet wird. Allerdings kann und sollte nicht auf eine kulturelle Veränderung gewartet werden.

Für die Einführung eines P4P-Systems in Deutschland – mit dem sicher gerechnet wird – werden folgende Fragestellungen aufgeworfen:

- Wie soll die Qualitätsmessung durchgeführt werden?
- Was sind valide Kennzahlen?
- Wer misst?
- Wann wird gemessen? (direkt nach der Entlassung zu früh, lieber erst nach 6 Wochen beim Patienten anrufen)
- Wie kann Qualität nachhaltig verbessert werden? (bspw. durch Etablierung von Peer-Reviews)

Derzeit wird eine Implementierung von P4P-Aspekten im selektivvertraglichen Bereich bevorzugt – gerade als erster Schritt. Gerade die Auswahl der Vertragspartner ist dafür relevant. Größere, flächendeckende IV-Verträge mit qualitätsorientierter, aber vor allem extra-budgetärer Vergütung würden positiv aufgenommen und könnten auch gute Qualitätsverbesserungen erzielen. Dann wären Investitionen in Qualitätsmaßnahmen rentabel – solange die IV-Verträge vor allem auf Rabatten aufbauen, lohnen sich diese nicht.

Als wichtigeres Thema als qualitätsorientierte Vergütung werden Risikomanagement und weitere Prozessoptimierungen eingeschätzt.

5.4.9.3. Auswertung anhand des Kriterienkatalogs

Eine Auswertung anhand des Kriterienkatalogs ist für diese Fallstudie nicht sinnvoll. Zwar gibt es einen IV-Vertrag, dieser beinhaltet jedoch keine Gewährleistungsklausel. Andere Verträge mit Gewährleistung sind in Diskussion, können aber noch nicht ausgewertet werden. Das Interesse der Fallstudie liegt vor allem an der Einschätzung zu P4P, die sehr konkret und detailliert ist. Bemerkenswert ist vor allem, dass trotz einer sehr positiven Einstellung zu P4P bislang keine Aktivitäten in Richtung qualitätsorientierter Vergütung oder Garantieklauseln durchgeführt wurden

5.5. Bewertung der Nutzbarkeit des Kriterienkatalogs

Der Kriterienkatalog wurde im Rahmen der Fallstudien fünfmal angewandt – sowohl an einzelnen IV-Verträgen als auch an Qualitätsmanagementsystemen. Jede Fallstudie wurde anhand der Systemelemente und Kriterienkategorien und zumindest einiger Einzelkriterien evaluiert. Dies ermöglicht eine Beschreibung der Nutzbarkeit des Katalogs.

Der Kriterienkatalog zeichnet sich zunächst durch eine schnelle Bewertbarkeit aus – gerade die Detailbeschreibungen der einzelnen Kriterien ermöglichen eine sehr konkrete, vergleichbare Auswertung. Dies ist umso mehr der Fall, je mehr Daten und Informationen vorliegen. Aufwendiger als die Bewertung ist daher die Datenerhebung durch qualitative Interviews. Eine Ausweitung von quantitativen Datensätzen sowie eine quantitativere Bewertungsskala wären daher langfristig erstrebenswert.

Doch auch bei weniger Daten ist eine Einschätzung der Kriterien gut möglich. Die Systemelemente sind immer bewertbar – in diesem Rahmen ebenso die Kriterienkategorien. Dadurch ist eine strukturierte Analyse von P4P-Systemen unabhängig von der derzeitigen Umsetzung stets möglich. Meist sind außerdem die Einzelkriterien wenigstens für die Qualitäts- oder Anreizteil des Kataloges durchgängig auswertbar. Das ist vor allem abhängig davon, ob die Qualitätsindikatoren oder die Vergütungsaspekte für die Beteiligten von höherer Relevanz waren. Allerdings ist die Bewertung beim Leistungserbringer deutlich schwieriger, da hier die Qualitätsaspekte der IV-Verträge nur selten als P4P-Systeme genutzt werden und daher kein bewertbares System im Rahmen von IV-Verträgen existiert.

Zudem helfen die Kriterien dabei, eine vergleichende Bewertung der Systeme zu erreichen. Gerade eine Abgrenzung von relativ ähnlichen Systemen ist möglich. Neben einer sehr detaillierten Analyse wie in Fallstudie 1 sind somit auch kürzere, vergleichende Analysen wie in Fallstudien 2 und 3 erreichbar.

Derzeit ist eine quantitative Gesamtbewertung allerdings noch nicht möglich. Durch die noch qualitativ bewerteten Kriterien muss auch die Zusammenfassung der Erkenntnisse qualitativ erfolgen.

5.6. Zwischenfazit: Die Anwendung des Kriterienkatalogs ermöglicht detaillierte Ergebnisse zur Umsetzung von P4P-Elementen in den Fallstudien

Anhand der Fallstudien können zum einen Aussagen zur derzeitigen Umsetzung von P4P-Aspekten in den betrachteten IV-Verträgen, zum anderen Erkenntnisse zur Nutzbarkeit des Kataloges getroffen werden:

- In keinem der untersuchten IV-Verträge mit Gewährleistungsklauseln wurden die darin enthaltenen, qualitätsorentierten Vergütungselemente vollständig umgesetzt – dies liegt vor allem an folgenden Aspekten (mit unterschiedlicher Gewichtung für die verschiedenen Fallstudien):
  - Die Gewährleistungen werden meist nicht als Form von P4P gesehen.

- Selektivverträge haben meist nur eine geringe Fallzahl und somit eine geringere Priorität bei strategischen Entscheidungen.
- Die IV-Verträge sind sehr individuell und unterschiedlich ausgestaltet, so dass keine hohe Vergleichbarkeit oder Skaleneffekte erreicht werden können.
- Die meisten Selektivverträge werden für Fallzahlsteigerung oder Rabatte abgeschlossen, Qualitätsverbesserungen sind für den Vertragsabschluss nur sekundär relevant.
- Durch hohe Relevanz der Beziehung zwischen den Vertragspartnern werden Konsequenzen oft nicht ausreichend durchgesetzt.
- Die Analyse der Gewährleistungen ist meist nicht automatisiert, in keiner Fallstudie konnte die genaue Anzahl von Gewährleistungsfällen der IV-Verträge mit einfachen Mitteln ermittelt werden (Krankenkasse B allerdings hat bereits eine recht hohe Automatisierung der Erkennung von Gewährleistungsfällen).

➢ Allerdings gibt es auch positive Ansätze zu P4P-Systemen:

- *Beim Kostenträger*: Zum einen gibt es bei den Krankenkassen insgesamt ein sehr hohes Interesse an P4P (bspw. auch sichtbar an Bonus-Verträgen im ambulanten Sektor), zum anderen zeigen zwei der Kassen bereits recht gute Ansätze für eine Gewährleistungsmessung und -erkennung. Beide planen außerdem kurzfristig eine Softwareweiterentwicklung.
- *Beim Leistungserbringer*: Zum einen existiert der Vorschlag eines Bonusvertrages, der eine qualitätsorientierte Vergütung anhand von klinischen Indikatoren etablieren soll. Zum anderen gibt es zwei detaillierte Ansätze für interne Qualitätsmanagementsysteme, die einige Aspekte von P4P aufgreifen.

➢ Zudem ist das Interesse an P4P insgesamt relativ hoch – gerade bei den Kostenträgern wird die Entwicklung von qualitätsorientierter Vergütung sehr geschätzt. Doch auch die Leistungserbringer mit etablierten Qualitätsmanagementsystemen gehen von einer mittelfristig stattfindenden Einführung von P4P-Aspekten in Deutschland aus. Bei zwei Krankenhäusern ist die Meinung zu P4P jedoch etwas gemischter.

➢ Die Nutzbarkeit des Kriterienkatalogs ist als recht hoch einzuschätzen: In den 5 Anwendungen war die Kriterienbewertung schnell und detailliert möglich – allerdings sind einige Aspekte von P4P-Systemen (also meist entweder die Qualitätsorientierung oder das Anreizsystem) nur wenig bewertbar, da keine ausreichende Umsetzung vorliegt.

➢ Die Kataloganwendung ermöglicht eine strukturierte, vergleichbare Auswertung von teils sehr unterschiedlich gestalteten Systemen der qualitätsorientierten Vergütung.

6. Diskussion der Erkenntnisse und weiterführende Anwendung

6.1. Bedeutung des Kriterienkatalogs: Durch den Kriterienkatalog können P4P-Systeme einheitlich und umfassend bewertet und entwickelt werden

Erste Erkenntnisse konnten bereits bei der Entwicklung der „working definition“ erarbeitet werden. Obwohl Qualität im stationären Sektor vielfach diskutiert und definiert wurde, konnten die gängigen Eingrenzungen (bspw. nach Donabedian) durch die Betrachtung von Definitionen aus anderen Fachgebieten deutlich erweitert werden – gerade eine stärkere „Kunden-“ und „Wertorientierung“ sind dadurch möglich. Außerdem wird gezeigt, dass trotz einer großen Vielfalt verschiedener Qualitätsinitiativen weiterhin Anreizdefizite in den derzeitigen Vergütungssystemen existieren, die unter anderem durch qualitätsorientierte Vergütung adressiert werden könnten.

Der in dieser Arbeit entwickelte und angewandte Kriterienkatalog besteht aus einem Qualitäts- und Anreizteil und ermöglicht eine getrennte Betrachtung beider Aspekte mit theoretischen Ansätzen aus der medizinischen Qualitätsforschung und der Betriebswirtschaftslehre. Im ersten Teil des Kriterienkataloges wurden die Ziele eines qualitätsorientierten Kriterienkatalogs dargestellt und die bisherigen Auflistungen von Anforderungen an Qualitätsmessung diskutiert. Diese sind unter anderem:

- Existierende Auflistungen von verschiedenen Anforderungen für die Entwicklung oder Bewertung von Qualitätsindikatoren[1], z. B. in Deutschland von QUALIFY (unter anderem in Reiter et al. 2007; Geraedts et al. 2002 und Zorn/Ollenschläger 1999; oder international Remus et al. 2004; Campbell 2002; JCAHO 2006; Mattke et al. 2006; Mainz 2003; Arah et al. 2006; Ten Asbroek et al. 2004; Vos et al. 2009 und Wollersheim et al. 2007)
- Bewertungsschemata für P4P-Systeme in – systematischen – Reviews (z. B. in Deutschland SVR 2007; Scheppach et al. 2011; Veit et al. 2012a[2], Schrappe/Gültekin 2010; Mehrotra et al. 2010a und Amelung/Zahn 2009 oder international Rosenthal et al. 2005; van Herck et al. 2010; Bufalino et al. 2007; Institute of Medicine 2007 und Kuhmerker/Hartman 2007)
- Anforderungskataloge für Qualitätssysteme in anderen Bereichen (z. B. in der ambulanten Versorgung oder für Ärzteportale (so unter anderem Thomeczek 2002; Diel 2009)

[1] Einige Anforderungen auch in Arbeiten zur Bewertung von Indikatorensystemen und systemischen Reviews

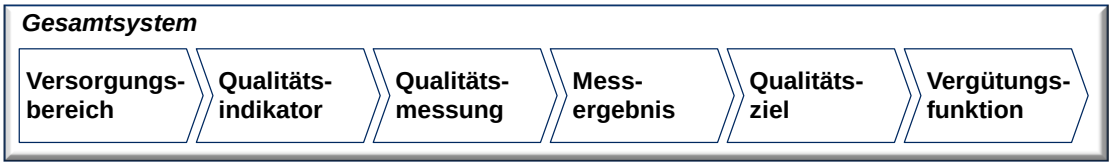
[2] Hier u. a. auch ein „Algorithmus“ zur Überprüfung der Kompatibilität und Eignung von Indikatoren für verschiedene P4P-Systeme, deren einzelne Anforderungen jeweils bei den Kriterien erwähnt werden (vgl. Veit et al. 2012e, S. 42ff. sowie Veit et al. 2012f, S. 69ff.)

- Anforderungen, die sich aus klinischen Praxisanwendungen entwickelt haben (vgl. u. a. Mansky 2010a).
- Anforderungskataloge an Anreizsysteme aus der betriebswissenschaftlichen Literatur (z. B. Becker 1990; Bleicher 1992; Seidel 1991; Bullinger 1996; Kerr 1997; Ortlieb 1993; Grewe 2000; Lehmann 2006; Laux 2006; Wünsch 2012)

Die Neuartigkeit des hier entwickelten Kriterienkatalogs besteht unter anderem darin, dass alle Aspekte von qualitätsorientierten Vergütungssystemen integriert betrachtet werden und nicht nur beispielsweise isoliert auf die Qualitätssicherung selbst oder auf das Anreizsystem eingegangen wird. Der Kriterienkatalog wurde speziell für die Bewertung von P4P-Systemen in der stationären Versorgung entwickelt – ein darauf ausgerichteter, spezieller Katalog existierte bis dato nicht. Eine in manchen Punkten ähnliche Konzeption für diesen Anwendungszweck ist allenfalls in einem kürzlich erschienen BQS-Gutachten zu finden (vgl. Veit et al. 2012a).

Zunächst wurde der Kriterienkatalog anhand von „Systemelementen“ strukturiert. Diese Strukturierung ermöglicht eine detaillierte Entwicklung der Kriterien entlang der Funktionssystematik von qualitätsorientierten Vergütungssytemen und stellt somit bereits eine Neuerung zu den existierenden Auflistungen dar. Die Systemelemente sind:

Abbildung 23: Systemelemente des Kriterienkatalogs



Für jedes dieser Systemelemente wurden anschließend Kriterien entwickelt, die die Anforderungen für ein qualitätsorientiertes Vergütungssystem beschreiben. Durch diese Strukturierung können die Kriterien für die einzelnen Elemente des qualitätsorientierten Vergütungssystems spezifischer und konkreter definiert werden, als dies bis dato der Fall war.

Der nächste Entwicklungsschritt bestand in der Definition von „Kriterienkategorien“, die pro Systemelement verschiedene Einzelkriterien zusammenfasst. Diese Zwischenebene verbessert die Nutzbarkeit des Katalogs, der insgesamt über 38 Einzelkriterien verfügt. Verschiedene dieser Kriterienkategorien sind neuartig gestaltet, so z. B. der explizite Fokus auf die Akzeptanz seitens der Akteure des Gesundheitswesens.

Vor allem bei den Einzelkriterien definiert der vorliegende Katalog vielfältige neue Aspekte und reicht über die früheren Auflistungen hinaus. Beispiele hierfür sind Aufteilung verschiedener Aspekte der

Relevanz des Indikators und des Versorgungsbereiches, einige Detaillierungsaspekte der Manipulationssicherheit oder der Grad der Integration in IT-Systeme. Die Einzelkriterien wurden alle detailliert entlang der folgenden Aspekte definiert: Bezug zu existierenden Auflistungen, Begründung, Anwendungsdefinition und -beispiele sowie Skala, Bewertungsarten und Informationsquellen. Dadurch ist ein sehr konkreter, detailliert ausgearbeiteter Kriterienkatalog entstanden, der in der Praxis schnell und verständlich anwendbar ist. Anwendungsmöglichkeiten sind neben der Analyse und Bewertung existierender qualitätsorientierter Vergütungssysteme auch die Entwicklung von neuen, umfassenden P4P-Systemen entlang der Kriterien.

Im zweiten Teil der Entwicklung des Kriterienkatalogs wurden zunächst die theoretischen Grundlagen von Anreizsystemen (v.a. aus der BWL) dargestellt. Durch die Anwendung des Prinzipal-Agent-Theorems konnten gerade für den stationären Sektor die Beziehungen der einzelnen Akteure beschrieben und das Anreizsystem konkretisiert werden. Weitere Erkenntnisse lieferten verschiedene verhaltenstheoretische Ansätze, wie der „Crowding-Out-Effekt“, der die Zusammenhänge zwischen intrinsischer und externer Motivation beschreibt. Außerdem existieren verschiedene Auflistungen von Anforderungen an Anreizsysteme aus der betriebswissenschaftlichen Literatur (so u. a. bei Becker 1990; Bleicher 1992; Grewe 2006; Lehmann 2006; Ortlieb 1993). Diese dienen als erste Anknüpfungspunkte für die Entwicklung des zweiten Teils des Kriterienkatalogs, müssen aber an die spezifische Situation eines medizinischen, qualitätsorientierten Anreizsystems in der stationären Versorgung angepasst werden.

Auch die Analyse der theoretischen Grundlagen von Zielen und Funktionsweisen von P4P ermöglicht erste Erkenntnisse. So werden P4P-Systeme bereits in vielfältiger Form und in verschiedenen Gesundheitssystemen angewandt, eine einheitliche Bewertungssystematik allerdings besteht nicht. Auch eine einheitliche Definition und Verwendung der Begrifflichkeiten wie „Pay-for-Performance“, „value-based-purchasing“ und vielen mehr existiert nicht. Die derzeitige P4P-Diskussionen thematisieren vor allem den Aufbau und die Gestaltung von P4P-Systemen (u. a. „Gestaltungsdimensionen“ oder „design choices“ genannt). Auch die existierenden Reviews beinhalten diese Thematik. Unter anderem werden folgende Aspekte diskutiert (vgl. u. a. SVR 2007; van Herck et al. 2010; Amelung/Zahn 2009; Emmert 2008; Scheppach et al. 2011; Schrappe/Gültekin 2010; Mehrotra et al. 2010a; Rosenthal/Dudley 2007; Bufalino 2006; Malzahn et al. 2010):

- *Bemessungsgrundlage*: Bewertung anhand von absoluten Vergleichen („absolut top“) oder relativen Verbesserungen („relativ top“)
- *Art der Vergütung*: Bonus- oder Maluszahlungen
- *Adressateneben*: Krankenhaus-, Fachabteilungs- oder Arztebene für die Qualitätsmessung und Anreizverteilung

- *Vergütungshöhe*: Verschiedene Adressatenebenen relevant, vermutlich wirksamer Anteil ab 5% der Gesamtvergütung

Klare Empfehlungen sind für diese Dimensionen häufig nur schwer möglich, da die Studienlage verschiedene Ergebnisse aufweist. Ziel dieser Arbeit ist aber weniger eine Definition von Gestaltungsdimensionen sondern vielmehr der wichtige Schritt davor: Eine Beschreibung der sich aus der Zielsetzung P4P ergebenden, notwendigen Anforderungen an qualitätsorientierte Vergütungssysteme. Auf die Gestaltungsdimensionen wird im Folgenden noch bei den Empfehlungen für ein mögliches, deutsches P4P-System eingegangen.

Speziell im zweiten Teil des Katalogs mit den Aspekten des Anreizsystems stellen viele Kriterien Neuerungen dar, weil bislang keine umfassenden Auflistungen von Anforderungen an die Vergütungssystematiken von P4P-Systemen existieren. Mehrere Einzelkriterien sind daher neuartig. Gerade die Anwendungsdefinitionen und -beispiele ermöglichen auch hier eine praxisrelevante Konkretisierung der Einzelkriterien, die eine direkte Anwendung für die Evaluation und Entwicklung von P4P-Systemen im stationären Sektor in Deutschland ermöglichen.

Abbildung 24 zeigt einen Überblick über den gesamten Kriterienkatalog.

Abbildung 24: Überblicksdarstellung des Kriterienkatalogs

Systemelement	Kriterienkategorie	#	Einzelkriterium
Gesamtsystem	Aussagekraft	1a	Abdeckung des Gesamtsystems
Versorgungs-bereich	Aussagekraft	2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet
		2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich
Qualitäts-indikator	Aussagekraft/ "Das Richtige messen"	3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten
		3b	Relevanz des Indikators für ein bestimmtes Fachgebiet
		3c	Sektorbezug vs. sektorübergreifende Nutzung
	Akzeptanz	4a	Medizinische Bedeutung in der medizinischen Fachliteratur
		4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften
		4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern
		4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit
		4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit
Messung	Wissenschaftlichkeit/ "Richtig messen"	5a	Reliabilität
		5b	Validität
		5c	Risikoadjustierung und Risikoadjustierbarkeit
	Messaufwand und Messsicherheit	6a	Objektiver Erhebungsaufwand und -komplexität
		6b	Manipulationssicherheit / Objektivität der Messung
		6c	Grad der Integration in automatische Prozesse / IT-Systeme
Ziele	Wirksamkeit/ Zielgerichtetheit	7a	Zielkonsistenz
		7b	Zielkongruenz
		7c	Reichweite
	Operationalisier-barkeit	8a	Erreichbarkeit
		8b	Transparenz
		8c	Akzeptanz
		8d	Langfristige Anreizwirkung der Ziele
	Vergleichbarkeit	9a	Gerechtigkeit
Vergütungs-funktion	Wirksamkeit/ Zielgerichtetheit	10a	Effizienz
		10b	Effektivität
		10c	Flexibilität
	Operationalisier-barkeit	11a	Anteil der Vergütung
		11b	Planbarkeit
Messergebnis	Operationalisier-barkeit	12a	Nutzung zur Steuerung
		12b	Aktualität
		12c	Kontrollierbarkeit
		12d	Praktische Umsetzung im Controlling
		12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit
	Vergleichbarkeit	13a	Zeitverlauf
		13b	Querschnittsvergleich/Ranking
		13c	Justiziabilität

Insgesamt kann der hier entwickelte Kriterienkatalog als umfassend und in vielen Teilen neuartig bezeichnet werden. Die Anwendbarkeit des Katalogs auf existierende qualitätsorientierte Vergütungssysteme in Deutschland wurde im nächsten Abschnitt überprüft.

Die Besonderheit der vorliegenden Arbeit besteht darin, dass sie die medizinische und betriebswissenschaftliche Sicht zur Entwicklung eines abgestimmten Kriterienkataloges integriert und ferner, anders als viele bisherige Ansätze, die Kriterienentwicklung aus Sicht der Notwendigkeiten und Zielsetzungen eines P4P-Systems umsetzt.

6.2. Erkenntnisse der Fallstudien: In Deutschland werden P4P-Systeme in der stationären Versorgung bei den hier untersuchten Verträgen derzeit nicht vollständig umgesetzt

In den Fallstudien konnte der Kriterienkatalog zum ersten Mal angewandt werden. Hierdurch sind Ergebnisse zum derzeitigen Stand von P4P im deutschen stationären Sektor anhand von Beispielen von IV-Verträgen mit Qualitätsgarantien entstanden, die aufgrund des qualitativen Charakters der empirischen Studie aber nicht verallgemeinbar sind.

Zusammenfassend kann vorab festgehalten werden, dass keine Fallstudie ein vollständig umgesetztes, funktionierendes System einer qualitätsorientierten Vergütung anhand von Gewährleistungen von IV-Verträgen enthält. In keiner der Fallstudien konnten alle Gewährleistungsfälle ausreichend standardisiert bzw. automatisiert erkannt werden[1]. Auch fehlten meist die zuverlässige Umsetzung der Vergütungskonsequenzen sowie ein standardisiertes Controlling der Gewährleistungsfälle. Die Umsetzungsprobleme sind vor allem mit zwei Grundelementen der Fallstudien zu erklären: Dem Charakter der Selektivverträge und den Gewährleistungsvereinbarungen. Bemerkenswert ist ferner, dass die Vereinbarung von Gewährleistungsklauseln von vielen Beteiligten nicht als eine Form von P4P bzw. qualitätsorientierter Vergütung eingeordnet wird.

<u>1. Probleme der Umsetzung von Qualitätselementen in Selektivverträgen</u>

Relative Größe der IV-Verträge

Eine der Kernschwierigkeiten, die sich in den Fallstudien zeigt, ist in der relativen Größe der untersuchten IV-Verträge zu sehen. Der größte hier untersuchte IV-Vertrag umfasst ca. 480 Fälle pro Jahr. In den Interviews wird deutlich, dass daher die Priorität von strategischen Entscheidungen (bspw. bezüglich Investitionen) bei der Regelversorgung liegt. Das gilt sowohl für Krankenkassen als auch für Krankenhäuser. Bei Kassen können sich dadurch beispielsweise die Umsetzungen von neuen Softwareprogrammierungen für die Abrechnung von Selektivverträgen verzögern bzw. diese haben nur eine sehr geringe Priorität. Auch im klinischen Bereich der Leistungserbringer sind IV-Verträge kaum relevant, da für die medizinische Behandlung (vielleicht glücklicherweise?) keine Unterscheidung zwischen Patienten des IV- oder Kollektivvertrags gemacht wird. Nur administrativ entsteht ein zusätzlicher Aufwand, beispielsweise für die Einschreibung in den IV-Vertrag oder die Umsetzung von zusätzlichen nicht-medizinischen Leistungen wie Wahlleistungsessen.

[1] Nur die Fallstudie 2 hat eine recht umfassende Erkennung von Gewährleistungsfällen, die durch weitere Softwareentwicklungen noch optimiert werden soll.

Unterschiedliche Ausgestaltung der Verträge

Eine weitere Schwierigkeit der Ausgestaltung der IV-Verträge liegt in den Unterschieden, die zwischen den hier analysierten Verträgen zu finden waren. Zwar waren große Ähnlichkeiten selbst in der Wortwahl von einigen Verträgen zu entdecken, was zur Annahme führt, dass es sich hierbei ggf. um Musterverträge eines Kassenverbandes handelt. Gleichzeitig sind die Verträge jedoch in den Details beispielsweise der Krankheitsdefinitionen so unterschiedlich, dass keine standardisierte Messung oder Datenerhebung möglich ist. Selbst innerhalb einer Kasse sind die Verträge so verschieden, dass die Falldefinitionen (für die betroffenen Vertragsfälle) und Analysen für Vorabevaluationen für jeden Vertrag neu entwickelt werden müssen. Auch beim Leistungserbringer steigt dadurch der Aufwand für die Patienteneinschreibung deutlich, da beispielsweise unterschiedliche Aufnahme- und Ausnahmeregelungen bestehen. Auf einer übergeordneten Ebene verhindern die verschiedenartigen Vertragsausgestaltungen vor allem die Vergleichbarkeit von Messergebnissen über Kassen und Leistungserbringer hinweg. Zusätzlich ist dadurch auch die Umsetzung der IV-Verträge (z. B. die veränderten Rechnungsprozesse) für Kasse und Leistungserbringer aufwendiger (vgl. auch Krüger-Brand/Osterloh 2011).

Gründe für die Vertragsschließung

Die Gründe für die Vertragsschließung wurden für jede Fallstudie diskutiert. Während für die Krankenkasse die Möglichkeit von Rabatten, Patientensteuerung und innovative Vertragsgestaltung im Vordergrund stehen, sind für den Leistungserbringer vor allem mögliche extra-budgetäre Vergütungen sowie Fallzahlsteigerung (durch Patientensteuerung) interessant. Dadurch entsteht in der IV-Vertragsevaluation ein starker Fokus auf Fallzahlcontrolling sowohl beim Leistungserbringer als auch den Kostenträgern. Gleichzeitig wurden von keinem der Befragten die Möglichkeiten von Qualitätsverbesserung als Hauptgrund der Vertragsschließung genannt.

Relevanz der Vertragsbeziehungen

In den Fallstudien wurde die Wichtigkeit einer guten Vertragsbeziehung zwischen Leistungserbringer und Krankenkasse hervorgehoben. Im Rahmen eines Selektivvertrages, der freiwillig eingegangen wird, sind die vertraglichen Beziehungen sehr relevant. Dies führt dazu, dass keine Aktivitäten durchgeführt werden, die zu einer deutlichen Verschlechterung der Vertragsbeziehung führen könnten. Daher wird ein Konsequenzenmanagement von „Strafen“ (wie bei den Gewährleistungen) nicht immer umgesetzt. Gleichzeitig werden auch vereinbarte Berichte nicht immer konsequent eingefordert.

2. Probleme der Umsetzung von Gewährleistungen

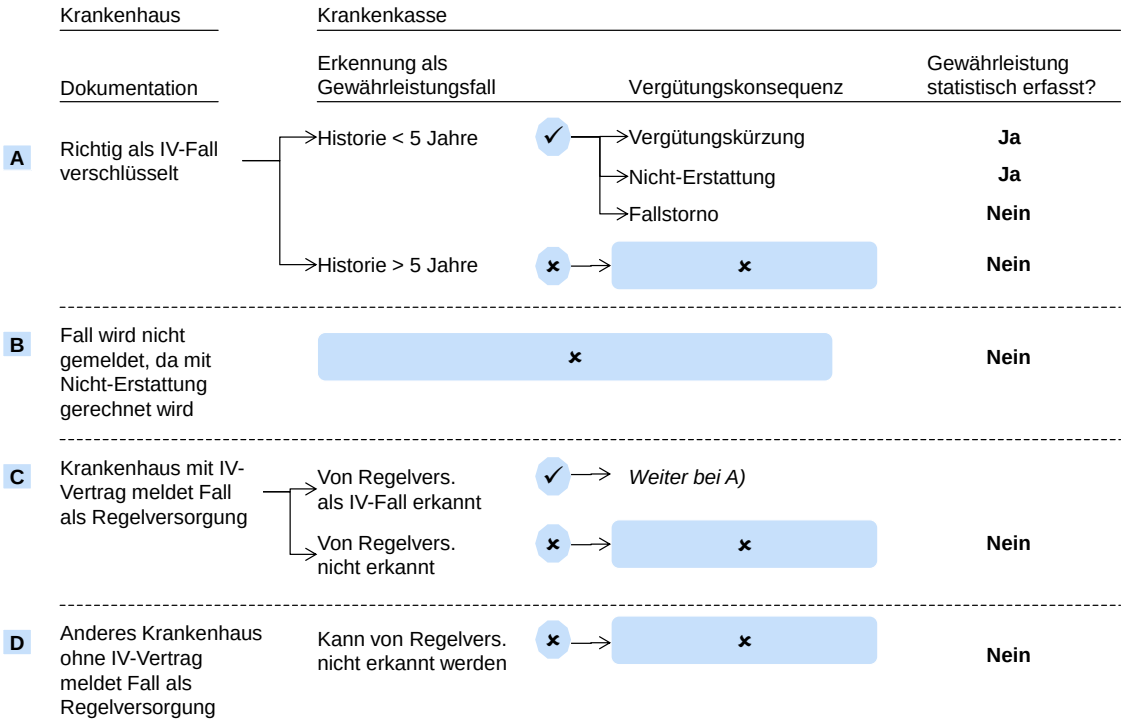
Geringe Anzahl von Gewährleistungsfällen

Die qualitätsvariable Vergütung über Gewährleistungen bzw. Gewährleistungsfälle bringt einige Schwierigkeiten mit sich. Zunächst amplifiziert dies das Problem der kleinen Fallzahlen – mit einer 5-Jahres-Revisionsrate von nur 4,7 % (Beispiel Hüftgelenksersatz; vgl. Heller 2010) ist selbst bei großen Verträgen nicht mit einer statistisch relevanten Anzahl von Gewährleistungsfällen zu rechnen. Da es sich hier also um Einzelfälle handelt, die bei Eintreten bei unscharfer Definition der Gewährleistungssachverhalte auch zu großem Diskussionsbedarf führen können, ist eine Qualitätsverbesserung aufgrund der variablen Vergütung zumindest kaum nachweisbar. Außerdem ist nach Praktikermeinung nicht damit zu rechnen, dass zusätzlich zur bereits vorhandenen Motivation der Ärzte, Revisionsoperationen zu vermeiden, ein weiterer wirksamer Verbesserungsanreiz durch die selektivvertraglichen Gewährleistungsverfahren hinzukommt, die nicht für alle Behandlungsfälle einer Art gelten.

Schwierigkeit der Analyse von Gewährleistungsfällen

Die Evaluation von Gewährleistungsfällen ist aufgrund der Abrechnungsrealität nur schwer möglich. Dies beginnt bereits bei der Verschlüsselung der potenziellen Gewährleistungsfälle. Es wird in der Regel von den Leistungserbringern erwartet, dass sie diese Fälle entsprechend angeben. Da hier weder kassen- noch bundesweit ein Standard etabliert ist, entsteht für den Leistungserbringer bei der Kodierung einerseits ein zusätzlicher Aufwand. Außerdem können Gewährleistungsfälle bei einigen Fallstudien in voller Höhe vergütet werden, indem sie (versehentlich oder mit Absicht) als Regelfall kodiert werden, da sie dann von der Kasse nur sehr schwer erkennbar sind. Andererseits bestehen seitens der Kassen kaum adäquate Verfahren zur Erkennung der Gewährleistungsfälle. Insgesamt stellt die Erkennung potenzieller Gewährleistungsfälle daher noch ein realistisches Hindernis dar – wie die Abbildung 25 zeigt, sind nur in einigen Fallkonstellationen die Gewährleistungen in Fallstudie 1 sicher erkennbar und statistisch auswertbar:

Abbildung 25: Möglichkeiten der Erkennung, Umsetzung und Messung von Gewährleistungsfällen, Fallstudie 1, Krankenkasse A



Einschränkungen der Gewährleistungsklauseln

Eine weitere Schwierigkeit stellen die deutlichen Einschränkungen der Gewährleistungsklauseln hinsichtlich Vorerkrankungen und Komplikationen dar. Dies erhöht den Umsetzungsaufwand bei der Krankenkasse, behindert oder verhindert sogar oft eine Automatisierung der Fallerkennung und verringert damit die Wahrscheinlichkeit einer Nicht-Erstattung und somit qualitätsvariablen Vergütung.

Beschränkung auf Malus

Gerade die Fallstudien bei den Leistungserbringern haben ergeben, dass die Beschränkung des Anreizsystems auf einen Malus nicht erwünscht ist. Die Möglichkeit, auch durch bessere Qualitätsergebnisse aufzufallen, so mehr Patienten anzuziehen und außerdem eine höhere interne Motivation zu erlangen, wird priorisiert.

Zwischenfazit

Die wichtigste Erkenntnis bleibt, dass Gewährleistungsklauseln als P4P-System derzeit in den Fallstudien nicht adäquat umgesetzt werden. In fünf von sieben Fallstudien gab es IV-Verträge mit Gewährleistungsklauseln. Aber nur bei den Kassenfallstudien wurden Gewährleistungsfälle tatsächlich identifiziert – und auch da nicht immer. Es ist aufgrund der zu erwartenden Häufigkeiten zu vermuten, dass nicht erkannte Gewährleistungsfälle bei den Leistungserbringern aufgetreten sind.

Erstaunlicherweise werden die Gewährleistungen von den Beteiligten meist nicht als Element einer qualitätsorientierten Vergütung gesehen. Mit einem Anreizeffekt wird seitens der Beteiligten nicht gerechnet.

3. Derzeitige Einstellungen zu P4P

Trotz der beschriebenen Umsetzungsprobleme sind einige positive Ansätze zu entdecken, die eine zukünftige Etablierung von P4P-Systemen ermöglichen können. Zum Ersten stehen besonders die Kassen der Idee des P4P sehr aufgeschlossen gegenüber. Auch einige Leistungserbringer erwarten mittelfristig die Einführung von qualitätsorientierter Vergütung in Deutschland – vor allem nach amerikanischem Beispiel. Zum Zweiten sind einige unternehmenskulturelle und -organisatorische Aspekte bereits so ausgerichtet, dass eine Etablierung von P4P gut möglich wäre.

Insgesamt zeigt sich, dass die Krankenkassen und Krankenhäuser bezüglich der Qualitätsorientierung unterschiedlich weit entwickelt sind. Bei den Krankenkassen wird dies anhand der unterschiedlich ausgeprägten Qualitätsmessungen und -evaluationen deutlich. Allen Kassen ist gemein, dass Qualität zwar als wichtiges Thema gesehen wird, aber dennoch Gewährleistungsfälle nicht voll-automatisch erkannt, umgesetzt oder ausgewertet werden. Allerdings ist bei einigen Kassen die Erkennung technisch einfacher und schneller umzusetzen. Positiv ist zu vermerken, dass bei zwei Kassen auch heute schon sehr detaillierte Vorabevaluationen zu möglichen Qualitätsverbesserungen individuell für jeden Vertrag durchgeführt werden. Gleichzeitig sind mehrere P4P-Elemente in einigen – hier nicht näher analysierten – Bonus-Verträgen im ambulanten Sektor bereits weiter entwickelt.

Bei den Leistungserbringern besteht der Unterschied in der Qualitätsorientierung vor allem in der Relevanz von internen Qualitätsmanagementsystemen. Je etablierter ein Managementsystem bereits ist, desto stärker ist der Qualitätsfokus auf allen Ebenen des klinischen und nicht-klinischen Managements sichtbar. Außerdem ist in diesem Fall die Entwicklung einer qualitätsorientierten Finanzierung deutlich akzeptierter. Wenn Qualität beim Leistungserbringer bislang vorwiegend auf Ebene der gesetzlichen Qualitätssicherung stattfand, werden dem Einsatz von P4P häufig zunächst die Schwierigkeiten bei der Qualitätsmessung entgegen gehalten. Bei zwei Befragten war die Einschätzung zu P4P nicht so positiv, da von einem Befragten eine nur vordergründig „qualitätsorientierte“ Gesamtvergütungskürzung seitens der Kostenträger befürchtet wird.

Insgesamt gehen jedoch die meisten Befragten davon aus, dass mittel- oder langfristig mit einer Entwicklung von qualitätsorientierten Vergütungsaspekten auch in Deutschland zu rechnen ist. Es werden zudem einige kurzfristige Initiativen erwartet bzw. im ambulanten Sektor bereits als Pilotprojekte umgesetzt. So hat beispielsweise eine Krankenkasse Gewährleistungsklauseln wieder in die IV-Verträge eingebunden und will anhand einer Softwareänderung eine automatische Erkennung integrieren. Auch neue Qualitätsindikatoren werden derzeit von einer Kasse gemeinsam mit einem

gesundheitsökonomischen Lehrstuhl entwickelt. Ferner wird erwartet, dass vor allem IV-Verträge mit Bonus-Systemen (wie beispielsweise im Exkurs der Fallstudie 5 dargestellt) weiterverbreitet werden.

Doch auch mit qualitätsorientierten Vergütungsaspekten für den Kollektivvertrag wird mittel- und langfristig gerechnet. Hierfür sei es zunächst notwendig, eine akzeptierte, valide, aufwandsarme Qualitätsmessung zu etablieren. Dies scheint die größte Hürde darzustellen. Außerdem gibt es einen klaren Wunsch der Leistungserbringer, die Vergütungsfunktion vor allem als Bonus einzuführen. Weitere Gestaltungsdimensionen werden in der Praxis bislang noch nicht diskutiert.

6.3. Schritte zu einer möglichen Umsetzung von P4P: Ausgewählte Empfehlungen für die Entwicklung eines P4P-Systems in Deutschland

An dieser Stelle sollen beispielhaft einige Aspekte eines qualitätsorientierten Vergütungssystems entwickelt werden, das die vorliegenden Kriterien möglichst weitgehend erfüllt. Die Entwicklung dieser Empfehlungen ergibt sich vor allem aus der Anwendung des Kriterienkatalogs. Dabei werden neben der detaillierten Beschreibung der Kriterien speziell die Erfahrungen der Fallstudien herangezogen und gleichzeitig die einiger amerikanischen Projekte genutzt. Nicht alle notwendigen Elemente eines P4P-Systems können hier hinreichend beschrieben werden. Der Fokus liegt bei den Gestaltungsdimensionen, die in dieser Arbeit mehrfach diskutiert wurden und eine besondere Bedeutung für die Kriterienerfüllung haben.

<u>Vertragsform</u>

Die Fragestellung „kollektiv vs. selektiv“ tauchte im Rahmen der empirischen Studien wiederholt auf. Die Entscheidung, ob ein P4P-System in Deutschland in einem Kollektivvertrag oder einzelnen Selektivverträgen zwischen Leistungserbringer und Kostenträger umgesetzt werden sollte, hat großen Einfluss auf die Anzahl der teilnehmenden Kostenträger, Leistungserbringer und vor allem Patienten/Versicherten und damit auf die einbezogenen Fälle. Verschiedene Vor- und Nachteile stehen sich hierbei gegenüber. So ermöglicht eine selektive Vertragsgestaltung den Wunsch der Vertragspartner nach innovativen Vertragsformen und Pilotprojekten. Außerdem kann ein solcher Vertrag für Leistungserbringer und Kostenträger als Wettbewerbsvorteil genutzt werden. Aus Qualitätsperspektive ist aber der Vorteil einer selektivvertraglichen Regelung vor allem die Möglichkeit, die Auswahl der Vertragspartner nach qualitativen Kriterien zu treffen. Dies hat bei den vorgestellten Fallstudien meist stattgefunden und könnte theoretisch zu einer Patientensteuerung führen, die für die Behandlungsqualität relevant sein kann. Leider findet diese Steuerung meist nicht ausreichend statt.

Diesen Vorteilen der selektivvertraglichen Gestaltung stehen einige wichtige Nachteile gegenüber. Der empirische Teil dieser Arbeit hat deutlich gezeigt, dass die untersuchten Qualitätsregelungen in Selektivverträgen keine ausreichende Relevanz für die beteiligten Vertragspartner haben (vgl. auch

Mansky/Seeger 2011, S. 202ff.; Veit et al. 2012e, S. 70f.). In allen Fallstudien wurde berichtet, dass die selektivvertraglichen Entscheidungen im Management sowohl der Leistungserbringer als auch der Kostenträger von relativ geringer Priorität seien und daher der Prozess der Vertrags- und/oder IT-Entwicklungen behindert werde. Zusätzlich seien die Qualitätsaspekte der Selektivverträge für den klinischen Alltag kaum relevant – vor allem weil die Behandlungsverfahren und damit die Behandlungsqualität innerhalb eines Leistungserbringers für alle Patienten weitgehend gleich ist und nicht vom Versicherer des Patienten, sondern eher von der Fachkompetenz des Leistungserbringers, der Prozessorganisation und dem Vorhandensein eines Qualitätsmanagementsystems abhängen. Medizinisch kann, soll und darf die Qualität der Behandlung auch nicht von der Krankenkasse des Patienten abhängen, d. h. es kann nicht versicherungsspezifische „Qualitäten“ bzw. Qualitätsdefinitionen geben. Ziel ist es in der medizinischen Qualitätssicherung besonders bei standardisierbaren Prozessen (wie z. B. Hüftgelenksersatz), einheitliche Behandlungspfade und Prozessabläufe sowie eine leitliniengerechte, dem Stand des Wissens entsprechende Behandlung sicherzustellen. Dies gewährleistet nach allen Erfahrungen bestmögliche Qualität (vgl. u. a. Bradley et al. 2006, S. 72ff.; Peterson et al. 2006, S. 1912ff.; Grimshaw/Russell 1993, S. 1317ff.; Komajda 2005, S. 165ff.; Menendez 2005, S. 757ff.). Aus diesem Ansatz wird aber auch erkennbar, dass die Qualitätskriterien nicht von der Versicherungsart des Patienten abhängen und auch nicht abhängen dürfen, sondern von seinen medizinischen Erfordernissen. In der klinischen Praxis sind ebenfalls unterschiedliche, von der Krankenkasse abhängige Qualitätsdefinitionen nicht umsetzbar. Gerade in den letzten Jahren haben die Krankenhäuser ihre Bemühungen verstärkt, ihre Prozesse im Sinne von Effizienz- und Qualitätssteigerung zu standardisieren. Kassenspezifische Qualitätsziele laufen dieser Entwicklung zuwider (im Extremfall sozusagen eine 200-Klassen-Medizin entsprechend der Zahl der Krankenkassen).

Ferner ergeben sich aus der Vielzahl von unterschiedlichen Verträgen Nachteile in der Umsetzung – sowohl bei der Messung als auch der operativen Umsetzung der Vergütungsaspekte. Bereits die Datenerhebung der Qualitätsaspekte (in den hier durchgeführten empirischen Fallstudien also die Erhebung der Gewährleistungsfälle) wird für die verschiedenen Verträge unterschiedlich definiert, so dass eine standardisierte Analyse der Qualitätsverbesserungen kaum möglich ist. Vor allem kassenübergreifend können die Erfolge von solchen Verträgen kaum evaluiert werden.

Es wird daher empfohlen, für P4P-Systeme in Deutschland zumindest kollektiv zu messen. Wenn die Datenerhebungen und Messungen für alle Beteiligten einheitlich und standardisiert sind, können zum einen die Ergebnisse solcher Projekte verglichen werden, zum anderen wird die Umsetzung sowohl für die Kostenträger als auch für die Leistungserbringer effizienter. Hierfür muss eine einheitliche Definition der betroffenen Versorgungsbereiche und der Qualitätsindikatoren entwickelt werden. Anschließend sollte eine Zusammenführung der Messergebnisse oder sogar der Daten angestrebt werden, um die Vergleichbarkeit der einzelnen Verträge zu erhöhen.

Eine Umsetzung der Qualitätsaspekte in kollektiv- oder selektivvertraglicher Vertragsformen beeinflusst unter anderem die Geschwindigkeit der Umsetzung. Während eine selektivvertragliche Umsetzung durch Einzelinitiativen in Pilotprojekten schnell möglich ist und bereits recht erfolgreich umgesetzt wird (vgl. Exkurs bei Fallstudie 5), müssen für eine kollektivvertragliche Regelung Gesetzesänderungen vorgenommen werden (z. B. zur Nutzung der notwendigen Daten; vgl. Mansky/Nimptsch 2010b; Veit et al. 2012e, S. 71). Da durch eine kollektivvertragliche Umsetzung die Kriterien der Reichweite, Akzeptanz und Relevanz für alle Beteiligten besser erfüllt werden, sollte eine Integration in den Kollektivvertrag langfristig angestrebt werden. Diese Entwicklungsschritte zeigen sich ebenso in den USA, wo nach erfolgreicher Umsetzung von verschiedenen, jeweils räumlich und/oder hinsichtlich der Teilnehmer beschränkten Pilotprojekten jetzt seitens der CMS ein umfassendes P4P-System für alle Medicare-Fälle eingeführt wurde (vgl. HHS - CMS 08.05.2011; Buhlman/Matthes 2012).

Zusammenfassung der Empfehlungen:

- Möglichst kollektives Messen durch einheitliche Definition der Versorgungsbereiche und Qualitätsindikatoren
- Übergangsweise zunächst selektives Kontrahieren fördern – langfristig Integration in Kollektivvertrag anstreben
- Im Rahmen bestehender Selektivverträge sollten Methoden zur Verbesserung des Vertragscontrollings gezielt gefördert bzw. eingefordert werden

<u>Art der Datenerhebung und Qualitätsmessung</u>

Für alle Befragten war die Messung der Qualität der Kernaspekt für die Akzeptanz und Umsetzbarkeit eines Systems von qualitätsorientierter Vergütung und wurde von einigen als Hauptschwierigkeit empfunden.

Wie bereits erwähnt, wird empfohlen, standardisierte Definitionen der Qualitätsindikatoren einzusetzen. Vor allem die Frage, welche Art von Datenerhebung und -messung genutzt werden soll, ist relevant. Aufgrund der Ergebnisse dieser Arbeit wird deutlich, wie wichtig es für die Beteiligten ist, dass die Messung aufwandsarm, einfach, valide und manipulationssicher durchgeführt werden kann. Gerade der geringe Aufwand ist vor allem für den Leistungserbringer im klinischen Alltag von großer Bedeutung. Daher wird die Nutzung von Routinedaten empfohlen, da zusätzliche Datenerhebungen bspw. durch Umfragen oder weitere Dokumentation meist nicht notwendig sind. Dies würde auch dem Wunsch der Krankenkassen entsprechen, die Qualitätsmessungen in den existierenden Datensätzen durchführen zu können.

Durch einen Fokus auf Parameter der Ergebnisqualität kann sowohl die Verständlichkeit für die interessierte Öffentlichkeit als auch die Akzeptanz bspw. bei medizinischen Fachgesellschaften erhöht werden (vgl. Kriterienbeschreibung 4b). Gerade für ein System der qualitätsorientierten Vergütung ist die Konzentration auf einen „Endpunkt" der Behandlungsqualität relevant – da Anreize geschaffen werden sollen, um das Gesamtergebnis zu verbessern und nicht nur isolierte Prozessschritte zu beeinflussen.

Parallel sollte für die Leistungserbringer eine Hilfestellung entwickelt oder gegeben werden, die Ergebnisverbesserung zu operationalisieren. Daher sollten auch eine enge Einbindung in ein Managementsystem (bspw. mit Peer-Review-Verfahren) bzw. die Analyse von Prozessen unterstützt werden. Wenn also die qualitätsabhängige Vergütung anhand der Ergebnisqualität (d. h. bspw. Revisionsrate nach Hüft-TEPs) variiert wird, wäre es sinnvoll, beispielsweise mittels eines Peer-Review Systems dem Leistungserbringer im Einzelfall Unterstützung anzubieten, wie bspw. die Revisionsrate gesenkt werden kann. Eine weitere Möglichkeit wäre die Auswertung und Kommunikation von Messergebnissen der Prozessqualität (bspw. Anzahl von Hüft-Operationen), um auch ohne ein internes Managementsystem auf Schwierigkeiten in den Behandlungsprozessen hinzuweisen und dem Leistungserbringer einen Vergleich mit anderen Kliniken zu ermöglichen. Diese flankierenden Maßnahmen müssen aber nicht Teil des P4P-Systems selbst sein. In diesem sollte die Vergütungshöhe möglichst vom Behandlungsergebnis abhängen.

Allen Befragten waren mehrere Indikatoren- und Messsysteme bekannt, trotzdem wurden in zwei Fallstudien eigene Qualitätsindikatoren entwickelt – dies verursacht einen hohen Entwicklungsaufwand für die beteiligten Akteure, dem diese auf lokaler Ebene nicht immer gewachsen sein können. Um erprobte, umfassende Indikatorensysteme zu nutzen, sollten die existierenden Systeme evaluiert und für die Nutzbarkeit in qualitätsorientierten Vergütungssystemen analysiert werden. Dazu kann der hier vorgestellte Kriterienkatalog wesentlich beitragen. Zusätzlich sollte das zu verwendende Indikatorensystem regelmäßig durch unabhängige Institute analysiert werden, da die Validität der Messung für viele Befragte ein relevantes Thema darstellt (vgl. auch AMA 2005).

Zusammenfassung der Empfehlungen:

- Nutzung von Routinedaten, um aufwandsarme, manipulationsresistente Messungen zu ermöglichen
- In der Vergütungsfunktion Fokus auf Ergebnisqualität mit flankierender Unterstützung durch Managementsysteme, die die Analyse und Verbesserung der Prozesse fördern
- Weiterentwicklung existierender Systeme und Überprüfung der Indikatorensysteme durch unabhängige Institute

Operationalisierung der variablen Vergütung

Während alle Befragten der empirischen Untersuchung die hohe Relevanz der Qualitätsmessung erwähnt haben, hatten sich nur einige über die Umsetzungsformen einer variablen Vergütung Gedanken gemacht. Vor allem bei den Leistungserbringern lag hierbei der Fokus auf der Frage, ob ein finanzielles Anreizsystem Bonus- oder Malusaspekte oder beides beinhalten sollte. Die Mehrzahl sprach sich für ein reines Bonus-System aus – auch mit dem Wissen, dass dieser Bonus durch Absenkung anderer Kosten finanziert werden müsste. Die Nutzung einer reinen Bonuszahlung gerade in der Einführungsphase eines neuen Systems wird auch durch wissenschaftliche Untersuchungen unterstützt (vgl. Custers et al. 2008, S. 6; Scheppach et al. 2011, S. 110; AMA 2005).

Zur Bewertungsart lassen sich ebenfalls einige Erkenntnisse darstellen: Anhand des Kriterienkatalogs wird deutlich, dass eine langfristige, als gerecht empfunden Anreizwirkung nur dann zu erwarten ist, wenn sowohl die absolut beste Qualität als auch die relative Qualitätsverbesserung belohnt werden – dies wird auch in verschiedenen Reviews herausgestellt (vgl. z. B. van Herck et al. 2010, S. 8; aber auch AMA 2005). Auch die Befragten der empirischen Fallstudie unterstützen eine Kombination von beiden Bewertungsaspekten, priorisieren aber langfristig eher eine Fokussierung auf die relative Verbesserung, um einen kontinuierlichen Verbesserungsprozess bei allen Beteiligten zu fördern.

Bezüglich einer notwendigen Mindest-Vergütungshöhe als Anteil der Gesamtvergütung kann aufgrund der Erkenntnisse dieser Arbeit nur eine grobe Einschätzung gegeben werden. Einerseits wird deutlich, dass die Wirksamkeit der variablen Vergütung davon abhängt, welche Beteiligten den Anreiz erhalten: In einer Fachabteilung ist ein relevanter Vergütungsanteil bei einem variablen Vergütungssystem beispielsweise von Hüft-TEPs vergleichsweise gut erreichbar. Bezogen auf die gesamte Klinik ist die Situation unterschiedlich. Beispielrechnungen zeigen, dass durch Garantieklauseln bei einer Fachklinik 0,5 % des Umsatzes qualitätsabhängig sein können. Bei einem Maximalversorger allerdings könnten bei sonst gleichen Fallzahlen im P4P-Bereich selbst aufgrund des höheren nichtbetroffenen Gesamtvolumens nur 0,02 % des Umsatzes betroffen sein (vgl. Seeger 2011). Der letztgenannte Anteil wäre meist nicht ausreichend, um seitens des Klinikmanagements als steuerungsrelevant angesehen zu werden. Doch auf beiden Ebenen (Fachabteilung und Management des Leistungserbringers) müssen die Anreizwirkungen ausreichend sein, um zum einen die klinische Prozesse zum anderen aber auch bspw. die Investitionsentscheidungen nachhaltig verändern zu können. Einige Studien gehen davon aus, dass dazu bis zu 5 % variabler Vergütungsanteil notwendig sind (vgl. Christianson et al. 2008, S. 12; Veit et al. 2012e, S. 32). Dies wird durch die Aussagen der Befragten in dieser Arbeit unterstützt.

Damit weiterhin mit Gewährleistungen gearbeitet wird, sollten Optimierungen vorgenommen werden. So sollte sich die Ergebnismessung z. B. entweder auf die Nutzung von mittleren Revisionsraten oder auf eine komplette Einzelfallgarantie, bei der die Verschuldensfrage von vornherein ausgeschlossen wird,

beziehen. Dadurch können langwierige, in der Praxis nicht führbare Diskussionen um das Zutreffen der Gewährleistungsregelung bei Einzelfällen und um die statistischer Aussagekraft umgangen werden (vgl. Mansky/Seeger 2011, S. 210f.). Eine noch bessere Reichweite und Anreizwirkung könnte erzielt werden, wenn anhand eines aggregierten Bepunktungssystems verschiedene Messergebnisse bspw. über eine qualitätsabhängige Anpassung des Basisfallwertes umgesetzt würden (vgl. Mansky 2010a). Ähnlich verfährt jetzt beispielsweise die CMS (vgl. HHS - CMS 08.05.2011, S. 26521).

Zusammenfassung der Empfehlungen:

- Fokussierung auf ein Bonussystem, gerade während der Einführungsphase
- Kombination von „absoluten“ und „relativen“ Bewertungsarten mit Schwerpunkt auf die langfristige Qualitätsverbesserung
- Hinreichend hoher Vergütungsanteil abhängig von der angestrebten Anreizebene
- Arten der Vergütung: Optimierte Garantieklausel ist möglich, langfristig sollten aber Bepunktungssysteme etabliert werden

Zusammenfassend wird beschrieben, wie die einzelnen Kriterien des Katalogs durch ein neues P4P-System in Deutschland bestmöglich erfüllt werden könnten.

Tabelle 38: Bewertungen P4P-Modell Systemelemente Gesamtsystem, Versorgungsbereich, Qualitätsindikator, Messung

Systemelement	Kriterienkategorie	#	Einzelkriterium	Anzustrebende Ausprägung für hohen Erfüllungsgrad
Gesamtsystem	**Aussagekraft**	1a	Abdeckung des Gesamtsystems	Möglichst hoher Anteil an Fallzahlen und Todesfällen im Indikatorensystem anzustreben
Versorgung-bereich	**Aussagekraft**	2a	Medizinische Bedeutung für das Versorgungssystem und das Fachgebiet	Auswahl von häufigen oder seltenen, aber relevanten Krankheitsbildern bzw. Eingriffen
		2b	Ökonomische Bedeutung für das Versorgungssystem oder einen Fachbereich	Auswahl von häufigen und/oder teuren Krankheitsbildern bzw. Eingriffen, möglichst Einbeziehung von Folgekosten (wie bspw. auch Erhöhung der Pflegestufe)
Qualitäts-indikator	**Aussagekraft/ "Das Richtige messen"**	3a	Medizinische Bedeutung für das Behandlungsergebnis des Patienten	Möglichst Nutzung von Ergebnisqualität, falls bei chronischen Erkrankungen nicht möglich, Auswahl von Prozessindikatoren für die ein Bezug zum Langzeitergebnis nachgewiesen wurde
		3b	Relevanz des Indikators für ein bestimmtes Fachgebiet	Auswahl des Indikators anhand von bspw. literaturbekannten Möglichkeiten zur Qualitätsverbesserung bzw. bekannten Qualitätsproblemen im jeweiligen Gebiet
		3c	Sektorbezug vs. sektorübergreifende Nutzung	Zunächst nur ein Sektor mit Möglichkeiten zur Ausweitung bspw. durch Ansätze wie QSR
	Akzeptanz	4a	Medizinische Bedeutung in der medizinischen Fachliteratur	Auswahl der Indikatoren vor allem anhand der wissenschaftlichen Literatur, insbesondere anhand der "Studienendpunkte" wissenschaftlicher Studien
		4b	Wahrgenommene medizinische Relevanz auf der Ebene der medizinischen Fachgesellschaften	Nutzung von Indikatoren, die auch von den Fachgesellschaften selbst als Qualitätskriterium genutzt bzw. empfohlen werden (z.B. Mortalitäten)
		4c	Wahrgenommene medizinische Relevanz/Nutzen bei den medizinischen Endanwendern	Nutzung von etablierten Indikatoren, die vom medizinischen Personal akzeptiert und wenn möglich auch für beeinflussbar gehalten werden
		4d	Verständlichkeit für Patienten und interessierte Öffentlichkeit	Indikatoren und Messergebnis sollten detailliert und laienverständlich erklärbar sein, möglichst einheitliche Nutzung und Darstellung auf verschiedenen Plattformen
		4e	Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit	Bezug auf Ergebnisse der Behandlung, die auch vom Patienten als wichtig verstanden werden (bspw. Sterblichkeiten, Komplikationsraten, Revisionsraten)
Messung	**Wissenschaft-lichkeit / "Richtig messen"**	5a	Reliabilität	Standardisierte und reproduzierbare Definition der Kennzahlen, möglichst weitgehende Vereinheitlichung der Dokumentationsweise, ggf. mit Stichprobenprüfungen
		5b	Validität	Einheitliche, standardisierte, nicht-interpretationsanfällige Definition der Messparameter, Bezug bzw. Beschränkung auf möglichst sicher erfassbare und erfasste Daten
		5c	Risikoadjustierung und Risikoadjustierbarkeit	Im ersten Schritt möglichst hohe medizinische Risikohomogenität des Indikators selbst, im zweiten Schritt Adjustierung der noch vorhandenen Risikounterschiede anhand möglichst sicher erfassbarer und reproduzierbarer Kriterien (bspw. Alter, Geschlecht, gut definierte Risikoscores), Vermeidung leicht beeinflussbarer, dokumentationsabhängiger Risikokriterien
	Messaufwand und -sicherheit	6a	Objektiver Erhebungsaufwand und -komplexität	Nutzung von Routinedaten, standardisierte, automatisierte Auswertung
		6b	Manipulationssicherheit / Objektivität der Messung	Nutzung von geprüften Abrechnungsdaten, Beschränkung auf möglichst sicher messbare Inhalte
		6c	Grad der Integration in automatische Prozesse / IT-Systeme	Standardisierte, algorithmenbasierte Messung, die eine vollständige Softwareintegration ermöglicht, möglichst Einbeziehung in kollektivvertragliche Regelungen um eine wirtschaftliche Umsetzung zu ermöglichen

Tabelle 39: Bewertungen P4P-Modell Systemelemente Ziele, Vergütungsfunktion, Messergebnis

Systemelement	Kriterienkategorie	#	Einzelkriterium	Anzustrebende Ausprägung für hohen Erfüllungsgrad
Ziele	Wirksamkeit/ Zielgerichtetheit	7a	Zielkonsistenz	Überprüfung der Konsistenz anhand von Einbeziehung der Beteiligten und ggf. Umfragen
		7b	Zielkongruenz	Gesamtes Zielsystem zeitgleich entwickeln und aufeinander abstimmen, Überprüfung der gleichzeitigen Erreichbarkeit und Widerspruchsfreiheit aller Ziele und Anreizwirkungen
		7c	Reichweite	Kollektive Messung und möglichst auch kollektivvertragliche Umsetzung, Entwicklung eines möglichst umfassenden Ziel- und Indikatorensystems
	Operationali-sierbarkeit	8a	Erreichbarkeit	Überprüfung der Erreichbarkeit anhand von Analysen bei beteiligten Leistungserbringern sowie ggf. auch anhand von Daten der Krankenkassen
		8b	Transparenz	Veröffentlichung der Zieldefinition, Übermittlung der Ergebnisse an alle beteiligten Akteure, abgestimmtes Konzept für Public Reporting (z.B. Suchportale)
		8c	Akzeptanz	Einbeziehung aller Beteiligten in den Entwicklungsprozess und transparente Kommunikation
		8d	Langfristige Anreizwirkung der Ziele	Kombination von "absoluten" und "relativen" Bewertungsaspekten, kontinuierliche Überprüfung und ggf. Anpassung der Ziele z.B. an den medizinischen Fortschritt, Auswahl von Zielen für die nach Datenanalysen mit Variationen der Versorgungsqualität zu rechnen ist, ggf. Streichung von Zielen wenn keine Versorgungsdifferenzen mehr erkennbar sind
	Vergleichbarkeit	9a	Gerechtigkeit	Ausreichende Risikoadjustierung, ggf. wahrgenommene Gerechtigkeit durch Umfragen überprüfen
Vergütungs-funktion	Wirksamkeit/ Zielgerichtetheit	10a	Effizienz	Analyse der Effizienz bei Kostenträgern und Leistungserbringern; Investitions- und Prozesskosten kostenseitig mit einbeziehen; einheitliche, kollektivvertragliche Umsetzung reduziert aufgrund von Skaleneffekten die Kosten für die beteiligten Einheiten
		10b	Effektivität	Evidenz anhand von möglichst randomisierten Studien im Rahmen von Pilotprojekten überprüfen, auch nach Einführung des Gesamtsystems Qualitätsverbesserungen evaluieren
		10c	Flexibilität	Vor allem durch relativen Vergütungsanteil zu erreichen, Anpassung an Inflation bspw. durch Bezug auf den Basisfallwert
	Operationali-sierbarkeit	11a	Anteil der Vergütung	Anteil für verschiedene Management- und Organisationsebenen betrachten, insgesamt sollte vermutlich mindestens ein Anteil von 5% der Gesamtvergütung bei maximaler Zielerreichung angestrebt werden, der faktisch wirksame Bereich sollte anhand von Häufigkeitsverteilungen überprüft werden
		11b	Planbarkeit	Indikatordefinitionen, Vergütungsfunktion und anzuwendene Verfahren sollten prospektiv kommuniziert werden
Messergebnis	Operationali-sierbarkeit	12a	Nutzung zur Steuerung	Möglichst aktuelle, verständliche und standardisierte Darstellung der Ergebnisse, damit die Beteiligten verstehen können, was von ihnen erwartet wird; Steuerung selbst muss aber von den Beteiligten umgesetzt werden; die Steuerbarkeit ergibt sich aus der Einhaltung der zuvor genannten Kriterien (Operationalisierbarkeit, Wahrgenommene Bedeutung, etc.)
		12b	Aktualität	Kurzfristergebnisse sollten wenn möglich monatlich darstellbar sein, Langzeitergebnisse sollten wenn möglich nicht älter als ca. 2 Jahre sein
		12c	Kontrollierbarkeit	Ergebnisse müssen beim Kostenträger und Leistungserbringer vorliegen und auswertbar sein, der Leistungserbringer sollte die Möglichkeit haben, die Ergebnisse auch auf Fallebene nachvollziehen zu können
		12d	Praktische Umsetzung im Controlling	Einbeziehung von Controlling-Aspekten in den Entwicklungsprozess, Automatisierbarkeit der Auswertungen, dazu u.a. notwendig standardisierte, algorithmische Definition der Indikatoren (siehe Kriterium 6d)
		12e	Verursachungsgerechte Zuordnung und Beeinflussbarkeit	Auswahl von Indikatoren, die fachlich möglichst gut zuordenbar sind, die eine nennenswerte Variationsbreite zeigen und deren Ergebnis medizinisch prinzipiell beeinflussbar ist (bspw. keine infausten Krankheitsbilder)
	Vergleichbarkeit	13a	Zeitverlauf	Daten im Jahresvergleich und -entwicklung darstellen; durch möglichst wenig Änderungen der Definitionen Vergleichbarkeit gewährleisten
		13b	Querschnittsvergleich/Ranking	Rankings im Vergleich zu anderen Kliniken und zum Bundesdurchschnitt ermöglichen, dadurch auch Transparenzerhöhung
		13c	Justiziabilität	Manipulationssicherheit, möglichst überprüfte Daten (z.B. Abrechnungs-prozess), klare gesetzliche Vorgaben zum Umgang mit geringen Mengen

Es zeigt sich, dass diese Kriterien auch im deutschen stationären Sektor gut umsetzbar wären. Außerdem kann der Katalog für einen ersten, groben Modellentwurf genutzt werden –für eine feinere Detaillierung können auch die Kriterienbeschreibungen und empirischen Ergebnisse dieser Arbeit wichtige Hilfestellungen liefern.

Resümee

Das erste zentrale Element der vorliegenden Arbeit ist die Entwicklung eines Kriterienkataloges für qualitätsorientierte Vergütungssysteme. Dabei standen folgende Aspekte im Mittelpunkt:

- Bisherige Auflistungen von Anforderungen betreffen entweder nur die Indikatorenentwicklung oder die Anreizsysteme. Ein umfassender Katalog, der alle Kriterien detailliert beschreibt, ist bislang nicht bekannt. Der hier entwickelte Katalog kombiniert Anforderungen aus der Qualitätsmessung der Medizin mit solchen aus Anreizsystemen der Betriebswirtschaftslehre. Die Gliederung anhand von sieben Systemelementen mit acht Kriterienkategorien und 38 Einzelkriterien ermöglicht es, das Vorgehen beim Aufbau von P4P-Systemen anhand der Erkenntnisse zur Funktionsweise von qualitätsorientierten Anreizsystemen zu operationalisieren und bietet andererseits einen verständlichen, praxisorientierten Aufbau des Katalogs.

- Der Katalog bietet damit eine umfassende Sicht auf alle System- und Funktionselemente von qualitätsorientierten Vergütungssystemen und definiert gleichzeitig sehr detailliert und praktisch anwendbar die wesentlichen Beurteilungskriterien. Jedes Kriterium wird anhand von Anwendungsbeispielen, Bewertungsmöglichkeiten, Skalen und Informationsquellen beschrieben.

- Wesentliche neue Aspekte, die in den Kriterien dieses Kataloges betrachtet werden, finden sich insbesondere im Hinblick auf die Manipulationssicherheit der Qualitätsmessung, die algorithmische Definition der Qualitätskriterien speziell auf der Basis verfügbarer Routinedaten, die Automatisierbarkeit der Messung und des damit verbundenen Controllings bei Leistungserbringern und Krankenkassen sowie die sich daraus ergebende stringente Umsetzung des Controllings, die Nutzung zur unternehmensinternen Steuerung und die Begründung der Relevanz von Qualitätsindikatoren anhand der medizinischen Fachliteratur. Viele der weiteren Kriterien, die auch in anderen Arbeiten schon berücksichtigt wurden, werden hinsichtlich der praktischen Anwendung deutlich präzisiert und weiterentwickelt.

- Eine zusätzliche, detaillierte Literaturdiskussion existierender Pilotprojekte von P4P in den USA und Deutschland bietet weitere Erkenntnisse für wichtige Gestaltungsdimensionen wie bspw. Vergütungshöhe, Bemessungsgrundlage und Adressatenebene. Hier werden vor allem mögliche Ausgestaltungen von P4P-Projekten beschrieben und die Bewertung von Qualitätsverbesserungen

diskutiert. Anforderungen an P4P-Systemen werden dabei nur am Rande betrachtet, da sie oft nicht Gegenstand dieser Veröffentlichungen sind.

Parallel zur theoretischen, deduktiven Entwicklung des Kriterienkatalogs wurden sieben qualitativ-empirische Fallstudien bei vier Leistungserbringern und drei Kostenträgern bezüglich der Umsetzung von P4P in IV-Verträgen mit Gewährleistungsklauseln durchgeführt. Diese Art von Verträgen stellen derzeit de facto eine der häufigste Umsetzungsformen von P4P im stationären Sektor in Deutschland dar. In den Fallstudien wurde der Kriterienkatalog einerseits iterativ weiter präzisiert, andererseits aber anschließend bereits auf die Projekte angewandt. Daraus lassen sich folgende Aussagen zur praktischen Implementierung der untersuchten IV-Verträge ableiten:

- In keinem der untersuchten IV-Verträge mit Gewährleistungsklauseln wurden die darin enthaltenen qualitätsorientierten Vergütungselemente vollständig umgesetzt. Oft scheitert die Umsetzung der getroffenen Vereinbarungen bereits daran, dass Gewährleistungsfälle nicht sicher erkannt werden. Hier wären auch im bestehenden System Verbesserungen möglich bzw. sind teilweise bereits geplant.

- Eine weitere Schwierigkeit der Umsetzung von P4P in IV-Verträgen ist die geringe Zahl von eingeschlossenen Fällen und somit auch daraus folgenden Gewährleistungsfällen. Die Befragten beschrieben eine recht geringe, strategische Relevanz solcher Verträge, so dass Investitionen in Prozessveränderungen und Software häufig nicht ausreichende wirtschaftliche Bedeutung erlangen und somit nicht durchgeführt werden. Außerdem ist durch die unterschiedliche Gestaltung von Selektivverträgen sowohl eine standardisierte Umsetzung und Evaluierung der P4P-Elemente mit Skaleneffekten als auch ein Vergleich verschiedener Systeme kaum möglich.

- Gleichzeitig zeigen aber die meisten Befragten in den Fallstudien ein großes Interesse an P4P und gehen davon aus, dass diese Vergütungsform mittel- oder langfristig in Deutschland eingesetzt wird. Hierfür wird eine valide, aussagekräftige, manipulationssichere, akzeptierte Messung von Qualität sowie eine Vergütung, die sich vor allem auf Bonus-Elemente konzentriert, erwartet.

- Außerhalb der untersuchten IV-Verträge haben einige der befragten Leistungserbringer und Kostenträger erste P4P-Elemete etabliert, bspw. in sehr detaillierten, etablierten Qualitätsmanagementsystemen oder Ideen für Bonusverträge mit klaren P4P-Aspekten bspw. im ambulanten Sektor. Auch im stationären Sektor gibt es erste Ansätze, die aber beim hier analysierten Leistungserbringer nicht umgesetzt wurden und daher nur am Rande analysiert werden.

- Der Katalog wurde insgesamt sechsmal angewandt, auch die anderen Fallstudien wurden anhand der Systemelemente und einzelnen Kriterien bewertet. Der Kriterienkatalog zeichnet sich zunächst durch eine schnelle Bewertbarkeit aus – gerade die Detailbeschreibungen der einzelnen Kriterien ermöglichen eine sehr konkrete, vergleichbare Auswertung. Dies ist naturgemäß umso mehr der Fall,

je mehr Daten und Informationen zu dem Projekt vorliegen. Aufwendiger als die Bewertung selbst war daher die Datenerhebung durch qualitative Interviews.

In der Arbeit werden außerdem erste Empfehlungen für eine mögliche Einführung eines bundesweiten P4P-Systems gegeben. Dazu wurden die wünschenswerten, jeweils optimalen Ausprägungen für jedes Element des neu entwickelten Kataloges in Kürze skizziert.

Im Ergebnis stellt sich die Umsetzung anhand eines kollektivvertraglichen Systems mit Fokus auf Ergebnisqualität, in dem die Erreichung der Kennzahlen anhand eines Gesamtscores vergütet wird, als ein möglicher, optimaler Lösungsansatz dar. Es verbleibt die Erkenntnis, dass eine qualitätsorientierte Vergütung in Deutschland umsetzbar wäre und langfristig die Qualitätsorientierung im stationären Sektor verbessern könnte.

Anhang: Interviewleitfäden

1.1. Interviewleitfaden – Krankenkasse: Controlling

1.2. Interviewleitfaden – Krankenkasse: Vertragsentwicklung

1.3. Interviewleitfaden – Krankenkasse: Sachbearbeiter

1.4. Interviewleitfaden – Krankenhaus: Geschäftsführung, ärztlicher Direktor, Qualitätsmanagement

Interviewleitfaden – Krankenkasse: Controlling

A. Inhalte des IV-Vertrages

1. Allgemeine Inhalte von IV-Verträgen

- Wie viele IV-Verträge liegen in Ihrem Verantwortungsbereich?
- Welche Art von Verträgen sind es (Indikationen, Sektoren, Größe, etc.)?
- Mit welchen Inhalten der IV-Verträge beschäftigen Sie sich für das Controlling besonders?

2. Definition von Qualität in den IV-Verträgen (allgemein)

- Welche Qualitätsindikatoren gibt es in den IV-Verträgen?
- Gibt es festgesetzte Qualitätsziele für die Leistungserbringer? Wie sind diese definiert?
- Werden die Qualitätsziele für die Mitarbeiter des Leistungserbringers und/oder Versicherte/Patienten publiziert?
- Sind die Ziele von allen Beteiligten akzeptiert oder im Nachhinein strittig?
- Gibt es Anpassungen der Qualitätsziele (z. B. an Qualitätsverbesserungen, Situation des Leistungserbringers)?

3. Qualitätsorientierte Vergütungsaspekte des IV-Vertrags

- Gibt es ein Konsequenzenmanagement für die Erreichung/Nicht-Erreichung der Qualitätsziele?
- Werden die Qualitätsziele/-ergebnisse finanziell vergütet / nicht vergütet?
- Wie funktioniert die „Vergütungsfunktion“ der Qualität?
- Wird diese Vergütungsfunktion regelmäßig überprüft oder angepasst?

4. Verbesserungsmöglichkeiten für die Inhalte der IV-Verträge

- Haben Sie Ideen für Verbesserungsmöglichkeiten der Inhalte der IV-Verträge?

B. Umsetzung des IV-Vertrages

5. Analyse und Konsequenzenmanagement von eigenen Qualitätsindikatoren

- Wie funktioniert die Datenerhebung und -analyse für die Messung der Qualitätsindikatoren?
- Wie sind die Messungen und Controllingaspekte in der EDV integriert?
- Gibt es Überprüfungsmechanismen für die automatische Erkennung von qualitätsrelevanten Fällen (Plausibilität, Vollständigkeit, Vollzähligkeit)?
- Welche Konsequenzen haben die Messergebnisse oder Qualitätsberichte?
- An wen werden die Messergebnisse und/oder Qualitätsberichte weitergeleitet?
- Gibt es eine monetäre Bewertung der Qualitätsindikatoren/-messung? Bezieht diese sich auf den Einzelfall oder auf die Gesamtvergütung?

6. Analyse und Konsequenzenmanagement von Gewährleistungsfällen

- Gibt es eine automatische Erkennung von Gewährleistungsfällen?
- Wenn nein: Wie werden die Fälle erkannt, wenn überhaupt?
- Ist die Erkennung eines Gewährleistungsfalles (weitgehend) unabhängig vom Leistungserbringer möglich?
- Was passiert, wenn ein möglicher Garantiefall erkannt wird?
- Gibt es einen standardisierten Prozess der Umsetzung von Gewährleistungsfällen?
- An wen werden diese potenziellen Gewährleistungsfälle im Streitfall weitergeleitet? (Eskalationsverfahren; MDK, Rechtsabteilung usw.)
- Werden die Informationen der Gewährleistung (automatisch) an den Leistungserbringer zurückgemeldet?
- Warum werden Gewährleistungsfälle ggf. nicht umgesetzt?

7. Controlling der IV-Verträge

- Gibt es regelmäßige Controllingberichte für die Qualitätsmessung?
- Was sind die Kerninhalte dieser Controllingberichte?
  - Sind die Ergebnisse der Qualitätsmessung im Controllingbericht integriert (sowohl Indikatoren des Vertrages als auch ggf. weitere eigene Indikatoren)?
- Werden eigene Messungen/Datenanalysen erstellt (u. a. um die Ergebnisse zu überprüfen)?
- Sind die monetären Vergütungsaspekte im Controllingbericht integriert?
- Wie werden diese ausgewertet?
- Wer sind die Adressaten der Controllingberichte?
- Welche Diskussion ergibt sich auf Grundlage der Berichte?
- Gibt es Controllingberichte, die dem Leistungserbringer zugestellt werden?
- Wie häufig werden die Controllingberichte erstellt?
- Müssen die Berichte weitgehend „manuell“ erstellt werden oder werden sie automatisch erzeugt?
- Wie aufwendig und kompliziert schätzen Sie die Erstellung der Controlling-Berichte ein?

8. Nutzung der Messergebnisse (z. B. zu Steuerungszwecken)

- Wie aktuell sind die Messergebnisse?
- Wann erfährt der Leistungserbringer ggf. von der Gesamtvergütung?
- Wird in der Qualitätsdokumentation auch der zeitliche Verlauf dargestellt?
- Werden verschiedene Leistungserbringer oder Fachabteilungen untereinander verglichen?
- Sind die Messergebnisse geeignet für die Justiziabilität des Vertrages?

9. Verbesserungsmöglichkeiten für die Umsetzung der IV-Verträge

- Haben Sie Ideen für Verbesserungsmöglichkeiten der Umsetzung des IV-Vertrages und für die Controllingberichte?

C. Bewertung und Weiterentwicklung des IV-Vertrages

10. Evaluationen der IV-Verträge

- Gibt es (zusätzlich zum Controllingbericht) regelmäßige Evaluationen der IV-Verträge?
- Werden die monetären Bewertungen der Qualitätsergebnisse in dieser Evaluation berücksichtigt?
- Werden die Messergebnisse von externen Stellen überprüft?

11. Evaluation und Weiterentwicklungen der IV-Verträge

- Wurde eine Qualitätsverbesserung festgestellt?
- Gibt es qualitative Vorteile von IV-Verträgen mit Qualitätsindikatoren vs. Vergleichsgruppen?
- Werden die Qualitätsvergütungen an ggf. veränderte Bedingungen angepasst?
- Werden die Qualitätsvergütungen an die Qualitätsverbesserungen angepasst?

12. Anreizsysteme in der Organisation

- Führt die Umsetzung der Qualitätsvereinbarungen zu Bonuszahlungen in der Kasse?
- Werden die Messergebnisse für eigene Steuerungsfunktionen und Vertragsweiterentwicklungen genutzt?

13. Einschätzungen zum Leistungserbringer

- Wie schätzen Sie die Bedeutung der Qualitätsthemen beim Leistungserbringer ein?
- Wie schätzen Sie die Anreizwirkung beim Leistungserbringer ein?
- Was könnte getan werden, um diese Anreizwirkung zu erhöhen?

14. Verbesserungsmöglichkeiten bei der Bewertung und Weiterentwicklung der IV-Verträge

- Welche Verbesserungsmöglichkeiten sehen Sie für die einzelnen Prozessschritte?

Interviewleitfaden – Krankenkasse: Vertragsentwicklung

A. Entwicklungsprozess und -beteiligte

1. Entwicklungsprozess der IV-Verträge

- Gibt es einen standardisierten Prozess für die Entwicklung von IV-Verträgen?
- Welche Schritte gibt es in diesem Prozess?
- Gibt es eine Projektstruktur und was sind die wichtigsten Meilensteine?
- Wie lange dauert in etwa der Entwicklungsprozess?

2. Beteiligte an der Entwicklung der IV-Verträge

- Wie viele Personen sind an der Entwicklung beteiligt?
- In welchen Abteilungen wird die Entwicklung hauptsächlich vorangetrieben und welche weiteren Abteilungen sind beteiligt?
- Wer ist an der Entwicklung von Seiten des/der Leistungserbringers beteiligt?
- Wie viele Gespräche gibt es mit Leistungserbringern?

3. Hauptziele der Entwicklung von IV-Verträgen

- Warum werden neue IV-Verträge, Musterverträge oder Leitlinien entwickelt?
- Was sind die Hauptziele bei der Entwicklung von IV-Verträgen (z. B. für die Vermarktung beim Versicherten, zur Abschöpfung der Anschubfinanzierung, verbesserte Zusammenarbeit mit Leistungserbringern, Qualitätsverbesserung, Einsparungen)?
- Welche quantitativen und qualitativen Aspekte werden in der Entwicklung des Vertrages berücksichtigt?
- Gibt es einen „Kriterienkatalog“ für die Entwicklung von IV-Verträgen?
- Wie wichtig sind Qualitätsaspekte in der Entwicklung der IV-Verträge?

4. Vorabevaluation (bzw. Erstellung von Business-Plänen) der IV-Verträge

- Gibt es eine "Vor-ab-Evaluation", wie. einen Business Plan für die IV-Verträge?
- Was beinhaltet diese Evaluation?
- Werden auch Qualitätsaspekte (z. B. eine mögliche Qualitätsverbesserung) in der Evaluation betrachtet?
- Werden Qualitätsaspekte für die Evaluation monetär bewertet und eine qualitätsorientierte Vergütung integriert?

5. Verbesserungsmöglichkeiten des Entwicklungsprozesses

- Haben Sie Ideen zu Verbesserungsmöglichkeiten beim Entwicklungsprozess?

B. Entwicklung der Qualitätsindikatoren

6. Auswahl der Indikation des Vertrages

- Wie wurde die Hauptindikation des IV-Vertrages ausgewählt?
- Welche Auswahlkriterien werden für die Wahl der Indikation herangezogen?
  - Fallzahl?
  - Umsatz?
  - Elektiv/Notfall?
  - Konkurrenzsituation?
  - (Weitere?)
- Inwieweit waren Überlegungen der Versorgungsqualität in der Bevölkerung und ihrem Versicherungskollektiv entscheidend für die Auswahl der Indikation?
- Welche Stellung hatte die mögliche Kosteneinsparung bei Qualitätsverbesserung bei der Wahl der Indikation?

7. Auswahl der Qualitätsindikatoren

- Wie und mit wem wurden die Qualitätsindikatoren entwickelt und ausgewählt?
- Welche Auswahlkriterien für die Qualitätsindikatoren gibt es?
- Wie wichtig waren die folgenden Aspekte bei der Definition der Qualitätsindikatoren:
  - Medizinische Bedeutung für das Behandlungsergebnis des Patienten (Indikatorevidenz)
  - Relevanz des Indikators für ein bestimmtes Fachgebiet
  - „Objektivierung/Quantifizierbarkeit“ des Ergebnisses, z. B. „saved lives, QALYs“ auch durch die Krankenkasse selbst
  - Sektorbezug und sektorübergreifende Nutzung
  - Wahrgenommene medizinische Bedeutung in der medizinischen Fachliteratur
  - Wahrgenommene medizinische Relevanz/Nutzen von med. Personal
  - Verständlichkeit für Patienten und interessierte Öffentlichkeit
  - Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit
  - Einsparpotenzial für die Krankenkasse
- Was waren die größten Herausforderungen bei der Auswahl der Indikatoren?
- Welche Themen mussten mit dem Leistungserbringer diskutiert werden?
- Wie lange hat in etwa die Auswahl der Indikatoren gedauert?

8. Entwicklung der Qualitätsmessung

- Wie und mit wem wurde die Messung der Indikatoren festgelegt?
- Welche Auswahlkriterien waren für die Entwicklung der Qualitätsmessung entscheidend?
- Wie wichtig waren die folgenden Aspekte für die Definition der Qualitätsmessung:
  - Wissenschaftlichkeit

    - Reliabilität
    - Validität
    - Sensitivität (Statistische Unterscheidungsfähigkeit)
    - Spezifität
    - Risikoadjustierbarkeit
  - Messaufwand und -sicherheit
    - Objektive Erhebungsaufwand und -komplexität
    - Manipulationssicherheit der Messung
    - Unabhängige Messung durch die Krankenkasse
    - Spezifizierbarkeit der Erfassungsqualität (gehört auch zu Manipulationssicherheit)
    - Vollständig- und Vollzähligkeit der Daten überprüfbar (gehört auch zu Manipulationssicherheit)
    - Grad der Integration in automatische Prozesse / IT-Systeme
- Welche Dokumentationen sind für die Messung notwendig?
  - 21er/301er Daten, wenn ja, welcher Anteil
  - BQS/AQUA
  - andere Zusatzerfassung im Krankenhaus
- Wer ist verantwortlich für die Messung und wer führt die Messung durch?
- Wurden Überprüfungsmechanismen für die Messergebnisse festgelegt (z. B. auf Vollständigkeit oder anhand von Routinedaten [QSR])?
- Gibt es einen Qualitätsbericht und wie oft wird dieser erstellt?

9. Entwicklung von Qualitätszielen

- Wurden konkrete Qualitätsziele definiert?
- Wie werden die Qualitätsziele festgelegt?
- Wie eng ist der Leistungserbringer in der Definition der Ziele eingebunden?
- Werden die Ziele von allen Beteiligten akzeptiert?
- Werden die Qualitätsziele an unterschiedliche Gegebenheiten beim Leistungserbringer angepasst?
- Werden die Qualitätsziele an Verbesserungen der Qualität angepasst?
- Werden die Qualitätsziele öffentlich bekanntgegeben (z. B. an Versicherte)?

10. Verbesserungsmöglichkeiten der Entwicklung von Qualitätsindikatoren

- Haben Sie Ideen für Verbesserungsmöglichkeiten bei der Entwicklung der Qualitätsindikatoren?

C. Entwicklung des IV-Vertrages (u. a. Vergütungsfunktion)

11. Gründe für die Entwicklung von Vergütungsaspekten

- Gibt es ein „Konsequenzenmanagement“ bei Nicht-Erreichung der Ziele oder eine qualitätsorientierte Vergütung?
- Wie und warum wurden diese Aspekte integriert?
- Welche wirtschaftlichen Effekte wurden kurz- und langfristig für die Kasse und den Leistungserbringer erwartet und wie wurden diese Effekte errechnet?
- Welche Gründe sehen Sie zukünftig für die Integration (bzw. Nicht-Integration) von qualitätsorientierten Vergütungsaspekten?

12. Ausgestaltung der Vergütungsfunktion

- Was für eine qualitätsorientierte Vergütung wurde im Vertrag vereinbart?
- Aufgrund welcher Kriterien und Aspekte wurden diese Vergütungsaspekte ausgewählt?
- Wie wichtig waren die folgenden Aspekte bei der Definition der Vergütungsfunktion:
  - Effizienz der Vergütung (kurz- und langfristig) für die Kasse und den Leistungserbringer
  - Effektivität der Vergütungsaspekte und gezielte Verbesserung der Qualität
  - Mindestanteil an Gesamtvergütung um Anreizwirkung für Leistungserbringer und Kasse zu gewährleisten
  - Flexibilität durch Anpassung an spezifische Gegebenheiten des Leistungserbringers und so „gerechte“ Darstellung
  - Langfristige Anreizwirkung durch Anpassung an neue Gegebenheiten
  - Planbarkeit der Vergütung für Kasse und Leistungserbringer
- Gibt es weitere Arten von Konsequenzenmanagement (z. B. Nicht-Erstattung bei Gewährleistungsklauseln) und wie werden diese umgesetzt?
- Wie relevant ist die qualitätsorientierte Vergütung in den kontinuierlichen Verhandlungen mit dem Leistungserbringer?

13. Weitere Nutzung der Messergebnisse (z. B. zu Steuerungszwecken)

- Können die Messergebnisse auch zur internen Steuerung der Leistungserbringer (und evtl. der Kasse) genutzt werden?
- Wie aktuell sind die Messergebnisse und wann erfährt der Leistungserbringer ggf. von der Gesamtvergütung?
- Wird in der Qualitätsdokumentation auch der zeitliche Verlauf dargestellt?
- Werden verschiedene Leistungserbringer oder FAs untereinander verglichen?
- Sind die Messergebnisse geeignet für Justiziabilität des IV-Vertrages?

14. Verbesserungsmöglichkeiten der Entwicklung von Vergütungsaspekten

- Welche Ideen haben Sie für Verbesserungsmöglichkeiten bei der Entwicklung von Vergütungsaspekten?

D. Weiterentwicklung des IV-Vertrages

15. Evaluationen der IV-Verträge

- Gibt es regelmäßige interne Evaluationen der IV-Verträge?
- Wird die Qualitätsveränderung der betroffenen Patientenkohorte ausgewertet (und ggf. sogar mit anderen Patientenkohorten verglichen)?
- Werden die monetären Bewertungen der Qualitätsergebnisse in der Evaluation berücksichtigt?
- Werden die Messergebnisse von externen Stellen überprüft?
- Gibt es wissenschaftliche Evaluationen der Verträge?

16. Weiterentwicklungen der IV-Verträge

- Werden die bestehenden Verträge regelmäßig weiterentwickelt?
- Wie oft werden neue Musterverträge oder Leitlinien zur Entwicklung von Verträgen erstellt?
- Wie oft werden die besprochenen Aspekte der Qualitätsindikatoren und Vergütungsfunktionen angepasst?
- Welche Aspekte sind bei dieser Weiterentwicklung besonders wichtig?
- Werden neue wissenschaftliche und fachspezifische Aspekte in die Weiterentwicklung der Qualitätsindikatoren mit einbezogen?

17. Verbesserungsmöglichkeiten der Evaluation und Weiterentwicklung

- Welche Ideen haben Sie für Verbesserungsmöglichkeiten bei der Evaluation und Weiterentwicklung der IV-Verträge?

Interviewleitfaden – Krankenkasse: Sachbearbeiter

A. Inhalte des IV-Vertrages

1. Allgemeine Inhalte von IV-Verträgen

- Wie viele IV-Verträge bearbeiten Sie derzeit in Ihrem Zuständigkeitsbereich?
- Kennen Sie diese Verträge inhaltlich (Indikationen, Sektoren, Größe, etc.)?
- Werden Ihnen von der Vertragsabteilung die Verträge vorgestellt und die Ihren Bereich betreffenden Umsetzungsfragen mit Ihnen erörtert?
- Werden Sie bei der Entwicklung der Verträge hinsichtlich der Sie betreffenden Umsetzungsfragen einbezogen?
- Welches sind die wesentlichen Problemen der IV-Verträge beschäftigen Sie sich in der täglichen Arbeit besonders?

2. Definition von Qualität in den IV-Verträgen (allgemein)

- Welche Qualitätsindikatoren verfolgen Sie in der Abrechnung?
- Gibt es festgesetzte Qualitätsziele für die Leistungserbringer, die Ihren Bereich betreffen? Wie sind diese definiert?
- Welche dieser Qualitätsziele verfolgen Sie in Ihrem Bereich?
- Gibt es während der Laufzeit der Verträge Anpassungen dieser Qualitätsziele?

3. Qualitätsorientierte Vergütungsaspekte des IV-Vertrags

- Haben die Qualitätsziele der IV-Verträge im Prinzip konkrete Auswirkungen auf die Einzelfallvergütung oder andere Abrechnungsaspekte?
- Welche Auswirkungen gibt es für die einzelnen Qualitätsziele der IV-Verträge?

B. Umsetzung des IV-Vertrages

4. Erkennung von IV-Vertragsfällen

- Wie werden Fälle der IV-Verträge in der normalen Sachbearbeitung erkannt?
- Wie sind diese „Ausgangsfälle“ definiert?
- Gibt es eine „automatische Erkennung“ von Fällen, die für die IV-Verträge relevant sind? (Markierung am Bildschirm, Ampel o. ä.)
  - falls ja: für welche Fälle?
  - falls nein: für welche Fälle nicht und warum nicht?
- Welche Daten werden bei der automatischen Erkennung angezeigt?
- Gibt es eine patientenbezogene Historie der Fälle bei der Erkennung?
- Wer unterstützt Sie bei Streitfällen hinsichtlich der Abrechnungsfragen?
- Benötigt die Bearbeitung der IV-Fälle gegenüber „Normalfällen“ zusätzliche Arbeitszeit? Wenn ja: Wie hoch schätzen Sie den mittleren zusätzlichen Zeitaufwand pro Fall ein? (in Minuten)

- Sind Sie der Meinung, dass die EDV-Unterstützung für die Umsetzung dieser Abrechnungsfragen in Ihrem Bereich zufriedenstellend ist?

5. Analyse und Verfahrensweise bei Gewährleistungsfällen

- Wie sind die potenziellen Gewährleistungsfälle definiert?
- Findet die Erkennung der Gewährleistungsfälle über ICDs/OPS statt? Kann sie somit automatisiert werden?
- Werden auch andere Fälle (z. B. Beatmungsfälle) - also weitere (voll- oder teilautomatisierbare) Definitionen – erkannt?
- Gibt es eine automatische "Warnung", sobald mögliche Gewährleistungsfälle auftauchen?
- Gibt es Überprüfungsmechanismen für die automatische Erkennung der Gewährleistungsfälle?
- Welche Informationen über den Gewährleistungsfall werden angezeigt?
- Was passiert im Garantiefall? Wird automatisch nicht-erstattet?
- Wird der Garantiefall in der EDV für Statistiken und Controlling markiert? (z. B. Fallart, Entgeltschlüssel o. a.)
- Was passiert, wenn das Krankenhaus den Eintritt der Garantie bestreitet? Wer unterstützt Sie bei Streitfällen hinsichtlich der Gewährleistungsfragen?
  - Wie viele Fälle (%) sind strittig?
  - Wie lange dauern erfahrungsgemäß die Auseinandersetzungen? (Schätzung Monate, evtl. Statistik?)
  - Wie viele Fälle (Anteil in %) werden zugunsten der Kasse entschieden?
- Wird der Versicherte über den Gewährleistungsfall informiert?
- Benötigt die Bearbeitung der Garantiefälle gegenüber „Normalfällen“ zusätzliche Arbeitszeit? Wenn ja: Wie hoch schätzen Sie den mittleren zusätzlichen Zeitaufwand pro Fall ein? (in Minuten)
- Sind Sie der Meinung, dass die EDV-Unterstützung für die Umsetzung dieser Abrechnungsfragen in Ihrem Bereich zufriedenstellend ist?

6. Übergeordnete Punkte und Verbesserungsvorschläge

- Falls nicht umgesetzt: Warum werden aus Ihrer Sicht die Garantievereinbarungen nicht umgesetzt?
- Glauben Sie, dass die Garantiefälle für die Krankenhäuser in Ihrem Zuständigkeitsbereich wichtig sind? Können Sie sich Änderungen der IV-Vereinbarungen vorstellen, die die Bedeutung für das Krankenhaus vergrößern würden?
- Werden Sie über Ergebnisse und Strategien bei den IV-Verträgen informiert? Welche Rückmeldungen erhalten Sie?
- Gibt es Bonus- / Malusvereinbarungen für den Abrechner?
- Was könnte man aus Ihrer Sicht bei der Abrechnung und Kontrolle der IV-Vertragsfälle (und der Garantiefälle) verbessern?

Interviewleitfaden – Krankenhaus: Geschäftsführung, ärztlicher Direktor, Qualitätsmanagement[51]

A. Qualitätsmessung und -indikatoren [GF, ÄD, QM]

1. Allgemeine Aspekte

- Welche Qualitätsmesssysteme gibt es bei Ihnen?
  - Gibt es Messungen, die über die gesetzlich vorgeschriebenen hinausgehen?
  - IQI, PSI, QSR?
  - Eigene?
- Wer hat die Entscheidung für die Entwicklung dieser Messsysteme getroffen?
  - Wann kam es zu dieser Entscheidung?
  - Wurden die Fachbereiche einbezogen?
  - Wurde die QM-Abteilung einbezogen?
- Welche wissenschaftlichen Erkenntnisse werden genutzt?

2. Auswahl der Versorgungsbereiche

- In welchen Fachabteilungen finden bei Ihnen zusätzliche Qualitätsmessungen statt?
- Welche Versorgungsbereiche/Krankheiten werden in der Messung betrachtet?
- Wie wurden diese Bereiche ausgewählt?
  - Fallzahl?
  - Umsatz?
  - Elektiv/Notfall?
  - Konkurrenzsituation?
  - (Weitere?)
- Inwieweit waren Überlegungen der Versorgungsqualität in der Bevölkerung und ihrem Versicherungskollektiv entscheidend für die Auswahl der Indikation?
- Welche Stellung hatte die mögliche Kosteneinsparung bei Qualitätsverbesserung bei der Wahl der Indikation?

3. Entwicklung der Indikatoren

- Wie und mit wem wurden die Qualitätsindikatoren entwickelt und ausgewählt?
- Welche Auswahlkriterien für die Qualitätsindikatoren gibt es?
- Wie wichtig waren die folgenden Aspekte bei der Definition der Qualitätsindikatoren:
  - Medizinische Bedeutung für das Behandlungsergebnis des Patienten

[51] Nutzung der Fragengruppen bei verschiedenen Personenkreisen im Unternehmen durch Kürzel gekennzeichnet (Geschäftsführung = GF, ärztlicher Direktor = ÄD, Qualitätsmanagement = QM)

- Relevanz des Indikators für ein bestimmtes Fachgebiet
- Sektorbezug und sektorübergreifende Nutzung
- Wahrgenommene medizinische Bedeutung in der medizinischen Fachliteratur
- Wahrgenommene medizinische Relevanz/Nutzen von med. Personal
- Verständlichkeit für Patienten und interessierte Öffentlichkeit
- Wahrgenommene medizinische Bedeutung durch interessierte Öffentlichkeit

➢ Was waren die größten Herausforderungen bei der Auswahl der Indikatoren?

➢ Welche Themen mussten mit dem Leistungserbringer diskutiert werden?

➢ Wie lange hat in etwa die Auswahl der Indikatoren gedauert?

4. Entwicklung der Qualitätsmessung

➢ Wie und mit wem wurde die Messung der Indikatoren festgelegt?

➢ Welche Auswahlkriterien waren für die Entwicklung der Qualitätsmessung entscheidend?

➢ Wie wichtig waren die folgenden Aspekte für die Definition der Qualitäts-messung:

- Wissenschaftlichkeit
  - Reliabilität
  - Validität
  - Risikoadjustierbarkeit
- Messaufwand und -sicherheit
  - Objektive Erhebungsaufwand und -komplexität
  - Manipulationssicherheit der Messung
  - Unabhängige Messung durch die Krankenkasse
  - Vollständig- und Vollzähligkeit der Daten überprüfbar
  - Grad der Integration in automatische Prozesse / IT-Systeme

➢ Welche Dokumentationen sind für die Messung notwendig?

- 21er/301er Daten, wenn ja, welcher Anteil
- BQS/AQUA
- andere Zusatzerfassung im Krankenhaus

➢ Wer ist verantwortlich für die Messung und wer führt die Messung durch?

➢ Wie aufwendig (in Zeit/Kosten) schätzen Sie die?

➢ Wurden Überprüfungsmechanismen für die Messergebnisse festgelegt (z. B. auf Vollständigkeit oder anhand von Routinedaten [QSR])?

➢ Gibt es einen Qualitätsbericht und wie oft wird dieser erstellt?

5. Entwicklung von Qualitätszielen

➢ Wurden konkrete Qualitätsziele definiert?

- Wie werden die Qualitätsziele festgelegt?
- Wie eng ist die Fachabteilung in der Definition der Ziele eingebunden?
- Gibt es Qualitätsziele für alle Mitarbeiterebenen (Chefärzte, Oberärzte, Assistenzärzte, Pflegedienst, sonstige Dienstarten)?
- Werden die Ziele von allen Beteiligten akzeptiert?
- Werden die Qualitätsziele an unterschiedliche Gegebenheiten beim Leistungserbringer angepasst?
- Werden die Qualitätsziele an Verbesserungen der Qualität angepasst?
- Werden die Qualitätsziele öffentlich bekanntgegeben (z. B. an Patienten)?

6. Verbesserungsmöglichkeiten der Indikatoren und Messung

- Haben Sie Ideen für Verbesserungsmöglichkeiten bei der Entwicklung der Qualitätsindikatoren?
- Welche Verbesserungsoptionen sehen Sie für Indikatoren?
- Sehen Sie Trends für die Weiterentwicklung von Indikatoren?

B. Managementsysteme und Public Reporting [GF, ÄD, QM]

7. Qualitätsmanagementsysteme

- Wie wird bei Ihnen die gesetzliche Vorgabe eines Qualitätsmanagementsystems umgesetzt?
- Gibt es Zertifizierungen/Akkreditierungen? (KTQ, EFQM, DIN EN ISO 9000, JCAHO, Procumcert?)
- Welche Managementsystemaspekte sind integriert – PDCA-Zyklus?
- Welche Aspekte der Managementsysteme sind im betrieblichen Alltag am wichtigsten?
- Wie werden die Systeme von den Mitarbeitern akzeptiert?

8. Zielvereinbarungssysteme

- Gibt es Zielvereinbarungen?
- Mit wem werden Zielvereinbarungen getroffen?
- Sind Qualitätsziele in diesen Vereinbarungen?
- Was sind die Konsequenzen bei Nicht-Erreichung?
- Gibt es variable Gehaltsanteile?
  - Wie hoch sind die variablen Anteil der Qualitätsziele?

9. Maßnahmen und Initiativen zur Qualitätsverbesserung

- Welche Initiativen gibt es zur Qualitätsverbesserung?
- Wie werden diese Maßnahmen umgesetzt?
- Gibt es messbare Ergebnisse dieser Initiativen?
  - Wer misst diese Ergebnisse?

- Welche Indikatoren gibt es?
- Gibt es Zielsetzungen für diese Maßnahmen?
- Gibt es Vergütungsaspekte für die Maßnahmen?

10. Maßnahmen des Public Reportings

- Welche Qualitätsberichte gibt es?
  - Werden zusätzliche Indikatoren veröffentlicht?
  - Gibt es Berichte der Zertifizierungen?
  - Werden Ergebnisse der Maßnahmen und Initiativen veröffentlicht?
- Wie werden die Qualitätsberichte eingeschätzt…
  - … von den Chefärzten?
  - … von den Pflegekräften?
  - … von weiteren Mitarbeitern?
  - … von den Patienten?
  - … von der Öffentlichkeit?
- Gibt es weitere Maßnahmen des Public Reportings?
  - Klinikführer
- Wie wichtig schätzen Sie diese Maßnahmen ein?
- Welchen Stellenwert hat das Qualitätsthema bei diesen Maßnahmen?

11. Verbesserungsmöglichkeiten für Managementsysteme

- Haben Sie Ideen zu Verbesserungsmöglichkeiten der Managementsysteme oder des Berichtswesens?

C. Anreizsysteme für Qualitätsverbesserungen / Vergütungsfunktionen

12. Gründe für die Entwicklung von qualitätsorientierter Vergütung

- Gibt es eine qualitätsorientierte Vergütung oder Anreize auf Basis von Qualitätszielen?
- Warum wurden diese Aspekte entwickelt?
  - Qualitätsaspekte
  - Marketingaspekte
  - Wirtschaftliche Aspekte

13. Ausgestaltung der Vergütungsfunktion

- Was für eine qualitätsorientierte Vergütung und Anreizsysteme existieren im Krankenhaus?
  - Variabler Anteil des Gehalts der Mitarbeiter? (nur bei außertariflichen)

- Integriert in Zielvereinbarungen
- Verteilungsschlüssel bspw. des Einkommens der Fachabteilungen anhand von Qualitätsparametern
- Gewährleitungsklauseln in IV-Verträgen

➢ Aufgrund welcher Kriterien und Aspekte wurden diese Vergütungsaspekte ausgewählt?

➢ Wie wichtig waren die folgenden Aspekte bei der Definition der Vergütungsfunktion:

- Effizienz der Vergütung (kurz- und langfristig) für die Kasse und den Leistungserbringer
- Effektivität der Vergütungsaspekte und gezielte Verbesserung der Qualität
- Mindestanteil an Gesamtvergütung um Anreizwirkung für Leistungserbringer und Kasse zu gewährleisten
- Flexibilität durch Anpassung an spezifische Gegebenheiten des Leistungserbringers und so „gerechte“ Darstellung
- Langfristige Anreizwirkung durch Anpassung an neue Gegebenheiten
- Planbarkeit der Vergütung für Kasse und Leistungserbringer

➢ Gibt es weitere Arten von Konsequenzenmanagement (z. B. Nicht-Erstattung bei Gewährleistungsklauseln) und wie werden diese umgesetzt?

➢ Wie relevant ist die qualitätsorientierte Vergütung in den kontinuierlichen Verhandlungen mit dem Leistungserbringer?

14. Weitere Nutzung der Messergebnisse (z. B. zu Steuerungszwecken)

➢ Können die Messergebnisse auch zur internen Steuerung genutzt werden?

➢ Wie werden die Ergebnisse den Fachabteilungen zugeordnet?

- Auf welche Mitarbeiterebene werden die Ergebnisse noch zugeordnet (Fachabteilung, Chefärzte, Oberärzte, Assistenzärzte, Pflegedienst?)

➢ Wie hoch schätzen Sie die Beeinflussbarkeit der Ergebnisse ein? Gibt es dazu Diskussionen?

➢ Wie aktuell sind die Messergebnisse und wann erfährt der Leistungserbringer ggf. von der Gesamtvergütung?

➢ Wird in der Qualitätsdokumentation auch der zeitliche Verlauf dargestellt?

➢ Werden verschiedene Leistungserbringer oder FAs untereinander verglichen?

➢ Sind die Messergebnisse geeignet für die Justiziabilität des IV-Vertrages?

15. Verbesserungsmöglichkeiten der Entwicklung von Vergütungsaspekten

➢ Welche Ideen haben Sie für Verbesserungsmöglichkeiten bei der Entwicklung von Vergütungsaspekten?

➢ Gibt es Ideen für Entwicklungen von Anreizsystemen?

➢ Wie würden Sie am liebsten die Anreizsysteme verbessern?

D. Nutzung von IV-Verträgen mit Qualitätsgarantien als P4P-System [QM, GF]

16. Hauptziele der Entwicklung von IV-Verträgen

- Warum werden neue IV-Verträge entwickelt?
- Was sind die Hauptziele bei der Entwicklung von IV-Verträgen (z. B. für die Vermarktung beim Patienten, verbesserte Zusammenarbeit mit Krankenkassen, Qualitätsverbesserung, Fallzahlsteigerung)?
- Welche quantitativen und qualitativen Aspekte werden in der Entwicklung des Vertrages berücksichtigt?
- Gibt es einen „Kriterienkatalog“ für die Entwicklung von IV-Verträgen?
- Wie wichtig sind Qualitätsaspekte in der Entwicklung der IV-Verträge?

17. Entwicklungsprozess der IV-Verträge

- Gibt es einen standardisierten Prozess für die Entwicklung von IV-Verträgen?
  - Welche Schritte gibt es in diesem Prozess?
  - Gibt es eine Projektstruktur und was sind die wichtigsten Meilensteine?
- Wie lange dauert in etwa der Entwicklungsprozess?
- Wie viele Personen sind an der Entwicklung beteiligt?
  - In welchen Abteilungen wird die Entwicklung hauptsächlich vorangetrieben und welche weiteren Abteilungen sind beteiligt?

18. Controlling der IV-Verträge mit Qualitätsgarantien [Medizin-Controlling/Controlling/Abrechnung]

- Gibt es IV-Verträge mit Qualitätsaspekten?
- Wie sind diese gestaltet?
  - Gewährleistungsklauseln
  - Qualitätsorientierte, variable Vergütung
- Wie werden die Qualitätsaspekte gemessen?
  - Welche Indikatoren?
  - Wie findet die Messung statt?
  - Welche Überprüfungen gibt es?
  - Wie oft gibt es Berichte?
- Was passiert beim Eintreten von Gewährleistungsfällen?
  - Werden sie immer der Kasse gemeldet?
  - Wird es anders kodiert?
  - Kommt es direkt zu einer Nicht-Erstattung?
  - Kam dies bereits vor?

19. Evaluation (bzw. Erstellung von Business-Plänen) der IV-Verträge

- Gibt es eine Evaluation, wie bspw. einen Business Plan für die IV-Verträge?
- Was beinhaltet diese Evaluation?
- Werden auch Qualitätsaspekte (wie z. B. eine mögliche Qualitätsverbesserung) in der Evaluation betrachtet?
- Werden Qualitätsaspekte für die Evaluation monetär bewertet und eine qualitätsorientierte Vergütung integriert?
- Wie wird mit den Ergebnissen umgegangen?

20. Verbesserungsmöglichkeiten der IV-Verträge

- Haben Sie Ideen zu Verbesserungsmöglichkeiten der IV-Verträge?

E. Vorschläge für P4P-Systeme [GF, ÄD, QM]

21. Einschätzung des Konzeptes von qualitätsorientierter Vergütung

- Was halten Sie vom Konzept einer qualitätsorientierten Vergütung?
- Wie schätzen Sie die derzeitige Entwicklung in den USA ein?
- Welche Aspekte empfinden Sie als hilfreich?
- Welche Aspekte empfinden Sie als nicht notwendig?

22. Einschätzung möglicher Umsetzung in Deutschland

- Welche zukünftigen Entwicklungen sehen Sie bezüglich stationärer Qualität in Deutschland?
- Welche Entwicklungen erwarten Sie bezüglich qualitätsorientierter Vergütung?
- Was glauben Sie sind die wichtigsten Elemente eines solchen Systems?
- Was sind wichtige Umsetzungsschwierigkeiten?
- Wie könnte man diese Schwierigkeiten adressieren?

23. Konkrete Umsetzung im Unternehmen

- Welche Entwicklungen erhoffen Sie sich für Ihr Unternehmen?
- Was sind wichtige Umsetzungsschwierigkeiten bei Ihnen?
- Wie könnte man diese Schwierigkeiten adressieren?

24. Einschätzungen zur Krankenkasse

- Wie schätzen Sie die Bedeutung der Qualitätsthemen bei der Krankenkasse ein?
- Wie schätzen Sie die Anreizwirkung bei der Krankenkasse ein?
- Was könnte getan werden, um diese Anreizwirkung zu erhöhen?

25. Sonstige Ideen

- Haben Sie noch weitere Ideen oder Vorstellungen zu dem Thema?

Literaturverzeichnis

AHRQ (2006) Inpatient Quality Indicators Overview. AHRQ Quality Indicators. Version 2.1. Hg. v. Agency for Healthcare Research and Quality. AHRQ: Rockville. MD.

Albers S, Klapper D, Konradt U, Walter A, Wolf J [Hsg.] (2006) Methodik der empirischen Forschung. Deutscher Universitätsvlg. Wiesbaden.

AMA (2005) Guidelines for Pay-for-Performance Programs. Hg. v. American Medical Associations. Online verfügbar unter http://www.ama-assn.org/resources/doc/psa/guidelines4pay62705.pdf, zuletzt geprüft am 02.02.2013.

Amelung VE, Zahn TP (2009) Pay-for-Performance (P4P) - Der Business Case für Qualität? Studie im Auftrag der DxCG Gesundheitsanalytik GmbH: Berlin.

Amelung VE, Zippan N, Mutegi CM, Zahn TP (2009) Monetäre Anreize gezielt einsetzen - das Konzept qualitätsorientierter Vergütung. *Monitor Versorgungsforschung* 2 (06): 25–30.

Andernach K (2006) Modell zur Bewertung und Steuerung der Qualitätsverbesserung im Rahmen von Qualitätsmanagementsystemen. Univ. Verl. d. TU Berlin.

AOK-Bundesverband, Forschungs- und Entwicklungsinstitut für das Sozial- und Gesundheitswesen Sachsen-Anhalt (FEISA), Helios Kliniken Gruppe, Wissenschaftliches Institut der AOK (WIdO) (2007) Qualitätssicherung der stationären Versorgung mit Routinedaten (QSR). Abschlussbericht. 1. Aufl. WIdO Wiss. Inst. der AOK: Bonn.

Appelros P, Högerås N, Terént A (2003) Case ascertainment in stroke studies: the risk of selection bias. *Acta Neurol Scand* 107 (2): 145–149.

AQUA-Institut GmbH (2010) Allgemeine Methoden. Im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach § 137a SGB V. Version 2.0. Hg. v. Gemeinsamer Bundesausschuss: Göttingen.

AQUA-Institut GmbH (2011a) AQUA-Qualitätsbericht 2010. Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH. Auftraggeber: Gemeinsamer Bundesausschuss: Göttingen.

AQUA-Institut GmbH (2011b) Konzept zur Sicherung der Dokumentationsqualität. Vor dem Hintergrund des Datenvalidierungsverfahrens in der sektorenübergreifenden Qualitätssicherung. Abschlussbericht. Hg. v. Gemeinsamer Bundesausschuss. Göttingen.

Arah OA, Westert GP, Hurst J, Klazinga NS (2006) A conceptual framework for the OECD Health Care Quality Indicators Project. *Int J Qual Health Care* 18 (Suppl 1): 5–13. doi: 10.1093/intqhc/mzl024.

ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) (2001) Beurteilung klinischer Messgrößen des Qualitätsmanagements. Qualitätskriterien und -Indikatoren in der Gesundheitsversorgung. *Z Arztl Fortbild Qualitatssich* 96 (5): 2–15.

ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) [Hg.] (2009) Qualitätsindikatoren. Manual für Autoren. Gemeinsame Einrichtung der Bundesärztekammer und der Kassenärztlichen Bundesvereinigung. ÄZQ Schriftenreihe. 36. Verlag Make a book: Berlin.

ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) (2010) Definitionen und Konzepte des Qualitätsmanagements. Online verfügbar unter http://www.aezq.de/aezq/kompendium_q-m-a/2-definitionen-und-konzepte-des-qualitaetsmanagements#6, zuletzt geprüft am 02.02.2013.

Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F (2002) Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. *J Psychosom Res* 53 (3): 737–740.

Bartels H (2009) Spezielle Gesichtspunkte postoperativer Komplikationen in der Viszeralchirurgie. *Chirurg* 80 (9): 780–789. doi: 10.1007/s00104-009-1690-3.

Barua B, Esmail N (2011) Hospital Report Card British Columbia 2011. Studies in Health Policy. The Fraser Institute. Toronto.

Bauer H (2010) BDC|Online - Information und Service für Deutsche Chirurgen. Sitzung "Pay for Performance" und künftige Kooperationsformen. 01.01.2010. Online verfügbar unter www.bdc.de, zuletzt geprüft am 25.01.2013.

Becker FG (1990) Anreizsysteme für Führungskräfte. Möglichkeiten zur strategisch-orientierten Steuerung des Managements. C.E. Poeschel: Stuttgart.

Beckman H (2006) Physician Reactions to Quantitative Individual Performance Reports. *American Journal of Medical Quality* 21 (3): 192–199. doi: 10.1177/1062860606287577.

Berthiaume JT, Chung RS, Ryskina KL, Walsh J, Legorreta AP (2006) Aligning financial incentives with quality of care in the hospital setting. *J Healthc Qual* 28 (2): 36-44, 51.

Berthiaume JT, Tyler PA, Ng-Osorio J, LaBresh KA (2004) Aligning financial incentives with "Get With The Guidelines" to improve cardiovascular care. *Am J Manag Care* 10 (7 Pt 2): 501–504.

Bestehorn K (2005) Medizinische Register: ein Beitrag zur Versorgungsforschung. *Med Klin* 100 (11): 722–728. doi: 10.1007/s00063-005-1098-x.

Bibel (1982) Die Bibel in heutigem Deutsch. Die Gute Nachricht des Alten und Neuen Testaments ohne die Spätschriften des Alten Testaments (Deuterokanonische Schriften/Apokryphen). 2. Aufl. Deutsche Bibelgesellschaft: Stuttgart.

Birkmeyer JD (2000) Should we regionalize major surgery? Potential benefits and policy considerations. *J Am Coll Surg* 190 (3): 341–349.

Black N (2010) Assessing the quality of hospitals. Comment on Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away.Lilford R, Pronovost P. BMJ. 2010 Apr 20; 340:c2016. Epub 2010 Apr 20. *BMJ* 340 (2): c2066. doi: 10.1136/bmj.c2066.

Black WC (2002) All-Cause Mortality in Randomized Trials of Cancer Screening. *CancerSpectrum Knowledge Environment* 94 (3): 167–173. doi: 10.1093/jnci/94.3.167.

Bleicher K (1992) Strategische Anreizsysteme. Flexible Vergütungssysteme für Führungskräfte. Schäffer-Poeschel Verlag; Verl. Neue Zürcher Zeitung: Stuttgart, Zürich.

Bleicher K (2002) Integriertes Management als Herausforderung. *In: Schwendt S,Funck D [Hg.] Integrierte Managementsysteme. Konzepte, Werkzeuge, Erfahrungen*. Heidelberg: Physica-Verlag: 1–24.

Blomqvist AG (1991) The doctor as double agent. Information asymmetry, health insurance and medical care. *J Health Econ* (Volume 10, Issue 4): 411–432. doi: 10.1016/0167-6296(91)90023-G.

Blumberg B, Cooper DR, Schindler PS (2008) Business research methods. 2. Aufl. McGraw-Hill Education: London.

Blumenstock G (2009) Entwicklung von Qualitätsindikatoren. *In: ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) [Hg.] Qualitätsindikatoren. Manual für Autoren*. Berlin: Verlag Make a book (ÄZQ Schriftenreihe, 36): 12–17.

Blum K, Offermanns M, Perner P MDK-Prüfungen – mehr als nur ein Einzelfall. *Das Krankenhaus* 2009 (2): 111–116.

BMG (Bundesministerium für Gesundheit) (14.11.2003) Gesetz zur Modernisierung der gesetzlichen Krankenversicherung. GKV-Modernisierungsgesetz - GMG. Berlin.

Böcken J, Braun B, Repschläger U [Hg.] (2013) Gesundheitsmonitor 2012. Bürgerorientierung im Gesundheitswesen - Kooperationsprojekt der Bertelsmann Stiftung und der BARMER GEK. Verlag Bertelsmann Stiftung: Gütersloh.

Bott J (2009) Perspectives on current uses of the AHRQ quality indicators and future development. *Dtsch Med Wochenschr* 134: S305. doi: 10.1055/s-0029-1242661.

Bottle A, Jarman B, Aylin P (2011) Strengths and weaknesses of hospital standardised mortality ratios. *BMJ* 342 (jan21 1): c7116. doi: 10.1136/bmj.c7116.

BQS-Institut (2007) Kompetenzen - Risikoadjustierung. Online verfügbar unter http://www.bqs-qualify.com/online/public/bqs/kompetenzen/komplinks/risikoadj, zuletzt aktualisiert am 16.01.2013.

Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK et al. (2006) Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *J Am Med Assoc (JAMA)* 296 (1): 72–78. doi: 10.1001/jama.296.1.72.

Braun B, Reiners H, Rosenwirth M, Schlette S (2012) Anreize zur Verhaltenssteuerung im Gesundheitswesen. Effekte bei Versicherten und Leistungsanbietern. Chartbook Bertelsmann Stiftung. Zentrum für Sozialpolitik: Gütersloh, Bremen.

Brühl R, Buch S (2006) Einheitliche Gütekriterien in der empirischen Forschung? Objektivität, Reliabilität und Validität in der Diskussion. ESCP-EAP Europ. Wirtschaftshochsch: Berlin.

Brush JE, Krumholz HM, Wright JS, Brindis RG, Cacchione JG, Drozda JP et al. (2006) American College of Cardiology 2006 Principles to Guide Physician Pay-for-Performance Programs. *J Am College Cardiol* 48 (12): 2603–2609. doi: 10.1016/j.jacc.2006.10.046.

Bufalino V (2006) Payment for Quality: Guiding Principles and Recommendations: Principles and Recommendations From the American Heart Association's Reimbursement, Coverage, and Access Policy Development Workgroup. *Circulation* 113 (8): 1151–1154. doi: 10.1161/circulationaha.105.171760.

Bufalino V, Peterson ED, Krumholz HM, Burke GL, LaBresh KA, Jones DW et al. (2007) Nonfinancial Incentives for Quality: A Policy Statement From the American Heart Association. *Circulation* 115 (3): 398–401. doi: 10.1161/circulationaha.106.180202.

Buhlman NW, Matthes N (2012) Using Metrics and Predictive Models to Succeed in CMS Hospital Value Based Payment Programs. PfP Summit: Los Angeles, 20.03.2012.

Buhr P, Klinke S (2006) Qualitative Folgen der DRG-Einführung für Arbeitsbedingungen und Versorgung im Krankenhaus unter Bedingungen fortgesetzter Budgetierung. Eine vergleichende Auswertung von vier Fallstudien. WZB: Veröffentlichungsreihe der Forschungsgruppe Public Health, Forschungsschwerpunkt Arbeit, Sozialstruktur und Sozialstaat, Wissenschaftszentrum Berlin für Sozialforschung: Berlin.

Bullinger H (1996) Erfolgsfaktor Mitarbeiter. Motivation - Kreativität - Innovation. Teubner: Stuttgart.

Bundesamt für Gesundheit [BAG] (2007) Qualitätsindikatoren der Schweizer Akutspitäler. Bericht über die Pilotstudie. Schweizerische Eidgenossenschaft. Bern.

Bundesamt für Gesundheit [BAG] (2010) Qualitätsindikatoren der Schweizer Akutspitäler 2007. Bericht über die Pilotstudie. Hg. v. BAG: Bern.

Bundesärztekammer (2011) (Muster-)Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte. Online verfügbar unter http://www.bundesaerztekammer.de/page.asp?his=1.100.1143#II, zuletzt geprüft am 21.06.2012.

Büscher G, Gerber A (2010) Gesundheitsökonomische Evaluation als Ansatz zur Steuerung der Ausgaben im Gesundheitswesen. *In: Lauterbach KW [Hg.] Gesundheitsökonomie, Management und Evidence based medicine. Handbuch für Praxis, Politik und Studium ;*. Stuttgart: Schattauer GmbH: 63–81.

Busse R (2006) Gesundheitsökonomie. *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz* 49 (1): 3–10. doi: 10.1007/s00103-005-1196-1.

Busse R, Geissler A, Quentin W, Wiley M [Hg.] (2011) Diagnosis-related groups in Europe. Moving towards transparency, efficiency and quality in hospitals. Open University Press: Maidenhead.

Busse R, Nimptsch U, Mansky T (2009) Measuring, monitoring, and managing quality in Germany's hospitals. *Health Aff* 28 (2): w294-304. doi: 10.1377/hlthaff.28.2.w294.

Busse R, Schreyögg J, Tiemann O [Hg.] (2010) Management im Gesundheitswesen. 2. Aufl. Springer: Heidelberg.

Campbell SM (2002) Research methods used in developing and applying quality indicators in primary care. *Quality and Safety in Health Care* 11 (4): 358–364.

Cannon MF (2007) Pay for Performance: Is Medicare a Good Candidate? The Cato Institute. *Yale J Health Policy Law Ethics* 7 (1): 1–38.

Chassin MR (1996) Quality improvement nearing the 21st century: prospects and perils. *Am J Med Qual* 11 (1): S4-7.

Chassin MR (2002) Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff* 21 (4): 40–51. doi: 10.1377/hlthaff.21.4.40.

Chassin MR, Loeb JM (2011) The ongoing quality improvement journey: next stop, high reliability. *Health Aff* 30 (4): 559–568.

Chien AT, Chin MH, Davis AM, Casalino LP (2007) Pay for performance, public reporting, and racial disparities in health care: how are programs being designed? *Med Care Res Rev* 64 (5 Suppl): 283S-304S.

Christianson JB, Leatherman S, Sutherland K (2008) Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. *Med Care Res Rev* 65 (6 Suppl): 5S-35S.

Coase R (1937) The Nature of the Firm. *Economica* Vol. 4 (16): 386–405.

Cochrane AL (1972a) Effectiveness and efficiency. The Royal Society of Medicine Press Ltd.: London.

Cochrane AL (1972b) The history of the measurement of ill health. *Int J Epidemiol* 1 (2): 89–92.

Codman EA (1996) At Study in Hospital Efficiency (Boston 1917). reprinted by the Joint Comission on Accreditation of Healthcare Organizations: Oakbrook Terrace, Illinois.

Collopy BT, Williams J, Rodgers L, Campbell J, Jenner N, Andrews N (2000) The ACHS Care Evaluation Program: a decade of achievement. Australian Council on Healthcare Standards. *J Qual Clin Pract* 20 (1): 36–41.

Comarow A (2006) Special report. Best hospitals. What it means to be best. *US News World Rep* 141 (2): 110.

Cots F, Chiarello P, Salvador X, Castells X, Quentin W (2011) DRG-based hospital payment: Intended and unintended consequences. *In: Busse R, Geissler A, Quentin W,Wiley M [Hg.] Diagnosis-related groups in Europe. Moving towards transparency, efficiency and quality in hospitals*. Maidenhead: Open University Press: 75–92.

Culter DM (1995) The Incidence of Adverse Medical Outcomes under Prospective Payment. *Econometrica* 63 (1): 29–50.

Custers T, Hurley J, Klazinga NS, Brown AD (2008) Selecting effective incentive structures in health care: A decision framework to support health care purchasers in finding the right incentives to drive performance. *BMC Health Serv Res* 8 (1): 66. doi: 10.1186/1472-6963-8-66.

Davidson G, Moscovice I, Remus D (2007) Hospital size, uncertainty, and pay-for-performance. *Health Care Financ Rev* 29 (1): 45–57.

Debatin JF, Ekkernkamp A, Schulte B, Baehr M [Hg.] (2010) Krankenhausmanagement. Strategien, Konzepte, Methoden. Med. Wiss. Verl.-Ges.: Berlin.

Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human behavior. Plenum Press: New York, NY.

DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB (1997) Comparing risk-adjustment methods for provider profiling. *Stat Med* 16 (23): 2645–2664.

Diel F (2009) Qualitätsindikatoren für eine Qualitätsorientierte Vergütung in der ambulanten Versorgung. Anforderungen und Erfahrungen, 2009, online verfügbar unter http://www.kvno.de/downloads/termine/focus/p4p09102009/focus_P4P_091009_Diel.pdf, zuletzt geprüft am 02.02.2013.

Dietrich M (2005) Qualität, Wirtschaftlichkeit und Erfolg von Krankenhäusern: Analyse der Relevanz marktorientierter Ansätze im Krankenhausmanagement. 1. Auflage. Deutscher Universitätsverlag/GWV Fachverlage GmbH: Wiesbaden.

Döbler K (2010) Vortrag Klaus Döbler, BQS: Bewertung von Qualitätsindikatoren für die sektorübergreifende Qualitätssicherung - QUALIFY, 22.01.2010.

Dolan E, Yanagihara D (2011) Measuring Total Cost of Care. Hg. v. Integrated Healthcare Association (Issue Brief, 3). Oakland.

Donabedian A (1966) Evaluating the quality of medical care. *Milbank Mem Fund Q* 44 (3): Suppl:166-206.

Donabedian A (1978) The quality of medical care. *Science* 200 (4344): 856–864. doi: 10.1126/science.417400.

Donabedian A (1980) Methods for deriving criteria for assessing the quality of medical care. *Med Care Rev* 37 (7): 653–698.

Donabedian A (1988) The quality of care. How can it be assessed? *J Am Med Assoc (JAMA)* 260 (12): 1743–1748.

Downs SH, Black N (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 52 (6): 377–384. doi: 10.1136/jech.52.6.377.

Drucker PF (1956) Praxis des Management. Ein Leitfaden f. d. Führungs-Aufgaben in d. modernen Wirtschaft. Econ-Verl: Düsseldorf.

Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A (2000) Selective referral to high-volume hospitals: estimating potentially avoidable deaths. *J Am Med Assoc (JAMA)* 283 (9): 1159–1166.

Eberlein-Gonska M (2011) Was ist an Qualitätsmanagement evidenzbasiert? *Bundesgesundheitsbl* 54 (2): 148–153. doi: 10.1007/s00103-010-1204-y.

Eichhorn S, Schmidt-Rettig B (1995) Mitarbeitermotivation im Krankenhaus. 1. Aufl. Bleicher: Gerlingen.

Eiff W v. (2000) Führung und Motivation in Krankenhäusern. Perspektiven und Empfehlungen für Personalmanagement und Organisation. Kohlhammer: Stuttgart.

Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. *Med Care* 36 (1): 8–27.

Emmert M (2008) Pay for Performance (P4P) im Gesundheitswesen. Ein Ansatz zur Verbesserung der Gesundheitsversorgung? HERZ, Health-Economics-Research-Zentrum: Burgdorf.

Emmert M, Maryschok M, Eisenreich S, Schöffski O (2009) Arzt-Bewertungsportale im Internet – Geeignet zur Identifikation guter Arztpraxen? *Gesundheitswesen* 71 (04): e18. doi: 10.1055/s-0028-1103288.

Endsley S, Kirkegaard M, Baker G, Murcko AC (2004) Getting rewards for your results: pay-for-performance programs. *Fam Pract Manag* 11 (3): 45–50.

Engesæter LB, Engesæter IØ, Fenstad AM, Havelin LI, Kärrholm J, Garellick G, Pedersen A, Overgaard S (2012) Low revision rate after total hip arthroplasty in patients with pediatric hip diseases. *Acta Orthop* 83 (5): 436–441. doi: 10.3109/17453674.2012.736171.

Ernst C (1999) Krankenhaus-Controlling und monetäre Anreize für leitende Ärzte. Dt. Univ.-Verl: Wiesbaden, Frankfurt (Main).

Fedderwitz J (2012) Parodontitistherapie: Zahnerhaltung als Königsweg. *Zahnarztl Mitt* 102 (3): 32–36.

Fetter RB (1999) DRGs and Quality Management in Hospitals. *In: Goldfield N [Hg.] Physician profiling and risk adjustment.* 2. Aufl. Gaithersburg, Md: Aspen: 101–116.

Flämig G, Schumann C, Roempp A, Marre R (2008) Möglichkeiten und Grenzen der risikoadjustierten Bewertung der Letalität bei der ambulant erworbenen Pneumonie mithilfe der § 21-Daten des Krankenhaus-Entgeltgesetzes (KHEntG). *Pneumologie* 62 (09): 562–568. doi: 10.1055/s-2008-1038186.

Fleßa S (2007) Grundzüge der Krankenhausbetriebslehre. Oldenbourg Verlag: München.

Flick U (2011) Qualitative Sozialforschung. Eine Einführung. Rowohlt-Taschenbuch-Verl: Reinbek bei Hamburg.

Flintrop J (2012) Mindestmengen in der Gesetzlichen Krankenversicherung: Warten auf das höchste Gericht. *Dtsch Arztebl* 109 (5): 11.

Fonarow GC, Abraham WT, Albert NM, Stough WG, Gheorghiade M, Greenberg BH et al. (2007) Association Between Performance Measures and Clinical Outcomes for Patients Hospitalized With Heart Failure. *J Am Med Assoc (JAMA)* 297 (1): 61–70. doi: 10.1001/jama.297.1.61.

Frey BS (1992) Tertium Datur: Pricing, Regulating and Intrinsic Motivation. *Kyklos* 45 (2): 161–184. doi: 10.1111/j.1467-6435.1992.tb02112.x.

Frey BS (2002) Managing Motivation. Wie Sie die neue Motivationsforschung für Ihr Unternehmen nutzen können. 2. Aufl. Gabler: Wiesbaden.

Frey BS (2005) Yes, Managers Should Be Paid Like Bureaucrats. *Journal of Management Inquiry* 14 (1): 96–111. doi: 10.1177/1056492604273757.

Frey BS, Benz M (2004) Anreizsysteme, ökonomische und verhaltenswissenschaftliche Dimension. *In: Schreyögg G,von Werder A [Hg.] Handwörterbuch Unternehmensführung und Organisation.* 4. Aufl. Stuttgart: Schäffer-Poeschel Verlag: 22–28.

Frølich A, Talavera JA, Broadhead P, Dudley RA (2007) A behavioral model of clinician responses to incentives to improve quality. *Health Policy* 80 (1): 179–193. doi: 10.1016/j.healthpol.2006.03.001.

Fuchs F (2010) A-IQI Austrian Inpatient Quality Indicators. Qualitätsindikatoren der Niederösterreichischen Landeskliniken-Holding. Definitionshandbuch, Version 1.0, Datenjahr 2010, Stand: 31. 05. 2010. Univ. Verl. d. TU Berlin: Berlin.

Funck D (2002) Konzeptionelle Anforderungen an Integrierte Managementsysteme. *In: Schwendt S,Funck D [Hg.] Integrierte Managementsysteme. Konzepte, Werkzeuge, Erfahrungen*. Heidelberg: Physica-Verlag: 25–44.

Gabler (1988) Wirtschaftslexikon. 12. Aufl. Gabler: Wiesbaden.

Garvin DA (1984) What Does 'Product Quality' Really Mean? *MIT Sloan Management Review* 26 (no. 1): 25–45.

Geiger W, Kotte W (2008) Handbuch Qualität. Grundlagen und Elemente des Qualitätsmanagements: Systeme - Perspektiven. 5., vollstandig uberarbeitete und erw. Friedr. Vieweg: Wiesbaden.

Geissler A, Quentin W, Scheller-Kreinsen D, Busse R (2011a) Introduction to DRGs in Europe. Common objectives across different hospital systems. *In: Busse R, Geissler A, Quentin W,Wiley M [Hg.] Diagnosis-related groups in Europe. Moving towards transparency, efficiency and quality in hospitals*. Open University Press: Maidenhead. 9–21.

Geissler A, Scheller-Kreinsen D, Quentin W, Busse R (2011b) Germany: Understanding G-DRGs. *In: Busse R, Geissler A, Quentin W,Wiley M [Hg.] Diagnosis-related groups in Europe. Moving towards transparency, efficiency and quality in hospitals*. Open University Press: Maidenhead: 243–272.

Gemeinsamer Bundesausschuss (23.06.2011) Regelungen zum Qualitätsbericht der Krankenhäuser, Qb-R, vom 19.03.2009. Online verfügbar unter http://www.g-ba.de/downloads/62-492-539/Qb-R_2011-05-19.pdf, zuletzt geprüft am 02.02.2013.

Gemeinsamer Bundesausschuss (01.01.2012) Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern. QSKH-RL, vom 15.08.2006. Online verfügbar unter http://www.g-ba.de/downloads/62-492-583/QSKH-RL_2011-10-20.pdf, zuletzt geprüft am 02.02.2013.

Geraedts M (2006) Qualitätsberichte deutscher Krankenhäuser und Qualitätsvergleiche von Einrichtungen des Gesundheitswesens aus Versichertensicht. *In: Böcken J, Braun B, Amhof R,Schnee M [Hg.] Gesundheitsmonitor 2006. Gesundheitsversorgung und Gestaltungsoptionen aus der Perspektive von Bevölkerung und Ärzten.* 1. Aufl.: Verlag Bertelsmann Stiftung. Gütersloh: 154ff.

Geraedts M (2009) Einsatz von Qualitätsindikatoren. *In: ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) [Hg.] Qualitätsindikatoren. Manual für Autoren*. Berlin: Verlag Make a book (ÄZQ Schriftenreihe, 36): 5–7.

Geraedts M, Auras S, Hermeling P, Cruppé W de (2009) Public Reporting – Formen und Effekte öffentlicher Qualitätsberichterstattung. *Dtsch Med Wochenschr* 134 (S 06): S232-S233.

Geraedts M, de Cruppe W (2006) Kontroverse Studienergebnisse zur Mindestmengenproblematik. *Z Arztl Fortbild Qualitatssich* 100 (2): 87–91.

Geraedts M, Selbmann H, Ollenschläger G (2002) Beurteilung der methodischen. Qualität klinischer Messgrößen. *Z Arztl Fortbild Qualitatssich* 96 (2): 91–96.

Glaser BG, Strauss AL (1998) Grounded theory. Strategien qualitativer Forschung. Verlag Hans Huber: Bern.

Golden BR (2006) The cost of pay-for-performance in healthcare: an alternative view. *Healthc Pap* 6 (4): 39-46; discussion 72-4.

Gorenoi V, Dinstios C, Hagen A (2005) Senkung der Restenoserate durch Einsatz beschichteter Stents bei koronarer Herzkrankheit. Schriftenreihe des Health Technology Assistants. Köln.

Gosden T, Forland F, Kristiansen IS, Sutton M, Leese B, Giuffrida A et al. (2001) Impact of payment method on behaviour of primary care physicians: a systematic review. *J Health Serv Res Policy* 6 (1): 44–55. doi: 10.1258/1355819011927198.

Greiner W (2007) Die Berechnung von Kosten und Nutzen. *In: Schöffski O [Hg.] Gesundheitsökonomische Evaluationen*. Berlin, Heidelberg, New York: Springer: 49–62.

Grenier-Sennelier C, Corriol C, Daucourt V, Michel P, Minvielle E (2005) Developing quality indicators in hospitals: the COMPAQH project. *Rev Epidemiol Sante Publique* 53 Spec No 1: 1S22-30.

Grewe A (2000) Implementierung neuer Anreizsysteme. Grundlagen, Konzept und Gestaltungsempfehlungen. Hampp: München.

Grewe A (2006) Implementierung neuer Anreizsysteme. Grundlagen, Konzept und Gestaltungsempfehlungen. 3., aktualisierte Aufl. Hampp: München.

Griebau D (2011) Der Behandlungsvertrag. § 10. *In: Ratzel R,Luxenburger B [Hg.] Handbuch Medizinrecht.* 2. Aufl. Bonn: Dt. Anwaltverl.: 607–642.

Grimshaw JM, Russell IT (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 342 (8883): 1317–1322.

Grothaus F (2009) Entwicklung der integrierten Versorgung in der Bundesrepublik Deutschland 2004 - 2008. Bericht gemäß $140d SGB V auf der Grundlage der Meldungen von Verträgen zur integrierten Versorgung. Bundesgeschäftsstelle Qualitätssicherung gGmbH: Düsseldorf.

Gucanin A (2003) Total Quality Management mit dem EFQM-Modell. Verbesserungspotenziale erkennen und für den Unternehmenserfolg nutzen. uni-edition: Berlin.

Häder M (2002) Delphi-Befragungen. Ein Arbeitsbuch. 1. Aufl. Westdeutscher Verlag: Wiesbaden.

Haeske-Seeberg H (2001) Handbuch Qualitätsmanagement im Krankenhaus. Strategien, Analysen, Konzepte. Kohlhammer: Stuttgart Berlin Köln.

Hannan EL, Racz MJ, Jollis JG, Peterson ED (1997) Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res* 31 (6): 659–678.

Hartweg H (2007) Die Entwicklung der integrierten Versorgung in Deutschland. Eine Analyse der Selektivvertragsmöglichkeiten gemäß der Regelungen der 140 a - d SGB V unter Aspekten der neuen Institutionenökonomie und der Evolutionsökonomie. Lit: Berlin.

Haubrock M (2002) Managementmethoden. *In: Haubrock M,Laux H [Hg.] Betriebswirtschaft und Management im Krankenhaus.* 3. Aufl. Bern: Verlag Hans Huber: 120–175.

Haubrock M, Laux H [Hg.] (2002) Betriebswirtschaft und Management im Krankenhaus. 3. Aufl. Verlag Hans Huber: Bern.

Hehrlein C, Brachmann J, Bode C, Kübler W (1998) Kathetergestützte Verfahren zur Verringerung der Restenoserate nach Koronarangioplastie: Ein Überblick zur Strahlentherapie. *Z Kardiol* 87 (2): s166-s170.

Helfferich C (2005) Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews. 2. Aufl. VS Verl. für Sozialwiss: Wiesbaden.

Helios Kliniken Gruppe [Hg.] (2005) Medizinischer Jahresbericht 2005: Fulda.

Heller G (2008) Risikoadjustierung von Ergebnisse - Vorgehensweise und Gefahren. Qualitätsmessung und Qualitätsmanagement mit Routinedaten. Wissenschaftliches Institut der AOK (WIdO): Potsdam, 16.09.2008. Online verfügbar unter http://www.qmr-kongress.de/download/heller.pdf, zuletzt geprüft am 16.01.2013.

Heller G (2010) Lebensdauer von Hüftgelenksimplantaten in Deutschland. *In: Klauber J, Geraedts M, Friedrich J,Wasem J [Hg.] Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb*. Stuttgart: Schattauer GmbH: 49–60.

Heller G, Günster C, Swart E (2008) Perspektiven der Qualitätssicherung mit Routinedaten (QSR). *In: Klauber J, Robra B,Schellschmidt H [Hg.] Krankenhaus-Report 2007:* Schattauer GmbH: 171–184.

Heller G, Schnell R (2007) Hospital Mortality Risk Adjustment Using Claims Data. *J Am Med Assoc (JAMA)* 297 (18): 1983. doi: 10.1001/jama.297.18.1984.

Henke K (2010) Erstellung eines Satellitenkontos für die Gesundheitswirtschaft in Deutschland. Forschungsprojekt im Auftrag des Bundesministeriums für Wirtschaft und Technologie. 1. Aufl. Nomos: Baden-Baden.

Henke K (2011) Volkswirtschaftliche Bedeutung der Gesundheitswirtschaft. Innovationen, Branchenverflechtung, Arbeitsmarkt ; auf der Grundlage eines Forschungsprojekts im Auftrag des Bundesministeriums für Wirtschaft und Technologie. 1. Aufl. Nomos: Baden-Baden.

Henke K, Göpffarth D (2010) Das Krankenhaus im System der Gesundheitsversorgung. *In: Hentze J,Kehres E [Hg.] Krankenhaus-Controlling. Konzepte, Methoden und Erfahrungen aus der Krankenhauspraxis.* 4. Aufl. Stuttgart: Kohlhammer: 35–52.

Henke K, Mackenthun B, Schreyögg J (2003) Gesundheit als Wirtschaftsfaktor. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 46 (8): 642–647. doi: 10.1007/s00103-003-0650-1.

Henke K, Martin K, Behrens C (1997) Direkte und indirekte Kosten von Krankheiten in der Bundesrepublik Deutschland 1980 und 1990. *J Public Health* 5 (2): 123–145.

Henke K, Rich RF, Steinbach A, Borchardt K (2004) Auf dem Wege zu einer integrierten Versorgung. Neue sozialrechtliche Rahmenbedingungen unter Berücksichtigung der Erfahrungen aus den USA und am Beispiel Berlins. Techn. Univ: Berlin.

Hertle D, Veit C (2012) Pay-for-Performance in Deutschland: Von der Option zur Umsetzung. *Monitor Versorgungsforschung* 5 (06): 40–46.

Herzberg F (1968) One more time: how do you motivate employees? *Harv Bus Rev* 81 (1): 87–96.

Heuss T (1951) Was ist Qualität? Zur Geschichte und zur Aufgabe des Deutschen Werkbundes. Verlag Rainer Wunderlich: Tübingen/Stuttgart.

HHS - CMS (08.05.2011) Medicare Program; Hospital Inpatient Value-Based Purchasing Program. 42 CFR Parts 422 and 480 76 (88): 26489–26546. Online verfügbar unter http://www.gpo.gov/fdsys/pkg/FR-2011-05-06/pdf/2011-10568.pdf, zuletzt geprüft am 13.01.2013.

Hildebrand R (2005) Qualitätsberichterstattung in Deutschland heute. *In: Klauber J, Robra B, Schnellschmidt H [Hg.] Krankenhausreport 2004. Schwerpunkt: Qualitätstransparenz - Instrumente und Konsequenzen.* 1. Aufl. Stuttgart [u.a.]: Schattauer GmbH: 27–47.

Hill CA, Winfrey KL, Rudolph BA (1997) "Best hospitals": a description of the methodology for the Index of Hospital Quality. *Inquiry* 34 (1): 80–90.

Himme A (2006) Gütekriterien der Messung: Reliabilität, Validität und Generalisierbarkeit. *In: Albers S [Hg.] Methodik der empirischen Forschung:* Deutscher Universitätsvlg: 383–400.

Himme A (2009) Gütekriterien der Messung: Reliabilität, Validität und Generalisierbarkeit. *In: Albers S, Klapper D, Konradt U, Walter A,Wolf J [Hg.] Methodik der empirischen Forschung:* Gabler Verlag: 485–500.

Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S (1995) Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. *J Clin Oncol.* 13 (5): 1249–1254.

Hoefert H [Hg.] (1997) Führung und Management im Krankenhaus. Verlag für Angewandte Psychologie: Göttingen.

Holcomb J (2000) The role of administrative data in measurement and reporting of quality of hospital care. *Tex Med* 96 (10): 48–52.

Horvath M (2008) Anforderungen an Anreizsysteme. Universität Wien. Online verfügbar unter http://othes.univie.ac.at/896/1/2008-07-15_0100597.pdf, zuletzt geprüft am 25.01.2013.

Hügler S (2010) Arztbewertungen im Internet: Kann man sie ignorieren? *Dtsch Med Wochenschr* 135 (27): 25.

Humburg S (2001) Mitarbeitermotivation im Krankenhaus. Das Organisationsklima und seine Bedeutung für das Qualitätsmanagement. Dt. Inst.-Verl: Köln, Bonn.

Hungenberg H (2006) Anreizsysteme für Führungskräfte — Theoretische Grundlagen und praktische Ausgestaltungsmöglichkeiten. *In: Hahn D,Taylor B [Hg.] Strategische Unternehmensplanung - strategische Unternehmensführung. Stand und Entwicklungstendenzen.* 9. überarbeitete Aufl. Berlin, Heidelberg, New York: Springer Verlag: 353–364.

Hurtado MP, Swift EK, Corrigan J (2001) Envisioning the national health care quality report. National Academy Press: Washington, D.C.

Hussey PS, Mattke S, Morse L, Ridgely MS (2007) Evaluation of the Use of AHRQ and Other Quality Indicators. Final Contract Report. Hg. v. RAND Health.

Iezzoni LI (1997) The risks of risk adjustment. *J Am Med Assoc (JAMA)* 278 (19): 1600–1607. doi: 10.1001/jama.278.19.1600.

Iezzoni LI (2003) Risk adjustment for measuring health care outcomes. 3. Aufl. Health Administration Press: Chicago.

Iezzoni LI, Foley SM, Heeren T, Daley J, Duncan CC, Fisher ES, Hughes J (1992) A method for screening the quality of hospital care using administrative data: preliminary validation results. *QRB Qual Rev Bull* 18 (11): 361–371.

Imai M (1993) Kaizen. Der Schlüssel zum Erfolg der Japaner im Wettbewerb. 3. Aufl. Ullstein: Berlin [u.a.].

InEK (2011) German Diagnosis Related Groups, Version 2012. Definitionshandbuch - Band 5. Institut für das Entgeltsystem im Krankenhaus GmbH: Siegburg.

Institute of Medicine, (2001) Crossing the quality chasm. A new health system for the 21st century. National Academy Press: Washington, D.C.

Institute of Medicine, (2007) Rewarding provider performance. Aligning incentives in Medicare. National Academy Press: Washington, DC.

Jäckel WM (2009a) Anforderungen an Qualitätsindikatoren. *In: ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) [Hg.] Qualitätsindikatoren. Manual für Autoren*. Berlin: Verlag Make a book (ÄZQ Schriftenreihe, 36): 8–11.

Jäckel WM (2009b) Definition und Typen von Qualitätsindikatoren. *In: ÄZQ (Ärztliches Zentrum für Qualität in der Medizin) [Hg.] Qualitätsindikatoren. Manual für Autoren,* 2-4. Berlin: Verlag Make a book (ÄZQ Schriftenreihe, 36).

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 17 (1): 1–12.

Jaeckel R (2009) Pay for Performance. Ein strategischer Ansatz für mehr Versorgungsqualität im Gesundheitswesen*. Gesellschaftspolitische Kommentare* 9 (09): 12–15.

Jakob J (2010) Qualität der BQS-Dokumentation am Beispiel intra- und postoperativer Komplikationen nach Cholezystektomien. Deutsche Gesellschaft für Chirurgie: Berlin.

JCAHO (2006) Attributes of Core Performance Measures and Associated Evaluation Criteria. Hg. v. Joint Commission on Accreditation of Healthcare Organizations, Illinois.

JCAHO (2010) Improving America's Hospitals. The Joint Commission's Annual Report on Quality and Safety: Oakbrook Terrace, Illinois.

Jena AB, Seabury S, Lakdawalla D, Chandra A (2011) Malpractice risk according to physician specialty*. N Engl J Med* 365 (7): 629–636.

Jha AK, Joynt KE, Orav EJ, Epstein AM (2012) The long-term effect of premier pay for performance on patient outcomes*. N Engl J Med* 366 (17): 1606–1615.

Johantgen M, Elixhauser A, Bali JK, Goldfarb M, Harris DR (1998) Quality indicators using hospital discharge data: state and national applications*. Jt Comm J Qual Improv* 24 (2): 88–105.

Juran JM (1991) Handbuch der Qualitätsplanung. 3., durchges. Aufl. Verl. Moderne Industrie: Landsberg/Lech.

Kaiser S (2008) Arzthaftungsrecht. §12. *In: Luxenburger B,Ratzel R [Hg.] Handbuch Medizinrecht:* Deutscher Anwaltverlag Gm: 601–701.

Kamke W, Breitkreuz B, Jereczek M, Dovifat C, Völler H (2002) Nachweis und Bedeutung von permanenten und transistorischen Schrittmacherfunktionsstörungen bei 180 Patienten mit Ein- und Zwei-Kammer-Aggregaten*. Herzschrittmacherther Elektrophysiol* 13 (4): 233–241.

Kassenärztliche Vereinigung Schleswig-Holstein Anlage 3. zu dem Vertrag zur Durchführung des Disease-Management-Programms nach § 137 f SGB V Asthma bronchiale auf der Grundlage des § 83 SGB V. Online verfügbar unter http://www.kvsh.de/KVSH/db2b/upload/downloads/ Anlage_3_sq_Krankenhaus_ DMP_asthma.pdf, zuletzt geprüft am 16.01.2013.

Kelsey T (2001) Fostering quality healthcare*. OECD Observer,* No. 229, November 2001, London.

Kerr J (1982) Assigning Managers on the Basis of the Life Cycle*. J Bus Strategy* 2 (4): 58–65. doi: 10.1108/eb038945.

Kerr S (1997) Ultimate rewards. What really motivates people to achieve. Harvard Business School Press: Boston, MA.

Khuri SF, Henderson WG (2005) The case against volume as a measure of quality of surgical care*. World J Surg* 29 (10): 1222–1229. doi: 10.1007/s00268-005-7987-6.

King N, Horrocks C (2010) Interviewing in qualitative research. SAGE: Los Angeles.

Klauber J, Geraedts M, Friedrich J, Wasem J [Hg.] (2010) Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb. Schattauer GmbH: Stuttgart.

Klemperer D (1995) Qualität in der Medizin. Der patientenzentrierte Qualitätsbegriff und seine Implikationen. *In: Damkowski W [Hg.] Patienten im Gesundheitssystem. Patientenunterstützung und -beratung ; Notwendigkeit Konzepte und Erfahrungen,* Mai 1995. Augsburg: Maro-Verl (Beiträge zur Sozialpolitik-Forschung, 10): 189–216.

Klusen N, Meusch A, Piesker J (2011) Pay for Performance. Weder Königs- noch Holzweg. *In: Qualitätsmanagement im Gesundheitswesen*. Baden-Baden: Nomos-Verl.-Ges: 89–116.

Knorr S, Weyermann M (2011) HSMR – Standardisierte Krankenhaus-Mortalitätsraten. Geeignet zum Qualitätsvergleich von Krankenhäusern? Routinedaten in der Versorgungsforschung, 17.03.2011.

Koek HL, Kardaun JW, Gevers E, Bruin A de, Reitsma JB, Grobbee DE, Bots ML (2007) Acute myocardial infarction incidence and hospital mortality: routinely collected national data versus linkage of national registers. *Eur J Epidemiol* 22 (11): 755–762. doi: 10.1007/s10654-007-9174-6.

Köhler A (2010) „Qualitätsorientierte Vergütung in der ambulanten Versorgung“. Statement von Dr. Andreas Köhler am 10.02.2010. Kassenärztliche Bundesvereinigung, vdek, Berlin.

Kohn LT, Corrigan JM, Donnaldson MS [Hg.] (1999) To err is human: Building a. Safer Health System. Institute of Medicine: Washington DC.

Komajda M (2005) Adherence to guidelines is a predictor of outcome in chronic heart failure: the MAHLER survey. *Eur Heart J* 26 (16): 1653–1659. doi: 10.1093/eurheartj/ehi251.

Kopp I (2011) Von Leitlinien zur Qualitätssicherung. *Bundesgesundheitsbl* 54 (2): 160–165. doi: 10.1007/s00103-010-1207-8.

Kostuj T, Smektala R (2010) Qualitätssicherung mithilfe von Routinedaten. *Unfallchirurg* 113 (12): 1047–1052. doi: 10.1007/s00113-010-1875-8.

Kovacs A (2002) Oberflächenmodifizierung von Dilatationskathetern zur Senkung der Restenoserate bei der perkutanen transluminalen Koronarangioplastie. Dissertation. Heidelberg.

Krahwinkel W, Rink O, Liebetrau M, Günther M, Schuler E, Kuhlen R (2011) 10 Jahre Peer Review - Verbesserung der medizinischen Behandlung durch Qualitätsindikatoren aus Routinedaten. *Dtsch Med Wochenschr* 136 (41): 2083–2088. doi: 10.1055/s-0031-1272577.

Kräkel M (2004) Prinzipal-Agenten-Ansatz. *In: Schreyögg G,von Werder A [Hg.] Handwörterbuch Unternehmensführung und Organisation.* 4. Aufl. Stuttgart: Schäffer-Poeschel Verlag: 1174–1181.

Kronsteiner O (2010) Mitarbeitermotivation versus Qualitätsmanagement. Widersprüche und Zusammenhänge am Beispiel der Raum- und Luftfahrtindustrie. 1. Aufl. GRIN Verlag: München.

Krüger-Brand HE, Osterloh F (2011) IT soll Umsetzung erleichtern. *Dtsch Arztebl* 108 (30): 1608.

Krumholz HM, Normand ST, Spertus JA, Shahian DM, Bradley EH (2007) Measuring Performance For Treating Heart Attacks And Heart Failure: The Case For Outcomes Measurement. *Health Aff* 26 (1): 75–85. doi: 10.1377/hlthaff.26.1.75.

Kuhlen R, Rink O, Zacher J [Hg.] (2011) Jahrbuch Qualitätsmedizin 2011. Initiative Qualitätsmedizin (IQM). MWV (Medizinisch Wissenschaftliche Verlagsgesellschaft): Berlin.

Kuhmerker K, Hartman T (2007) Pay-for-Performance in State Medicaid Programs. A Survey of State Medicaid Directors and Programs. IPRO: Lake Success, N.Y.

Kwaan MR (2006) Incidence, Patterns, and Prevention of Wrong-Site Surgery. *Arch Surg* 141 (4): 353.

Lamnek S (1995) Qualitative Sozialforschung. 3., korrigierte Aufl. Beltz: Weinheim.

Lamnek S (2005) Qualitative Sozialforschung. Lehrbuch. 4. Aufl. Beltz PVU: Weinheim.

Langer A (2010) Transkribieren. Grundlagen und Regeln. *In: Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim [u.a.]: Juventa-Verl: 515–526.

Laux H (2006) Unternehmensrechnung, Anreiz und Kontrolle. Die Messung, Zurechnung und Steuerung des Erfolges als Grundprobleme der Betriebswirtschaftslehre ; mit 37 Tabellen. 3. Aufl. Springer: Berlin, Heidelberg, New York.

Le Grand J (2003) Motivation, agency, and public policy. Of knights and knaves pawns and queens. Oxford Univ. Press: Oxford.

Leber WD (2009) Qualitätsmessung und qualitätsorientierte Vergütung aus Sicht des GKV-Spitzenverbandes. *Dtsch Med Wochenschr* 134 (S 08): S316. doi: 10.1055/s-0029-1242672.

Leber WD, Malzahn J, Wolff J (2007) Elektiv wird selektiv: Ein Vorschlag für einen nach Krankenhausleistungen differenzierenden Ordnungsrahmen ab 2009. Award Krankenhausfinanzierung 2009. Hg. v. B. Braun Stiftung, Melsungen.

Lehmann M (2006) Möglichkeiten und Grenzen der Ausgestaltung von Anreizsystemen für freie Mitarbeiter. Univ. Duisburg-Essen Fak. 3 Wirtschaftswiss: Duisburg.

Lessing C (2009) Risikomanagement und Patientensicherheit. *Unfallchirurg* 112 (6): 610–612. doi: 10.1007/s00113-009-1643-9.

Lester H, Schmittdiel J, Selby J, Fireman B, Campbell S, Lee J et al. (2010) The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ* 340 (may11 1): c1898. doi: 10.1136/bmj.c1898.

Lilford R, Pronovost P (2010) Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 340 (apr19 2): c2016. doi: 10.1136/bmj.c2016.

Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW (2007) Public Reporting and Pay for Performance. *N Engl J Med* 356;5: 486–496. doi: 10.1056/NEJMsa064964

Lindert K (2001) Anreizsysteme und Unternehmenssteuerung. Eine kritische Reflexion zur Funktion, Wirksamkeit und Effizienz von Anreizsystemen. Hampp: München [u.a.].

Litzkow P (2008) Anreizkompatible Entlohnungsformen bei deutschen Klinikärzten- Möglichkeiten und Grenzen. GRIN Verlag.

Locke RG, Srinivasan M (2008) Attitudes toward pay-for-performance initiatives among primary care osteopathic physicians in small group practices. *J Am Osteopath Assoc* 108 (1): 21–24.

Lohr KN (1990) Medicare. A strategy for quality assurance. [1st ed. 1990, repr.]. National Academy Press: Washington, DC.

Lowrie EG, Teng M, Lacson E, Lew N, Lazarus JM, Owen WF (2001) Association between prevalent care process measures and facility-specific mortality rates. *Kidney Int* 60 (5): 1917–1929. doi: 10.1046/j.1523-1755.2001.00029.x.

Luft HS, Bunker JP, Enthoven AC (1979) Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med* 301 (25): 1364–1369.

Lüngen M, Rath T (2011) Analyse und Evaluierung des QUALIFY Instruments zur Bewertung von Qualitätsindikatoren anhand Analyse und Evaluierung des QUALIFY Instruments zur Bewertung von Qualitätsindikatoren anhand eines strukturierten qualitativen Interviews. *Z Evid Fortbild Qual Gesundhwes* (105): 38–43.

Luxenburger B, Ratzel R [Hg.] (2008) Handbuch Medizinrecht. Deutscher Anwaltverlag, Bonn.

Maass C, Schleiz W, Weyermann M, Drösler S (2011) Krankenhaus-Routinedaten zur externen Qualitätssicherung? *Dtsch Med Wochenschr* 136 (09): 409–414. doi: 10.1055/s-0031-1274523.

Mainz J (2003) Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 15 (6): 523–530. doi: 10.1093/intqhc/mzg081.

Maiwald S (2010) Qualitätsmanagement bei Low Cost Airlines. GRIN Verlag: München.

Malchau H, Herberts P, Eisler T, Garellick G, Söderman P (2002) The Swedish Total Hip Replacement Register. *J Bone Joint Surg Am* 84-A Suppl 2: 2–20.

Malzahn J (2010) Pay for Performance. Herausforderungen und Lösungen. AOK-Bundesverband: Düsseldorf, Vortrag 17.04.2010.

Malzahn J, Heyde K, Fahlenbrach C (2010) Pay for Performance - Rahmenbedingungen für ein konkretes Modell im Bereich der Endoprothetik. *In: Klauber J, Geraedts M, Friedrich J,Wasem J [Hg.] Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb*. Stuttgart: Schattauer GmbH: 131–146.

Mansky T (2008) Neue Methoden der Qualitätsmessung und des Qualitätsmanagements. *In: Kurth B [Hg.] Monitoring der gesundheitlichen Versorgung in Deutschland. Konzepte, Anforderungen, Datenquellen*. Unter Mitarbeit von C. Fuchs, B.-M Kurth und P.C Scriba. Köln: Deutscher Ärzte-Verlag (Report Versorgungsforschung): 149–170.

Mansky T (2009) Qualitätsmessung und Qualitätsmanagement mit Routinedaten. [2. Kongress Qualitätsmessung und Qualitätsmanagement mit Routinedaten]. Thieme Georg Verlag: Stuttgart.

Mansky T (2010a) Qualitätsorientierte Vergütung: Was wäre machbar? Neue Entwicklungen bei der Messung der Ergebnisqualität eröffnen neue Optionen in der qualitätsorientierten Vergütung. *Chirurg* (BDC 1): 6–9.

Mansky T (2010b) Stand und Perspektive der stationären Qualitätssicherung in Deutschland. *In: Klauber J, Geraedts M, Friedrich J,Wasem J [Hg.] Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb*. Stuttgart: Schattauer GmbH: 19–37.

Mansky T (2013) Was erwarten die potenziellen Patienten vom Krankenhaus? *In: Böcken J, Braun B,Repschläger U [Hg.] Gesundheitsmonitor 2012. Bürgerorientierung im Gesundheitswesen - Kooperationsprojekt der Bertelsmann Stiftung und der BARMER GEK*. Gütersloh: Verlag Bertelsmann Stiftung: 136–159.

Mansky T, Nimptsch U (2010a) Kennzahlengestütztes ergebnisorientiertes Qualitätsmanagement. *In: Debatin JF, Ekkernkamp A, Schulte B,Baehr M [Hg.] Krankenhausmanagement. Strategien, Konzepte, Methoden*. Berlin: Med. Wiss. Verl.-Ges.: 419–425.

Mansky T, Nimptsch U (2010b) Notwendigkeit eines ungehinderten Zugangs zu sozial- und krankheitsbezogenen Versichertendaten für die Bundesärztekammer und andere ärztliche Körperschaften sowie wissenschaftliche Fachgesellschaften zur Optimierung der ärztlichen Versorgung. Expertise im Rahmen der Förderinitiative zur Versorgungsforschung der Bundesärztekammer. TU Berlin.

Mansky T, Nimptsch U, Winklmair C, Vogel K, Hellerhoff F (2011) G-IQI - German Inpatient Quality Indicators Version 3.1. Definitionshandbuch der Version 3.1 für das Datenjahr 2010. Stand: 18.03.2011: Berlin.

Mansky T, Seeger DG (2011) Gestaltung einer qualitätsorientierten Vergütung aus betriebswirtschaftlicher Sicht. *In: Repschläger U [Hg.] BARMER Gesundheitswesen aktuell 2011. [Beiträge und Analysen]*. Barmer GEK. Wuppertal: 194–215.

Mant J (2001) Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 13 (6): 475–480.

Marotzki W (2003) Leitfadeninterview. *In: Bohnsack R, Marotzki W, Meuser M,Uwe [Hg.] Hauptbegriffe Qualitativer Sozialforschung. Ein Wörterbuch*. Opladen: Leske + Budrich: 114.

Marshall M, Harrison S (2005) It's about more than money: financial incentives and internal motivation. *Qual Saf Health Care* 14 (1): 4–5. doi: 10.1136/qshc.2004.013193.

Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory. Oxford Univ. Press: New York.

Maslow AH (1943) A Theory of Human Motivation. *Psychological Review* 50: 370–396.

Matiaske W, Weller I (2006) Kann weniger mehr sein? Theoretische Überlegungen und empirische Befunde zur These der Verdrängung intrinsischer Motivation durch externe Anreize. *In: Rapp R, Sedlmeier P,Zunker-Rapp G [Hg.] Perspectives on cognition. A Festschrift for Manfred Wettler*. Lengerich: Pabst: 113–132.

Mattke S, Epstein AM, Leatherman S (2006) The OECD Health Care Quality Indicators Project: history and background. *Int J Qual Health Care* 18 Suppl 1: 1–4. doi: 10.1093/intqhc/mzl019.

Maynard A (2012) The powers and pitfalls of payment for performance. *Health Econ* 21 (1): 3–12. doi: 10.1002/hec.1810.

Mayring P (2002) Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken. 5., überarb. und neu ausgestattete Aufl. Beltz: Weinheim.

Mayring P (2010a) Qualitative Inhaltsanalyse. *In: Mey G,Mruck K [Hg.] Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS Verl. für Sozialwiss: 601–613.

Mayring P (2010b) Qualitative Inhaltsanalyse. Grundlagen und Techniken. 11. Aufl. Beltz: Weinheim.

McAuliffe WE (1978) Studies of process-outcome correlations in medical care evaluations: a critique. *Med Care* 16 (11): 907–930.

McDonald L (1998) Florence Nightingale: Passionate Statistician. *Journal of Holistic Nursing* 16 (2): 267–277.

McDonald R, Harrison S, Checkland K, Campbell SM, Roland M (2007) Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study. *BMJ* 334 (7608): 1357. doi: 10.1136/bmj.39238.890810.BE.

McDonald R, White J, Marmor TR (2009) Paying for performance in primary medical care: learning about and learning from "success" and "failure" in England and California. *J Health Polit Policy Law* 34 (5): 747–776. doi: 10.1215/03616878-2009-024.

McGlynn EA (1998) Choosing and evaluating clinical performance measures. *Jt Comm J Qual Improv* 24 (9): 470–479.

McIntyre D, Rogers L, Heier EJ (2001) Overview, History, and Objectives of Performance Measurement. *Health Care Financing Review* 22 (3): 7-21.

MDK (2012) Der Medizinische Dienst in Zahlen: online verfügbar: http://www.mdk.de/314.htm (zuletzt geprüft 16.01.2013).

Mehrotra A, Damberg CL, Sorbero ME, Teleki SS (2008) Pay for Performance in the Hospital Setting: What Is the State of the Evidence? *Am J Med Qual* 24 (1): 19–28. doi: 10.1177/1062860608326634.

Mehrotra A, Damberg CL, Sorbero ME, Teleki SS, Mattke S (2010a) Pay for Performance im Krankenhaus: Erfahrungen in den USA. *In: Klauber J, Geraedts M, Friedrich J,Wasem J [Hg.] Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb*. Stuttgart: Schattauer GmbH: 117–130.

Mehrotra A, Sorbero ME, Damberg CL (2010b) Using the lessons of behavioral economics to design more effective pay-for-performance programs. *Am J Manag Care* 16 (7): 497–503.

Mehta RH, Liang L, Karve AM, Hernandez AF, Rumsfeld JS, Fonarow GC, Peterson ED (2008) Association of Patient Case-Mix Adjustment, Hospital Process Performance Rankings, and Eligibility for Financial Incentives. *J Am Med Assoc (JAMA)* 300 (16): 1897–1903. doi: 10.1001/jama.300.16.1897.

Menendez R (2005) Guidelines for the Treatment of Community-acquired Pneumonia: Predictors of Adherence and Outcome. *Am J Respir Crit Care Med* 172 (6): 757–762. doi: 10.1164/rccm.200411-1444OC.

Meterko M, Young GJ, White B, Bokhour BG, Burgess JF, Berlowitz D et al. (2006) Provider Attitudes toward Pay-for-Performance Programs: Development and Validation of a Measurement Instrument. *Health Serv Res* 41 (5): 1959–1978. doi: 10.1111/j.1475-6773.2006.00582.x.

Meuser M, Nagel U (2009) Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage. *In: Pickel S, Pickel G, Lauth H,Jahn D [Hg.] Methoden der vergleichenden Politik- und Sozialwissenschaft*. Wiesbaden: VS Verl. für Sozialwiss: 465–479.

Meyer-Hofmann B, Schillhorn K, Weng N, Eichwald D (2012) Teil E zum Gutachten zur Ermittlung des nationalen Sachstandes im Bereich "Pay-for-Performance im Gesundheitswesen". Rechtsgutachten. *In: Veit C, Hertle D, Bungar S, Trümner A, Ganske V, Meyer-Hofmann B [Hg.] Pay-for-Performance im Gesundheitswesen. Sachbestandsbericht zur Evidenz und Realisierung sowie Darlegung der Grundlagen für eine künftige Weiterentwicklung*. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit. Düsseldorf.

Millenson ML (2004) Pay for performance: the best worst choice. *Qual Saf Health Care* 13 (5): 323–324. doi: 10.1136/qshc.2004.011668.

Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Muhlbaier LH et al. (1994) Using Medicare claims for outcomes research. *Med Care* 32 (7 Suppl): JS38-51.

Moehring R, Anderson D (2011) "But my patients are different!": risk adjustment in 2012 and beyond. *Infect Control Hosp Epidemiol* 32 (10): 987–989. doi: 10.1086/662202.

Möller K, Dahm F, Bäune S (2008) Neue Versorgungsformen (MVZ, integrierte Versorgung). § 8. *In: Luxenburger B,Ratzel R [Hg.] Handbuch Medizinrecht:* Deutscher Anwaltverlag Gm: 479–834.

Müller M (1999) Zielorientiertes Management - Management by Objectives. *In: Eckardstein D von, Mayrhofer W,Kasper H [Hg.] Management. Theorien - Führung - Veränderung*. Stuttgart: Schäffer-Poeschel Verlag: 271–283.

Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JR (2006) Cost-Effectiveness of Hospital Pay-for-Performance Incentives. *Med Care Res Rev* 63 (1 suppl): 49S. doi: 10.1177/1077558705283629.

National Research Council (1997) Preparing for the 21st Century: Focusing on Quality in a Changing Health Care System. Hg. v. The National Academies Press. Washington DC.

Newhouse JP (1998) Risk adjustment: where are we now? *Inquiry* 35 (2): 122–131.

Newhouse JP, Buntin MB, Chapman JD (1997) Risk adjustment and Medicare: taking a closer look. *Health Aff* 16 (5): 26–43.

Nicholl J (2007) Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 61 (11): 1010–1013. doi: 10.1136/jech.2007.061747.

Nightingale F (1863) Notes on hospitales. 3. Aufl. Longman: London.

Nimptsch U, Mansky T (2012a) Krankheitsspezifische Versorgungsmerkmale in Deutschland: Analyse anhand der Bundesauswertung der German Inpatient Quality Indicators. *Dtsch Med Wochenschr* 137 (28-29): 1449–1457. doi: 10.1055/s-0032-1305086.

Nimptsch U, Mansky T (2012b) Trends in der akutstationären Schlaganfallversorgung in Deutschland. Eine Beobachtungsstudie mit Krankenhausabrechnungsdaten von 2005-2010. *Dtsch Arztebl* 109 (51-52): 885–892.

Norbert R, Martin H (2008) Klinikführer Rheinland: Qualität im Vergleich. *Das Krankenhaus* 9: 900–906.

OECD (2011) Gesundheitsdaten 2010. Deutschland im Vergleich, Berlin.

Ollenschläger G, Thomeczek C (2002) Qualitätsmanagement im Gesundheitswesen: Fehlerprävention und Umgang mit Fehlern in der Medizin. *Med Klin* 97 (9): 564–570. doi: 10.1007/s00063-002-1197-x.

Ortlieb P (1993) Qualitätsmanagement und betriebliches Anreizsystem. Centaurus: Pfaffenweiler.

Or Z, Häkkinen U (2011) DRG and quality: For better or worse? *In: Busse R, Geissler A, Quentin W,Wiley M [Hg.] Diagnosis-related groups in Europe. Moving towards transparency, efficiency and quality in hospitals*. Maidenhead: Open University Press: 115–129.

Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Raiha P, Karja-Koskenkari P et al. (2005) The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil* 12 (2): 132–137.

Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S (2006) Does pay-for-performance improve the quality of health care? *Ann Intern Med* 145 (4): 265–272.

Petersen T (1989) Optimale Anreizsysteme. Betriebswirtschaftliche Implikationen der Prinzipal-Agenten-Theorie. Gabler: Wiesbaden.

Petersen U, Kaminski A, Jackson A (2007) Analyse aktueller Klinikführer: Was bringen sie dem Patienten? *Das Krankenhaus* 2007 (11): 1112-1117.

Peterson ED, Roe MT, Mulgund J, DeLong ER, Lytle BL, Brindis RG et al. (2006) Association between hospital process performance and outcomes among patients with acute coronary syndromes. *J Am Med Assoc (JAMA)* 295 (16): 1912–1920. doi: 10.1001/jama.295.16.1912.

Pfitzinger E (2011) Qualitätsmanagement nach DIN EN ISO 9000 ff. im Gesundheitswesen. Mit Excel-Tabellen zur Selbsteinschätzung. 1. Aufl. Beuth: Berlin.

Picot A, Dietl H, Franck E (2002) Organisation. Eine ökonomische Perspektive. 3. Aufl. Schäffer-Poeschel Verlag: Stuttgart.

Pigou AC (1924) The economics of welfare. 2. Aufl. MacMillan: London.

Pira A (2000) Umfassendes Qualitätsmanagement im Spital. Das EFQM-Modell als Basis. vdf Hochschulverl. an der ETH: Zürich.

Porter ME (2010) What is value in health care? *N Engl J Med* 363 (26): 2477–2481. doi: 10.1056/NEJMp1011024.

Porter ME, Guth C (2012) Redefining the German Healthcare System. 1. Aufl. Springer: Berlin.

Porter ME, Teisberg EO (2006) Redefining health care. Creating value-based competition on results. Harvard Business School Press: Boston, Mass.

Raspe H, Wissing F (2010) Versorgungsforschung in Deutschland. Stand - Perspektiven - Förderung ; Stellungnahme ; Standpunkte. Wiley-VCH: Weinheim.

Rau F (2002) DRG-Einführung in Deutschland: Ziele, Problemfelder und Perspektiven auf der Grundlage des Fallpauschalengesetzes aus Sicht des BMG. *Z Arztl Fortbild Qualitatssich* 96 (8): 498–504.

Reiter A, Fischer B, Kötting J, Geraedts M, Jäckel WM, Barlag H, Döbler K (2007) QUALIFY: Ein Instrument zur Bewertung von Qualitätsindikatoren. Hg. v. BQS gGmbh.

Reiter A, Fischer B, Kötting J, Geraedts M, Jäckel WM, Döbler K (2008) QUALIFY: Ein Instrument zur Bewertung von Qualiätsindikatoren. *Z Arztl Fortbild Qualitatssich* (101): 683–688.

Reiter KL, Nahra TA, Alexander JA, Wheeler JR (2006) Hospital responses to pay-for-performance incentives. *Health Serv Manage Res* 19 (2): 123–134. doi: 10.1258/095148406776829086.

Remus D Ph.D., N R, Fraser I Ph.D. (2004) Guidance for Using the AHRQ Quality Indicators for Hospital-level Public Reporting or Payment. U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality: Rockville, Maryland (AHRQ Publication No.04-0086-EF).

Riha O (2010) Kodifizierung ärztlicher Ethik. Vom Hippokratischen Eid zum Genfer Gelöbnis ; [Vortrag gehalten am 10.10.2008]. Hirzel: Stuttgart.

Rink O (2011) Erste Ergebnisse trägerübergreifender Reviews bei IQM 2010. *In: Kuhlen R, Rink O,Zacher J [Hg.] Jahrbuch Qualitätsmedizin 2011*. Berlin: MWV (Medizinisch Wissenschaftliche Verlagsgesellschaft): 89–110.

Robinson JC, Williams T, Yanagihara D (2009) Measurement Of And Reward For Efficiency In California's Pay-For-Performance Program. *Health Aff* 28 (5): 1438–1447. doi: 10.1377/hlthaff.28.5.1438.

Roland M (2004) Linking physicians' pay to the quality of care. a major experiment in the United Kingdom. *N Engl J Med* 351 (14): 1448–1454. doi: 10.1056/NEJMhpr041294

Roland M (2006) Pay-for-performance: too much of a good thing? A conversation with Martin Roland. Interview by Robert Galvin. *Health Aff* 25 (5): w412-9. doi: 10.1377/hlthaff.25.w412.

Rosenstiel L v. (1980) Grundlagen der Organisationspsychologie. Basiswissen und Anwendungshinweise. 2. Aufl. Pöschel: Stuttgart.

Rosenthal MB (2008) Beyond Pay for Performance — Emerging Models of Provider-Payment Reform. *In: New England Journal of Medicine,* September 18: 1197–1200.

Rosenthal MB, Dudley RA (2007) Pay-for-Performance: Will the Latest Payment Trend Improve Care? *J Am Med Assoc (JAMA)* 297 (7): 740–744. doi: 10.1001/jama.297.7.740.

Rosenthal MB, Frank RG, Li Z, Epstein AM (2005) Early Experience With Pay-for-Performance. From Concept to Practice. *J Am Med Assoc (JAMA)* 294 (14): 1788–1793. doi:10.1001/jama.294.14.1788.

Rosenthal MB, Landon BE, Howitt K, Ryu Song H, Epstein AM (2008) Climbing Up The Pay-For-Performance Learning Curve: Where Are The Early Adopters Now? *Health Aff* 26 (6): 1674–1682.

Rosenthal M, Frank R (2006) What Is the Empirical Basis for Paying for Quality in Health Care? *Med Care Res Rev* 63 (2): 135-157. doi: 10.1177/1077558705285291.

Rubin HR (2001) The advantages and disadvantages of process-based measures of health care quality. *Int J Qual Health Care* 13 (6): 469–474. doi: 10.1093/intqhc/13.6.469.

Ryan A, Blustein J (2012) Making the best of hospital pay for performance. *N Engl J Med* 366 (17): 1557–1559. doi: 10.1056/NEJMp1202563.

Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (2001) Bedarfsgerechtigkeit und Wirtschaftlichkeit III: Über-, Unter- und Fehlversorgung. Gutachten 2000/2001. Bundesanzeiger-Verl: Bonn.

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312 (7023): 71–72.

Sautter KM, Bokhour BG, White B, Young GJ, Burgess JF, Berlowitz D, Wheeler JR (2007) The early experience of a hospital-based pay-for-performance program. *J Healthc Manag* 52 (2): 95-107.

Sax A (2010) Qualität bedeutet Zuwendung. Interview mit Mathias Binswanger, Ökonomieprofessor und Autor eines Buches über sinnlose Wettbewerbe. *Care Management* 3 (6): 2–3.

Schäffer U, Pelster C (2007) Zur Relevanz des Controllability-Prinzips für die Unternehmenspraxis. *Z Control Manag* 51 (6): 422–436.

Schanz G [Hg.] (1991) Handbuch Anreizsysteme. Poeschel: Stuttgart.

Schär (2002) Ziele und Zielsystem. *In: Haubrock M,Laux H [Hg.] Betriebswirtschaft und Management im Krankenhaus.* 3. Aufl. Bern: Verlag Hans Huber: 97–103.

Scheppach M, Emmert M, Schöffski O (2011) Pay for Performance (P4P) im Gesundheitswesen. Leitfaden für eine erfolgreiche Einführung. HERZ, Health-Economics-Research-Zentrum: Burgdorf.

Schmeisser W (2008) Finanzorientierte Personalwirtschaft. Oldenbourg.

Schneeweiss S, Sangha O (2000) Weiterentwicklung von Krankenhausbetriebsvergleichen: Wie wichtig ist Risiko-Adjustierung für den Krankenhausvergleich? *In: Sieben G,Litsch M [Hg.] Krankenhausbetriebsvergleiche*. Berlin: Springer Verlag: 131–147.

Schnell R, Hill P, Esser E (2005) Methoden der empirischen Sozialforschung. Oldenbourg Verlag: München.

Schöffski O [Hg.] (2007a) Gesundheitsökonomische Evaluationen. Springer: Berlin, Heidelberg, New York.

Schöffski O (2007b) Grundformen gesundheitsökonomischer Evaluation. *In: Schöffski O [Hg.] Gesundheitsökonomische Evaluationen*. Berlin, Heidelberg, New York: Springer: 65–94.

Schopenhauer A (1851) Parerga und Paralipomena. Kleine philosophische Schriften. 2. Teil Aphorismen zur Lebensweisheit. Hayn: Berlin.

Schott H (2005) Arztrecht Kodex Hammurapi. *Dtsch Arztebl* 102 (30): A2101.

Schramm F (2006) Rechtliche Möglichkeiten und Grenzen der Integrierten Versorgung in der Kardiologie. *Clin Res Cardiol* 95 (S2): ii32. doi: 10.1007/s00392-006-1210-0.

Schrappe M, Gültekin N (2010) Pay for Performance (P4P): Auswirkungen auf die Qualität und Abgrenzung von der Einzelleistungsvergütung. *In: Klauber J, Geraedts M, Friedrich J,Wasem J [Hg.] Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb*. Stuttgart: Schattauer GmbH: 105–113.

Schrappe M, Gültekin N (2011) Pay for Performance (P4P). *Bundesgesundheitsbl* 54 (2): 166–170. doi: 10.1007/s00103-010-1203-z.

Schreyögg G, von Werder A [Hg.] (2004) Handwörterbuch Unternehmensführung und Organisation. 4. Aufl. Schäffer-Poeschel Verlag: Stuttgart.

Schreyögg J (2010) Kundenmanagement im Gesundheitswesen. Einführung und methodische Grundlagen. *In: Busse R, Schreyögg J,Tiemann O [Hg.] Management im Gesundheitswesen.* 2. Aufl. Heidelberg: Springer: 152–155.

Schuh R, Neumann D, Rauf R, Hofstaetter J, Boehler N, Labek G (2012) Revision rate of Birmingham Hip Resurfacing arthroplasty: comparison of published literature and arthroplasty register data. *Int Orthop* 36 (7): 1349–1354. doi: 10.1007/s00264-012-1502-0.

Schwartz A (1997) Informations- und Anreizprobleme im Krankenhaussektor. Eine institutionenökonomische Analyse. Dt. Univ.-Verl: Wiesbaden.

Schwendt S, Funck D [Hg.] (2002) Integrierte Managementsysteme. Konzepte, Werkzeuge, Erfahrungen. Physica-Verlag: Heidelberg.

Seeger DG (2011) Nur Marketing? Mindestanforderungen an „P4P"-Verfahren. *Dtsch Med Wochenschr* 136 (S 02): S62. doi: 10.1055/s-0031-1286091.

Seidel E (1991) Anreize zu ökologisch verpflichtetem Wirtschaften. *In: Schanz G [Hg.] Handbuch Anreizsysteme*. Stuttgart: Poeschel: 171–190.

Sens B, Wenzlaff P, Pommer G, Hardt H v. d. (2010) Auswirkungen der DRG-Einführung: Die Qualität hat nicht gelitten. *Dtsch Arztebl* 107 (1-2): A 25–7.

Shahian DM, Normand ST (2003) The volume-outcome relationship: from Luft to Leapfrog. *Ann Thorac Surg* 75 (3): 1048–1058.

Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand ST (2010) Variability in the measurement of hospital-wide mortality rates. *N Engl J Med* 363 (26): 2530–2539. doi: 10.1056/NEJMsa1006396.

Shaw RE, Anderson H, Brindis RG, Krone RJ, Klein LW, McKay CR et al. (2002) Development of a risk adjustment mortality model using the American College of Cardiology–National Cardiovascular Data Registry (ACC–NCDR) experience: 1998–2000. *J Am Coll Cardiol* 39 (7): 1104–1112. doi: 10.1016/S0735-1097(02)01731-X.

Shen Y (2003) The effect of financial pressure on the quality of care in hospitals. *J Health Econ* 22 (2): 243–269.

Silber S (2006) Argumente für die Integrierte Versorgung als Regelversorgung in der Kardiologie. *Clin Res Cardiol* 95 (S2): ii37. doi: 10.1007/s00392-006-1212-y.

Sorbero M, Damberg C, Shaw R, Teleki S, Lovejoy S, Decristofaro A et al. (2006) Assessment of Pay-For-Performance Options für Medicare Physician Services. Final Report. prepared for the Assistant Secretary for Planning and Evaluation, US Department of health and Human Services. Washington DC.

Spindler J (2010) Fallpauschalenbezogene Krankenhausstatistik Diagnosen und Prozeduren der Krankenhauspatienten auf Basis der Daten nach §21 Krankenhausentgeltgesetz. *In: Klauber J, Geraedts M, Friedrich J,Wasem J [Hg.] Krankenhausreport 2011. Schwerpunkt Qualität durch Wettbewerb*. Stuttgart: Schattauer GmbH: 349–378.

Staehle WH, Conrad P, Sydow J (1999) Management. Eine verhaltenswissenschaftliche Perspektive. 8. Aufl. Vahlen: München.

Sturm H (2002) Krankenhausreformen und Personalwirtschaft. Voraussetzungen und Auswirkungen der Ökonomisierung im Krankenhauswesen. 1. Aufl. Dt. Univ.-Verl: Wiesbaden.

Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M (2012) Reduced Mortality with Hospital Pay for Performance in England. *N Engl J Med* 367 (19): 1821–1828. doi: 10.1056/NEJMsa1114951.

SVR (2005) Gutachten des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen. Koordination und Qualität im Gesundheitswesen. Bundesanzeiger-Verl: Köln.

SVR (2007) Kooperation und Verantwortung. Voraussetzungen einer zielorientierten Gesundheitsversorgung. Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen: Bonn.

SVR (2012) Sondergutachten 2012. Wettbewerb an der Schnittstelle zwischen ambulanter und stationärer Gesundheitsversorgung. Drucksache 17/10323. Hg. v. Deutscher Bundestag – 17. Wahlperiode. Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen. Bonn.

Ten Asbroek AH, Arah OA, Geelhoed J, Custers T, Delnoij DM, Klazinga NS (2004) Developing a national performance indicator framework for the Dutch health system. *Int J Qual Health Care* 16 (Suppl 1): i65-i71. doi: 10.1093/intqhc/mzh020.

Thölen F Dr. med., Becker A Prof. Dr. med., Aufenanger J Prof. Dr. med. (2010) Qualitätsmanagement in der Transfusionsmedizin. *Das Krankenhaus* (10): 937–941.

Thomeczek (2002) Checkliste zur Bewertung von Qualitätsmanagement-Systemen in der ambulanten Versorgung. Hg. v. ÄZQ (Ärztliches Zentrum für Qualität in der Medizin), Berlin.

Thüsing C (2005) Qualitätsmanagement im Krankenhaus. *Med Klin* 100 (3): 149–153. doi: 10.1007/s00063-005-1012-6.

Transplantationsskandal (2013) Transplantationsskandal: Strafen sind fraglich. *Dtsch Arztebl* 2013, 11.01.2013. Online verfügbar unter http://www.aerzteblatt.de/nachrichten/52991, zuletzt geprüft am 15.01.2013.

Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, Laupacis A, Kapral MK (2004) Impracticability of Informed Consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 350 (14): 1414–1421. doi: 10.1056/NEJMsa031697.

Valuck T (2008) Medicare Value-Based Purchasing: Non-Payment for Selected Hospital-Acquired Conditions. *Dtsch Med Wochenschr* 133 (S 05): S140. doi: 10.1055/s-0028-1085587.

Van Emmerich C, Metzinger B (2010) Qualitätssicherung mit Routinedaten aus Sicht der Deutschen Krankenhausgesellschaft. *Das Krankenhaus* 12: 1177–1182.

van Herck P, Smedt D de, Annemans L, Remmen R, Rosenthal MB, Sermeus W (2010) Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res* 10 (247). doi: 10.1186/1472-6963-10-247.

Veit C, Hertle D, Bungar S, Trümner A, Ganske V, Meyer-Hofmann B [Hg.] (2012a) Pay-for-Performance im Gesundheitswesen. Sachbestandsbericht zur Evidenz und Realisierung sowie Darlegung der Grundlagen für eine künftige Weiterentwicklung. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit. BQS-Institut, Düsseldorf.

Veit C, Hertle D, Bungar S, Trümner A, Ganske V, Meyer-Hofmann B (2012b) Teil A - Zusammenfassung, Einführung, Projekte. *In: Veit C, Hertle D, Bungar S, Trümner A, Ganske V,Meyer-Hofmann B [Hg.] Pay-for-Performance im Gesundheitswesen. Sachbestandsbericht zur Evidenz und Realisierung sowie*

Darlegung der Grundlagen für eine künftige Weiterentwicklung. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit, Düsseldorf.

Veit C, Hertle D, Bungar S, Trümner A, Ganske V, Meyer-Hofmann B (2012c) Teil B - Wissenschaftliche Evidenz. *In: Veit C, Hertle D, Bungar S, Trümner A, Ganske V,Meyer-Hofmann B [Hg.] Pay-for-Performance im Gesundheitswesen. Sachbestandsbericht zur Evidenz und Realisierung sowie Darlegung der Grundlagen für eine künftige Weiterentwicklung*. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit, Düsseldorf.

Veit C, Hertle D, Bungar S, Trümner A, Ganske V, Meyer-Hofmann B (2012e) Teil C - P4P-Elemente, Methoden und Umsetzung. *In: Veit C, Hertle D, Bungar S, Trümner A, Ganske V,Meyer-Hofmann B [Hg.] Pay-for-Performance im Gesundheitswesen. Sachbestandsbericht zur Evidenz und Realisierung sowie Darlegung der Grundlagen für eine künftige Weiterentwicklung*. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit, Düsseldorf.

Veit C, Hertle D, Bungar S, Trümner A, Ganske V, Meyer-Hofmann B (2012f) Teil F - Literaturverzeichnis, Anhänge. *In: Veit C, Hertle D, Bungar S, Trümner A, Ganske V,Meyer-Hofmann B [Hg.] Pay-for-Performance im Gesundheitswesen. Sachbestandsbericht zur Evidenz und Realisierung sowie Darlegung der Grundlagen für eine künftige Weiterentwicklung*. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit, Düsseldorf.

Velasco-Garrido M, Busse R (2004) Förderung der Qualität in deutschen Krankenhäusern? Eine kritische Diskussion der ersten Mindestmengenvereinbarung. *Gesundheits- und Sozialpolitik* 58 (5-6): 10–20.

Volpp KG, Pauly MV, Loewenstein G, Bangsberg D (2009) P4P4P: An Agenda For Research On Pay-For-Performance For Patients. *Health Aff* 28 (1): 206–214. doi: 10.1377/hlthaff.28.1.206.

Vos M de, Graafmans W, Kooistra M, Meijboom B, van der Voort P, Westert G (2009) Using quality indicators to improve hospital care: a review of the literature. *International Journal for Quality in Health Care* 21 (2): 119–129. doi: 10.1093/intqhc/mzn059.

Wälchli A (1995) Strategische Anreizgestaltung. Modell eines Anreizsystems für strategisches Denken und Handeln des Managements. Verlag Haupt: Bern, Stuttgart, Wien.

Wang TY, Fonarow GC, Hernandez AF, Liang L, Ellrodt G, Nallamothu BK et al. (2009) The dissociation between door-to-balloon time improvement and improvements in other acute myocardial infarction care processes and patient outcomes. *Arch Intern Med* 169 (15): 1411–1419. doi: 10.1001/archinternmed.2009.223.

Wegscheider K (2004) Medizinische Register. *Bundesgesundheitsbl* 47 (5): 416–421. doi: 10.1007/s00103-004-0828-1.

Weiner BJ, Alexander JA, Shortell SM, Baker LC, Becker M, Geppert JJ (2006) Quality Improvement Implementation and Hospital Performance on Quality Indicators. *Health Serv Res* 41 (2): 307–334. doi: 10.1111/j.1475-6773.2005.00483.x.

Wells TS (1864) Some Causes of Excessive Mortality after Surgical Operations. *BMJ* 2 (196): 384–388.

Weltärztebund [Hg.] (1994) Deklaration von Genf - Erstverabschiedung 1948. Genfer Gelöbnis: Stockholm.

Wennberg JE, Roos N, Sola L, Schori A, Jaffe R (1987) Use of claims data systems to evaluate health care outcomes. Mortality and reoperation following prostatectomy. *J Am Med Assoc (JAMA)* 257 (7): 933–936.

Wennberg J, Gittelsohn A (1973) Small Area Variations in Health Care Delivery: A population-based health information system can guide planning and regulatory decision-making. *Science* 182 (4117): 1102–1108.

Werner RM, Bradlow ET (2006) Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates. *J Am Med Assoc (JAMA)* 296 (22): 2694–2702. doi: 10.1001/jama.296.22.2694.

Werner RM, Kolstad JT, Stuart EA, Polsky D (2011) The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Aff* 30 (4): 690–698. doi: 10.1377/hlthaff.2010.1277.

WHO (2006) Gesundheit im Schlaglicht: Deutschland 2004. Weltgesundheitsorganisation Regionalbüro für Europa: Kopenhagen.

WHO (2010) Germany: Health Profil. Weltgesundheitsorganisation Regionalbüro für Europa.

Wicke L, Blenk L (1993) Umweltökonomie. Eine praxisorientierte Einführung. 4. Aufl. Vahlen: München.

Wiese H (2010) Mikroökonomik. Eine Einführung. 5. Aufl. Springer: Berlin, Heidelberg.

Wild J (1973) Organisation und Hierarchie. *Zeitschrift für Organisation* 42: 45–54.

Williams SC, Watt A, Schmaltz SP, Koss RG, Loeb JM (2006a) Assessing the reliability of standardized performance indicators. *Int J Qual Health Care* 18 (3): 246–255. doi: 10.1093/intqhc/mzi098.

Williams TR, Raube K, Damberg CL, Mardon RE (2006b) Pay for performance: its influence on the use of IT in physician organizations. *J Med Pract Manage* 21 (5): 301–306.

Wirth C, Mutschler W, Bischoff HP, Püschmann H, Neu J (2009) Komplikationen in Orthopädie und Unfallchirurgie: Vermeiden; erkennen; behandeln. Thieme Georg Verlag, Stuttgart.

Wiswede G (1980) Motivation und Arbeitsverhalten. Organisationspsychologische und industriesoziologische Aspekte der Arbeitswelt. Reinhardt: München.

Witzel A (2000) Das problemzentrierte Interview. *In: Forum Qualitative Sozialforschung,* 1 Art. 22 (1).

Wollersheim H, Hermens R, Hulscher M, Braspenning J, Ouwens M, Schouten J et al. (2007) Clinical indicators: development and applications. *Neth J Med* 65 (1): 15–22.

Wrona T (2005) Die Fallstudienanalyse als wissenschaftliche Forschungsmethode. ESCP-EAP European School of Management: Berlin.

Wünsch A (2012) Anreizkompatible Entlohnungsfunktionen unter Berücksichtigung der kumulativen Prospect-Theorie. Herleitung und Analyse. 1. Aufl. Callsen-Bracker Verlag: Berlin, Technische Universität Berlin.

Wynia MK (2009) The risks of rewards in health care: how pay-for-performance could threaten, or bolster, medical professionalism. *J Gen Intern Med* 24 (7): 884–887. doi: 10.1007/s11606-009-0984-y.

Yin RK (2003) Case study research. Design and methods. 3. Aufl. Sage Publications: Thousand Oaks, Calif.

Zacher J (2011) IQM-Indikatoren Version 2.0. *In: Kuhlen R, Rink O,Zacher J [Hg.] Jahrbuch Qualitätsmedizin 2011*. Berlin: MWV (Medizinisch Wissenschaftliche Verlagsgesellschaft): 1–30.

Zacher J, Rohn C (2010) Messung und Verbesserung medizinischer Ergebnisqualität. *Deutsche Gesellschaft für Unfallchirurgie e.V.: Mitteilungen und Nachrichten* 32. (62): 96–97.

Zahnd D (2011) IQM-Version 2.0 Adaption auf die Schweiz: Die CH-IQI. *In: Kuhlen R, Rink O,Zacher J [Hg.] Jahrbuch Qualitätsmedizin 2011*. Berlin: MWV (Medizinisch Wissenschaftliche Verlagsgesellschaft): 55–66.

Zertifikatehandel (2012) Krankenkasse bringt Zertifikatehandel für OP-Leistungen ins Spiel. *Dtsch Arztebl* 2012, 30.07.2012. Online verfügbar unter http://www.aerzteblatt.de/nachrichten/51065/ Krankenkasse-bringt-Zertifikatehandel-fuer-OP-Leistungen-ins-Spiel , zuletzt geprüft am 02.02.2013.

Zhan C, Miller MR (2003) Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 12 Suppl. 2: ii58-63. doi: 10.1136/qhc.12.suppl_2.ii58.

Zimmermann H, Henke K, Broer M (2009) Finanzwissenschaft. Eine Einführung in die Lehre von der öffentlichen Finanzwirtschaft. 10. Aufl. Vahlen: München.

Zollondz H (2002) Grundlagen Qualitätsmanagement. Einführung in Geschichte, Begriffe, Systeme und Konzepte. Oldenbourg Verlag: München.

Zorn U, Ollenschläger G (1999) Qualitätsbestimmung in der medizinischen Versorgung--ein universelles Entwicklungsschema für Qualitätsindikatoren. *Z Arztl Fortbild Qualitatssich* 93 (2): 123–128.

Zweifel P, Breyer F, Kifmann M (2009) Health Economics. Springer Verlag: Berlin, Heidelberg.