

Institut für Weltwirtschaft
an der Universität Kiel
Düsternbrooker Weg 120
24105 Kiel

Kieler Arbeitspapier Nr. 1222

**Die Erstellung medizinischer Leistungen –
entscheidungstheoretische Grundlagen und
gesundheitspolitischer Handlungsbedarf**

von

Michael Stolpe

Juli 2004

Für den Inhalt der Kieler Arbeitspapiere sind die jeweiligen Autorinnen und Autoren verantwortlich, nicht das Institut. Da es sich um Manuskripte in einer vorläufigen Fassung handelt, wird gebeten, sich mit Anregungen und Kritik direkt an die Autorinnen und Autoren zu wenden und etwaige Zitate mit ihnen abzustimmen.

Die Erstellung medizinischer Leistungen – entscheidungstheoretische Grundlagen und gesundheitspolitischer Handlungsbedarf

This paper surveys the decision-theoretic foundations of health care production and derives policy implications. The priority at the beginning of the 21st century must be to accelerate the diffusion of new medical knowledge, a public good, so that the efficiency in the adoption of new technology is enhanced. Neither central planning, nor an extension of intellectual property rights can be expected to meet this challenge. Instead, medical practice needs to be organised as an open system – similar to the “Republic of Science” or the open source community in software development. The paper describes specific reforms in ambulatory, stationary and managed care that would help to enhance the innovative capacity and efficiency of health systems.

Dieser Aufsatz stellt die entscheidungstheoretischen Grundlagen der Produktion von Gesundheitsleistungen dar und leitet daraus Schlussfolgerungen für die Gesundheitspolitik ab. Die bedeutendste Herausforderung am Beginn des 21. Jahrhunderts liegt darin, die Diffusion neuen medizinischen Wissens – eines öffentlichen Gutes – zu beschleunigen und so die Effizienz in der Adoption neuer Technologien zu erhöhen. Weder zentral gesteuerte Planungsinstanzen, noch eine Ausweitung geistiger Eigentumsrechte können dieser Herausforderung gerecht werden. Stattdessen muss der Innovationsprozess in der medizinischen Praxis als offenes System organisiert werden – ähnlich wie das naturwissenschaftliche Forschungssystem oder die Open-Source-Bewegung in der Software-Entwicklung. Der Aufsatz diskutiert anschließend, wie der ambulante Sektor, der stationäre Sektor und integrierte Versorgungsstrukturen zu größerer Innovationsfähigkeit und Produktionseffizienz im Gesundheitswesen beitragen können.

Schlagworte: Medizinische Diagnose, Bayes-Theorem, Medizinische Praxisvariationen, Integrierte Versorgung

JEL Klassifikation: I10, I11, I18

Dr. Michael Stolpe

Institut für Weltwirtschaft

24105 Kiel, Germany

Telefon: +49/431/8814-246

Fax: +49/431/8814-500

E-mail: mstolpe@ifw.uni-kiel.de

Inhalt

1. Motivation	1
2. Der organisatorische Kontext medizinischer Entscheidungen	5
2.1 Die Praxis der Differentialdiagnose	6
2.2 Klassische Statistik und Bayesianische Entscheidungstheorie	10
2.3 Theoretische Motivation des Bayes-Theorems bei medizinischen Entscheidungen.....	16
3. Die normative Analyse medizinischer Entscheidungen	20
3.1 Die Interpretation diagnostischer Tests	21
3.2 Entscheidungen auf Basis des Erwartungswertes	23
3.3 Entscheidung über die Durchführung eines diagnostischen Tests	26
3.4 Auswahl zwischen alternativen Tests.....	31
4. Gesundheitspolitische Implikationen medizinischer Praxis- variationen.....	36
4.1 Wohlfahrtswirkungen	37
4.2 Öffentliche Güter in der Absorption medizinischen Wissens.....	50
4.3 Die Rolle von Behandlungsleitlinien	70
4.4 Steuerungsmöglichkeiten der Gesundheitspolitik.....	83
5. Der ambulante Sektor: anbieterinduzierte Nachfrage und ärztliche Vergütungssysteme	95
5.1 Informationsökonomische Grundlagen	97
5.2 Angebotsinduzierte Nachfrage und Nachfrageüberhang	109

5.3	Ärztliche Vergütungssysteme.....	113
6.	Der stationäre Sektor: Skalenerträge und Spezialisierung im Krankenhaus.....	116
6.1	Alternative Formen der Krankenhausfinanzierung.....	117
6.2	Möglichkeiten und Grenzen des Wettbewerbs.....	118
6.3	Ausblick.....	123
7.	Integrierte Versorgung: Gatekeeper und medizinische Leitlinien.....	126
7.1	Managed Care.....	126
7.2	Entwicklung und Anwendung medizinischer Leitlinien.....	133
8.	Schlussbemerkung.....	136

Schaubilder

Schaubild 1:	Diagnose als Zuordnungsproblem	148
Schaubild 2a:	A priori und a posteriori Verteilung	149
Schaubild 2b:	Vennsches Mengendiagramm.....	149
Schaubild 3:	Baumdarstellung des Bayes Theorems.....	150
Schaubild 4:	Risikofunktionen alternativer Entscheidungsfunktionen	151
Schaubild 5:	Diagnostischer Test mit positivem Ergebnis.....	152
Schaubild 6:	Diagnostischer Test mit negativem Ergebnis.....	153
Schaubild 7:	Medizinischer Entscheidungsbaum	154
Schaubild 8:	Schwellenwahrscheinlichkeit für eine Behandlung und das Kosten-Nutzen-Verhältnis.....	155
Schaubild 9:	Entscheidung über die Durchführung eines diagnostischen Tests	156
Schaubild 10:	Die Abhängigkeit der Durchführung eines Tests von seiner Sensitivität und der Rate falscher positiver Ergebnisse.....	157
Schaubild 11:	Die Interdependenz von Test- und Behandlungs- entscheidung	158
Schaubild 12a:	Uninformativer Test	159
Schaubild 12b:	Perfekter Test.....	159
Schaubild 12c:	Typischer Test für eine Reihenuntersuchung.....	160
Schaubild 12e:	Receiver Operating Characteristics (ROC) für zwei unterschiedliche medizinische Testverfahren	161

Schaubild 13a:	Effizienzverluste durch medizinische Praxisvariationen	162
Schaubild 13b:	Medizinische Praxisvariationen und Moral Hazard	163
Schaubild 13c:	Medizinische Praxisvariationen bei unzureichender Diffusion einer neuen Medizintechnologie	164
Schaubild 13d:	Medizinische Praxisvariationen und Monopolpreis- bildung	165
Schaubild 14:	Angebot und Nachfrage neuer medizinischer Information	166
Schaubild 15:	Anbieterinduzierte Nachfrage.....	167
Schaubild 16:	Nachfrageüberhang.....	168
Schaubild 17:	Kostenerstattung	169
Schaubild 18:	Einzeleistungsvergütung.....	170

1. Motivation

Die Erstellung medizinischer Leistungen ist im Kern ein Entscheidungsprozess unter Unsicherheit. Während sein Ergebnis in der Regel einem einzelnen Patienten wie ein *privates* Gut zugute kommt, ist die wichtigste Ressource, die der Arzt bei der Lösung des Entscheidungsproblems verwendet, eine Art *öffentliches* Gut: der jeweils aktuelle Stand des medizinischen Wissens. Seit dem frühen 20. Jahrhundert hat sich dieses Wissen geradezu explosionsartig vermehrt. Die medizinische Praxis ist daher einem rasanten technologischen Wandel unterworfen, der oft eine zweiseitige Wirkung hat: einerseits können neue Diagnosemöglichkeiten helfen, Unsicherheit zu reduzieren, andererseits können sie in Verbindung mit neuen therapeutischen Möglichkeiten und durch die Aufdeckung bislang unbekannter Krankheitsbilder die Komplexität medizinischer Entscheidungen und die subjektive Unsicherheit des Arztes sogar noch vergrößern.

In diesem Aufsatz werden medizinische Entscheidungen als Ergebnis interaktiver stochastischer Prozesse erklärt, die durch lokal differenziertes Bayesianisches Lernen entstehen. Die lokale Diffusion des dabei gewonnenen Wissens stellt einen empirisch bedeutsamen Rückkoppelungsmechanismus her, der den medizinischen Entscheidungsprozess abhängig von dem jeweiligen Pfad des lokalen Lernens macht. Über die einzelne Diagnoseentscheidung hinaus hat die Hypothese der Pfadabhängigkeit grundlegende Bedeutung für die Adoption medizinischer Technologien und für die organisatorische Gestaltung des Gesundheitswesens. Dabei geht es zum Beispiel um die optimale Arbeitsteilung zwischen ambulanten und stationären Versorgungsformen und um die Frage, ob und in welchem Maße integrierte Versorgungsstrukturen die Effizienz der medizinischen Leistungserstellung verbessern können. Die Vernachlässi-

gung dieser Zusammenhänge kann zu falschen gesundheitspolitischen Entscheidungen und erheblichen Wohlfahrtsverlusten führen.

Nicht-Mediziner, insbesondere die Konsumenten medizinischer Leistungen, nehmen die fundamentale Unsicherheit vieler medizinischer Entscheidungen oft erst wahr, wenn medizinische Fehler Todesfälle zur Folge haben. Schnurrer und Frölich (2003) schätzen, dass allein in den internistischen Abteilungen deutscher Krankenhäuser jedes Jahr ca. 58.000 Menschen aufgrund unerwünschter Arzneimittelwirkungen versterben. Als wichtigste Gründe nennt die Studie eine oft unzureichende Anpassung der Arzneimitteltherapie – zum Beispiel an die Nieren- und Leberfunktion, die Größe, das Gewicht und das Alter des einzelnen Patienten – und eine oft unzureichende Beachtung möglicher Interaktionen zwischen verschiedenen Arzneimitteln, die der Patient gleichzeitig einnimmt. Schnurrer und Frölich (2003) schätzen, dass die Hälfte der arzneimittelbedingten Todesfälle mit dem bereits veröffentlichten pharmakologischen Wissen verhindert werden könnte, wenn dieses bei allen ärztlichen Entscheidungen über den Einsatz von Arzneimitteln auch tatsächlich berücksichtigt würde.

Für die USA hat die National Academy of Sciences (1999) einen einflussreichen Bericht mit dem Titel „To Err is Human“ vorgelegt, nach dem im Jahre 1984 bei 3,7 Prozent aller Krankenhauseinweisungen im Bundesstaat New York behandlungsinduzierte Gesundheitsschäden auftraten, die in mehr als zwei Drittel der analysierten Fälle auf vermeidbare medizinische Fehlentscheidungen zurückzuführen waren. Für die USA insgesamt wurde die Zahl von Todesfällen aufgrund medizinischer Fehler auf 50.000 bis 100.000 pro Jahr geschätzt. Offensichtlich weisen diese Zahlen auf erhebliche Effizienzprobleme in der medizinischen Leistungserstellung und gleichzeitig auf ein großes Potential zur Verringerung kostenträchtiger Fehlentscheidungen hin. So hat dem Bericht zufolge allein die experimentelle Beteiligung von Pharmazeuten bei

ärztlichen Visiten in ausgewählten amerikanischen Krankenhäusern die Zahl der Medikamentierungsfehler von 10,4 pro 1000 Patiententage um 66 Prozent auf 3,5 reduziert. Die volkswirtschaftlichen Kosten aller Medikamentierungsfehler schätzt der Bericht auf 37,6 Milliarden US-Dollar (\$) pro Jahr, wovon 17 Milliarden \$ den vermeidbaren Fehlern zuzurechnen sind.

Um die Handlungsmöglichkeiten der Gesundheitspolitik im Hinblick auf das Effizienzziel zu erkennen, bedarf es einer entscheidungstheoretisch fundierten Analyse der medizinischen Leistungserstellung, wie ich sie im Folgenden darstellen werde. Diese Analyse zeigt, dass eine zentrale Aufgabe der Gesundheitspolitik im Setzen von Anreizen zur Bereitstellung, Diffusion, Bewertung und Adoption neuen Wissens liegt. Das relevante medizinische Wissen wird nicht allein durch klinische Forschung gewonnen, sondern entsteht zu einem großen Teil bei der Entwicklung neuer Technologien und durch Erfahrungen in der medizinischen Praxis. Um die Effizienz der medizinischen Leistungserstellung zu gewährleisten, muss die Gesundheitspolitik den Einfluss des technologischen Wandels auf die Ergebnisse medizinischer Entscheidungsprozesse erkennen und ihre Steuerungsmöglichkeiten durch die Gestaltung der organisatorischen Rahmenbedingungen nutzen. Dies setzt eine Theorie voraus, die die Informations- und Anreizsituation des Arztes unter sich ändernden Ressourcenrestriktionen beachtet. Die Gesundheitspolitik muss deshalb mikroökonomisch fundiert sein.

Empirische Länderquerschnittsuntersuchungen zu den Determinanten von Gesundheitsausgaben, wie sie in Gerdtham und Jönsson (2000) besprochen werden, beruhen meist auf einer aggregierten neoklassischen Produktionsfunktion, die ein einheitliches Outputmaß für das gesamte Gesundheitswesen auf die Einsatzmengen standardisierter Inputs zurückführt. Auf dieser Grundlage lässt sich jedoch der Einfluss organisatorischer Gestaltungsalternativen auf die Effizienz in der medizinischen Leistungserstellung nicht analysieren. Denn

nur auf disaggregierter Ebene lässt sich empirisch erkennen, in welchem Maße ein beobachteter Zuwachs medizinischer Leistungen durch eine bessere Zuordnung der Patienten zur jeweils optimalen therapeutischen Strategie erreicht wurde und nicht auf einen gleichzeitigen Zuwachs der Menge oder Qualität medizinischer Inputs bei konstanter Qualität der Zuordnung zurückzuführen ist.

Die Zuordnung der Patienten zu alternativen therapeutischen Strategien ist der Kern eines notwendigerweise dezentralen Entscheidungsprozesses, dessen Ergebnis die Veränderung individueller Gesundheitszustände ist. Um die Produktionseffizienz medizinischer Leistungserstellung zu untersuchen, ist also nicht nur zwischen unterschiedlich zu behandelnden Krankheiten, sondern auch zwischen Patientengruppen aus verschiedenen epidemiologischen Milieus zu unterscheiden. Zwar kann ein direkter Vergleich des Ressourceneinsatzes und des Zielerreichungsgrades die Kosteneffektivität einzelner medizinischer Maßnahmen im Hinblick auf ein durch die Art der Erkrankung definiertes Ziel empirisch ermitteln. Die Produktionseffizienz der medizinischen Leistungserstellung insgesamt hängt aber auch davon ab, dass die Art der Erkrankung in möglichst vielen Fällen zutreffend erkannt wird.

Im Folgenden werde ich die wesentlichen Elemente einer mikroökonomischen Theorie der medizinischen Leistungserstellung nicht nur abstrakt darstellen, sondern auch ihre Implikationen für die Organisation des ambulanten und des stationären Sektors sowie für die Arbeitsteilung und Koordination zwischen diesen diskutieren. Abschnitt 2 stellt den organisatorischen Kontext medizinischer Entscheidungen dar, der sich in unterschiedlichen medizinischen Milieus widerspiegelt und analysiert seinen Einfluss auf den Prozess der Entscheidungsfindung. Abschnitt 3 beschreibt die Interpretation diagnostischer Tests, die Organisation mehrstufiger und sequentieller Entscheidungen sowie die Auswahl diagnostischer Tests aus normativer Sicht. Abschnitt 4 diskutiert die gesund-

heitspolitischen Implikationen medizinischer Praxisvariationen. Abschnitt 5 behandelt die besonderen Probleme und Reformmöglichkeiten des ambulanten Sektors, Abschnitt 6 die Probleme und Reformmöglichkeiten des stationären Sektors. Abschnitt 7 lotet die Möglichkeiten und Grenzen integrierter Versorgungsstrukturen aus. Abschnitt 8 fasst die wichtigsten Schlussfolgerungen zusammen.

2. Der organisatorische Kontext medizinischer Entscheidungen

Ein Gesundheitswesen ist effizient organisiert, wenn es die Bedingungen schafft und Anreize setzt, die effiziente medizinische Entscheidungen gewährleisten. Im Kern geht es dabei um einzelne medizinische Diagnosen. Effiziente Entscheidungen führen zwar nicht in jedem Einzelfall zum optimalen Ergebnis, aber zu einer im Mittel kosteneffektiven Behandlung, die entweder den Behandlungserfolg bei gegebenem Ressourceneinsatz maximiert oder den Ressourceneinsatz zur Erreichung eines vorgegebenen Behandlungserfolgs minimiert. Die Kosteneffektivität einzelner Maßnahmen ist dabei lediglich eine notwendige, keine hinreichende Bedingung. Um die Häufigkeit von Fehlentscheidungen zu reduzieren, muss nicht nur das jeweils aktuelle medizinische Wissen über Ursachen, Diagnose- und Behandlungsmöglichkeiten, sondern auch die im Einzelfall relevante Information über die Wahrscheinlichkeit bestimmter Krankheiten und die Erfolgsaussichten einzelner medizinischer Interventionen am Ort der Entscheidung zur Verfügung stehen. Alle diese Informationen sind aus ökonomischer Sicht *öffentliche* Güter, weil ihre Verwendung in einer Entscheidung die Verwendung in anderen Entscheidungen weder ausschließt, noch in irgendeiner Weise beeinträchtigt.

Die optimale Organisation medizinischer Entscheidungen kann sich mit dem medizinisch-technischen Fortschritt ändern, denn dieser schafft ständig neue therapeutische Möglichkeiten und neue Diagnoseverfahren. Ein Beispiel hierfür

ist die Entstehung des modernen Krankenhauses, das – wie Abschnitt 6 näher erläutert – eine ganz andere Rolle spielt als die Pflegeanstalten des 19. Jahrhunderts. Erst die Entwicklung aufwändiger Diagnoseverfahren und die rasante Ausweitung und Vertiefung des medizinischen Wissens haben die Spezialisierung der Ärzte auf Teilgebiete der Medizin erforderlich gemacht und die Frage einer effizienten Arbeitsteilung und Koordination wichtig werden lassen. Auch Allgemeinärzte sind heute als Spezialisten anzusehen – als Spezialisten für Allgemeinmedizin, wie es die neue Berufsbezeichnung „Facharzt für Allgemeinmedizin“ unterstreicht. Sie können gleichzeitig eine wichtige Rolle als Gatekeeper spielen, indem sie eine Vorauswahl der Patienten treffen, die sie an niedergelassene oder stationäre Fachärzte überweisen.

Wenn neue Technologien und neue medizinische Erkenntnisse und Erfahrungen ungleichmäßig im medizinischen System oder im geografischen Raum verteilt sind, entstehen Praxisvariationen, die sich nicht mit den relevanten Determinanten optimaler medizinischer Entscheidungen erklären lassen. Sie offenbaren erhebliche Effizienzprobleme, die letztlich aus der fundamentalen Unsicherheit vieler medizinischer Entscheidungen resultieren. Neuere empirische Schätzungen, die Abschnitt 4 erläutert, lassen sogar vermuten, dass der Wohlfahrtsverlust aus medizinischen Praxisvariationen in einem modernen Gesundheitswesen insgesamt größer ist als der Wohlfahrtsverlust aus den Problemen adverser Selektion und Moral Hazard, die typischerweise bei asymmetrisch verteilten Informationen in der Finanzierung medizinischer Leistungen durch Krankenversicherungsverträge auftreten können.

2.1 Die Praxis der Differentialdiagnose

Das Grundproblem medizinischer Entscheidungen ist, dass sie unter Unsicherheit getroffen werden müssen, die der ärztliche Diagnoseprozess in aller Regel nur teilweise beseitigen kann. Einerseits müssen ärztliche Entscheidungen unter

Zeitdruck und allgemeiner Ressourcenknappheit gefällt werden. Andererseits ist die Komplexität des Problems nicht zu unterschätzen. Nach der internationalen Klassifikation von Krankheiten (Deutsches Institut für medizinische Dokumentation und Information 1994) gibt es etwa 10.000 verschiedene Diagnosen, denen in Deutschland mehr als 40.000 zugelassene Medikamente sowie eine Vielzahl von Operationsverfahren und anderen therapeutischen Möglichkeiten gegenübersteht.

Wendet sich ein Patient mit einem oder mehreren Symptomen an einen Arzt, scheidet eine vollständige Untersuchung aller theoretisch denkbaren Krankheiten in der Regel aus. Vielmehr muss der Arzt in kurzer Zeit eine kleine Auswahl der möglichen Krankheiten festlegen, auf die er seine Untersuchungen vorrangig konzentrieren will. Der Arzt bedient sich dabei der Methode der Differentialdiagnose: Er berücksichtigt mehrere mögliche Ursachen einer Erkrankung, bevor er sich für eine Diagnose entscheidet und den Patienten einer bestimmten Therapie zuordnet. Dass dieser Prozess nicht immer linear abläuft, verdeutlicht Schaubild 1.¹

Am Anfang des medizinischen Entscheidungsprozesses stehen die Beratungsursachen (A, B, C, D oder E), die den Besuch beim Arzt auslösen. Nicht immer ist die vorliegende Beratungsursache als Symptom einer Krankheit zu deuten (Fall B). Es können andererseits auch mehrere Symptome (Fall D) hinter einer Beratungsursache stecken. Verschiedene Beratungsursachen (Fall A und E) können mit ein- und demselben Symptom (Fall b) verknüpft sein. Und in manchen Fällen wird der Arzt direkt aufgrund eines beobachteten Symptoms (Fall a) eine therapeutische Intervention veranlassen. In Fall c wird dem

¹ Für Beispiele aus der Praxis, die noch konkreter sind, siehe den Beitrag von Estes und Mandel (1995) über die Diagnose sexuell übertragbarer Krankheiten mit mehr als 20 Diagnosemöglichkeiten, den Beitrag von Knoper (1995) über die Diagnose bei Husten mit mehr als zehn Diagnosemöglichkeiten oder den Beitrag von Wendt (1995) über die Diagnose bei Kopfschmerz mit mehr als 15 Diagnosemöglichkeiten.

Symptom kein behandelbarer Krankheitswert zugeordnet, so dass der Diagnoseprozess abgebrochen werden kann.

In den meisten Fällen wird der Arzt diagnostische Hypothesen zunächst testen, um eine gesicherte Diagnose zu erhalten, aus der dann die therapeutischen Handlungen abgeleitet werden. Einzelne Testergebnisse können eindeutig zu gesicherten Diagnosen führen (Fall β) oder mehrdeutig sein (Fall α). Viele Diagnosen können erst bei Vorliegen der Ergebnisse mehrerer unterschiedlicher Tests als gesichert gelten (Fall Γ und E). Aus einer gesicherten Diagnose folgt manchmal eine einzige therapeutische Handlung (Fall A), manchmal ein Bündel von Maßnahmen (Fall Δ). Erfolgreiche therapeutische Handlungen generieren ihrerseits neue Informationen, die zu einer Revision der Diagnose führen können (Fall \mathcal{B}). Im Fall \mathcal{D} treten zusätzlich Nebenwirkungen auf, die ein eigenes Symptom darstellen (Fall e) und einen neuen Diagnoseprozess auslösen.

Bei der *Differentialdiagnose*, wie sie zum Beispiel Sox et al. (1988) ausführlich beschreiben, ist die Suche nach neuen Informationen und die Auswahl diagnostischer Tests stets abhängig von den diagnostischen Hypothesen, die der Arzt testen will. Dabei berücksichtigt der Arzt zu jedem Zeitpunkt nur eine begrenzte Zahl von Hypothesen, wird aber neue Hypothesen formulieren, wenn er eine untersuchte Hypothese verwerfen muss. Zu Beginn einer Differentialdiagnose muss der Arzt entscheiden, welche diagnostischen Hypothesen vorrangig untersucht werden sollen. Ausschlaggebend hierfür ist die Wahrscheinlichkeit einer vermuteten Krankheit und die Wichtigkeit oder Dringlichkeit der medizinischen Intervention. Bei der Auswahl diagnostischer Hypothesen geht es also um die drei folgenden Fragen: Erstens, welche Krankheiten sind bei den vorliegenden Symptomen wahrscheinlich? Zweitens, welche möglichen Krankheiten sind lebensbedrohlich oder verursachen andere irrever-

siblen Schäden, wenn sie nicht rasch behandelt werden? Drittens, welche Krankheiten sind überhaupt behandelbar?

Die untersuchungsleitenden Hypothesen können durch neue Informationen, die zum Beispiel aus Testergebnissen gewonnen wurden, entweder bestätigt oder falsifiziert werden. Nach der Wissenschaftstheorie von *Karl Popper* kommt es in erster Linie darauf an, Hypothesen zu falsifizieren. Die Bestätigung einer Hypothese ist dagegen wenig aussagekräftig, denn sie braucht nur ein zufälliges Ergebnis der jeweiligen Stichprobeninformation zu sein. Manchmal dokumentiert sie auch einfach nur das Bemühen des Forschers oder diagnostizierenden Arztes, sein eigenes Vorurteil mit „Evidenz“ zu stützen. Erst wenn eine Hypothese ernsthafte Falsifikationsversuche übersteht, kann sie als vorläufig akzeptiert gelten.

Diese Vorgehensweise wird in der Medizin „*Screening and Branching*“ genannt. Der Arzt sucht dabei nach einem Befund, der fast immer mit der vermuteten Krankheit verbunden ist. Lässt sich der Befund dann nicht finden, verwirft der Arzt die diagnostische Hypothese und richtet seine weitere Untersuchung auf die Möglichkeit einer anderen Krankheit, einer anderen diagnostischen Hypothese aus. Dies ist in vielen Fällen effizient, weil der Arzt seinen Patienten so in relativ kurzer Zeit auf viele mögliche Krankheiten hin untersuchen kann. Allerdings ist diese Vorgehensweise nicht ohne Risiko, insbesondere wenn eine mögliche Krankheit lebensgefährlich ist oder andere bleibende Schäden zur Folge haben kann.

Zur direkten Bestätigung einer hypothetischen Krankheit kommen überhaupt nur Befunde in Frage, die ausschließlich bei dieser Krankheit auftreten. So etwas ist in der Medizin jedoch sehr selten. Oft ist es unmöglich, vollkommene Sicherheit über die Ursache einer Erkrankung zu erhalten. Zum Glück ist das meistens auch nicht notwendig. Denn der diagnostische Prozess kann im

Allgemeinen spätestens dann beendet werden, wenn zusätzliche Information die Wahl der therapeutischen Strategie nicht mehr ändern würde. In der Regel sollte der Prozess sogar schon früher abgebrochen werden, denn Effizienz verlangt, dass auf jeder Stufe die Risiken und Kosten weiterer Untersuchungen gegen den potentiellen Wert zusätzlicher Information abgewogen werden.

2.2 Klassische Statistik und Bayesianische Entscheidungstheorie

In der wissenschaftlichen Forschung werden neue Informationen oft mithilfe der klassischen Statistik gewonnen, die auf einem frequentistischen Wahrscheinlichkeitsbegriff beruht und Schlussfolgerungen mit dem Gesetz der großen Zahl begründet. Im Gegensatz zur reinen Wissenschaft, die Hypothesen auf der Basis großer Stichproben testen und so das wahre Modell der Welt identifizieren kann, steht der Medizin bei ihrer Suche nach der Ursache einer Erkrankung immer nur der einzelne Patient zur Verfügung. Die Ergebnisse diagnostischer Tests sind deshalb oft selbst mit großer Unsicherheit behaftet. Das *Bayes-Theorem* beschreibt, wie a priori-Wahrscheinlichkeiten revidiert werden, wenn neue, aber unsichere Informationen bekannt werden, die für die Bewertung gegebener Entscheidungsalternativen relevant sind. Jede a posteriori-Wahrscheinlichkeit ist daher eine *bedingte* Wahrscheinlichkeit.

Schaubild 2a stellt beispielhaft dar, wie sich die a posteriori Verteilung von Wahrscheinlichkeiten für einen kontinuierlichen Parameter von der a priori Verteilung unterscheidet; ihr Mittelwert hat sich nach rechts verschoben und ihre Varianz hat sich verringert. Natürlich gibt es andere Beispiele, in denen sich der Parameter nach links verschiebt; die Varianz wird jedoch immer verringert, wenn neue Information die subjektive Unsicherheit des Arztes verringert. Mit Tönies (1993), S. 67, kann das Bayes-Theorem für die medizinische Anwendung durch die einfache Formel: $WD(S) = WS(D)WD/WS$ dargestellt werden. Darin ist $WD(S)$ die Wahrscheinlichkeit einer Diagnose bei Vorliegen

eines Symptoms, $WS(D)$ die Wahrscheinlichkeit eines Symptoms, falls eine bestimmte Diagnose zutrifft, WD die Wahrscheinlichkeit einer Diagnose im gegebenen epidemiologischen Milieu und WS die Wahrscheinlichkeit eines Symptoms im selben Milieu. WS und WD leitet der Arzt aus den Ergebnissen epidemiologischer Forschung und aus seiner eigenen beruflichen Erfahrung ab. Angaben über $WS(D)$ werden aus der medizinischen Forschung gewonnen.

Die *a priori-Wahrscheinlichkeiten* werden wesentlich durch die institutionelle Zuordnung eines Arztes im medizinischen System und durch den zeitlichen und örtlichen Kontext der Diagnose bestimmt. Dabei ist zwischen Querschnittsdiagnosen, die zu einem Zeitpunkt gemacht werden, und individuellen Längsschnittdiagnosen zu unterscheiden, die den Krankheitsverlauf eines einzelnen Patienten über einen längeren Zeitraum berücksichtigen. Letztere werden vor allem von Allgemein- oder Hausärzten gemacht, die einen langjährigen Kontakt zu ihren Patienten pflegen, so dass die *a priori-Wahrscheinlichkeiten* die individuelle Krankengeschichte widerspiegeln. Fachärzte dagegen führen fast immer Querschnittsdiagnosen durch, weil sie in der Regel mit neuen oder nur im Einzelfall überwiesenen Patienten zu tun haben.

Dabei hängen die *a priori-Wahrscheinlichkeiten* für einzelne Arten von Krankheiten in erster Linie von der fachlichen Spezialisierung des Facharztes ab. Ein Lungenarzt beispielsweise wird die Ursache einer Erkrankung eher in der Lunge suchen als in anderen Organen, ein Hautarzt wird eher an Hautkrankheiten denken als ein Internist, ein Orthopäde wird zuerst an die Knochen denken, ein Urologe an die Nieren, ein Proktologe an den Enddarm, usw. Der *zeitliche* Kontext spielt vor allem bei ansteckenden Krankheiten eine wichtige Rolle, die in Epidemien auftreten. Der *örtliche* Kontext einer Diagnose spielt eine Rolle, wenn relevante Informationen über die *a priori-Wahrscheinlichkeiten* bestimmter Krankheiten oder über die relativen Erfolgchancen therapeutischer Alternativen lediglich lokal verbreitet sind und

sich örtliche „Schulen“ mit vorherrschenden medizinischen Lehrmeinungen herausgebildet haben.

Für die Herleitung und Anwendung des Bayes-Theorems sind zunächst einige formale Grundbegriffe zu klären. Die Notation und Annahmen des Bayes Theorems lassen sich am besten anhand des Vennschen Mengendiagramms verdeutlichen, das in Schaubild 2b dargestellt ist. Ausgangspunkt für die formale Darstellung ist die bedingte Wahrscheinlichkeit $\pi(A|B)$, definiert als $\pi(A|B) = \pi(A \text{ und } B) / \pi(B)$. Das Bayes Theorem liefert ja nichts anderes als die bedingte Wahrscheinlichkeit für eine vermutete Krankheit bei Vorliegen eines positiven oder negativen Testergebnisses.² Dabei macht man sich zu nutze, dass die bedingte Wahrscheinlichkeit eines Testergebnisses $\pi(R|D)$ oder $\pi(R|\text{nicht } D)$ unabhängig von vorab verfügbarer Information ist. Die bedingte Wahrscheinlichkeit eines Testergebnisses ist deshalb für alle a priori-Wahrscheinlichkeiten gleich. Es gilt, *erstens*, $\pi(D|+)$: = die Wahrscheinlichkeit der Krankheit D bei positivem Testergebnis, *zweitens*, $\pi(+|D)$: = die Likelihood oder die Wahrscheinlichkeit eines positiven Testergebnisses bei Vorliegen der Krankheit D und *drittens*, $\pi(+|\text{nicht } D)$: = die Likelihood oder die Wahrscheinlichkeit eines positiven Testergebnisses bei Nicht-Vorliegen der Krankheit D. Die gesamte Wahrscheinlichkeit eines positiven Testergebnisses ist

$$\pi(+)=\pi(+ \text{ und } D)+\pi(+ \text{ und nicht } D).$$

Um die bedingte Wahrscheinlichkeit für das Vorliegen einer vermuteten Krankheit D bei positivem Testergebnis zu ermitteln, ist die Wahrscheinlichkeit für das gemeinsame Vorliegen der Krankheit und eines positiven Testergebnisses durch die Summe der gemeinsamen Wahrscheinlichkeiten für ein positives

² Bei vielen medizinischen Tests liegen die Ergebnisse zunächst in quantitativer Form als kontinuierliche Meßgröße vor und werden dann in binäre Form überführt, indem geprüft wird, ob der Wert der Test-Statistik einem vorgegebenen Grenzwert über- oder unterschreitet.

Testergebnis und das Vorliegen der Krankheit einerseits und für ein positives Testergebnis und das Nicht-Vorliegen der Krankheit andererseits zu dividieren: $\pi(D|+) = \pi(D \text{ und } +) / (\pi(+ \text{ und } D) + \pi(+ \text{ und nicht } D))$. Die Wahrscheinlichkeiten für das gemeinsame Vorliegen zweier Ereignisse A und B können durch das Produkt aus der bedingten Wahrscheinlichkeit für A bei Eintritt des Ereignisses B und der Wahrscheinlichkeit für das Ereignis B ersetzt werden. Daher liefert das Bayes Theorem die folgende Formel für die bedingte Wahrscheinlichkeit einer Krankheit bei einem positiven Testergebnis:

$$\pi(D|+) = \frac{\pi(D) \pi(+|D)}{(\pi(D) \pi(+|D)) + (\pi(\text{nicht } D) \pi(+|\text{nicht } D))}$$

Dieser Formel zufolge wird die revidierte Wahrscheinlichkeit bei einer gegebenen a priori-Wahrscheinlichkeit $\pi(D)$ aus den sogenannten *Likelihoods*, den bedingten Wahrscheinlichkeiten eines positiven Testergebnisses bei Vorliegen und bei Nicht-Vorliegen der Krankheit D, abgeleitet. Um das Bayes-Theorem für den Fall formal darzustellen, dass ein Test negativ ausgefallen ist, müssen in der obigen Formel lediglich die vier Plus-Zeichen durch Minus-Zeichen ersetzt werden. Dabei werden die bedingten Wahrscheinlichkeiten eines negativen Testergebnisses bei Vorliegen und Nicht-Vorliegen der Krankheit D wiederum als Likelihoods bezeichnet.

In Schaubild 3 ist das Bayes-Theorem anhand eines Beispiels als Entscheidungsbaum dargestellt. In einer Testpopulation mit auffälligem Röntgenbefund haben 30 Prozent der Menschen Krebs, 70 Prozent keinen Krebs. Dies ist ein exogenes Ereignis. Zur Entscheidung wird ein Test durchgeführt, der bei den an Krebs Erkrankten 180 wahre positive Ergebnisse und 120 falsche negative Ergebnisse liefert. Dies entspricht den Feldern a und c in Tabelle 1. Bei den nicht Erkrankten liefert der Test 28 falsche positive Ergebnisse und 672 wahre negative Ergebnisse. Dies entspricht den Feldern b und d in Tabelle 1.

Tabelle 1: Wahre und falsche Testergebnisse

Diagnostisches Testergebnis	Untersuchte Krankheit		
	vorhanden	nicht vorhanden	
positiv	a	b	a+b
negativ	c	d	c+d
	a+c	b+d	

Neben den Fehlern erster Art ($c/(a+c)$), deren Wahrscheinlichkeit als *Rate der falschen negativen Testergebnisse* (RFN) $\pi(-|D)=1-\pi(+|D)$ definiert ist, und den Fehlern zweiter Art ($b/(b+d)$), deren Wahrscheinlichkeit als *Rate der falschen positiven Testergebnisse* (RFP) $\pi(+|\text{nicht } D)$ definiert ist, illustriert Tabelle 1 die wichtigsten in der Praxis gebräuchlichen Definitionen der „Güte“ eines diagnostischen Tests. Von einem guten Test ist einerseits zu erwarten, dass er „sensitiv“ ist und beim Vorliegen der vermuteten Krankheit überwiegend positive Ergebnisse liefert; andererseits, dass er „spezifisch“ ist und möglichst *ausschließlich* bei der vermuteten Krankheit anschlägt, bei allen anderen Krankheiten aber negativ ausfällt. Daher wird erstens $\pi(+|D)$, die Rate der wahren positiven Testergebnisse, als *Sensitivität* eines Tests (SEN) definiert; sie ist gleich dem Quotienten $a/(a+c)$ in Tabelle 1. Und zweitens wird $\pi(-|\text{nicht } D)$, die Rate der wahren negativen Testergebnisse, als *Spezifität* eines Tests (SPE) bezeichnet. Die Spezifität ist also durch $d/(b+d)$ definiert. Als Testeigenschaften sind die Sensitivität und Spezifität zwar von den a priori-Wahrscheinlichkeiten unabhängig, nicht aber von der besonderen Zusammensetzung einer Testpopulation, die sich zum Beispiel hinsichtlich des Durchschnittsalters der Testpersonen unterscheiden kann, und auch nicht von

der Höhe des Grenzwertes, der im Falle einer kontinuierlichen Meßgröße als Teststatistik ein positives von einem negativen Ergebnis abgrenzt.

Um die Fähigkeit eines Tests anzugeben, zwischen erkrankten und nicht-erkrankten Patienten zu unterscheiden, werden Likelihood-Ratios verwendet, wobei $LR_+ = \pi(+|D)/\pi(+|\text{nicht } D) = \text{SEN}/(1-\text{SPE})$ die Likelihood-Ratio für ein positives Testergebnis und $LR_- = \pi(-|D)/\pi(-|\text{nicht } D) = (1-\text{SEN})/\text{SPE}$ die Likelihood-Ratio für ein negatives Testergebnis ist. Die Likelihood-Ratio für einen positiven Test ist umso größer, je größer die Rate der wahren positiven Ergebnisse im Verhältnis zur Rate falscher positiver Ergebnisse ist. Für einen negativen Test ist die Likelihood-Ratio umso kleiner, je größer die Rate der wahren negativen Ergebnisse im Verhältnis zur Rate falscher negativer Ergebnisse ist. Ein perfekt diskriminierender Test zeichnet sich durch eine Sensitivität und Spezifität von nahezu Eins aus, so dass die a priori-Wahrscheinlichkeiten praktisch keinen Einfluss mehr auf das a posteriori-Wahrscheinlichkeitsverhältnis haben.

Für den Arzt ist es vor allem wichtig, die Zusammenhänge des Bayes-Theorems mit Hilfe der vorab bekannten Eigenschaften eines Tests darstellen zu können. Um mit Hilfe der Likelihood-Ratio zum Beispiel die relative Wahrscheinlichkeit einer Krankheit bei positiv getesteten Personen auszurechnen, wird in der Praxis oft die folgende Schreibweise des Bayes-Theorem verwandt:

$$\frac{\pi(D|+)}{1-\pi(D|+)} = \frac{\pi(D)}{1-\pi(D)} \cdot LR_+$$

$\pi/(1-\pi)$ ist dabei das Verhältnis der Wahrscheinlichkeiten, dass ein Ereignis stattfindet, zum Beispiel eine Erkrankung bei Vorliegen eines positiven Tests Krankheit, oder nicht stattfindet, zum Beispiel das Nicht-Vorliegen der Krankheit trotz eines positiven Tests. Diese Version des Bayes-Theorems lässt sich ebenfalls anhand von Tabelle 1 verdeutlichen: Die a priori-Wahrschein-

lichkeit ist $WD = (a + c)/(a + b + c + d)$. Das a priori-Wahrscheinlichkeitsverhältnis ist $PR = WD/(1 - WD)$. Das a posteriori-Wahrscheinlichkeitsverhältnis ist $PO = (LR +) \cdot WD/(1 - WD) = PR \cdot LR +$. Und die a posteriori-Wahrscheinlichkeit ist $WD(S) = PO/(PO + 1)$.

Manchmal sind nicht die Likelihood-Ratios, sondern nur die elementarsten Testeigenschaften, die Sensitivität und Spezifität sowie die Rate der falschen positiven Testergebnisse, bekannt. Für den Fall eines *positiven* Testergebnisses lässt sich dann das Bayes-Theorem durch die Formel $\pi(D|+) = \pi(D) \cdot SEN / (\pi(D) \cdot SEN + (1 - \pi(D)) \cdot RFP)$ und für den Fall eines negativen Testergebnisses durch $\pi(D|-) = \pi(D) \cdot (1 - SEN) / (\pi(D) \cdot (1 - SEN) + (1 - \pi(D)) \cdot SPE)$ zum Ausdruck bringen. Für die Wahrscheinlichkeit eines positiven Testergebnisses gilt demnach: $\pi(+)=\pi(D) \cdot \pi(+|D) + (1 - \pi(D)) \pi(+|\text{nicht } D) = \pi(D) \cdot SEN + (1 - \pi(D)) \cdot RFP$. Für die Wahrscheinlichkeit eines negativen Testergebnisses, $\pi(-)=1 - \pi(+)$, gilt Entsprechendes.

2.3 Theoretische Motivation des Bayes-Theorems bei medizinischen Entscheidungen

Da medizinische Entscheidungen aus Hypothesen abgeleitet werden, besteht das eigentliche Entscheidungsproblem darin, die zur Auswahl stehenden Hypothesen zu testen.³ Die klassische Statistik hat erkannt, dass dabei zwei Arten von Fehlern auftreten können – die irrtümliche Ablehnung einer richtigen Hypothese (*Fehler erster Art*) und die Annahme einer falschen Hypothese (*Fehler zweiter Art*) – und dass diese Fehler nicht gleich schwerwiegend sein müssen. Die Kriterien zur Auswahl eines guten Testverfahrens müssen die Schwere der möglichen Fehler berücksichtigen. Aber die klassische Statistik, die auf einem frequentistischen Wahrscheinlichkeitsbegriff beruht, hat keinen allgemeingülti-

³ Diese Darstellung folgt Silvey (1975), Kap. 11 – An Introduction to Decision Theory.

gen Weg gefunden, dieses Prinzip in der Praxis zu verwirklichen. Es stellt sich daher die Frage, ob die Anwendung des Bayes-Theorems die Auswahl des optimalen Testverfahrens gewährleisten kann.

Die Auswirkungen der Fehler einer Entscheidung können in einer *Verlustfunktion* $L(\delta, \theta)$ zusammengefasst werden. Die Entscheidungsregel soll den erwarteten Verlust minimieren, und zwar für jeden beliebigen Wert des unbekanntes Parameters, über den es eine Hypothese zu testen gilt. Es geht also darum, eine Entscheidungsfunktion, eine Abbildung des Stichprobenraumes X in den Entscheidungsraum E , zu finden, die den Verlust infolge der Entscheidung d minimiert. Dabei ist zu berücksichtigen, dass die Höhe des erwarteten Verlustes auch von dem wahren Wert des unbekanntes Parameters abhängt.

Für jede Entscheidungsfunktion δ lässt sich dies in einer *Risikofunktion* $R_\delta(\theta) = E_\theta \{L(\delta, \theta)\}$ zum Ausdruck bringen, die den erwarteten Verlust in Abhängigkeit der möglichen wahren Werte des Parameters θ angibt. Man möchte nun eigentlich eine Entscheidungsfunktion δ^* suchen, die den erwarteten Verlust für jeden beliebigen Wert des unbekanntes Parameters θ minimiert. In der Praxis existiert eine solche Entscheidungsfunktion jedoch nur in ganz seltenen Fällen, weil der Verlauf der Verlustfunktion über den zulässigen Wertebereich für θ für unterschiedliche Entscheidungsfunktionen im Allgemeinen sehr unterschiedlich ist. Schaubild 4 stellt einige der in Silvey (1975), S. 164, analysierten Verlustfunktionen dar. Dabei wird deutlich, dass keine der Funktionen über den gesamten Wertebereich für θ stets kleinere Werte annimmt als alle anderen Funktionen.

Die *Lösung von Bayes* zeigt, dass diese Schwierigkeiten und Widersprüche der Entscheidungstheorie verschwinden, wenn man eine a priori-Verteilung über den Parameterraum zulässt, die für alle möglichen Werte des Parameters θ

Wahrscheinlichkeiten angibt. Für jede Entscheidungsfunktion kann dann das erwartete Risiko, \bar{R}_δ , in Bezug auf die a priori-Verteilung berechnet werden und ist somit unabhängig von dem tatsächlichen Wert des unbekanntem Parameters. Die beste Entscheidungsfunktion ist die, die das erwartete Risiko in Bezug auf die a priori-Verteilung minimiert. Das Risiko dieser Entscheidungsfunktion heißt Bayes-Risiko.

Diese Vorgehensweise lässt sich anhand eines für die Medizin typischen *Beispiels* verdeutlichen, in dem es nur zwei Zustände und zwei Entscheidungsalternativen gibt.⁴ Der Parameterraum Θ enthält nur zwei Elemente, θ_1 und θ_2 , so dass auch der Entscheidungsraum nur zwei Elemente hat, d_1 (θ_1 ist wahr; der Patient ist gesund) und d_2 (θ_2 ist wahr; der Patient ist krank). Zwischen diesen beiden Hypothesen gilt es zu entscheiden. Die Verlustfunktion sei $L_{ij} = L(d_i, \theta_j)$. Weil der Verlust aus der korrekten Entscheidung über den wahren Parameterwert geringer ist als der Verlust aus einer falschen Entscheidung, gilt $L_{11} < L_{21}$ und $L_{22} < L_{12}$.

Die a priori-Wahrscheinlichkeit seien π_1 und π_2 , und x sei das Ergebnis eines Tests. Dann gilt: Wenn $\delta(x) = d_1$ gewählt wird, ist der erwartete a priori-Verlust proportional zu $\pi_1 \cdot p(x|\theta_1) \cdot L_{11} + \pi_2 \cdot p(x|\theta_2) \cdot L_{12}$. Wenn $\delta(x) = d_2$ gewählt wird, ist der erwartete a priori-Verlust proportional zu $\pi_1 \cdot p(x|\theta_1) \cdot L_{21} + \pi_2 \cdot p(x|\theta_2) \cdot L_{22}$. Nach der Bayesianischen Theorie ist also $\delta(x) = d_1$ zu wählen, wenn $\pi_2 \cdot p(x|\theta_2) \cdot (L_{12} - L_{22}) < \pi_1 \cdot p(x|\theta_1) \cdot (L_{21} - L_{11})$. Bei umgekehrtem Ungleichheitszeichen ist $\delta(x) = d_2$ zu wählen.

Die Bayesianische Entscheidungsregel stellt somit einen *Likelihood Ratio-Test* dar. Der Annahmehbereich für θ_2 ist $\left\{ x : \frac{p(x|\theta_2)}{p(x|\theta_1)} > \frac{\pi_1(L_{21} - L_{11})}{\pi_2(L_{12} - L_{22})} \right\}$. Der Annahme-

bereich des Tests wird also durch die a priori-Wahrscheinlichkeiten und durch die relative Größe der möglichen Verluste, die Verluststruktur, bestimmt, nicht aber durch die absolute Größe der Verluste. Da im Allgemeinen nur die Festlegung einer a priori-Wahrscheinlichkeitsverteilung die Abhängigkeit der optimalen Entscheidungsfunktion von dem wahren Wert des unbekanntem Parameters aufheben kann, ist die Anwendung des Bayes-Theorems das einzige allgemein akzeptable Verfahren für rationale medizinische Entscheidungen. Zugleich ermöglicht es, Besonderheiten des einzelnen Falles, einschließlich besonderer Risiken, in konsistenter Weise zu berücksichtigen. Zwei Beispiele hierfür sind das gleichzeitige Auftreten mehrerer Krankheiten, aus dem veränderte a priori-Wahrscheinlichkeiten folgen können, und besondere Präferenzen des Patienten, etwa prohibitive Ängste vor einer Operation, die sich in einer veränderten Verluststruktur widerspiegeln.

Damit ist auch klar, wie die Anwendungsregel für einen medizinischen Test mit der Likelihood-Ratio $LR+$ lauten muss. Das a posteriori-Wahrscheinlichkeitsverhältnis für eine Krankheit D – das Produkt des a priori-Wahrscheinlichkeitsverhältnisses $\pi(D)/(1-\pi(D))$ und der Likelihood-Ratio für einen positiven Test ($LR+$) – muss größer sein als das Verhältnis aus dem Verlust der irrtümlichen Behandlung eines Gesunden zu dem Verlust aus der unterlassenen Behandlung eines Kranken: $\frac{\pi(D)}{1-\pi(D)} \cdot LR+ > \frac{U(D-A+) - U(D-A-)}{U(D+A-) - U(D+A+)}$. Dabei ist

$U(D+A+)$ der Nutzen des erkrankten Patienten aus Behandlung A , $U(D-A-)$ der Nutzen des Nicht-Erkrankten bei Nichtbehandlung, $U(D-A+)$ der Nutzen des Nicht-Erkrankten bei Anwendung der Behandlung A und $U(D+A-)$ der Nutzen des Erkrankten bei unterlassener Behandlung. Mit diesem Ansatz lässt sich auch erklären, wie der Annahmebereich einer Krankheitshypothese bei

⁴ Für eine ausführlichere Darstellung und für weitere Beispiele, siehe Silvey (1975), S. 168 ff.

einem gegebenen Test variiert, wenn ein neues Medikament die Heilungsaussichten im Falle der Krankheit verbessert – ohne dass größere Nebenwirkungen auftreten.

3. Die normative Analyse medizinischer Entscheidungen

Aus der Anwendung des Bayes-Theorems lässt sich eine normative Theorie medizinischer Entscheidungen ableiten, die erklärt, wie die Ergebnisse diagnostischer Tests zu interpretieren sind, wie mehrstufige Entscheidungen zu strukturieren sind, und nach welchen Kriterien über die Auswahl und Durchführung diagnostischer Tests zu entscheiden ist. Auf Pauker und Kassirer (1980) geht die Einsicht zurück, dass Kosten-Nutzen-Überlegungen die Grundlage medizinischer Entscheidungen bilden sollten. Ein Test ist danach nur sinnvoll, wenn daraus mit hinreichender Wahrscheinlichkeit eine *Verbesserung* des erwarteten Kosten-Nutzen-Verhältnisses der Behandlungsentscheidung folgt, die die Kosten des Tests rechtfertigt.

Die Komplexität dieser Zusammenhänge lässt sich am besten reduzieren, wenn das Entscheidungsproblem schrittweise, und zwar rekursiv, erklärt wird. Dieser Abschnitt beschreibt daher als Erstes, wie ein gegebenes Testergebnis zu interpretieren ist, als Zweites, wie die Entscheidung über eine therapeutische Intervention bei gegebenem Befund von ihren erwarteten Kosten und Nutzen abhängt, als Drittes, wie über die Durchführung eines gegebenen Tests zu entscheiden, und als Viertes, nach welchen Kriterien zwischen alternativen Tests auszuwählen ist. Grundlegend ist dabei die Bayesianische Einsicht, dass die Interpretation neuer Information, die zum Beispiel durch einen Test gewonnen wird, stets davon abhängt, was zuvor bereits über den Patienten bekannt war. Allerdings kann dieser normative Ansatz letztlich nicht erklären, wie ein Arzt seine a priori-Wahrscheinlichkeitsannahmen begründen sollte.

3.1 Die Interpretation diagnostischer Tests

Die praktische Bedeutung des Bayes-Theorems für die Interpretation eines gegebenen Testergebnisses lässt sich anhand von Schaubildern verdeutlichen, die das Zusammenspiel unterschiedlicher a priori-Wahrscheinlichkeiten einer Krankheit mit den statistischen Eigenschaften des Tests bei der Bestimmung der a posteriori-Wahrscheinlichkeiten darstellen. Wenn eine Diagnose schon fast sicher ist, hat ein bestätigender Test kaum noch Einfluss auf die a posteriori-Wahrscheinlichkeit. Die Widerlegung einer diagnostischen Hypothese dagegen kann die a posteriori-Wahrscheinlichkeit einer Krankheit ganz erheblich reduzieren. Von Bedeutung bei der Auswahl und Interpretation alternativer Testverfahren sind daher vor allem die Rate falscher positiver Ergebnisse (Eins minus die Spezifität eines Tests) und die Sensitivität (die Rate wahrer positiver Ergebnisse) der bereitstehenden Testverfahren.

Bei gleichen Risiken und Kosten bevorzugt ein Arzt immer den Test mit der höchsten Rate wahrer positiver Ergebnisse und der niedrigsten Rate falscher positiver Ergebnisse. Er sucht also den Test mit der höchsten Likelihood-Ratio. Leider sind diese Tests oft sehr teuer oder sogar gesundheitsschädlich. Häufig verlangen die besten Tests erhebliche körperliche Eingriffe, zum Beispiel das Aufschneiden des Patienten oder das Einführen eines radioaktiven Kontrastmittels für Röntgenaufnahmen. Es stellt sich daher in der Praxis oft die Frage, ob die Verwendung eines besseren Tests überhaupt zu einer klinisch bedeutsamen Änderung der a posteriori-Wahrscheinlichkeit führt.

Die Rate falscher positiver Testergebnisse (RFP) hat einen großen Einfluss auf die Interpretation eines positiven Testergebnisses: Je niedriger die Rate falscher positiver Testergebnisse, desto höher ist die a posteriori-Wahrscheinlichkeit der Krankheit. Gleichzeitig gilt jedoch: Wenn die a priori-Wahrscheinlichkeit einer Krankheit sehr niedrig ist, kann ein positives Testergebnis die Diagnose oft

nicht beeinflussen, denn in diesem Fall ist das Gewicht falscher positiver Ergebnisse besonders groß. Wenn andererseits die a priori-Wahrscheinlichkeit einer Krankheit sehr hoch ist, spielt die Rate falscher positiver Ergebnisse eines Tests keine große Rolle. Außerdem gilt generell, dass die Rate falscher positiver Testergebnisse keinen Einfluss auf die Interpretation eines negativen Testergebnisses hat.

Hinter diesen Aussagen steht die folgende Überlegung: Alle nicht erkrankten Patienten können ein falsches positives Testergebnis auslösen. Ist die a priori-Wahrscheinlichkeit niedrig, gibt es relativ viele nicht-erkrankte Patienten und die Zahl falscher positiver Testergebnisse wird die Zahl wahrer positiver Testergebnisse übertreffen. Ist die a priori-Wahrscheinlichkeit hoch, wird die Zahl der wahren positiven die der falschen positiven Testergebnisse übertreffen und die Rate der falschen positiven Testergebnisse hat nur noch einen kleinen Einfluss auf die Testinterpretation.

Schaubild 5 stellt die a posteriori-Wahrscheinlichkeit als Funktion der a priori-Wahrscheinlichkeit in Abhängigkeit eines positiven Testergebnisses bei unterschiedlichen Raten falscher positiver Testergebnisse dar. Bei gegebener a priori-Wahrscheinlichkeit einer Erkrankung ist die a posteriori-Wahrscheinlichkeit umso höher, je niedriger die Rate falscher positiver Testergebnisse (RFP) ist.

Die Sensitivität eines Tests ist vor allem dann wichtig, wenn es um die Falsifikation einer diagnostischen Hypothese geht: Die Rate wahrer positiver Testergebnisse hat – bei gegebener Rate falscher positiver Testergebnisse – einen großen Einfluss auf die Interpretation negativer Testergebnisse. Je höher die Rate wahrer positiver Testergebnisse, desto niedriger ist die a posteriori-Wahrscheinlichkeit einer Krankheit, weil dann der Fehler erster Art relativ klein ist. Ist jedoch die a priori-Wahrscheinlichkeit einer Krankheit hinreichend hoch,

schließt ein negatives Testergebnis die Krankheit nicht aus. Andererseits spielt bei einer niedrigen a priori-Wahrscheinlichkeit einer Krankheit, die Rate wahrer positiver Testergebnisse keine große Rolle, weil die Ablehnung der Diagnose dann schon vor dem Test nahezu feststeht. Generell gilt, dass die Rate wahrer positiver Testergebnisse nur einen kleinen Einfluss auf die Interpretation eines positiven Testergebnisses hat.

Schaubild 6 stellt die a posteriori-Wahrscheinlichkeit als Funktion der a priori-Wahrscheinlichkeit in Abhängigkeit eines negativen Testergebnisses bei unterschiedlichen Raten wahrer positiver Testergebnisse dar. Bei gegebener a priori-Wahrscheinlichkeit einer Erkrankung ist die a posteriori-Wahrscheinlichkeit umso niedriger, je höher die Rate wahrer positiver Testergebnisse (WP) ist.

3.2 Entscheidungen auf Basis des Erwartungswertes

Schon wenn mehr als ein diagnostischer Test zur Auswahl bereit steht, hat der Arzt ein zweistufiges Entscheidungsproblem zu lösen, das auf der ersten Stufe die Testauswahl und auf der zweiten Stufe die Festlegung der Therapie erfordert. Für mehrstufige Entscheidungsprobleme, wie sie in der Medizin oft vorkommen, sind *Entscheidungsbaume* die in der Praxis bevorzugte Methode zur Abbildung, Strukturierung und Evaluation der Entscheidungsalternativen. Dabei werden aus Kosten-Nutzen-Vergleichen Schwellenwahrscheinlichkeiten abgeleitet, um zu entscheiden, ob zusätzliche Informationen die Änderung einer gewählten therapeutischen Strategie begründen können.

Die wichtigsten Schritte zur Erstellung eines Entscheidungsbaumes sind die Definition des Entscheidungsproblems, die Identifikation der Entscheidungsalternativen, die Auflistung der möglichen klinischen Ergebnisse, die Verknüpfung der Entscheidungs- und Ereignisknoten, die Festlegung des Zeithorizontes, die Ermittlung der Wahrscheinlichkeit jedes Ereignisses und die Bestimmung des Wertes aller möglichen Ergebnisse. Die Berechnung des

Erwartungswertes erfolgt dabei für jede Entscheidungsalternative rekursiv. Es werden zunächst die Erwartungswerte der zuletzt zu treffenden Entscheidungen berechnet und jeweils als Auszahlungswerte für die Alternativen der vorher zu treffenden Entscheidungen eingesetzt. Dabei werden die Werte der einzelnen Ereignisse bei jedem Ereignisknoten mit ihren bedingten Wahrscheinlichkeiten gewichtet und aufsummiert. Abschließend sollte eine Sensitivitätsanalyse durchgeführt werden, um den Einfluss alternativer a priori-Annahmen über die Wahrscheinlichkeiten der Ereignisse, Bewertungen und kausalen Zusammenhänge auf die zu treffende Entscheidung zu quantifizieren.

Schaubild 7 stellt die wesentlichen Elemente eines medizinischen Entscheidungsbaumes dar. An dem quadratischen Entscheidungsknoten ist über eine Operation zu entscheiden. An den beiden runden Ereignisknoten zeigt sich dann mit jeweils unterschiedlichen Wahrscheinlichkeiten, ob der Patient überlebt (L) oder stirbt (†).

Schwellenwahrscheinlichkeit für eine Entscheidungsalternative. Die Schwellenwahrscheinlichkeit π^* (für eine Behandlung) ist definiert als die Erkrankungswahrscheinlichkeit, bei der ein Arzt indifferent zwischen Behandlung und Nicht-Behandlung ist. Pauker und Kassirer (1980) haben die Schwellenwahrscheinlichkeit als das Verhältnis der Kosten zur Summe der Kosten und Nutzen einer Behandlung definiert. Je höher die Kosten und je geringer der Nutzen, desto sicherer muss die Diagnose der Krankheit sein. Sowohl die Kosten als auch die Nutzen einer Behandlung können sich von Patient zu Patient deutlich unterscheiden, selbst wenn genau die gleiche Krankheit vorliegt. Da das Vorliegen einer bestimmten Erkrankung zum Zeitpunkt der Diagnose fast immer nur mit einer Wahrscheinlichkeit unter Eins festgestellt werden kann, sind der erwartete Nutzen aus Behandlung A

$EU(A+) = \pi U(D+A+) + (1-\pi)U(D-A+)$ und der erwartete Nutzen aus dem Verzicht auf Behandlung A $EU(A-) = \pi U(D+A-) + (1-\pi)U(D-A-)$ jeweils

kleiner als die Nutzenniveaus, die unter Sicherheit erreichbar wären. Indifferenz impliziert demnach

$$\pi U(D + A -) + (1 - \pi) U(D - A -) = \pi U(D + A +) + (1 - \pi) U(D - A +) \text{ bzw.}$$

$$U(D - A -) + \pi(U(D + A -) - U(D - A -)) = U(D - A +) + \pi(U(D + A +) - U(D - A +)).$$

Wird diese Indifferenz-Bedingung in jedem Fall erfüllt sein? Im Allgemeinen ja, denn die Erwartungsnutzen bei Behandlung und Nicht-Behandlung sind jeweils lineare Funktionen der a priori Erkrankungswahrscheinlichkeit, die sich in der Regel bei einem a priori Wahrscheinlichkeitswert in dem offenen Intervall zwischen Null und Eins schneiden. Die Existenz eines Schnittpunktes folgt daraus, dass der Nutzen aus Nicht-Behandlung bei voller Gesundheit in aller Regel mindestens so hoch ist wie der Nutzen aus Behandlung, während der Nutzen aus Nicht-Behandlung im Falle einer Erkrankung niedriger sein wird als der Nutzen aus Behandlung.

Die Schwellenwahrscheinlichkeit für eine Behandlung folgt dann aus der Gleichung, die den Schnittpunkt der beiden Funktionen definiert:

$$\pi^* (U(D + A +) - U(D + A -)) = (1 - \pi^*) (U(D - A -) - U(D - A +)) \text{ bzw.}$$

$$\frac{\pi^*}{1 - \pi^*} = \frac{U(D - A -) - U(D - A +)}{U(D + A +) - U(D + A -)}, \text{ und die sich nach } \pi^* \text{ auflösen lässt.}$$

Definiert man $C \equiv U(D - A -) - U(D - A +)$ als die Netto-Kosten der Behandlung gesunder Patienten und $B \equiv U(D + A +) - U(D + A -)$ als den Netto-Nutzen der Behandlung kranker Patienten, dann gilt $\pi^*/(1 - \pi^*) = C/B \Rightarrow \pi^*(B + C)/B = C/B \Rightarrow \pi^* = C/(C + B)$. Die Schwellenwahrscheinlichkeit einer Behandlung ist also umso *größer*, je größer die Kosten für gesunde Patienten und je geringer der Nutzen für kranke Patienten sind. Und sie ist umso *niedriger*, je größer der

Nutzen für kranke Patienten und je geringer die Kosten für gesunde Patienten sind.

Schaubild 8 stellt die Schwellenwahrscheinlichkeit als Funktion des Kosten-Nutzen-Verhältnisses einer Behandlung dar. Die Kurve hat eine positive, aber im Verlauf abnehmende Steigung, weil die Schwellenwahrscheinlichkeit unterproportional zum Kosten-Nutzen-Verhältnis zunimmt und asymptotisch höchstens gegen Eins gehen kann.

3.3 Entscheidung über die Durchführung eines diagnostischen Tests

In der Praxis ist die Frage, ob bei einem Patienten ein Test durchgeführt werden soll, oft verknüpft mit der Frage, welcher von mehreren alternativen Tests gewählt werden soll. Zunächst ist dann zwischen den Testalternativen auszuwählen, wobei vor allem zwei grundsätzliche Fragen zu klären sind: Welche Tests sind für welche Art von diagnostischen Hypothesen vorzuziehen? Und: Geht es um die Bestätigung oder die Widerlegung einer diagnostischen Hypothese? Erst danach kann die Frage analysiert werden, ob der bevorzugte Test überhaupt durchgeführt werden soll. Zur Einführung wird in diesem Unterabschnitt der einfachste Fall angenommen, in dem von Anfang an nur ein Test in Frage kommt. Die Alternativen sind dann, *erstens*, sofortige Behandlung, *zweitens*, Durchführung eines Tests, sowie *drittens*, kein Test und keine Behandlung. Die schwierigere Frage, wie zwischen alternativen Tests auszuwählen ist, wird erst im nachfolgenden Unterabschnitt behandelt.

Der erwartete Nutzen eines Tests ist die Summe der Nutzen im Falle (i) eines positiven Tests bei Vorliegen der untersuchten Krankheit, (ii) eines negativen Tests bei Vorliegen der untersuchten Krankheit, (iii) eines positiven Tests bei Nicht-Vorliegen der untersuchten Krankheit und (iv) eines negativen Tests bei Nicht-Vorliegen der untersuchten Krankheit, sowie (v) des direkten Nutzens aus der Testanwendung selbst, $U(T)$:

$$EU(\text{Test}) = \pi(D)\pi(+|D)U(D + A +) + \pi(D)\pi(-|D)U(D + A -) + (1 - \pi(D))\pi(+|\text{nicht } D)U(D - A +) + (1 - \pi(D))\pi(-|\text{nicht } D)U(D - A -) + U(T).$$

Der direkte Nutzen aus der Testanwendung ist in der Regel negativ, weil viele medizinische Tests mit erheblichem Zeitaufwand, Schmerzen oder sogar realen Gesundheitsrisiken verbunden sind. Die Nutzenniveaus der vier möglichen Kombinationen von Testergebnis und Gesundheitszustand unterscheiden sich, weil sie die Entscheidung des Arztes über die Behandlung oder Nicht-Behandlung seines Patienten widerspiegeln, die in Abhängigkeit des Testergebnisses unterschiedlich ausfällt. Die Nutzenniveaus werden jeweils mit der Wahrscheinlichkeit ihres Eintretens, also mit der Wahrscheinlichkeit zu erkranken bzw. gesund zu bleiben, gewichtet und dann zum Erwartungsnutzen aufsummiert.

Schaubild 9 stellt den Erwartungsnutzen des Tests als Funktion der a priori-Wahrscheinlichkeit einer Erkrankung dar. Bei einer a priori-Wahrscheinlichkeit von 0 oder 1 liegt der Erwartungsnutzen des Tests notwendigerweise zwischen den jeweiligen Erwartungsnutzen bei sofortiger Behandlung und bei Verzicht auf Behandlung, ohne einen Test durchzuführen. Nur wenn es im mittleren Bereich ein Intervall von a priori-Wahrscheinlichkeiten gibt, in dem der Erwartungsnutzen des Tests sowohl den Erwartungsnutzen aus sofortiger Behandlung als auch den Erwartungsnutzen aus Nicht-Behandlung übersteigt, kann es überhaupt eine Nachfrage nach dem Test geben. Jeder Test, der diese Bedingung erfüllt, hat zwei Schwellenwahrscheinlichkeiten. Die Schwelle zwischen Nicht-Behandlung und Test ist

$$\pi_1 = \frac{\pi(+|\text{nicht } D) \cdot C - U(T)}{\pi(+|\text{nicht } D) \cdot C + \pi(+|D) \cdot B} = \frac{RFP \cdot C - U(T)}{RFP \cdot C + SEN \cdot B} = \frac{(1 - SPE) \cdot C - U(T)}{(1 - SPE) \cdot C + SEN \cdot B}.$$

Die Schwelle zwischen Test und sofortiger Behandlung ist

$$\pi_2 = \frac{(1 - \pi(+|\text{nicht D})) \cdot C + U(T)}{(1 - \pi(+|\text{nicht D})) \cdot C + (1 - \pi(+|D)) \cdot B} = \frac{(1 - \text{RFP}) \cdot C + U(T)}{(1 - \text{RFP}) \cdot C + (1 - \text{SEN}) \cdot B} = \frac{\text{SPE} \cdot C + U(T)}{\text{SPE} \cdot C + (1 - \text{SEN}) \cdot B} .$$

Diese formalen Ausdrücke machen deutlich, dass die Höhe der beiden Schwellenwahrscheinlichkeiten nicht nur vom Verlauf der Erwartungsnutzenfunktionen abhängt, sondern auch von den Testeigenschaften Sensitivität und Spezifität. Die Schaubilder 10 und 11 stellen die Interdependenz zwischen der Behandlungsentscheidung und der Entscheidung zur Durchführung eines Tests mithilfe sogenannter Testindikationskurven (Bernstein 1997) graphisch dar, wobei die Rate falscher positiver Tests (Eins minus die Spezifität des Tests) als FPR (statt RFP) abgekürzt ist.

Schaubild 10 stellt die Abhängigkeit der Schwellenwahrscheinlichkeiten für die Durchführung eines Tests von den Testeigenschaften Sensitivität und Rate der falschen positiven Ergebnisse (FPR) dar, wie sie erstmals von Pauker und Kassirer (1980) abgeleitet wurde. Die beiden Kurven unterhalb der Hauptdiagonalen geben auf der Abzisse für zwei Tests mit unterschiedlicher Sensitivität ($\alpha > \beta$) jeweils die Schwellenwahrscheinlichkeit für die Entscheidung zwischen dem Test und sofortiger Behandlung an, die aus einer auf der Ordinate abgebildeten unmittelbaren und insofern „testunabhängigen“ Schwellenwahrscheinlichkeit für die Behandlungsentscheidung folgt.

Diese „testunabhängige“ Schwellenwahrscheinlichkeit ist für die Behandlungsentscheidung massgeblich, wenn entweder gar kein Test zur Verfügung steht oder der Test bereits durchgeführt wurde. Sie ist umso größer, je schlechter das Kosten-Nutzen-Verhältnis der Behandlung ist. Schaubild 10 berücksichtigt also die Kosten und Nutzen einer Behandlung, indem es die Schwellenwahrscheinlichkeiten für den Test als Funktion der Schwellenwahrscheinlichkeiten der unmittelbaren Behandlungsentscheidung darstellt. Das Schaubild

zeigt, dass die a priori-Wahrscheinlichkeit, ab der auf einen Test verzichtet und sofort behandelt werden kann, umso größer ist, je höher einerseits die „testunabhängige“ Schwellenwahrscheinlichkeit für die Behandlung und je höher andererseits die Sensitivität des zur Verfügung stehenden Tests ist.

Die Schwellenwahrscheinlichkeit für den Verzicht auf Test und Behandlung hängt von der Rate falscher positiver Testergebnisse (FPR) oder von Eins minus der Spezifität des Tests ab. Je kleiner die Rate falscher positiver Ergebnisse, desto geringer ist bei gegebener „testunabhängiger“ Schwellenwahrscheinlichkeit der Behandlung die Schwellenwahrscheinlichkeit für die Durchführung des Tests. Dies wird durch die beiden Kurven oberhalb der Hauptdiagonalen für die Raten falsch positiver Tests γ und δ dargestellt.

Die statistische Qualität des angebotenen Tests, die „testunabhängige“ Schwellenwahrscheinlichkeit der Behandlung und die a priori-Wahrscheinlichkeit der Erkrankung determinieren also gemeinsam, ob die Durchführung eines Tests rational ist. Im Allgemeinen wird ein Arzt dann einen Test durchführen, wenn sowohl die „testunabhängige“ Schwellenwahrscheinlichkeit als auch die a priori-Wahrscheinlichkeit der Krankheitshypothese im mittleren Bereich liegen.

Schaubild 11 nutzt die von Bernstein (1997) entwickelten Testindikationskurven, um die Schwellenwertanalyse von Pauker und Kassirer (1980) für einen gegebenen Test noch besser zu verdeutlichen. Ausgangspunkt ist wiederum die „testunabhängige“ Schwellenwahrscheinlichkeit für die Behandlung des Patienten, die als horizontale Linie I beispielhaft in Höhe von 0,5 eingezeichnet ist. Um daraus Schwellenwahrscheinlichkeiten für die Anwendung des Tests abzuleiten, machen sich die Testindikationskurven zunutze, dass die funktionale Beziehung zwischen der a posteriori-Wahrscheinlichkeit und der a priori-Wahrscheinlichkeit bei gegebener Sensitivität und Spezifität eines Tests

unabhängig von den besonderen Gesundheitsproblemen und Präferenzen des einzelnen Patienten ist.

Die Testindikationskurven bilden einfach alle Lösungen des Bayes-Theorems für die gesamte Menge der möglichen a priori-Wahrscheinlichkeiten einer Krankheit ab. Die Schnittpunkte der Linie I mit den beiden Testindikationskurven definieren die Schwellenwahrscheinlichkeiten II für den Verzicht auf Test und Behandlung (im Falle von a priori-Wahrscheinlichkeiten links von II und unterhalb von I, also im Rechteck A_1) und III für die sofortige Behandlung (die bei allen Kombinationen von a priori- und a posteriori-Wahrscheinlichkeiten in dem Rechteck C_1 angezeigt ist). Nur im Bereich mittlerer a priori-Wahrscheinlichkeiten hat das Ergebnis des Tests einen Einfluss darauf, ob die Behandlung durchgeführt wird (B_1) oder nicht (B_0).

In einem anwendungsorientierten Schaubild hat Glasziou (1991) die Ergebnisse von Pauker und Kassirer (1980) mithilfe des von Fagan (1975) entwickelten „Nomogram“ dargestellt. Das „Nomogram“ ist ein einfaches Instrument zur praktischen Berechnung von a posteriori-Wahrscheinlichkeiten. Es besteht aus drei (senkrechten) Skalen, auf denen rechts die a priori-Wahrscheinlichkeit $P(D)$ zwischen 0,99 und 0,001, links die a posteriori-Wahrscheinlichkeit $P(D|T)$ zwischen 0,001 und 0,99 sowie in der Mitte die Likelihood-Ratio $P(T|D)/P(T|\bar{D})$ zwischen 0,001 und 1000 abgebildet ist. Die Skalen für die a priori- und die a posteriori-Wahrscheinlichkeiten verlaufen also in umgekehrter Richtung.

Die Likelihood-Ratio ist so skaliert, dass der Arzt mithilfe eines Lineals, das die rechte und die mittlere Skala jeweils in Höhe der gegebenen a priori-Wahrscheinlichkeit und der gegebenen Likelihood-Ratio schneidet, automatisch auf der linken Skala die a posteriori-Wahrscheinlichkeit findet. Um die Schwellenwahrscheinlichkeiten für die Testentscheidung abzuleiten, wird der Spieß einfach umgedreht. Von einer gegebenen „testunabhängigen“ Schwellen-

wahrscheinlichkeit auf der linken Skala aus wird eine Gerade durch die Likelihood-Ratio für einen positiven Test und eine andere Gerade durch die Likelihood-Ratio für einen negativen Test gezogen. Die Schnittpunkte dieser Geraden mit der rechten Skala kennzeichnen die gesuchten Schwellenwahrscheinlichkeiten. Auch dabei ist das Ziel, diagnostische Tests nur noch bei jenen Patienten durchzuführen, die wirklich einen Vorteil davon haben, weil die optimale Behandlungsentscheidung tatsächlich von dem Ergebnis des Tests abhängt.

3.4 Auswahl zwischen alternativen Tests

Nachdem klar ist, dass die Anwendung diagnostischer Tests situationsabhängig ist, müssen Kriterien entwickelt werden, nach denen die Vorteilhaftigkeit alternativer Tests verglichen werden kann. Richter und Lange (1997) unterscheiden in diesem Zusammenhang zwischen den technischen und den diagnostischen Grundlagen, wobei die technischen Voraussetzungen Präzision, Richtigkeit und Reproduzierbarkeit, die sie unter dem Oberbegriff Zuverlässigkeit aufführen, lediglich notwendige, aber nicht hinreichende Bedingungen für einen validen diagnostischen Test sind. Letztlich entscheidend sind die diagnostischen Kriterien der „Güte“ eines Tests. Dabei müssen die Sensitivität und Spezifität, die die Fähigkeit eines Tests beschreiben, korrekt zwischen erkrankten und nicht erkrankten Patienten zu unterscheiden, durch zusätzliche Kennzahlen ergänzt werden, um die *Effizienz* eines Tests zu beurteilen. So kombiniert der prädikative Wert eines Tests die Eigenschaften Sensitivität, Spezifität und die Prävalenz der Krankheit, um die Bedeutung von Testergebnissen bei einer vorgegebenen Schwellenwahrscheinlichkeit für die Behandlungsentscheidung zu bestimmen: Bei einem positiven Testergebnis beschreibt der prädikative Wert den Anteil der wahren positiven an allen positiven Testergebnissen – $a/(a+b)$ in Tabelle 1 –, bei einem negativen Test den Anteil der wahren negativen an allen negativen Testergebnissen – $d/(c+d)$ in Tabelle 1.

Eine Verallgemeinerung ist das Konzept der statistischen Effizienz, die als der Anteil wahrer Testergebnisse an allen Testergebnissen definiert ist, aber auch stets nur bei einer vorgegebenen Schwellenwahrscheinlichkeit für die Behandlungsentscheidung und Prävalenz der Erkrankung gilt: $(a+d)/(a+b+c+d)$ in Tabelle 1. Da die statistische Effizienz gewissermaßen den prädikativen Wert positiver und negativer Tests kombiniert, teilt sie als Gütekriterium bei der Testauswahl sowohl die Stärken als auch die Schwächen des prädikativen Wertes: Wenn die Kontextfaktoren – die Prävalenz der Krankheit und die Schwellenwahrscheinlichkeit der Behandlungsentscheidung – feststehen, bietet der prädikative Wert zwar mehr Information als die Sensitivität und die Spezifität eines Tests. In einzelnen Situationen aber kann der prädikative Wert irreführend sein, weil er die fallspezifischen Kosten falscher positiver Ergebnisse und falscher negativer Ergebnisse nicht berücksichtigt.

Um diese Kosten explizit zu berücksichtigen, ist auf Grundlage der statistischen Entscheidungstheorie das Konzept der Receiver Operating Characteristics (ROC) entwickelt worden, das bei Zweig und Campbell (1993) ausführlich erläutert wird. Die ROC beschreiben die Güte eines Tests durch seine diagnostische Genauigkeit – die Fähigkeit, Patienten klinisch relevanten Untergruppen zu zuordnen. Dies ist die grundlegendste Eigenschaft eines Tests. Ihre graphische Darstellung kann dazu benutzt werden, einen ökonomisch sinnvollen Test auszuwählen, der durch eine bestimmte Kombination von Sensitivität und Spezifität gekennzeichnet ist. Um den Trade-off zwischen der Rate wahrer positiver und der Rate falscher positiver Ergebnisse transparent zu machen, bildet die sogenannte ROC-Kurve alle paarweisen Kombinationen von Sensitivität und Spezifität ab, die aus einer kontinuierlichen Variation der Schwellenwahrscheinlichkeit für die Behandlungsentscheidung resultieren. Der bei einem gegebenen Testverfahren unvermeidliche Trade-off zwischen der Sensitivität und der Spezifität wird einerseits durch die relativen Kosten

falscher positiver und falscher negativer Entscheidungen sowie der Nutzen aus korrekten Entscheidungen und andererseits durch die relative a priori-Wahrscheinlichkeit der beiden Gesundheitszustände determiniert, zwischen denen der Test diskriminieren soll.

Bei vielen medizinischen Tests überlappen sich die Verteilungen der Teststatistik für erkrankte und nicht erkrankte Testpersonen zumindest teilweise. Wenn sie sich vollständig überlappen, wie es in Schaubild 12a dargestellt ist, hat der Test gar keinen Aussagewert. Wenn sie sich überhaupt nicht überlappen, kann der Test perfekt zwischen erkrankten und nicht erkrankten Testpersonen unterscheiden (Schaubild 12b). Die Rate falscher positiver Testergebnisse sowie die Rate falscher negativer Testergebnisse hängt unmittelbar von der Größe des Überlappungsbereichs ab. Wenn ein Überlappungsbereich existiert, muss ein Grenzwert für den Test festgelegt werden, der gesunde von kranken Testpersonen abgrenzt. Hierbei ist der bereits erwähnte Trade-off relevant. Um zu verhindern, dass Gesunde irrtümlich als krank klassifiziert werden, muss man hinnehmen, dass dann nicht alle erkrankten Testpersonen als krank erkannt werden. Umgekehrt gilt: Will man alle erkrankten Testpersonen als krank erkennen, muss man eine höhere Rate falscher positiver Ergebnisse in Kauf nehmen. Die Variation des Grenzwertes für einen solchen Test ist tendenziell eine Art Nullsummenspiel, wie es die beiden Schaubilder 12c und 12d illustrieren.

Um mithilfe der ROC-Kurve den jeweils optimalen Test auszuwählen, müssen zunächst die Rate wahrer positiver Testergebnisse und die Rate falscher positiver Testergebnisse für jeden möglichen Grenzwert des Tests berechnet werden. Wie in Schaubild 12e wird die Rate wahrer positiver Tests (SEN) üblicherweise auf der Ordinate und die Rate falscher positiver Tests (FPR) üblicherweise auf der Abzisse abgetragen. Die 45-Grad-Linie definiert den geometrischen Ort für alle Testergebnisse, bei denen die Rate wahrer positiver

Ergebnisse genau gleich der Rate falscher positiver Ergebnisse ist. Da die Likelihood-Ratio in diesem Fall 1,0 ist, haben die Ergebnisse eines solchen Tests keinen Aussagewert bei der Bestimmung der a posteriori-Wahrscheinlichkeit für eine Krankheit. Für alle Testergebnisse oberhalb der 45-Grad-Linie ist die Likelihood-Ratio *größer* als 1,0 und der Test wird die a posteriori-Wahrscheinlichkeit der Krankheit erhöhen. Für alle Testergebnisse unterhalb der 45-Grad-Linie ist die Likelihood-Ratio *kleiner* als 1,0, so dass diese Testergebnisse die a posteriori-Wahrscheinlichkeit der Krankheit im Vergleich zur a priori-Wahrscheinlichkeit verringern.

Zwei oder mehr zur Auswahl stehende Testverfahren können mithilfe der ROC-Kurve direkt verglichen werden: Das Testverfahren mit der größten Fläche unterhalb der ROC-Kurve ist der bevorzugte Test. Um eine ökonomisch effiziente Definition des Grenzwertes zwischen negativem und positivem Testbefund festzulegen, müssen als fallspezifische Charakteristika die a priori-Wahrscheinlichkeiten für die zu testenden Hypothesen und die relativen Kosten falscher negativer und falscher positiver Testergebnisse berücksichtigt werden. Der Grenzwert für einen positiven Testbefund wird dabei so festgelegt, dass bei seinem Überschreiten die Behandlung des Patienten auf die getestete Krankheit gerade vorteilhaft wird.

Wenn zum Beispiel die relativen Kosten falscher positiver Ergebnisse sehr viel größer sind als die Kosten falscher negativer Ergebnisse, ist die Spezifität des Tests höher zu gewichten als die Sensitivität. Aber das *optimale* Verhältnis von Sensitivität und Spezifität kann erst festgelegt werden, wenn die Prävalenz der Krankheit bekannt ist, denn sie bestimmt gemeinsam mit der Sensitivität und Spezifität die tatsächliche Wahrscheinlichkeit falscher positiver und falscher negativer Ergebnisse in einer gegebenen Testpopulation. Die Steigung einer Tangentialkurve, die die Kosten falscher positiver Ergebnisse, die Kosten falscher negativer Ergebnisse und die Prävalenz der Erkrankung widerspiegelt,

ist durch die Formel $m = (C/B) \cdot (1 - \pi(D)) / \pi(D)$ definiert und legt im Tangentialpunkt auf der ROC-Kurve den Schwellenwert für die Testentscheidung fest.

Der jeweils optimale Grenzwert für ein positives Testergebnis wird also dadurch gefunden, dass die ROC-Kurve tangential zu einer Geraden verlaufen muss, deren Steigung *positiv* von den Netto-Kosten der Behandlung Gesunder, $C \equiv U(D - A -) - U(D - A +)$ (den Kosten falscher positiver Ergebnisse), multipliziert mit der Wahrscheinlichkeit des Nicht-Vorliegens der Krankheit, abhängt und *negativ* von den Netto-Nutzen der Behandlung Erkrankter, $B \equiv U(D + A +) - U(D + A -)$ (den Kosten falscher negativer Ergebnisse), multipliziert mit der Wahrscheinlichkeit für das Vorliegen der Krankheit. Die Steigung dieser Tangente ist umso größer, je kleiner die a priori-Wahrscheinlichkeit der Krankheit und je größer die Kosten der Behandlung Nicht-Erkrankter im Verhältnis zu den Netto-Nutzen der Behandlung erkrankter Testpersonen sind. Jeder Grenzwert wird dabei wie ein separater Test mit seiner eigenen Likelihood-Ratio betrachtet.⁵

Bei einer Reihenuntersuchung zur Früherkennung einer Krankheit unter asymptomatischen Testpersonen (Schaubild 12c) ist der Grenzwert für einen positiven Testbefund im Allgemeinen sehr viel höher als bei einem Test, der ausschließlich auf symptomatische Patienten in der Akutmedizin angewandt wird (Schaubild 12d). Denn die Likelihood-Ratio muss bei einem Früherkennungstest sehr viel höher sein, damit ein positives Testergebnis die niedrige a priori-Wahrscheinlichkeit der Krankheit bei asymptomatischen Testpersonen in eine hinreichend große a posteriori-Wahrscheinlichkeit transformiert, um die Testentscheidung bei einer gegebenen Schwellenwahrscheinlichkeit für die Behandlung zu rechtfertigen. Dafür wird oft eine sehr viel höhere Rate falscher negativer Testergebnisse in Kauf genommen, als es in der Akutmedizin akzeptabel

⁵ Vergleiche hierzu die Ableitung der Bayesianischen Entscheidungsregel in Abschnitt 2.3.

wäre. In der Akutmedizin sind die Kosten der Nicht-Behandlung einer Krankheit typischerweise sehr viel höher als die Kosten des Nicht-Erkennens eines asymptomatisch Erkrankten in der sekundären Prävention.

4. Gesundheitspolitische Implikationen medizinischer Praxisvariationen

Eine Reihe von empirischen Studien⁶ aus verschiedenen Ländern mit ganz unterschiedlichen Gesundheitssystemen hat gezeigt, dass es bei vielen medizinischen Leistungen auffällige Unterschiede in der relativen Anwendungshäufigkeit zwischen Regionen gibt. Diese Unterschiede lassen sich nur zum Teil durch die Varianz der relevanten sozialen und ökonomischen Einflussgrößen – Gesundheitsstatus, Bildungsniveau, Einkommen, Versicherungsumfang und andere Faktoren – erklären, die als Determinanten der Nachfrage des einzelnen Patienten nach medizinischen Leistungen gelten können. Die Evidenz spricht vielmehr dafür, dass die Unsicherheit der Ärzte über die korrekte Diagnose und die beste Behandlungsstrategie für einen großen Teil der auffälligen interregionalen Varianz in der Anwendungsrate einzelner medizinischer Leistungen verantwortlich ist. Demnach haben sich bei medizinischen Leistungen, die nach einer relativ unsicheren Diagnose zum Einsatz kommen oder über deren Wirksamkeit die Meinungen weit auseinandergehen, unterschiedliche ärztliche Ansichten und Praktizierstile herausgebildet.⁷

Idiosynkratische Praktizierstile und medizinische „Denkschulen“ können durch die oft intensive lokale Kommunikation mit den Kollegen innerhalb einer Region verstärkt werden. Damit ist eine empirisch überprüfbare Hypothese zu

⁶ Unter anderem McPherson et al. (1981), Roos (1984), Vayda (1973) und Wennberg et al. (1987).

⁷ Der Anteil der in Deutschland angewandten diagnostischen und therapeutischen Verfahren, die lediglich auf unsystematischen Erfahrungen beruhen, wird vielfach als hoch eingeschätzt. Berger et al. (1997) zitieren Autoren, denen zufolge nur etwa 10 bis 30 Prozent aller Patienten in Übereinstimmung mit Ergebnissen wissenschaftlicher Forschung behandelt werden.

der Frage eingeführt, woher die a priori-Wahrscheinlichkeitsannahmen über die relative Häufigkeit einzelner Krankheiten und über die Erfolgsaussichten alternativer therapeutischer Methoden stammen. Im Folgenden werde ich als Erstes die theoretische Analyse der Wohlfahrtswirkungen medizinischer Praxisvariationen und die wichtigsten Ergebnisse empirischer Schätzungen darstellen, als Zweites die Rolle öffentlicher Güter in der Absorption medizinischen Wissens herausarbeiten, als Drittes die Rolle von Behandlungsleitlinien bei der Diffusion medizinischer Innovationen verdeutlichen und schließlich als Viertes die Steuerungsmöglichkeiten der Gesundheitspolitik im Überblick diskutieren.

4.1 Wohlfahrtswirkungen

Um die Wohlfahrtswirkungen regionaler Praxisvariationen zu quantifizieren und gesundheitspolitische Schlussfolgerungen zu begründen, müssen die theoretischen Erklärungsansätze dieses Phänomens präzisiert und gegeneinander getestet werden. Den bislang überzeugendsten Ansatz bieten Phelps und Mooney (1993). Sie haben einen Bayesianischen Lernprozess formal modelliert, der die Entstehung regional begrenzter „Schulen“ konsistent erklären kann. Dabei bildet sich die Meinung des einzelnen Arztes über die „korrekte“ Anwendungsrate einer medizinischen Leistungen aus dem in der Ausbildung erworbenen Wissen, aus praktischer Erfahrung und aus anderen Informationsquellen, vor allem von Kollegen vor Ort. Die Anpassung der einzelnen Meinung an die vorherrschende „Schule“ einer Region ergibt sich in dem Modell aus der rationalen Minimierung einer Verlustfunktion, deren Argument die Differenz zwischen mitgebrachter Auffassung und der korrekten Anwendungsrate ist.

Die vorherrschende „Schule“ in einer Region kann als soziales Kapital interpretiert werden, das die Kosten der Kommunikation und der Koordination zwischen den Ärzten und zwischen Ärzten und den anderen Beteiligten im Prozess der medizinischen Leistungserstellung, einschließlich der Patienten,

senkt.⁸ Ohne ein solches Sozialkapital müsste bei einer Vielzahl einzelner Kooperationsvorgänge zwischen Fachärzten und Allgemeinärzten, Krankenhäusern und Pflegepersonal jeweils explizit vereinbart werden, welche medizinische Lehre im Detail gelten soll. Die Einhaltung dieser Vereinbarungen müsste dann im Interesse der Patienten kontinuierlich überwacht werden. Es ist klar, dass die Kosten detaillierter Vereinbarungen, die sich vor allem aus den Kosten der Vertragsformulierung und den Kosten der Überwachung zusammensetzen, der Arbeitsteilung Grenzen setzen können. Bei einem hinreichend großen Sozialkapital reichen oft *unvollständige* Verträge aus, so dass sich die Kosten detaillierter Vereinbarungen reduzieren oder ganz vermeiden lassen.

Je niedriger aber die Kosten der Koordination, desto weiter kann die Arbeitsteilung voranschreiten und desto stärker können sich Ärzte spezialisieren und Lernerträge auf einem eng abgegrenzten Gebiet realisieren. Zusätzlich zu den statischen Effizienzgewinnen aus verstärkter Arbeitsteilung kann auch die Vertiefung des medizinischen Wissens durch Spezialisierung die Produktivität des Gesundheitswesens in einer Region insgesamt erhöhen. Die Gesundheitspolitik muss daher beachten, dass die soziale Interaktion im Bayesianischen Lernen der Ärzte ökonomische Komplementaritäten entstehen lässt, die von großer Bedeutung für langfristige Investitionsentscheidungen und für die dynamische Effizienz der regionalen Gesundheitsversorgung sein können.

Bayesianisches Lernen in einem binomialen Ereignismodell. Phelps und Mooney (1993) nehmen an, dass die Ärzte eine quadratische Verlustfunktion minimieren, deren Argument die Differenz zwischen den eigenen a priori Annahmen und dem korrekten Parameter θ ist. Um das Modell möglichst einfach zu halten, postulieren sie für die subjektiven Annahmen der Ärzte über die korrekte Anwendungsrate θ eine Beta-Verteilung:

⁸ Siehe Becker und Murphy (2000) für eine ausführliche theoretische Analyse der Entstehung und der Auswirkungen von Sozialkapital in einer Volkswirtschaft.

$h(\theta) = \Gamma(\alpha + \beta) \theta^{\alpha-1} (1 - \theta)^{\beta-1} / (\Gamma(\alpha)\Gamma(\beta))$ mit $0 < \theta < 1$. Die Parameter α und β bilden dabei gewissermaßen das Lehrbuchwissen oder die in der Ausbildung erworbene Erfahrung eines Arztes ab, so dass $\alpha/(\alpha + \beta)$ die angenommene korrekte Anwendungsrate darstellt. Neue Erfahrungen, die von α und β abweichen, werden den einzelnen Arzt veranlassen, seine Annahmen sukzessive an die Norm seiner Umgebung anzupassen. Denn ein Großteil neuer Erfahrungen spiegelt die medizinische Praxis seiner Kollegen wider, die eine relativ kostengünstige Quelle zusätzlicher Informationen darstellt.

Hat ein junger Arzt bis zu einem bestimmten Zeitpunkt beispielsweise N Anwendungen seiner Kollegen, X_1, \dots, X_N , beobachtet, so dass $y = \sum_N X_i$ mit $X_i = 1$ oder $X_i = 0$, dann ist die den erwarteten Verlust minimierende

Anwendungsrate $w(y) = (\alpha + y)/(\alpha + \beta + N) = \frac{(\alpha + \beta)}{(\alpha + \beta + N)} \frac{\alpha}{\alpha + \beta} + \frac{N}{(\alpha + \beta + N)} \frac{y}{N}$. Der

Arzt passt seine eigenen Annahmen über die korrekte Anwendungsrate einer medizinischen Intervention umso schneller an die Norm seiner Umgebung an, je stärker sich diese Norm von dem in der Ausbildung Gelernten unterscheidet und je umfangreicher die Erfahrung ist, die der Arzt während seiner medizinischen Praxis sammelt. Zwar werden stets auch andere Quellen neuer Information einen gewissen Einfluss haben, aber gerade die zuverlässigsten Informationsquellen – wissenschaftliche Zeitschriften und Konferenzen – erfordern vom einzelnen Arzt erhebliche Investitionen, vor allem seiner knappen Zeit.

Das binomiale Ereignismodell ist leicht einsichtig, wenn ärztliche Entscheidungen diskret sind, wenn es also darum geht, ob eine Behandlung durchgeführt wird oder nicht. Man spricht auch von der extensiven Marge. Auf der intensiven Marge dagegen, bei der es um die Intensität einer Behandlung geht, sind die Entscheidungen über einen kontinuierlichen Bereich definiert. Auch diesen Fall kann aber ein Bayesianische Lernmodell nach geringfügigen Modifikationen abbilden, wie sie Phelps und Mooney (1993), S. 157, erläutern.

Ursachen unterschiedlicher Lehrmeinungen in der Medizin. Phelps und Mooney (1993) diskutieren eine Reihe von möglichen Gründen, warum schon die universitäre Ausbildung von Medizinern durch unterschiedliche Schulen geprägt ist. Der wichtigste Grund scheint darin zu liegen, dass relativ wenige medizinische Interventionen, die im Alltag praktiziert werden, wirklich einer sorgfältigen empirischen Untersuchung ihrer Wirksamkeit unterzogen wurden. Ethisch gilt es in der Regel als nicht vertretbar, einer Kontrollgruppe von Patienten eine bereits allgemein akzeptierte medizinische Leistung vorzuenthalten, um daraus Schlüsse über deren Wirksamkeit ziehen zu können. Zudem gibt es keine privaten Eigentumsrechte an dem Wissen, das durch systematische empirische Untersuchungen der Wirksamkeit medizinischer Intervention erzeugt würde. Der finanzielle Anreiz, solche Untersuchungen durchzuführen und ihre Ergebnisse zu veröffentlichen, ist daher gering. Hinzu kommt, dass es auch keine juristische Verantwortung für die systematische Anwendung falscher medizinischer Praktiken gibt, denn ärztliches Fehlverhalten muss jeweils im Einzelfall von dem betroffenen Patienten nachgewiesen werden, der gegen einen Arzt vor Gericht ziehen will. Gerichte entscheiden niemals grundsätzlich über die Zulässigkeit oder Vorteilhaftigkeit einer allgemein akzeptierten medizinischen Praxis, sondern ausschließlich darüber, ob einem einzelnen Patienten schuldhaft geschadet wurde.

Regional unterschiedliche Ausprägungen der medizinischen Praxis müssen nicht zwangsläufig auf die Präsenz einer medizinischen Fakultät an einer regionalen Universität zurückzuführen sein. Zur Entstehung lokaler „Schulen“ trägt auch die zunehmende Verbreitung der ärztlichen Gemeinschaftspraxis und das Verhalten der Patienten bei. Wenn ein junger Arzt sich als Mitglied einer Gemeinschaftspraxis niederlässt, kommt es automatisch zu einem Auswahlprozess, bei dem die a priori-Annahmen über die korrekte Anwendung medizinischer Praktiken eine wichtige Rolle spielen. Kein Arzt wird in einer

Gemeinschaftspraxis mit Kollegen zusammenarbeiten wollen, die bei wichtigen medizinischen Fragen konträre Auffassungen vertreten.

Auch das Verhalten der Patienten ist oft durch Einstellungen und Vorurteile geprägt, die die Entstehung regionaler „Schulen“ begünstigen. Viele Patienten suchen vor größeren Interventionen eine zweite Meinung von einem anderen Arzt in der Nachbarschaft und entwickeln dabei besonderes Misstrauen gegen Ärzte, deren Auffassung stark von der Norm in der Umgebung abweicht. Dies wiederum setzt die Ärzte in ihrem Wettbewerb um Patienten unter einen gewissen Anpassungsdruck. Ein Beispiel hierfür ist die Homöopathie, die vermutlich nicht nur von echten Homöopathen, sondern auch von Ärzten angeboten wird, die selbst nicht von ihrer Wirkung überzeugt sind, die aber den Weggang von Patienten fürchten, die daran glauben.⁹ Auf Seiten der Patienten sind vielfältige Formen der sozialen Interaktion vorstellbar, die die Herausbildung von kollektiven Präferenzen im Sinne der Theorie von Becker und Murphy (2000) erklären können.

Empirische Wohlfahrtsschätzungen. Neben der Arbeit an theoretischen Erklärungsansätzen haben Ökonomen auch versucht, die Wohlfahrtsverluste aus regionalen Praxisvariationen ökonometrisch zu quantifizieren und grob mit den Wohlfahrtsverlusten aus anderen Formen des Marktversagens im Gesundheitswesen zu vergleichen. Phelps und Parente (1990) haben dazu ein einfaches Modell entwickelt, in dem medizinische Leistungen zu konstanten Grenzkosten angeboten werden, während der Grenznutzen bei einer zunehmenden Anwendungsrate einen fallenden Verlauf hat. Diese Annahme ist plausibel, weil nicht nur die individuelle Bewertung eines gegebenen Heilerfolgs gestreut ist, sondern auch die Verträglichkeit einer Maßnahme und die Schwere möglicher

⁹ Linde (1999) stellt fest, dass ein „überzeugender Wirksamkeitsnachweis durch unabhängig reproduzierte, methodisch hochwertige Untersuchungen“ für homöopathische Therapieansätze nicht vorliegt.

Nebenwirkungen von Patient zu Patient variiert. Zur Vereinfachung kann der Grenznutzen einer medizinischen Leistung mit der marginalen Zahlungsbereitschaft gleich gesetzt werden.

Eine Zwei-Regionen-Version dieses Modells mit linearen Grenznutzenkurven ist in Schaubild 13a dargestellt. Der Schnittpunkt der Kurve der wahren Grenznutzen $V^*(x)$ mit den Grenzkosten GK bestimmt die korrekte Anwendungsrate der medizinischen Leistung in beiden regional abgegrenzten Bevölkerungsgruppen. In einer der Regionen aber schätzen die Ärzte die Grenznutzen zu hoch (hier durch $V_2(x)$ dargestellt), in der anderen zu niedrig ($V_1(x)$) ein und wenden daher die Leistung zu oft beziehungsweise zu selten an. Das Dreieck unterhalb von $V^*(x)$ und oberhalb von GK im Bereich der Anwendungsraten zwischen X_1 und X^* stellt den Wohlfahrtsverlust aufgrund der zu geringen Anwendungsrate $V_1(x)$ in Region 1 dar; Patienten, deren marginale Zahlungsbereitschaft über den Grenzkosten liegt, werden nicht behandelt. Analog stellt das Dreieck unterhalb von GK und oberhalb von $V^*(x)$ den Wohlfahrtsverlust aus der zu hohen Anwendungsrate $V_2(x)$ in Region 2 dar; Patienten, deren marginale Zahlungsbereitschaft unter den Grenzkosten liegt, werden dennoch behandelt.

Ausgehend von einem linearen Verlauf der Grenznutzenkurven haben Phelps und Parente (1990) den Wohlfahrtsverlust durch die Formel

$W = 0,5 \cdot M(x) \cdot CV^2 / |\eta(x)|$ approximiert. Darin bezeichnet CV den Variationskoeffizienten der nicht-korrekten Anwendung, $\eta(x)$ den absoluten Betrag der Preiselastizität der Nachfrage (also der Grenznutzenfunktion) und $M(x)$ die gesamten Ausgaben für die betrachtete medizinische Leistungsart x über alle Regionen eines Landes oder Untersuchungsraumes. Der Variationskoeffizient $CV = \sigma(r)/E(r)$ ist definiert als Quotient aus Standardabweichung und Mittelwert, wobei diese Werte aus den Residuen r einer Regression der Anwendungsrate

auf die relevanten sozialen und ökonomischen Erklärungsvariablen errechnet wurden. Die Formel zeigt, dass der Wohlfahrtsverlust mit dem Variationskoeffizienten und den Gesamtausgaben steigt, aber bei einer höheren Preiselastizität der Nachfrage abnimmt. Dies ist in der graphischen Darstellung leicht zu erkennen, wenn man sich klar macht, wie sich in Schaubild 13a die Größe der beiden Dreiecke unterhalb und oberhalb von $v^*(x)$ bei einer Verschiebung oder Drehung der Grenznutzenkurve ändert.

Phelps und Mooney (1990) nehmen an, dass die im Durchschnitt beobachtete Anwendungsrate einer medizinischen Leistung die korrekte ist, und haben den Wohlfahrtsverlust für die 25 verlustreichsten medizinischen Leistungen des Jahres 1987 in den USA auf über 6 Milliarden Dollar (\$) geschätzt: 1,2 Prozent der US-Gesundheitsausgaben in Höhe von insgesamt \$ 500 Milliarden. Im Jahre 2000 belief sich der entsprechende Wohlfahrtsverlust aus medizinischen Praxisvariationen einer Schätzung von Phelps (2000) folgend auf ca. \$ 11 Milliarden.

Diese Schätzungen sollten mit Vorsicht betrachtet werden. In erster Linie dienen sie dazu, die Größenordnung des Problems zu verdeutlichen. Die zugrundeliegende Methodik ließe sich in mehrererlei Hinsicht weiterentwickeln und verfeinern. Phelps und Mooney (1993) selbst weisen darauf hin, dass empirische Studien den Wohlfahrtsverlust aus regionalen Praxisvariationen tendenziell überschätzen, wenn die Varianz der Anwendungsrate gemessen wurde, *bevor* alle exogenen Einflüsse auf regionale Abweichungen von der durchschnittlichen Anwendungsrate herausgerechnet wurden. Andererseits würde der Wohlfahrtsverlust unterschätzt, wenn es nicht nur *interregionale*, sondern auch *intra*regionale Praxisvariationen geben sollte. Gerade in Zeiten eines raschen Technologiewandels in der Medizin werden sich das Adoptionsverhalten und die daraus resultierenden Ergebnisse im Bayesianischen Lernprozess auch innerhalb einzelner Regionen unterscheiden, etwa von Krankenhaus zu Krankenhaus oder von Arzt zu Arzt.

Dies verdeutlichen Phelps und Mooney (1993), indem sie die Wohlfahrtsverluste aus *interregionalen* Variationen in der Rate der Krankenhauseinweisungen, der extensiven Marge, und aus *intraregionalen* Variationen in der von verschiedenen Krankenhäusern praktizierten Behandlungsintensität, der intensiven Marge, getrennt schätzen und dann addieren. Für das Krankheitsbild Herzinfarkt finden sie allein im Bundesstaat New York Wohlfahrtsverluste von über 50 Millionen Dollar jährlich, die sowohl auf die in verschiedenen Counties unterschiedlich hohen Einweisungsraten als auch auf unterschiedliche Behandlungsintensitäten zurückzuführen sind. Für die USA insgesamt haben Phelps und Mooney (1993), S. 167 f., den Wohlfahrtsverlust des Jahres 1990 allein aus den beobachteten Variationen bei der Rate der Krankenhauseinweisungen auf \$ 33 Milliarden geschätzt, wobei sie zwischen 110 Diagnosefallgruppen unterschieden haben. Demgegenüber schätzen Phelps und Mooney (1993) ebenfalls für die USA den Wohlfahrtsverlust allein aus Variationen bei der Einweisung in Krankenhäuser auf \$ 130 pro Patient.

Da Praxisvariationen nicht die einzigen Ursachen von Wohlfahrtsverlusten im Gesundheitswesen sind, ist es für die Gesundheitspolitik unerlässlich, die empirische Größenordnung mit der Größenordnung anderer Arten von Wohlfahrtsverlusten zu vergleichen, insbesondere mit dem Mehrkonsum medizinischer Leistungen, die durch Dritte bezahlt werden, und mit der größeren Nachfrage von „schlechten“ Risiken nach Versicherungsschutz, wenn die Versicherungsnehmer einen Informationsvorsprung bezüglich ihres Risikos besitzen. Diese beiden Arten von Wohlfahrtsverlusten hängen ursächlich mit dem Drittzahlerprinzip zusammen und werden als Moral Hazard und adverse Selektion bezeichnet. Zum Vergleich berichtet Phelps (2000), S. 256, von Schätzungen des Wohlfahrtsverlustes aus dem ex post Moral Hazard durch

Krankenversicherungsschutz in Höhe von \$ 265 pro Person im Jahre 1986.¹⁰ Der gesamte Wohlfahrtsverlust aus medizinischen Praxisvariationen dürfte also den ungefähr gleich hohen Wohlfahrtsverlust aus Moral Hazard und adverser Selektion weit überschreiten. Andererseits ist zu vermuten, dass der Wohlfahrtsverlust aus medizinischen Praxisvariationen auch jene Kosten übertrifft, die voraussichtlich entstehen würden, wenn man alle Ärzte über die korrekte Anwendung einzelner medizinischer Leistungen umfassend informieren wollte.

Wennberg et al. (2002) schätzen, dass allein im amerikanischen Medicare-System jedes Jahr 29 Prozent der Gesamtausgaben oder 40 Milliarden Dollar eingespart werden könnten, wenn in allen Regionen die medizinische Praxis Anwendung fände, die in den jeweils preisgünstigsten Regionen praktiziert wird. Natürlich stellt dies nur die Obergrenze der Einsparmöglichkeit dar, denn es ist ja nicht ohne weiteres klar, dass die jeweils preisgünstigste Medizin die beste ist, wenn auch der Nutzen der Patienten in den Vergleich einbezogen wird. Gleichwohl belegen diese Zahlen die Größe der potentiellen Effizienzgewinne im Gesundheitswesen. Wennberg et al. (2002) machen eine Reihe von Vorschlägen, wie die Qualität und die Effizienz des Medicare-Systems verbessert werden kann. Im Kern geht es darum, die medizinische Praxis mit umfassender Information über die jeweils korrekte Anwendung medizinischer Leistungen zu versorgen. Warum dies bislang nicht gelungen ist, stellt gerade in einem modernen, durch Drittzahler finanzierten Gesundheitswesen eine zentrale gesundheitspolitische Frage dar, weil die Theorie vermuten lässt, dass der Wohlfahrtsverlust aus regionalen Praxisvariationen bei einem gleichzeitigen Auftreten von ex post Moral Hazard noch höher ausfällt.

¹⁰ Wie Nyman und Maude-Griffin (2001) in einem anderen Zusammenhang gezeigt haben, müssen wir heute davon ausgehen, dass dieser Wert um bis zu zwei Drittel zu hoch geschätzt worden ist. Denn der positive Einkommenseffekt aus Versicherungsschutz auf die beobachtete Zunahme der Nachfrage nach medizinischen Leistungen ist unberücksichtigt geblieben.

Praxisvariationen bei Krankenversicherungsschutz. Die Schätzung des Wohlfahrtsverlustes muss bei ex post Moral Hazard nach oben korrigiert werden, weil die im Durchschnitt beobachtete Anwendungsrate medizinischer Leistungen größer sein wird als die korrekte Anwendungsrate. Phelps und Mooney (1993) argumentieren, dass der ex post Moral Hazard bei Krankenversicherungsschutz einen zusätzlichen Wohlfahrtsverlust in Höhe von $0,5(M - X^*)^2 dV/dX$ für jede regionale Beobachtungseinheit entstehen lässt. Diese Formel geht jedoch von der nicht allgemein gültigen Annahme aus, dass die korrekte Anwendungsrate bei einem Anbieter-Preis von Null größer ist als die größte tatsächlich beobachtete Anwendungsrate in einer Region: nur dann wird der Wohlfahrtsverlust aufgrund übermäßiger Anwendung einer medizinischen Leistung durch ein vollständiges Dreieck – analog zur Fläche II in Schaubild 13a – beschrieben.

Schaubild 13b stellt die von Phelps und Mooney (1993) diskutierte Vergrößerung des Wohlfahrtsverlustes aus regionalen Praxisvariationen bei gleichzeitigem ex post Moral Hazard als Folge von Krankenversicherungsschutz graphisch dar. In diesem Schaubild ist V_1 wieder die marginale Zahlungsbereitschaft für die medizinische Leistung in der Region 1, V_2 die marginale Zahlungsbereitschaft in der Region 2 und V_a die durchschnittliche marginale Zahlungsbereitschaft in beiden Regionen. V^* beschreibt die marginale Zahlungsbereitschaft, die zu beobachten wäre, wenn alle Betroffenen korrekt über die gesundheitlichen Wirkungen der medizinischen Maßnahmen informiert wären.

Da diese vier Kurven annahmegemäß parallel verlaufen, ist die korrekte Anwendungsrate der medizinischen Maßnahme bei jedem beliebigen Preis kleiner als die im Durchschnitt der beiden Regionen beobachtete Anwendungsrate. Die korrekte Anwendungsrate ist aber höher als die in der Region 1 beobachtete Anwendungsrate. Das Dreieck, das die Flächen A und B umfasst, überschätzt den Wohlfahrtsverlust aus der zu niedrigen Anwendungsrate in Region 1, wäh-

rend das Dreieck C den Wohlfahrtsverlust aus der zu hohen Anwendungsrate in der Region 2 überschätzt. Der Wohlfahrtsverlust ist um die Flächen B, B', D und D' zu korrigieren. Dabei gilt $D = D'$ und $B = B'$; der Überschätzung des Wohlfahrtsverlusts in Region 1 in Höhe von B steht eine Unterschätzung des Wohlfahrtsverlusts in Region 2 in Höhe von B' gegenüber. Der zusätzliche Wohlfahrtsverlust in Folge von Moral Hazard bei gleichzeitigen Praxisvariationen entspricht somit den beiden Dreiecken D und D'. Bei genauer Betrachtung ist allerdings festzustellen, dass das Dreieck D gleich dem Wohlfahrtsverlust ist, den der ex post Moral Hazard auch ohne die Praxisvariationen verursachen würde. Daher ist der *zusätzliche* Wohlfahrtsverlust aus den regionalen Praxisvariationen strenggenommen nur durch die Fläche D' gegeben.¹¹

Auswirkungen auf die Innovationsanreize. Schaubild 13c stellt den zusätzlichen Wohlfahrtsverlust aus regionaler Praxisvariation dar, der sich bei unvollständiger Diffusion einer neuen medizinischen Technologie ergibt. In diesem Fall ist davon auszugehen, dass die korrekte Anwendungsrate höher ist als die im Durchschnitt der beiden Regionen beobachtete Anwendungsrate. Wieder gilt $D = D'$, aber gleichzeitig gilt nun $B > B'$, denn $A = D' + B' + C$. Daraus folgt: $B = B' + D + D'$. Die Summe der beiden Dreiecke D und D' ist demnach genau gleich der Differenz zwischen B und B' und stellt den zusätzlichen Wohlfahrtsverlust bei regionaler Praxisvariation und unvollständiger Diffusion einer neuen Medizintechnologie dar. Gäbe es in diesem Fall keine regionalen Praxisvariationen, wäre der Wohlfahrtsverlust aus der unzureichenden Diffusion der medi-

¹¹ Einschränkung muss gesagt werden: Wo die Annahme nicht gilt, dass die korrekte Anwendungsrate bei einem Anbieter-Preis von Null größer ist als die größte tatsächlich in einer Region beobachtete Rate, ist der zusätzliche Wohlfahrtsverlust kleiner als D'. Wenn zudem die durchschnittlich beobachtete Anwendungsrate bei einem Anbieter-Preis von Null kleiner ist als die größte in einer Region beobachtete Anwendungsrate, kann der zusätzliche Wohlfahrtsverlust sogar in einen Wohlfahrtsgewinn gegenüber dem auf Basis der durchschnittlichen Anwendungsrate gemessenen Wohlfahrtsverlust umschlagen. Aber natürlich heißt das nicht, dass der ex post Moral Hazard insgesamt zu einem Wohlfahrtsgewinn führt; die Größe des Dreiecks D, das den eigentlichen Wohlfahrtsverlust aus ex post Moral Hazard beschreibt, bleibt unverändert.

zinischen Technologie allein durch die Summe $B + D$ definiert, wobei $B > D$ gelten würde. Da das Auftreten regionaler Praxisvariationen in diesem Fall den Wohlfahrtsverlust aus der unzureichenden Diffusion einer medizinischen Technologie eher zu verringern scheint, stellt sich der Gesundheitspolitik die Frage, ob regionale Praxisvariationen einen Beitrag leisten können, die privaten Innovationsanreize zu verbessern.¹²

Die Antwort hierauf ist negativ. Wie Schaubild 13d zeigt, verringern regionale Praxisvariationen den Monopolpreisbildungsspielraum bei der Einführung einer patentgeschützten medizinischen Technologie. Es wird angenommen, dass keine Preisdiskriminierung zwischen den Regionen möglich ist und der Anbieter der neuen Technologie nur einen einheitlichen Preisaufschlag auf seine Grenzkosten erheben kann. Dies ergibt sich nicht nur aus der Annahme, dass der Monopolist vor Einführung seiner Technologie nicht weiß, in welchen Regionen die Nachfrage zu groß oder zu klein sein wird, sondern auch aus der Möglichkeit von Arbitragegeschäften. Das Schaubild 13d zeigt, dass der optimale Monopolpreis höher wäre, wenn der Monopolist sich nicht an der durchschnittlichen Zahlungsbereitschaft der beiden Regionen V_a , sondern an V^* orientieren würde – der Zahlungsbereitschaft, die die korrekte Information über die medizinische Wirksamkeit der neuen Technologie widerspiegelt.

Die These, dass umfassende Informationen über die Wirksamkeit einer neuen medizinischen Technologie in der Praxis oft fehlen, wird empirisch unter anderem durch den regelmäßig überragenden Einfluss von (meist weiblichen) Pharma-Referenten auf die Verkaufszahlen neuer Medikamente belegt. Viele Ärzte verschreiben offenbar bevorzugt Medikamente, die sie entweder schon lange kennen oder über die sie kürzlich neue Informationen erhalten haben.

¹² Grundsätzlich sind die Innovationsanreize bei einer suboptimalen Technologiediffusion, die zum Beispiel auf einen monopolistischen Preisaufschlag ohne perfekte Preisdiskriminierung zurückzuführen sein kann, zu gering.

Allerdings führen die Informationen der Pharma-Referenten natürlich im Allgemeinen nicht zur korrekten Anwendungsrate, wie sie von Patienten nachgefragt würde, die nicht nur über ein einzelnes – von einer Pharma-Referentin beworbenes – Medikament, sondern über die komparativen Vor- und Nachteile *aller* therapeutischen Möglichkeiten in einem vorliegenden Krankheitsfall vollständig informiert sind. Kein gewinnorientierter Pharma-Anbieter per se hat einen Anreiz zur korrekten Information in diesem Sinne. Der Anbieter, der es dennoch versuchte, wäre aus Sicht der Ärzte und Konsumenten noch nicht einmal glaubwürdig.¹³

Bei *dynamischer* Betrachtung ist von einem positiven Rückkoppelungsprozess auszugehen, weil die unzureichende Adoption ein im Vergleich zur erstbesten Situation langsames Lernen aus der Erfahrung mit der neuen Technologie impliziert. Es ist deshalb zu vermuten, dass die privaten Innovationsanreize durch regionale Praxisvariationen tendenziell geschmälert werden. Unabhängig davon bleibt jedoch die Hypothese relevant, dass Krankenversicherungsschutz durch das Auftreten von ex post Moral Hazard zu einer generellen Intensivierung des Technologieeinsatzes führt und die privaten Innovationsanreize insgesamt verstärkt. Natürlich können auch dabei Fehlanreize auftreten. Insbesondere kann dynamischer Moral Hazard, wie er von Goddeeris (1980) analysiert wurde, zu einer qualitativen Verzerrung in Richtung auf „Me Too“-Innovationen führen, so dass die Produktvielfalt unnötig eingeschränkt wird und die sozialen Erträge aus dem Bayesianischen Lernen geschmälert werden.¹⁴

¹³ Aus diesem Dilemma lässt sich eine Begründung zur Regulierung von Produktinformationen, insbesondere über mögliche Nebenwirkungen, ableiten. Siehe hierzu Neumann et al. (2000).

¹⁴ Zeitreihen-Evidenz, dass die Rate von Preissteigerungen bei einzelnen medizinischen Leistungen negativ mit dem Anteil der finanziellen Selbstbeteiligung des Patienten korreliert ist, stützt die These des dynamischen ex post Moral Hazard. Siehe Newhouse (1981).

4.2 Öffentliche Güter in der Absorption medizinischen Wissens

Angesichts der vorliegenden empirischen Evidenz zu den Wohlfahrtswirkungen regionaler Praxisvariationen kann die Gesundheitspolitik nicht länger die Frage vernachlässigen, wie sich das Angebot aktueller Informationen über die korrekte Anwendung medizinischer Leistungen verbessern und die Diffusion und Anwendung dieser Informationen in der Praxis fördern lässt.¹⁵ Je größer die technologische Absorptionskapazität eines Gesundheitssystems, desto größer ist auch der Umfang des allgemein verfügbaren medizinischen Wissens und desto niedriger sind die Kosten der Koordination zwischen Spezialisten. Bei abnehmenden Koordinationskosten ermöglicht eine zunehmende Spezialisierung die Vertiefung der Arbeitsteilung. Je tiefer die Arbeitsteilung, umso höher ist die insgesamt erreichbare Produktivität. Diese Einsicht ist Becker und Murphy (1992) zufolge von zentraler Bedeutung für wirtschaftliches Wachstum generell. Im Gesundheitswesen verursachen fehlende öffentliche Güter demnach nicht nur statische Effizienzverluste, sondern beeinträchtigen auch das Wachstum der Produktivität.

Wer aber investiert in die Bereitstellung der für medizinische Entscheidungen relevanten öffentlichen Güter? Wer sorgt für eine rasche Diffusion und korrekte Anwendung neuen medizinischen Wissen und neuer Technologien? Ist eine unzureichende Versorgung des Gesundheitswesens in erster Linie auf Probleme der Angebotsseite – insbesondere die Schwierigkeit, für neu erstelltes Wissen

¹⁵ Eine Ausnahme ist Smythe (2002), der die Frage untersucht, wie öffentliche Information über den wahren Wert einer medizinischen Innovation den optimalen Adoptionszeitpunkt aus Sicht eines Arztes beeinflusst, wenn dessen Reputation irreversibel von seiner Technologiewahl abhängt. Je größer die Zuverlässigkeit der öffentlichen Information – gemessen durch die Varianz des relevanten Fehlerterms –, desto höher ist im Allgemeinen die Adoptionsrate nach dem Bekanntwerden der öffentlichen Information. Dies bedeutet jedoch nicht unbedingt, dass die Adoptionsrate vor dem Bekanntwerden der öffentlichen Information sinkt. Dies wäre zwar anzunehmen, wenn es allein darum ginge, die Adoption zurückzustellen, um auf das zuverlässige Signal zu warten und Fehlinvestitionen zu vermeiden. Aber gleichzeitig kann die größere Zuverlässigkeit der erwarteten Information bedeuten, dass der Reputationsgewinn höher eingeschätzt und die Adoption aus diesem Grund vorgezogen wird.

einen angemessenen Preis durchzusetzen – oder auf Probleme der Nachfrage-seite zurückzuführen – eine fehlende Zahlungsbereitschaft für klinisch relevantes Wissen, zum Beispiel aufgrund unzureichender Qualitätskontrollen für medizinische Leistungen? Welche Anforderungen sind an eine Informationsinfrastruktur zu stellen, die Angebot und Nachfrage zusammenführt? Und wie ist ein geeignetes System von Anreizen zu gestalten, das eine möglichst vollständige Erfassung neuer Informationen, ihre systematische (monetäre) Bewertung, das Herausfiltern relevanter Empfehlungen für die Praxis und die regelmäßige Fortbildung der Anwender dieses Wissens – vor allem der Ärzte und Pflegekräfte – gewährleistet?

Probleme auf der Angebotsseite medizinischer Information. Im Einzelnen sind auf der Angebotsseite die folgenden fünf logischen Schritte zu bewältigen. Als *Erstes* sind Informationsquellen zu erschließen und Informationen zu sammeln, die nicht nur aus der aktuellen medizinischen Forschung, sondern auch aus der stationären und ambulanten medizinischen Praxis stammen sollten. Als *Zweites* sind diese Informationen – zunächst ungefiltert – an Informationsintermediäre weiterzugeben, um von diesen als *Drittes* hinsichtlich ihrer Qualität bewertet zu werden. Dabei kann ein Reputationswettbewerb der Informationsanbieter nützlich sein. Als *Viertes* sind die Informationen zu aggregieren und so aufzubereiten, dass sie ohne allzu große Friktionskosten in die Praxis integriert werden können. Hierbei ist die Frage nach der optimalen Aktualisierungshäufigkeit und nach der optimalen Größe der jeweils angebotenen Informationsmenge zu stellen. Im *fünften* und letzten Schritt sind die neuen Informationen für die unterschiedlichen Zielgruppen in geeigneter Weise aufzubereiten und darzustellen. Neben den Ärzten gehören vor allem auch die Patienten, die Krankenversicherungen und Regulierungsbehörden zu den Zielgruppen neuer medizinischer Informationen.

Die direkte Regulierung medizinischer Informationen ist eine Möglichkeit, einem unzureichenden Informationsangebot zu begegnen. Dabei gibt es jedoch einen Trade-off, der im Zusammenhang mit der Zulassung neuer Pharmaka empirisch dokumentiert ist. Je höher die Informationsanforderungen an klinische Tests, die die Wirksamkeit und Unbedenklichkeit eines neuen Medikamentes beweisen sollen, desto höher sind die zu erwartenden Kosten der Markteinführung und desto länger werden die Intervalle zwischen zwei neu auf den Markt kommenden Pharmaka sein. So hat die Dauer des Entwicklungsprozesses eines neuen patentgeschützten Arzneimittels von vier Jahren anfangs der 60er Jahre, bevor die Contergan-Katastrophe zu einer weltweiten Verschärfung der Zulassungsbedingungen und Informationsanforderungen führte, auf etwa zwölf Jahre Ende der 90er Jahre zugenommen. DiMasi et al. (2003) belegen, dass sich die jährliche Wachstumsrate der klinischen Forschungsausgaben für ein neu zugelassenes Arzneimittel von 6,1 Prozent in den 70er Jahren auf 11,8 Prozent in den 80er Jahren deutlich erhöht hat, was vermutlich zu einem erheblichen Teil auf verschärfte Informationsanforderungen zurückzuführen ist. Die Regulierung der Informationsanforderungen bei der Zulassung neuer Medikamente hat daher einen erheblichen Einfluss auf die Häufigkeit und den Innovationsgrad neuer Pharmaka. Eine theoretische Analyse dieser Zusammenhänge bieten Neumann et al. (2000).

Wie groß das Potential medizinischer Information – insbesondere auch aus der medizinischen Praxis – am Beginn des 21. Jahrhunderts ist, wird durch die Arbeit des amerikanisch-isländischen Unternehmens deCODE Genetics Inc. illustriert, das sich bereits nach wenigen Jahren als innovativer Informationsanbieter und Partner der Pharma-Industrie einen Namen gemacht hat. Die Arbeit dieses Unternehmens ist auf seiner Homepage unter www.decode.is dokumentiert. Der Hauptfokus von deCODE ist die Identifizierung der genetischen Ursachen allgemein verbreiteter Krankheiten und die Anwendung dieser

Informationen in der Entwicklung neuer Pharmaka und Diagnostika, die besser in der Lage sind, gezielt die biologischen Ursachen von Krankheiten und nicht nur ihre Symptome zu bekämpfen. Dazu müssen nicht nur Diagnoseinstrumente entwickelt werden, die die genetischen Ursachen gegebener Krankheitsfälle aufdecken können, sondern auch solche, die im Rahmen der prädikativen genetischen Diagnostik Risikokonstellationen erkennen, so dass die Medizin frühzeitig präventiv oder therapeutisch intervenieren kann.

Propping und Aretz (2004) beschreiben die Bedeutung der modernen Human-genetik für die Therapie und Vorbeugung von Erbkrankheiten. Ob eine prädikative genetische Diagnostik sinnvoll ist, müssen Experten für jeden Genort und jede Genvariante einzeln festlegen und dabei das Nutzen-Kosten-Verhältnis einer Behandlung im Krankheitsfall, die Häufigkeit und Schwere der Krankheit sowie die Vorbeugungs- und Behandlungsmöglichkeiten berücksichtigen. Schlaganfall, Herzkrankheiten und Alzheimer sind Beispiele für Volkskrankheiten, bei denen das Zusammenwirken mehrerer Gene und schädlicher Umweltfaktoren wesentlich zu ihrer Entstehung beizutragen scheint. Um den komplexen Ursachen auf die Schliche zu kommen, müssen Korrelationen zwischen dem Auftreten einzelner Krankheiten und genetischen Variationen in einer möglichst großen Stichprobe der Bevölkerung empirisch ermittelt werden. Um dies effizient zu bewerkstelligen, ist es erforderlich, eine genaue und möglichst umfassende genealogische Datenbank zu verwenden, die dokumentiert, wie die genetischen Komponenten einzelner Krankheiten von Generation zu Generation weitergegeben werden.

Hier setzt die Firma deCODE an, die in Island über einen umfangreichen Datensatz verfügt, der genetische, krankheitsbezogene und genealogische Informationen kombiniert und es daher ermöglicht, effektiv nach krankheitsbezogenen Genen zu suchen. Dazu werden alle Patienten einer bestimmten Krankheit Großfamilien zugeordnet. Ziel ist es dabei nicht nur, Gene zu

isolieren, die in einzelnen Großfamilien bestimmte Krankheiten verursachen, sondern auch die genaue Funktion dieser Gene und ihre Interaktion mit anderen Genen zu untersuchen. Aus den Ergebnissen dieser Forschung können diagnostische „Marker“ und Hinweise für die individuelle Auswahl und Dosierung von Pharmaka abgeleitet werden.

Zur Durchführung dieses Projektes ist eine genealogische Datenbank aufgebaut worden, die nach Angaben von deCODE alle lebenden Isländer umfasst und bis zur Besiedelung der Insel vor 1100 Jahren zurückreicht. Die Datenbank enthält zudem DNA-Proben und detaillierte Krankheitsdaten von 100.000 freiwilligen Teilnehmern in mehr als 50 auf einzelne Krankheiten bezogenen Teilprojekten. Hunderte von Allgemeinärzten und Fachärzten wirken an dem Projekt mit. Die Bereitschaft, Daten zur Verfügung zu stellen, scheint sehr groß zu sein. Nach Unternehmensangaben haben mehr als 90 Prozent der angesprochenen Isländer ihre Beteiligung zugesagt.

Der Datenschutz ist nach Angaben des Unternehmens deCODE gewahrt, so dass die Identität einzelner Patienten von den beteiligten Wissenschaftlern nicht nachvollzogen werden kann. Die patientenbezogenen Daten sind über eine Identifikationsnummer verschlüsselt, die ausschließlich der isländischen Datenschutzbehörde bekannt ist. Gleichwohl bleiben die datenschutzrechtlichen Aspekte dieses Forschungsprojektes umstritten. Obwohl nur Blutproben analysiert werden sollen, die freiwillig gegeben wurden, schließt das Unternehmen nicht aus, genealogische Daten auch von Menschen zu verwenden, die ihre Zustimmung nicht ausdrücklich erteilt haben. Alle bereits vor Beginn des Forschungsvorhabens verstorbenen Isländer konnten ohnehin ihre Zustimmung nicht erteilen, weil das Vorhaben ja zu ihren Lebzeiten noch gar nicht bekannt war. Aber auch die genealogischen Daten noch lebender Isländer sollen ausgewertet werden, solange der Einzelne nicht ausdrücklich widersprochen hat.

Probleme auf der Nachfrageseite medizinischer Information. Auch auf der Nachfrageseite sind mehrere logische Schritte zu bewältigen. Als *Erstes* sind dort Anreize zur Artikulation der Informationsbedürfnisse zu setzen. Auf die Informationsnachfrage der Ärzte hat vor allem das Vergütungssystem einen maßgeblichen Einfluss: Das Interesse an neuen Technologien beispielsweise wird größer sein, wenn ein Arzt für jede Einzelleistung extra vergütet wird, als wenn er eine Kopfpauschale oder ein festes Gehalt bezieht. Aus ähnlichen Gründen dürfte auch die Informationsnachfrage der Patienten größer sein, wenn diese eine finanzielle Selbstbeteiligung zu leisten haben. Im *zweiten* Schritt sind die nachgefragten Informationen zu bewerten, um ihre Priorität festzulegen, wobei zwischen ihrem sozialen und privaten Wert zu unterscheiden ist. Für den Arzt ist die Nachfrage nach Informationen zunächst ein Kostenfaktor. Es ist deshalb wichtig, dass regelmäßig über eine Anpassung der Qualitätsanforderungen und des Vergütungssystems nachgedacht wird, um die Anreize der Ärzte immer wieder in Einklang mit den sozialen Anreizen zu bringen, Informationen nachzufragen, dem Wert dieser Informationen für die Volkswirtschaft insgesamt. Dies kann insbesondere bei einem prospektiven Vergütungssystem wichtig sein, dessen Vergütungssätze angepasst werden müssen, wenn neue medizinische Technologien die zu erwartenden Kosten eines Behandlungsfalles erhöhen oder verringern. Als *Drittes* ist dafür zu sorgen, dass die Ärzte einen Anreiz haben, in ihre Absorptionskapazität für neue medizinische Informationen zu investieren, zum Beispiel durch die Spezialisierung auf Teilgebiete der Medizin, durch die Teilnahme an Fortbildungsveranstaltungen und durch die Anschaffung von Computersystemen und Software zur Entscheidungsunterstützung. Im *vierten* Schritt sind die Ärzte zu vernetzen, damit die individuelle Informationsnachfrage zeitnah zu den jeweiligen Informationsanbietern weitergeleitet werden kann.

Die empirische Untersuchung von Reng et al. (2003) hat den dominierenden Einfluss des Preises auf die Auswahl von Informationsquellen durch deutsche Ärzte bestätigt. In einer Umfrage, an der sich Anfang 2002 knapp 16.000 in Klinik oder Praxis tätige Ärzte beteiligten, wurde dem Gespräch mit Kollegen im Vergleich zur Informationsbeschaffung durch Fachbücher, Fachzeitschriften, Kongresse, Informationen der Industrie und den elektronischen Medien (wie CD-ROMs, Online-Datenbanken und Medizinportale im Internet) das mit Abstand beste Preis-Leistungs-Verhältnis zugebilligt. Auch im Hinblick auf die Einzelkriterien Schnelligkeit und praktische Nutzbarkeit schnitt das Gespräch mit den Kollegen am besten ab. Allein bei der Qualität und Vertrauenswürdigkeit der Informationen wurden Fachbücher und Fachzeitschriften von den deutschen Ärzten höher bewertet. Bei allen vier Einzelkriterien schnitten dagegen die elektronischen Medien nur knapp besser ab als die (am schlechtesten bewerteten) Informationen aus der Pharma-Industrie, was nach Ansicht von Reng et al. (2003) die Notwendigkeit begründet, in „die Entwicklung nachvollziehbarer Strukturen der Informationsvermittlung sowie besonders in die Qualität und Nachhaltigkeit der verfügbar gemachten Inhalte zu investieren“.

Das individuelle Informationsnachfrageverhalten des Arztes wird stark durch die Regulierung des Gesundheitswesens beeinflusst. Ein wichtiges Element ist dabei das Haftungsrecht. So argumentiert Heimpel (2003), dass eine Leitlinie, die außerhalb der eigenen Klinik entwickelt wurde, bei den Ärzten schon deshalb auf wenig Interesse stoße, weil der Produzent der Leitlinie nicht für das Handeln eines Arztes haftet, der sich auf die in der Leitlinie enthaltenen Informationen verlässt. Rychlik (1999) dagegen meint, dass Leitlinien gleichwohl die Feststellung eines Behandlungsfehlers erleichtern, wenn sie einen ärztlichen Handlungskorridor mit einer Untergrenze im Sinne eines

Mindeststandards definieren, auf den sich Gutachter im Streitfall stützen können.

Die haftungsrechtlichen Implikationen sind sicherlich ein Grund, systemische Interdependenzen zwischen dem Informationsnachfrageverhalten verschiedener Ärzte zu vermuten. Ein weiterer Grund weist auf die regionale Dimension solcher Interdependenzen hin: häufig ist es notwendig, dass Ärzte untereinander und mit anderen Akteuren des Gesundheitswesens eng zusammenarbeiten, zum Beispiel bei chronischen Krankheiten. Aus der Notwendigkeit der Zusammenarbeit entsteht ein Risiko des Koordinationsversagens bei der Aktualisierung medizinischer Informationen über neue Behandlungsmöglichkeiten. Medizinische Behandlungsleitlinien, die für alle Beteiligten gelten und von Zeit zu Zeit an neue wissenschaftliche Erkenntnisse angepasst werden, stellen einen Ansatz dar, dieses Koordinationsproblem auf bessere Weise als durch tradierte medizinische „Schulen“ zu lösen.¹⁶

Neben der inhaltlichen Koordination geht es dabei auch um die zeitliche Dimension: Sollen medizinische Behandlungsleitlinien in festen Zeitintervallen aktualisiert werden oder soll eine Aktualisierung nur in Reaktion auf wissenschaftliche Durchbrüche erfolgen, deren wahre Bedeutung allerdings oft erst im Nachhinein erkennbar ist? Wie groß sind die Anpassungskosten der medizinischen Praxis an neue Behandlungsleitlinien, und wie sollten sie auf der Nachfrageseite berücksichtigt werden? Welchen Einfluss haben Anpassungskosten insbesondere auf die optimale Häufigkeit und Informationsdichte der Aktualisierung von Behandlungsleitlinien?

Optimale Frequenz medizinischer Wissenstransfers. Weil die unvermeidlichen Anpassungskosten in der medizinischen Praxis zumindest teilweise Fixkosten sind, deren Höhe von der jeweiligen Informationsmenge unabhängig ist, würde

¹⁶ Vgl. Abschnitt 4.1.

es selbst bei einem kostenlosen Angebot neuer Informationen keine kontinuierliche Aktualisierung von Behandlungsleitlinien geben. Vielmehr wird die Aktualisierung in diskreten Zeitintervallen erfolgen und muss folglich jedes Mal eine diskrete Informationsmenge umfassen.

Ein einfaches Modell der Koordination von Angebot und Nachfrage und der optimalen Häufigkeit der Aktualisierung medizinischer Informationen ist in Schaubild 14 dargestellt. Darin ist auf der Abzisse die Zeit t abgebildet, während die Ordinate einerseits die Bereitstellungskosten medizinischer Informationen und andererseits den Nettowert neuer Informationen in der medizinischen Praxis abbildet. Zur Vereinfachung nehme ich an, dass es einen konstanten wissenschaftlichen Fortschritt gibt und der Stand des medizinischen Wissens mit exogen vorgegebener Rate wächst.

Die Frage ist, in welchen Zeitintervallen eine medizinische Behandlungsleitlinie aktualisiert und welche Menge neuer Informationen bei einer Aktualisierung verarbeitet werden sollte. Die Antwort hierauf hängt einerseits davon ab, wie sich der Nettowert einer Aktualisierung im Zeitablauf und damit die Nachfrage nach neuem medizinischen Wissen verändert, und andererseits davon, wie sich die Bereitstellungskosten dieses Wissens im Zeitablauf entwickeln. Eine Aktualisierung sollte dann erfolgen, wenn der Nettowert neuer Informationen in der medizinischen Praxis gerade gleich den Bereitstellungskosten ist.

Schaubild 14 beschreibt diese Zusammenhänge. Nachdem eine Aktualisierung auf den neuesten Stand des medizinischen Wissens erfolgt ist, wird der Nettowert einer weiteren Aktualisierung zunächst negativ sein, weil die fixen Anpassungskosten den Grenzwert weiterer Informationen übersteigen. Aufgrund des exogenen wissenschaftlichen Fortschritts nimmt die Informationsmenge im Zeitablauf zu, die im Falle einer Aktualisierung in eine Behandlungsleitlinie einfließt. Der Nettowert dieser neuen Informationen in der medizini-

schen Praxis hat daher einen mit t linear ansteigenden Verlauf, von A über B nach C beziehungsweise von D über E nach F.

Auf der Angebotsseite bestimmt die Informationsmenge den im Prinzip zeitunabhängigen Kostenblock der Informationsbereitstellung zur Aktualisierung einer Behandlungsleitlinie, während die Elastizität des Informationsangebotes gewissermaßen die Zunahme der Grenzkosten der Bereitstellung bei einem abnehmenden Intervall seit der letzten Aktualisierung bestimmt. Je kürzer die zur Verfügung stehende Zeit, desto höher sind die Kosten der Bereitstellung einer gegebenen Menge an Information, was bei einem zunehmenden Anteil paralleler statt sequentieller Forschung und Informationsaggregation sofort plausibel ist. Die Gesamtkosten der Informationsbereitstellung setzen sich additiv aus den zeitunabhängigen Kosten, die ich hier der Einfachheit halber „Fixkosten“ nenne, und aus den zeitabhängigen „variablen“ Kosten zusammen. Ich nehme an, dass diese Gesamtkosten einen u-förmigen Verlauf haben, wie er durch die Kurven k_1 und k_2 in Schaubild 14 gezeigt wird. Bei einer Vergrößerung des Zeitintervalls seit der letzten Aktualisierung dominiert zunächst die Verringerung der Grenzkosten, danach die Zunahme der „Fixkosten“ – aufgrund der größeren akkumulierten Informationsmenge, die in die Aktualisierung einfließt.

Im Idealfall erfolgt die Aktualisierung der Behandlungsleitlinie in Punkt C, dem Minimum der gesamten Bereitstellungskosten neuer medizinischer Information. In diesem Punkt ist der Nettowert der neuen Information gerade gleich den gesamten Bereitstellungskosten. Diese Bedingung könnte aber – bei einem anderen Verlauf der Nettowert-Kurve – in der Praxis auch links von C oder rechts von C auf der k_1 -Kurve erfüllt sein. Dann wäre die Informationsnachfrage zeitlich schlecht mit der Entwicklung der Angebotsbedingungen koordiniert. Zudem macht Schaubild 14 auf ein potentiell noch gravierenderes Problem aufmerksam: Wenn die Nettoanpassungskosten sehr groß sind, wie in

H, muss es gar keinen Schnittpunkt zwischen der Nettowertkurve und der Kurve der Gesamtbereitstellungskosten medizinischer Informationen geben. Dies ist durch die gestrichelte Linie angedeutet, die ihren Ursprung in H hat und durch den Punkt I läuft.

Eine schlechte zeitliche Koordination oder zu große Anpassungskosten in der Praxis können ein Hinweis auf Investitionsbedarf sein, der die Situation auf der Angebotsseite oder bei den Anwendern medizinischer Information verbessert. Solche Investitionen könnten zum Ziel haben, einerseits die Anpassungskosten der Praxis an neue Behandlungsleitlinien strukturell und dauerhaft zu senken, andererseits den Verlauf der Grenzkosten in der Informationsbereitstellung zu verändern oder die Fixkosten der Informationsaktualisierung für eine Behandlungsleitlinie zu senken. Derartige Investitionen sollte sich aus sozialer Sicht durchaus rentieren, wenn dadurch ein effizientes Intervall der Informationsaktualisierung erreicht wird. Die k_2 -Kurve zeigt, wie sich zum Beispiel eine Senkung der zeitunabhängigen „Fixkosten“ der Informationsaktualisierung im Vergleich zur k_1 -Kurve auswirkt.

Nachdem eine Aktualisierung der Behandlungsleitlinie in Punkt C vorgenommen wurde, dominieren auf der Nachfrageseite wieder die Anpassungskosten, so dass der Nettowert einer weiteren Aktualisierung zunächst wieder negativ ist (Punkt D). Aufgrund des exogenen wissenschaftlichen Fortschritts wird sich aber der Nettowert einer weiteren Aktualisierung der Behandlungsleitlinie kontinuierlich erhöhen, bis die Aktualisierung in F stattfindet. Der Punkt F liegt auf der Bereitstellungskostenkurve k_2 , die insgesamt niedriger liegt als k_1 , weil ich annehme, dass die „Fixkosten“ der Bereitstellung einer durch die Zeit gegebenen Informationsmenge infolge von Investitionen in die Informationsverarbeitungskapazität gesenkt wurden. Dadurch verkürzt sich auch die Länge des Zeitintervalls seit der letzten Aktualisierung: $t_2 - t_1 < t_1 - 0$. Investitionen, durch die sich die Elastizität des Informationsangebots im Zeitverlauf

verbessert, schlagen sich in einem flacheren Verlauf der gesamten Bereitstellungskostenkurve vor ihrem Minimum nieder. Auch dies würde das optimale Intervall seit der letzten Aktualisierung verkürzen.

Es sind weitere Variationen der Annahmen denkbar, deren Auswirkungen ebenfalls in Schaubild 14 dargestellt werden können. Zum Beispiel ist eine zeitinvariante Erhöhung des Nettowertes neuer medizinischer Information durch das Auftreten einer neuartigen Epidemie denkbar, wie im Falle von AIDS. Je mehr Menschen die Epidemie erfasst, desto wertvoller ist die Aktualisierung von Behandlungsleitlinien, die die Effektivität medizinischer Interventionen verbessern helfen.

Elemente einer effizienten Informationsinfrastruktur. Der Schwerpunkt beim Aufbau einer medizinischen Informationsinfrastruktur liegt heute nicht mehr in erster Linie bei der Hardware, die durch das Internet und eine weitreichende Computerisierung bereits stark ausgebaut ist, sondern vor allem bei der Entwicklung geeigneter Software, also im ökonomischen Sinne bei Institutionen, die Angebot und Nachfrage neuen Wissens effizient zusammenbringen. Ziel sollte es dabei nicht sein, private Initiative zu verdrängen, sondern dort zu ergänzen, wo ein privates Angebot nicht oder nur unzureichend zustande kommt. Die technologische Revolution der 90er Jahre in der relevanten Hardware, insbesondere die Adoption des Internet als universelle Datenautobahn, hat die Chancen für den Aufbau einer effizienten medizinischen Informationsinfrastruktur entscheidend verbessert. So nutzten Anfang 2002 bereits 84,8 Prozent der von Reng et al. (2003) befragten deutschen Ärzte am Arbeitsplatz einen Computer und 82,3 Prozent verfügten über einen Zugang zum Internet. Gleichwohl ist der Erfolg der neuen Medien in der Medizin kein Selbstläufer. Vielmehr sollten Erfahrungen aus anderen Bereichen genutzt werden, um die Entwicklung der relevanten Institutionen so zu gestalten, dass sie keine unerwünschten Verhaltensreaktionen bei den Anbietern oder Nach-

fragern medizinischer Informationen auslösen. Die Notwendigkeit einer zuverlässigen Software wird erkannt. Fast neun von zehn der in Reng et al. (2003) befragten Ärzte gaben an, dass elektronisch verfügbare Medizininformationen aus „klar erkennbaren und kontinuierlich aktualisierten Quellen“ stammen sollten.

Zeitnahe Informationen aus der medizinischen Praxis werden längst nicht mehr nur bei der Auswahl vorgegebener Behandlungsmöglichkeiten für einzelne Patienten gebraucht, sondern auch, um klinische Leitlinien für die Versorgung auszuarbeiten, Entscheidungen über die Erstattungsfähigkeit medizinischer Leistungen und pharmazeutischer Produkte vorzubereiten, klinische Tests zu planen und ganz allgemein die Allokation von Ressourcen im Gesundheitswesen zu steuern.

Wennberg et al. (2002) argumentieren, dass das Fehlen einer Informationsinfrastruktur, die die ärztliche Compliance mit den jeweils aktuellsten evidenzbasierten Standards der Medizin ermöglicht, eine wesentliche Ursache für Effizienzverluste durch medizinische Praxisvariationen darstellt. Den Ertrag aus der Beseitigung dieser medizinischen Praxisvariationen schätzt er allein für die USA auf 40 Milliarden Dollar (\$). Der Aufbau der Informationsinfrastruktur verspricht daher sehr hohe soziale Erträge.

Für die USA beschreibt Detmer (2003), wie sich eine integrierte Informationsinfrastruktur für das gesamte Gesundheitswesen entwickeln könnte. Die Informationsinfrastruktur sollte sowohl unmittelbar den einzelnen Patienten dienen, als auch die Entscheidungsgrundlagen in der privaten und öffentlichen Versorgung verbessern. Gleichzeitig sollte sie der medizinischen und pharmazeutischen Forschung – unter Beachtung des Datenschutzes – zeitnahe Informationen aus der klinischen und ambulanten Praxis liefern. Hierzu muss die Informationsinfrastruktur Standards setzen, die den Austausch von Informa-

tionen über institutionelle und nationale Grenzen ermöglichen und die Zusammenarbeit zwischen dem privaten und öffentlichen Sektor fördern.

In den privaten Versorgungsstrukturen muss Detmer (2003) zufolge die Priorität auf der Förderung evidenzbasierter Medizin liegen, damit sich die Sicherheit, die Qualität und die Zugangsmöglichkeiten der Patienten zu medizinischen Leistungen verbessern. Im öffentlichen Gesundheitswesen kommt es darauf an, die Möglichkeiten der Überwachung und des Schutzes der Gesundheit in größeren Bevölkerungsgruppen zu gewährleisten. In der Forschung geht es darum, dass Informationen nicht nur aus Laborexperimenten, sondern auch aus der medizinischen Praxis Einfluss auf die Ausrichtung von Forschungsvorhaben bekommen. Insbesondere sollten Erfahrungen mit neuen medizinischen Technologien, etwa mit Pharmaka, zeitnah in die Forschung zurückfließen. Dadurch ließen sich nicht nur künftige Forschungsarbeiten zielgenauer planen, sondern auch klinische Studien besser fokussieren. Das Potential, die Produktivität der weiteren Forschung durch ein kontinuierliches Feedback zu verbessern, sollte nach Abschluss der genetischen Revolution, die in wichtigen Bereichen eine individuelle Anpassung der Pharmaka an den einzelnen Patienten ermöglicht, besonders groß sein.

Ein kontinuierliches Feedback aus der Anwendung in der Entwicklung neuer Technologien entspricht vom Ansatz her einem „Concurrent Engineering“, wie es seit den späten 80er Jahren in weiten Teilen des verarbeitenden Gewerbes zur Beschleunigung des Innovationsprozesses praktiziert wird. Daher muss neben dem Aufbau einer technologischen Infrastruktur auch eine Lernkultur etabliert werden, die die Kommunikation zwischen allen Beteiligten fördert, den Austausch von Daten erleichtert und eine effiziente Diffusion neuen Wissens gewährleistet. Kein wissensbasierter Wirtschaftszweig kann auf Dauer ohne eine moderne Infrastruktur zum Austausch von Informationen wachsen, die mit den Anforderungen einer offenen Lernkultur kompatibel sind. Es lohnt sich

daher, einen näheren Blick auf die Elemente zu werfen, die bei der Entwicklung einer technologischen Infrastruktur im verarbeitenden Gewerbe bewährt haben und zum Beispiel in Tasse (1992) ausführlich beschrieben werden.

Mit Tasse (1992) können wir die technologische Infrastruktur allgemein als die Summe des wissenschaftlichen, ingenieurtechnischen und angewandten Wissens definieren, das Unternehmen des verarbeitenden Gewerbes zur Verfügung steht. Die technologische Infrastruktur umfasst daher einerseits generische Technologien und andererseits die sogenannten Infratechnologien, zu denen neben Einzelinformationen über organisatorische Praktiken und Techniken der Produktion vor allem Standards für die Schnittstellen zwischen einzelnen Produktsystemen, Testverfahren zur Leistungsmessung und Qualitätssicherungssysteme gehören. Zusätzlich umfasst die technologische Infrastruktur eines Wirtschaftsraumes auch die wichtigsten Grundlagen strategischer Planungsmethoden, Foren für gemeinsame Planungen und Kollaborationen über Unternehmensgrenzen hinweg sowie die jeweils gültigen Zuordnungsregeln geistiger Eigentumsrechte. Ein wesentliches Charakteristikum aller Elemente, die als quasi öffentliche Güter zur technologischen Infrastruktur gehören, ist die Breite ihrer Anwendung und die begrenzte Varianz der Intensität ihrer Nutzung bei den einzelnen Akteuren in einem Wirtschaftsraum. Allerdings sind nicht alle Elemente notwendigerweise charakteristisch für ein ganzes Land; eine regionale Differenzierung kann in manchen Technologiebereichen sogar von Vorteil sein, wenn die Entstehung effizienter Cluster dadurch begünstigt wird.

Die genannten Begriffe haben unmittelbar Relevanz für die Produktion medizinischer Leistungen. Als generische Technologie können auch hier die Ergebnisse der ersten – nicht auf den Erwerb geistiger Eigentumsrechte ausgerichteten – Phase in der Entwicklung neuer Technologien bezeichnet werden. Diese Phase endet im Allgemeinen mit einem Prototyp, der im Labor

getestet werden kann. Tasse (1992) vermutet, dass ein einzelnes privates Unternehmen aufgrund fehlender Eigentumsrechte selbst keine ausreichenden Anreize hat, in generische Technologien zu investieren. Es ist offensichtlich, dass dies auf einen Großteil der medizinischen Technologie zutrifft. Viele generische Technologien der Medizin, wie zum Beispiel die Methoden der Organtransplantation, werden in den Universitätskliniken entwickelt und getestet. Auch die Entwicklung der biotechnologischen Grundlagen für die Schrittmacherinnovation der modernen pharmazeutischen Industrie gehört zum Bereich der generischen Technologie. Eine zentrale Rolle spielt dabei die Ausbildung junger Wissenschaftler in den Forschungslabors der Universitäten.

Der Begriff Infratechnologie subsumiert *erstens* die Summe der wissenschaftlichen und ingenieurtechnischen Daten, die in der Durchführung von Forschung und Entwicklung und in der Kontrolle von Produktionsprozessen benötigt werden, *zweitens* die Summe der Mess- und Testmethoden, die zur Qualitätssicherung in Forschung und Entwicklung, in der Kontrolle der laufenden Produktion und – bei vielen Produkten der Spitzentechnik – auch zur Anpassung und Wartung am Ort des Kunden gebraucht werden, und *drittens* all jene organisatorischen Praktiken und Techniken, die eine effiziente Nutzung komplementärer Technologien in ein und demselben Produktionsprozess ermöglichen. In der Medizin obliegt es traditionell in erster Linie den Universitätskliniken, die Infratechnologie bereitzustellen, die für die Entwicklung und Anwendung neuer Medizintechnologien gebraucht wird.

Wenn der Innovationsprozess in der Medizin durch die Entwicklung einer geeigneten technologischen Infrastruktur weiter beschleunigt und qualitativ verbessert werden soll, ist in Zukunft verstärkt in die Entwicklung der entsprechenden Managementmethoden zu investieren, die als ein die ärztlichen Praxisgrundsätze ergänzender Teil der Infrastruktur des Gesundheitswesens gelten könnte. Im Grunde geht es um nichts anderes als eine medizinspezifische

Variante des „Concurrent Engineering“, der simultanen Planung von Produkten und Produktionsprozessen mit dem Ziel, gleichzeitig die Qualität des Endproduktes und die Effizienz des Produktionsprozesses zu optimieren.

Bevor ich im Einzelnen die Anforderungen einer effizienten Organisation der medizinischen Leistungserstellung darstelle, soll die Rolle der technologischen Infrastruktur kurz anhand der Entwicklung eines ähnlich informationsintensiven Wirtschaftszweiges genauer analysiert werden. Baetjer (1998) bietet eine solche Analyse am Beispiel der Softwareindustrie. Er sieht die größte Herausforderung der Softwareentwicklung in der Koordination verteilten Lernens. Kein einzelner Programmierer kann die komplexe Entwicklungsarbeit eines modernen Softwareprogramms alleine leisten; es ist unvermeidlich, die Aufgaben auf viele Köpfe zu verteilen, die nicht an einem Ort zusammenkommen, ja sich noch nicht einmal persönlich kennen müssen.

Damit dies gelingt, muss das Programmiervorhaben nach genau zu definierenden Kriterien modularisiert werden, so dass die einzelnen Module komplementär zueinander sind. Einen konsistenten Ansatz bietet die normative Theorie der objektorientierten Programmierung. Ihr Anliegen ist die Gestaltung von Programmieraufgaben als ein sozialer Lernprozess. Deshalb zielen die einzelnen Prinzipien der objektorientierten Programmierung darauf ab, die Kommunikation zwischen Programmierern zu erleichtern, klar abgegrenzte Teilaufgaben zu zuweisen und allen Beteiligten das übergeordnete Projektziel deutlich zu machen. Ein Standardlehrbuch zur objektorientierten Programmierung ist Booch (1994).

Zu den wesentlichen Prinzipien gehören die Einkapselung einzelner Programmbestandteile (Objekte), die strukturierte Nachrichtenweitergabe zwischen diesen, der Polymorphismus (die unterschiedliche Reaktion verschiedener Objekte auf ein und dieselbe Information), das Verbergen von Informationen

innerhalb von Objekten, die Klassenstruktur (die Objekte gleicher Art und gleicher Reaktionsweise zusammenfasst) und das Vererbungsprinzip (das die Beziehung zwischen Klassen mit teilweise gleicher Reaktionsweise festlegt). Diese Prinzipien sind in ihrer Wirkung komplementär und ermöglichen nicht nur neue Formen der Interaktion zwischen den an einem Projekt beteiligten Programmierern, sondern auch zwischen Softwarenutzern und Softwareentwicklern sowie zwischen den Entwicklern und ihren Werkzeugen. Aber das eigentliche Ziel der Modularität ist, wie Baetjer (1998) zu Recht betont, die Evolutionsfähigkeit der Software. Evolutionsfähigkeit wird daran gemessen, wie gut ein Softwareprodukt sich an geänderte Spezifikationen anpassen und wie leicht es sich mit anderen Softwareprodukten kompatibel machen lässt.

Aus diesen Anforderungen lassen sich Regeln für eine gute Softwareentwicklung ableiten: Anstatt ein neues Softwareprodukt für die speziellen Umweltbedingungen einer einzelnen Anwendung zu optimieren, sollte eine möglichst allgemeingültige Lösung gesucht und modular umgesetzt werden. Denn erst damit wird gewährleistet, dass andere Programmierer in der Zukunft die Möglichkeit zur Weiterentwicklung haben. Um ein Maximum an Flexibilität in die Software zu implementieren, sollte stets die beste verfügbare objektorientierte Vorgehensweise gewählt werden. Wenn gleichzeitig darauf geachtet wird, dass einzelne Komponenten in anderen Zusammenhängen wieder verwendet werden können, entsteht aus der Vielzahl der Entwicklungsprojekte fast automatisch eine technologische Infrastruktur und eine Art Markt für Software-Komponenten, die aufgrund standardisierter Schnittstellen und Qualitätsanforderungen reibungslos zusammenarbeiten.

Die Open-Source-Bewegung hat viele dieser Ideen aufgenommen und umgesetzt. Sie hat gezeigt, dass ein Markt für Software-Komponenten auch entstehen kann, wenn den einzelnen Transaktionen dabei kein monetärer Zahlungsfluss gegenüber steht, sondern die Entwickler neuer Software-Komponenten durch

nicht-pekuniäre Erträge entlohnt werden. Lerner und Tirole (2002) argumentieren, dass die Anreize, die Open-Source-Entwickler motivieren, vor allem in der persönlichen Anerkennung durch andere Programmierer und in der Eröffnung neuer beruflicher Chancen liegen und als Signalling-Strategie im Sinne der ökonomischen Allokationstheorie bei asymmetrischen Informationen interpretiert werden können. Diese Anreize sind umso stärker, je größer die qualitative Überlegenheit von Open-Source-Software gegenüber vergleichbarer konventioneller Software ist. Die privaten Anreize sind daher direkt mit dem *sozialen* Ertrag aus der Entwicklung von Open-Source-Software verknüpft.

Ähnliche Anforderungen wie in der Software-Entwicklung sind auch an die Entwicklung der Informationsinfrastruktur für das Gesundheitswesen zu stellen. Auch hier ist eine enge Kooperation zwischen Forschung und Praxis bei der Entwicklung komplexer Systemtechnologien erforderlich. Private Unternehmen sind zwar in vielen Wirtschaftszweigen die erfolgreicherer Innovatoren, aber Detmer (2003) nennt drei überzeugende Gründe, warum der öffentliche und der private Sektor gemeinsam die Informationsinfrastruktur für das Gesundheitswesen entwickeln sollten. *Erstens* wächst die Informationsmenge im Gesundheitswesen exponentiell, so dass eine privat finanzierte Infrastruktur schnell an ihre Kapazitätsgrenze stoßen würde und Anbietern mit monopolistischer Preissetzungsmacht dann sehr starke Anreize geben würde, die Preise für den Zugang zu Information zu erhöhen, anstatt in die Ausweitung des Zugangs zu investieren. Über jeden einzelnen Patienten werden immer mehr Daten gesammelt und verfügbar gemacht – einerseits, weil die Menschen länger leben und dabei mehr chronische Krankheiten erleben, andererseits weil die diagnostischen Möglichkeiten sich erheblich verbessert haben und viel mehr Tests durchgeführt werden können. Auch in der wissenschaftlichen Forschung und in der pharmazeutischen Industrie haben sich die Möglichkeiten der Informationsverarbeitung vervielfacht. Durch die biotechnologische Revolution

können die Informationen heute auch dort viel systematischer ausgewertet werden.

Der *zweite* Grund liegt darin, dass Standards gesetzt werden müssen, um den Austausch von Informationen zu erleichtern und die Kosten der Informationsverarbeitung zu senken. Dies erfordert unter anderem eine kontinuierliche Weiterentwicklung der medizinischen Terminologie, die ohne Beteiligung der Universitäten und staatlichen Regulierungsbehörden kaum zu bewältigen wäre. Berger et al. (1997) erwarten, dass insbesondere die internationalen Qualitätsstandards der evidenzbasierten Medizin bei der Auswertung empirischer Forschungsarbeiten über therapeutisch relevante Fragestellungen zu grundlegenden Veränderungen in der medizinischen Publikationspraxis führen, weil die Autoren an der Aufnahme ihrer Arbeiten in die entsprechenden Datenbanken interessiert seien. Berger et al. (1997), S. 350, schreiben: „Wenn mit einer Steigerung der Qualität gleichzeitig auch eine Verminderung der Quantität medizinischer Publikationen einherginge – es gibt weltweit ca. 25000 medizinische Zeitschriften – wäre das äußerst wünschenswert.“

Der *dritte* Grund liegt darin, dass eine technologische Plattform benötigt wird, die den Zugang aller Organisationen im Gesundheitswesen zur Informationsinfrastruktur gewährleistet. Es ist klar, dass die Informationsinfrastruktur eine Art öffentliches Gut darstellt, von dem viele Beteiligte und vor allem zukünftige Generationen profitieren werden. Detmer (2002) weist zu Recht darauf hin, dass in einem wettbewerblich organisierten Gesundheitswesen die privaten Akteure unzureichende Anreize haben, in den Aufbau einer solchen Informationsinfrastruktur zu investieren. Zudem haben die Vergütungssysteme in der Vergangenheit kaum Anreize gesetzt, Investitionen in zukünftige Effizienzverbesserungen zu tätigen. Mit dem Aufbau der Informationsinfrastruktur sollte daher eine Prioritätenverschiebung einhergehen, die künftig die Qualität der medizinischen Leistungen stärker honoriert.

Öffentliche Information über die Qualität einzelner Anbieter ist eine wichtige Voraussetzung dafür, dass die Nachfrage nach medizinischen Leistungen durch eine korrekte Bewertung gesteuert wird. Dabei ist jedoch zu beachten, dass die Veröffentlichung qualitätsbezogener Informationen die Leistungsanbieter zu einer sozial unerwünschten „positiven“ Risikoselektion unter den Patienten veranlassen könnte. Durch das Vermeiden und Abwimmeln „schwieriger“ Patienten, bei denen der angestrebte Heilerfolg nur mit überdurchschnittlichem Aufwand zu erreichen ist, könnte ein Anbieter sich im Qualitätsvergleich zu Wettbewerbern besser stellen, wenn in diesem Vergleich die durchschnittliche Fallschwere nicht oder nur unvollständig berücksichtigt wird. Aus theoretischer Sicht besteht hier staatlicher Regulierungsbedarf. Allerdings ist die empirische Evidenz nicht eindeutig. Cutler et al. (2004) beispielsweise berichten in einer Studie über die Auswirkungen eines Qualitätsvergleichs zu Herzoperationen in New York, dass die als schlecht eingestuften Krankenhäuser relativ gesunde Patienten an konkurrierende Krankenhäuser verloren haben, aber die Anzahl ihrer Operationen an „schwierigen“ Patienten nicht oder nur geringfügig reduzierten.

4.3 Die Rolle von Behandlungsleitlinien

Behandlungsleitlinien sind Empfehlungen, die die Entscheidungsfindung in der medizinischen Praxis verbessern sollen. Die meisten Entscheidungen, die Ärzte in ihrem Berufsalltag zu treffen haben, sind zu kompliziert, um jedes Mal vollständig neu vorbereitet und ausgeführt zu werden. Hierzu müsste ein Arzt nicht nur alle denkbaren Behandlungsoptionen identifizieren und im Hinblick auf die verfügbare wissenschaftliche Evidenz bewerten, sondern auch ihre Konsequenzen müsste er umfassend berücksichtigen und mit den erwarteten Kosten vergleichen, wobei jedes Mal eine Vielzahl von logistischen, wirtschaftlichen, juristischen, sozialen und personenbezogenen Faktoren zu berücksichtigen wäre.

Der Zweck von Behandlungsleitlinien ist es aber nicht nur, jene Entscheidungen, die sich in ähnlicher Form wiederholen, zu vereinfachen, sondern auch ihre Qualität und die Kosten-Effektivität des Ressourceneinsatzes zu verbessern. Behandlungsleitlinien richten sich daher in erster Linie auf häufig auftretende und chronische, weniger auf seltene Krankheiten mit kurzen Verlaufszeiten. Bei letzteren ist eine vollständig individualisierte Entscheidungsfindung praktisch unvermeidlich. Die Integration von Leitlinien in die Behandlung abgegrenzter Krankheitsbilder ist vor allem in amerikanischen Health Maintenance Organisationen (HMOs) weit fortgeschritten, die dem einzelnen Vertragsarzt oft sogar finanzielle Anreize zur Anwendung zentral vereinbarter Behandlungsleitlinien bieten. Es ist klar, dass die HMOs auch selbst von der erfolgreichen Implementierung ihrer Behandlungsleitlinien profitieren, weil dies nicht nur die Kosten-Effektivität, sondern auch die Kalkulierbarkeit medizinischer Behandlungsepisoden verbessern.

Eddy (1990b) beschreibt das Potential von Behandlungsleitlinien, die medizinische Praxis auch über organisatorische Grenzen hinweg einer direkten Steuerung zugänglich zu machen. Leitlinien stellen ein Vehikel dar, mit dem sich Informationen über neue medizinische Technologien, die Ergebnisse wissenschaftlicher Forschung und wichtige Erkenntnisse aus klinischer Erfahrung in ein Pool des Wissens einführen lassen, das allen Ärzten offen steht. Je nach dem Grad ihrer Verbindlichkeit, können Leitlinien den Ärzten auch Anreize setzen, sich aktiv um das neue Wissen zu bemühen; sie können also gleichzeitig das Angebot und die Nachfrage medizinischer Informationen verbessern.

Mit Eddy können wir die Gesamtheit der Praxisgrundsätze („Practice Policies“), zu denen die Behandlungsleitlinien gehören, als das „zentrale Nervensystem der medizinischen Praxis“ bezeichnen. Praxisgrundsätze beeinflussen die medizinische Praxis auf vielfältige Weise. Sie beeinflussen die Auswahl von Methoden in der Zukunft und die Bewertung von durchgeführten Behandlungen in der

Vergangenheit. Als wichtige Determinante der tatsächlichen medizinischen Praxis sind sie zudem ein unverzichtbares Element in jeder Strategie, die Qualität der medizinischen Versorgung zu verbessern. Nach Eddy (1990b), S. 878, sind drei Arten von „Practice Policies“ zu unterscheiden:

1. *Standards*, die auf jeden Fall eingehalten werden müssen, aber in der Regel nicht so umfassend sind, dass sie die Behandlung eines Krankheitsfalles vollständig beschreiben. Solche Standards werden im deutschen Kontext auch als *Richtlinien* bezeichnet.
2. *Leitlinien* („Guidelines“), die hinreichend flexibel sein müssen, um im Einzelfall eine abweichende Behandlung zu erlauben. Die Abweichung ist aber zu begründen. Die Weltgesundheitsorganisation definiert Leitlinien als „systematische Feststellungen“, um „Entscheidungen von Ärzten und Patienten über eine Gesundheitsversorgung in spezifischen medizinischen Situationen zu unterstützen“.¹⁷
3. *Optionen*, die lediglich alternative Behandlungsmöglichkeiten darstellen. Wenn eine Behandlungsmöglichkeit als Option klassifiziert ist, schließt dies andere Optionen nicht aus. Die Optionen sind aber von den Nicht-Optionen abzugrenzen, so dass sie das Spektrum der Möglichkeiten doch einschränken und dem Arzt einen Teil der Entscheidungsfindung abnehmen. Solche Optionen werden daher im deutschen Kontext auch als *Empfehlungen* bezeichnet.

Bei der Formulierung von Behandlungsleitlinien lassen sich vor allem zwei Ansätze unterscheiden: die globale subjektive Bewertung und die evidenzbasierte Medizin, zu der auch die von Eddy (1990c) gesondert diskutierten ergebnisorientierten und präferenzbasierten Leitlinien gehören. Bei dem ersten Ansatz wird ein Panel von Experten gebeten, alternative medizinische Optionen

zu bewerten und eine konsensuale Empfehlung für die Behandlung eines abgegrenzten Krankheitsbildes zu geben. Dieser auch „konsensus-orientiert“ genannte Ansatz ist vergleichsweise billig.

Beim evidenzbasierten Ansatz wird alle verfügbare Evidenz beschrieben und zur Begründung von Handlungsempfehlungen verwendet. Mit Eddy (1990c) lassen sich verschiedene Stufen dieses Ansatzes unterscheiden, die je nach Datenlage und Zielsetzung zum Einsatz kommen. Allerdings werden die möglichen Ergebnisse und Risiken einzelner Optionen nicht immer ökonomisch bewertet und miteinander verglichen. Erst der ergebnisorientierte Ansatz ergänzt die Einsichten der evidenzbasierten Medizin um eine Bewertung auf der Basis empirischer Ergebnisforschung („Outcomes Research“). Dabei werden die Ergebnisse mit rein wissenschaftlichen Methoden analysiert und quantifiziert. Dieser Ansatz kann schon sehr teuer sein. Bei dem präferenzbasierten Ansatz schließlich wird zusätzlich versucht, mithilfe psychologischer Methoden die Präferenzen der Patienten abzubilden und in die Bewertung von Ergebnissen sowie therapeutischen und diagnostischen Verfahren einzubeziehen.

Nach Eddy (1990d) muss die Entwicklung ergebnisorientierter Leitlinien mindestens die folgenden sechs Elemente umfassen: *Erstens* die Identifikation möglicher Ergebnisse und die davon getrennte Beurteilung ihrer jeweiligen Vor- und Nachteile. *Zweitens* die Abschätzung der Wirkung einer vorgeschlagenen Intervention auf alle Ergebnisse zu, die für den Patienten wichtig sind. *Drittens* die Einschätzung, wie die Ergebnisse von Charakteristika der Patienten abhängen, die ein Arzt in seiner Praxis tatsächlich identifizieren kann und auf die er durch eine individuelle Anpassung der Therapie reagieren kann. *Viertens* Ergebnisschätzungen, die sich so weit wie möglich auf experimentelle wissenschaftliche Evidenz („Randomised Controlled Trials“) stützen. *Fünftens*

¹⁷ Zitiert nach Heimpel (2003), S. 226.

die Verwendung möglichst gut gesicherter Methoden bei der Schätzung der Ergebnisse. Und *sechstens* die Berücksichtigung von Präferenzen der Patienten auf der Grundlage empirischer Methoden.

Die Formulierung von Behandlungsleitlinien erfordert weitaus mehr Aufwand als die isolierte Behandlungsentscheidung für einen einzelnen Patienten. Denn die Konsequenzen *falscher* Behandlungsleitlinien können wegen der großen Zahl potentieller Betroffener weitaus schwerwiegender sein als die einer einzelnen Fehlentscheidung. Weil medizinische Entscheidungen oft so unsicher sind, ist es unvermeidlich, dass einzelne Ärzte Fehler machen. Im Umkehrschluss heißt das aber auch: Soweit strikt befolgte Behandlungsleitlinien eine Vereinheitlichung der medizinischen Praxis herbeiführen, können sie das Lernen aus praktischer Erfahrung behindern. Diese Gefahr erscheint besonders akut, wenn bei der Umsetzung einer Behandlungsleitlinie die empirische Varianz relevanter Patientencharakteristika ignoriert wird.

Um die nötige Flexibilität zu erhalten, sollten medizinische Praxisgrundsätze nicht nur aus Standards bestehen, die auf jeden Fall angewendet werden müssen. Vielmehr müssen sie auch Behandlungsleitlinien enthalten, die sich an die besonderen Umstände des einzelnen Falles anpassen lassen, ohne dass dem Arzt Fehlverhalten vorgeworfen werden kann. Ein wichtiger Beitrag zur Flexibilität sind darüber hinaus die Behandlungsoptionen, die keine explizite Handlungsempfehlung für den Arzt beinhalten.

Nach Eddy (1990g) sind medizinische Interventionen als *Behandlungsoptionen* zu klassifizieren, wenn entweder die zu erwartenden medizinischen Ergebnisse oder aber die Präferenzen der Patienten bezüglich dieser Ergebnisse unbekannt sind. Dies kann darauf zurückzuführen sein, dass die Ergebnisse von unbekanntem Kontextfaktoren abhängen oder dass die Patientenpräferenzen sehr uneinheitlich ausgeprägt sind. Wenn die Konsequenzen einer medizinischen

Intervention unbekannt sind, sollte natürlich zunächst zusätzliche Forschung durchgeführt werden, um die Ergebnisse besser zu definieren und zu quantifizieren. Lassen sich keine allgemeinen Patientenpräferenzen messen, sollte der Arzt die möglichen Ergebnisse einer medizinischen Intervention mit seinem Patienten besprechen, um die individuellen Präferenzen im Einzelfall festzustellen. Empirische Forschung hat gezeigt, dass die Compliance der Patienten mit einer therapeutischen Strategie besser ist, wenn der Arzt vorher alle Optionen mit seinem Patienten besprochen hat. Die Umsetzung von Behandlungsleitlinien schließt nicht aus, dass die Patienten aktiv an der Entscheidungsfindung in ihrem Krankheitsfall beteiligt werden.¹⁸

Ein weiteres Problem ist die oftmals mangelhafte Compliance der Ärzte mit medizinischen Behandlungsleitlinien, an deren Entstehung sie nicht selbst beteiligt waren (Hasenbein et al. 2003 und Heimpel 2003). Nach Erdmanns (2003) Einschätzung werden in vielen Teilgebieten der Medizin allenfalls 30 Prozent der Patienten im Einklang mit nationalen und internationalen Leitlinien behandelt. Das Compliance Problem stellt sich verschärft, wenn die Zahler für medizinische Leistungen Behandlungsleitlinien nutzen wollen, um medizinische Entscheidungen an dem Kriterium der Kosten-Effektivität auszurichten, während die Ärzte – wie die Patienten – im Allgemeinen an einer Behandlung interessiert sind, deren Nutzen die Risiken für den einzelnen Patienten übersteigt. Dieser Konflikt lässt sich nicht dadurch lösen, dass man die Formulierung von Leitlinien allein den Ärzten vor Ort, etwa in einem Krankenhaus oder in einer abgegrenzten Region, überlässt, auch wenn solche „internen Leitlinien“ – wie Heimpel (2003) argumentiert – in der Praxis eher akzeptiert

¹⁸ In Deutschland werden medizinische Leitlinien in mehreren Versionen veröffentlicht, von denen eine die Empfehlungen der Leitlinie speziell für Patienten zusammenfasst, während eine andere die klinisch relevanten Informationen für den behandelnden Arzt in einer übersichtlichen Form darstellt. Die gesamte Leitlinie, die zusätzliche Informationen zur Methodik ihrer Entwicklung und einen Überblick über den aktuellen Stand der medizinischen Wissenschaft enthält, steht zwar als Hintergrunddokument zur Verfügung, wird aber in der Praxis selten konsultiert.

werden. Vielmehr muss die Formulierung von Behandlungsleitlinien auf wissenschaftlichen Kosten-Effektivitätsanalysen beruhen. Dabei sollte jedoch nicht einfach die Perspektive des Drittzahlers für die Leistungen, zum Beispiel einer Krankenversicherung, eingenommen werden, sondern die Perspektive der Volkswirtschaft insgesamt. Damit ist ein wichtiger Grund genannt, warum die Formulierung und Umsetzung medizinischer Behandlungsleitlinien nicht allein den Ärzten oder den Versicherungen überlassen werden kann, sondern eine gesundheitspolitische Aufgabe ist, die ohne ein Mindestmaß an staatlicher Unterstützung und Regulierung kaum zu bewältigen ist.

In Deutschland hat es sich die *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften* (AWMF) zur Aufgabe gemacht, die Entwicklung von Leitlinien für die Diagnose und Therapie einzelner Krankheiten zu koordinieren und die Ergebnisse über das Internet zu verbreiten. Albert et al. (2003) beschreiben das methodische Vorgehen bei einer Evidenz-, konsensus- und ergebnisbasierten Leitlinienentwicklung anhand des Beispiels der Brustkrebs-Früherkennung. Brustkrebs ist eine von vier Krankheiten, für die sich Leistungserbringer, Kostenträger und politisch Verantwortliche im deutschen Gesundheitswesen auf die beispielgebende Entwicklung von Disease-Management-Programmen geeinigt haben. Ziel ist dabei nicht die isolierte Beschreibung fachmedizinischer Teilbereiche, sondern die Darstellung der gesamten Versorgungskette und die qualitätssichernde Abstimmung der notwendigen interdisziplinären Kooperation, wie sie für das "Disease-Management" charakteristisch ist. Die Qualitätssicherung anhand reproduzierbarer Indikatoren soll sich daher ebenfalls auf das Gesamtergebnis der Versorgungskette – weniger auf fachliche Teilbereiche, sondern vor allem auf die Qualität der Schnittstellen in der Versorgungskette – beziehen, so dass typische Abstimmungsprobleme wie zum Beispiel psychisch oft belastende Doppeluntersuchungen vermieden werden. Beteiligt an der Entwicklung der

Leitlinieninhalte zur Brustkrebsbehandlung waren Vertreter aller Gruppen, auf deren Akzeptanz es bei der Anwendung der Leitlinie ankommt: neben den medizinisch-wissenschaftlichen Fachgesellschaften auch Berufsverbände, Patientenverbände und andere Interessengruppen.

Evidence-based medicine. Bislang fließt in die Leitlinienentwicklung noch zu wenig Feedback aus der Praxis ein. Bis Anfang der 90er Jahre fehlte im Grunde eine geeignete Infratechnologie, eine Art Software, um die Informationen aus der Praxis systematisch auszuwerten und gezielt weiterzugeben. Seit Mitte der 90er Jahren hat sich die evidenzbasierte Medizin zur wichtigsten Infratechnologie des Gesundheitswesens entwickelt. Ausgehend von Kanada, wo eine von Gordon Guyatt geleitete Forschergruppe der McMaster University den Startschuss gab, hat sie sich im angelsächsischen Raum bereits fest etabliert.¹⁹

Sackett et al. (2000), S. 1, beschreiben evidenzbasierte Medizin als „die Integration des aktuellen Standes der empirischen Forschung mit klinischer Expertise und den Wertvorstellungen des Patienten“. Der aktuelle Stand der empirischen Forschung ergibt sich aus allen für die klinische Praxis relevanten Studien, insbesondere aus klinischer Forschung, die Beobachtungen an Patientengruppen auswertet, um die Präzision diagnostischer Tests und deren Prognoseeigenschaften sowie die Effektivität und Sicherheit therapeutischer, rehabilitierender und vorbeugender Maßnahmen zu untersuchen. Mit klinischer Expertise ist die Fähigkeit des Arztes gemeint, seine speziellen Fähigkeiten und Erfahrungen zu nutzen, um den Gesundheitszustand seines Patienten zu diagnostizieren und individuelle Risiken und Nutzen aus potenziellen therapeutischen Handlungen sowie die persönlichen Wertvorstellungen des Patienten zu erkennen. Aus den Wertvorstellungen des Patienten lassen sich seine Erwar-

¹⁹ Als Geburtsstunde der evidenzbasierten Medizin kann die Veröffentlichung von Evidence-Based Medicine Working Group (1992) gelten.

tungen an die Medizin ableiten, die in der medizinischen Entscheidungsfindung zu berücksichtigen sind, wenn dem Wohl des Patienten gedient werden soll.

Berger et al. (1997), S. 350, bringen die Radikalität dieser so selbstverständlich klingenden Zielsetzung zum Ausdruck, wenn sie schreiben: „Die *evidence-based medicine* macht mit ihrer *external evidence* ein naturwissenschaftlich rationales Handwerkszeug für die handelnde Medizin verfügbar und steht damit im Kontrast zu der als ärztliche Kunst verniedlichten, an der traditionellen Medizinhierarchie ausgerichteten Irrationalität vergangener Epochen“. Wie stark diese Irrationalität noch heute nachwirkt, macht Erdmann (2003) in seiner kritischen Würdigung des im Jahr 2003 vorgelegten Referentenentwurfes für ein deutsches Gesetz über die Verordnungsfähigkeit von Arzneimitteln in der vertragsärztlichen Versorgung deutlich: „Hier werden unwirksame, nicht geprüfte und obskure Pharmaka auf Kosten der Allgemeinheit in Zukunft mit dem Rückhalt einer 'Positivliste' als 'verordnungsfähig' gläubigen Kranken gegenüber dargestellt. (...) Liest man, was hier als 'verordnungsfähig' aufgelistet wird, so glaubt man sich in das frühe Mittelalter versetzt: Hodenextrakte, lapis albus (weißer Stein), mucosa oculi suis (Schleimhaut des Schweineauges), (...) die Haut weiblicher Rinderfeten, Potenzholz, Kohle, Gunpowder, ja sogar anus bovis (A...loch der Kuh) sind also jetzt als verordnungsfähig deklariert? Man kann testes juveniles bovis (Jungstierhoden) verordnen!“

Sackett et al. (2000), S. 2, nennen *vier Gründe*, warum die Grundsätze der evidenzbasierten Medizin in der Vergangenheit allenfalls eine geringe Rolle in der medizinischen Praxis gespielt haben: *Erstens*, die Nachfrage nach valider Information über Diagnose, Prognose, Therapie und Prävention war im ärztlichen Alltag einfach zu groß, um zeitnah befriedigt zu werden. *Zweitens*, die konventionellen Quellen entsprechender Informationen waren oftmals veraltet (Lehrbücher) oder falsch (Experten), ineffektiv (medizinische Fortbildung) oder wegen eines zu großen und zu anspruchsvollen Informations-

volumens bei gleichzeitig fragwürdiger Validität für die praktische Anwendung kaum zu gebrauchen (medizinische Fachzeitschriften). *Drittens* klaffte bei vielen Ärzten eine große Lücke zwischen ihren diagnostischen Fähigkeiten und ihrem klinischem Urteilsvermögen, welches mit Erfahrung zunimmt, *einerseits* und dem Umfang des eigenen medizinischen Wissens im Verhältnis zu dem jeweils neuesten Stand der medizinischen Wissenschaft *andererseits*. Diese Lücke lässt nicht nur die klinischen Erfolge bei vielen Ärzten mit dem Alter abnehmen, sondern macht es auch immer schwieriger, aktuelle Informationsquellen effektiv zu nutzen. *Viertens* haben viele Ärzte oft zu wenig Zeit für den einzelnen Patienten, um die aktuellste Evidenz zu suchen und in ihren diagnostischen Entscheidungen zu integrieren.

Berger et al. (1997) beschreiben anhand von Beispielen, wie die nicht-evidenzbasierte Medizin noch bis heute bestimmte Vorgehensweisen in Diagnose und Therapie „ritualisiert“, obwohl es sich dabei vielfach um Trugschlüsse handelt, die für die Patienten teils nutzlos und teils sogar schädlich sind. Allein die Verordnung von Medikamenten ohne gesicherte Wirkung verschlingt Berger et al. (1997) zufolge in Deutschland jedes Jahr mehrere Milliarden Euro. Hinzu kommen Schäden, die durch Therapien entstehen, deren einzelne Elemente zwar isolierte Wirkungen haben mögen, die aber nicht als Gesamtstrategie durch methodisch einwandfreie Studien abgesichert sind, deren Massstab der Gesundheitszustand am Ende aller Behandlungsschritte ist.

Sackett et al. (2000), S. 3, nennen fünf *neuere Entwicklungen*, die es seit Mitte der 90er Jahre immer mehr Ärzten erlauben, die genannten Probleme zu überwinden: *Erstens* sind Strategien zur effizienten Auswertung neuer Evidenz über therapeutische Möglichkeiten entwickelt worden, insbesondere im Hinblick auf ihre Validität (den Wahrheitsgehalt) und ihre Relevanz (die Größe des zu erwartenden Effektes und die Anwendbarkeit in der klinischen Praxis).

Berger et al. (1997) weisen darauf hin, dass sich erst in den letzten zwei Jahrzehnten die Methodik der „Randomised Controlled Clinical Trials“ (RCT) zum Standard der Evaluierung medizinischer Alternativen bei medikamentösen und chirurgischen Behandlungen entwickelt hat; in jüngster Zeit wird diese Methode, die einen qualitativen Quantensprung in der medizinischen Forschung bewirkt hat, auch zur Evaluierung diagnostischer Strategien eingesetzt. *Zweitens* gibt es heute systematische Übersichten über die empirischen Wirkungen medizinischer Interventionen, zum Beispiel durch die sogenannte Cochrane-Collaboration. *Drittens* gibt es heute Fachzeitschriften speziell zur evidenzbasierten Medizin, die jene zwei Prozent aller medizinischen Aufsätze aus anderen Zeitschriften enthalten, die von unmittelbarem klinischem Anwendungsnutzen sind, weil sie therapeutisch relevante Fragen behandeln und valide Ergebnisse berichten. *Viertens* sind durch die Fortschritte der Informationstechnologie elektronisch vernetzte Datenbanken entstanden, die die genannten Informationen dem einzelnen Arzt innerhalb von Sekunden zur Verfügung stellen können. *Fünftens* sind Strategien des lebenslangen Lernens entwickelt worden, mit denen Ärzte ihre klinische Erfolgsrate kontinuierlich verbessern oder zumindest aufrecht halten können.

Die von der Cochrane-Collaboration herausgegebene Cochrane-Library ist eine elektronische Publikation mit dem Anspruch, zu jeder therapeutischen Fragestellung die beste verfügbare Evidenz anzubieten. Dazu werden systematische Übersichtsarbeiten („Reviews“) erstellt, die eine Vielzahl relevanter Studien in strukturierter Form nach einem vorab festgelegten Protokoll auswerten und darstellen. Bei einem Cochrane-Review handelt es sich also im Idealfall um eine formal standardisierte Zusammenfassung aller zu der jeweiligen Fragestellung relevanten klinischen Studien, die vorab festgelegten Qualitätskriterien genügen soll. Sobald eine hinreichende Zahl neuer Studien erschienen ist, wird

das Verfahren zur Überarbeitung des jeweils relevanten Cochrane-Reviews eingeleitet.

Dabei werden die zugrundeliegenden Originalarbeiten nach der Zuverlässigkeit der wissenschaftlichen Evidenz auf sechs Stufen bewertet, die von der amerikanischen *Agency of Health Care Policy and Research* festgelegt wurden: Meta-Analysen, die eine Vielzahl von RCT-Studien nach bestimmten Regeln zusammenfassen, gelten als die höchste Stufe der wissenschaftlichen Evidenz (Ia), gefolgt von Evidenz aus einzelnen RCT-Studien (Ib), aus einzelnen gut angelegten und durch eine Kontrollgruppe gesicherten, aber nicht randomisierten Studien (IIa), aus quasi-experimentellen Studien ohne Kontrollgruppe (IIb), aus deskriptiven Studien (III) und schließlich aus Berichten von Expertengruppen oder anerkannten medizinischen Autoritäten (IV). Das Deutsche Institut für Medizinische Dokumentation und Information (www.dimdi.de) hält die „Cochrane Database of Systematic Reviews“ zur Recherche im Internet bereit. Weitere Quellen für methodisch valide und klinisch relevante Arbeiten zu einzelnen therapeutischen Fragestellungen sind die beiden Zeitschriften „Evidence Based Medicine“ und „ACP Journal Club“, die in keiner Klinik-Bibliothek fehlen sollten.

Es ist noch zu früh, den Einfluss der evidenzbasierten Medizin auf die Produktionseffizienz in der Erstellung medizinischer Leistungen aus volkswirtschaftlicher Perspektive empirisch zu untersuchen. Aber erste Beobachtungen sprechen offenbar für einen positiven Einfluss. Wennberg et al. (2002) weisen auf die amerikanischen Erfahrungen von freiwilligen Initiativen regional kooperierender Leistungsanbieter hin, die eine Forschungs- und Verwaltungsinfrastruktur implementiert haben, um systematisch die jeweils beste medizinische Praxis zur Anwendung zu bringen. Diese Experimente haben die regionale Mortalität und Sterblichkeit aufgrund medizinischer Fehler

offenbar reduziert. Als ein Beispiel nennen Wennberg et al. (2002) die *Northern New England Cardiovascular Study Group*.

Die Grenzen der evidenzbasierten Medizin werden da erreicht, wo Krankheiten zu selten auftreten, um valide klinische Studien zu allen Behandlungsalternativen durchzuführen, die bei einem gegebenen Stand der medizinischen Technologie theoretisch möglich sind. Ohnehin gibt es nur relativ wenige Krankheiten, die so häufig vorkommen, dass es für private Unternehmen lohnt, im großen Stil in die Entwicklung neuer therapeutischer Technologien zu investieren. Je häufiger eine Krankheit auftritt, desto eher lässt sich eine hinreichend große Zahl von Patienten finden, die zur experimentellen Anwendungen neuartiger Therapieformen bereit sind, durch die sich die hohen Anforderungen an RCT-Studien erfüllen lassen.

Klinische Studien im Vorfeld der Zulassung neuer Medikamente sind schon heute für einen wesentlichen Teil der hohen Forschungs- und Entwicklungskosten verantwortlich, die sich oft nur bei hohen Verkaufszahlen amortisieren. Die meisten Krankheiten kommen so selten vor, dass sich kaum genug Patienten für Wirksamkeits- und Nebenwirkungsuntersuchungen nach dem Standard der RCT-Studien finden lassen würden. Wenn staatliche Regulierungen die Erstellung von RCT-Studien zu einer Zulassungsvoraussetzung machen, die der jeweilige Anbieter auf eigenes Risiko zu erfüllen hat, würde dies die finanziellen Anreize zur Entwicklung spezieller Pharmaka zur Behandlung seltener Krankheiten noch weiter schwächen. Allerdings könnte das von Berger et al. (1997), S. 350, prognostizierte Umdenken in der medizinischen Publizistik dazu führen, dass sich künftig ein größerer Teil der in diesem Bereich tätigen Wissenschaftler auf die Erstellung von RCT-Studien konzentriert, so dass dann auch mehr wissenschaftliche Ressourcen für die systematische Untersuchung von Behandlungsalternativen bei seltenen Krankheiten zur Verfügung stehen.

4.4 Steuerungsmöglichkeiten der Gesundheitspolitik

Zielsetzungen. Zunächst ist zu fragen, welche grundlegenden Weichenstellungen erforderlich sind, um das Gesundheitswesen eines Landes mit dem öffentlichen Gut neuen medizinischen Wissens zu versorgen. Erfordert dieses Ziel eine Zentralisierung des Gesundheitswesens oder kann es im Wesentlichen durch ein geeignetes System von Anreizen für autonome Leistungsersteller erreicht werden? Ein geeignetes Anreizsystem könnte sich am Beispiel des wissenschaftlichen Systems orientieren, das im Allgemeinen keine direkte finanzielle Belohnung für Erfolge in der Forschung, sondern Anreize zu einer möglichst frühen Veröffentlichung neuen Wissens setzt, indem es bei Beförderungs- und Berufungsentscheidungen vor allem Originalität und zeitliche Priorität belohnt.

Um die organisatorischen Anforderungen eines effizienten Innovationsprozesses im Gesundheitswesen zu verstehen, bietet sich die Unterscheidung zwischen einem im Wesentlichen statischen medizinischen Wissen vor Einführung universellen Krankenversicherungsschutzes und der heutigen Situation an, in der das medizinische Wissen rasch expandiert, wozu sicherlich die durch den Krankenversicherungsschutz verbesserten Diffusionsbedingungen beigetragen haben. Bei statischem medizinischen Wissen hatte ein Arzt im Prinzip unbegrenzt Zeit, sich durch einen Bayesianischen Lernprozess an das durchschnittliche Verhalten seiner Kollegen anzupassen, so dass in diesem Prozess auch interregionale Unterschiede mit Sicherheit früher oder später eingeebnet wurden.

Heute dagegen ist ein ständiger Wandel zu steuern, der durch einen fast kontinuierlichen Fluss medizinischer Innovationen charakterisiert ist, so dass der Bayesianische Lernprozess des einzelnen Arztes niemals wirklich genug Zeit hat, zu einem Langfrist-Gleichgewicht zu konvergieren. Die beobachtete

Informationsdiffusion wird daher zu jedem beliebigen Zeitpunkt unvollständig sein. Offensichtlich ist dies zum Teil eine unvermeidliche Konsequenz der hohen technologischen Innovationsrate in der modernen Medizin; medizintechnische und pharmazeutische Innovationen sind eine wichtige Quelle regional unterschiedlich verteilter Information und immer wieder neuer Unsicherheit in der medizinischen Praxis. Wie in anderen Wirtschaftszweigen ist der Diffusionsverlauf neuer Technologien im Gesundheitswesen in der räumlichen und zeitlichen Dimension oft recht ungleichmäßig, obwohl in vielen Ländern ein de facto universeller Krankenversicherungsschutz die früher sehr großen Unterschiede in der Zahlungsbereitschaft und -fähigkeit für medizinische Leistungen eingeebnet hat.

Diese stilisierten Beobachtungen machen deutlich, dass die Produktionseffizienz eines Gesundheitswesens heute nicht mehr statisch, sondern nur noch dynamisch definiert werden kann. Um ihre verbliebenen Steuerungsmöglichkeiten realistisch einzuschätzen, muss die Gesundheitspolitik die Trägheit der relevanten Institutionen beachten, wenn sie versucht, durch die Gestaltung Einfluss auf die Produktions- und Innovationsbedingungen zu nehmen. Für einen effizienten Innovationsprozess im Gesundheitswesen sind vor allem die Skalenerträge bei der Entwicklung von Behandlungsleitlinien zu nutzen, Barrieren der Diffusion neuen medizinischen Wissens zu beseitigen und Anreize zur Anwendung der jeweils aktuellsten Behandlungsleitlinien zu setzen.

Der grundlegende Wandel lässt sich mithilfe von Metaphern aus dem verarbeitenden Gewerbe näher beschreiben: längst hat ein Übergang vom medizinischen „Handwerk“, das dem Arzt den individuellen Spielraum einer dezentralen Werkstattproduktion gab, zur arbeitsteiligen „Industrieproduktion“

medizinischer Leistungen stattgefunden.²⁰ In diesem Entwicklungsprozess steht das Gesundheitswesen heute vor der Herausforderung, ein Äquivalent zum „Concurrent Engineering“ zu finden, das die Geschwindigkeit und Qualität von Produktinnovationen in vielen Bereichen des verarbeitenden Gewerbes seit den frühen 90er Jahren erheblich gesteigert hat. Zwar müssen Skalenerträge in der Entwicklung von Leitlinien und im Aufbau einer medizinischen Informationsinfrastruktur möglichst vollständig genutzt werden. Aber dies darf nicht auf Kosten einer notwendigen Differenzierung und flexiblen Anpassung an veränderte Bedingungen im Zeitablauf geschehen.

Die Entwicklung von Behandlungsleitlinien sollte daher ähnlich organisiert werden, wie die Arbeitsteilung zwischen der Konstruktionsabteilung und der laufenden Produktion in einem innovativen Industriebetrieb, wo die Fertigungsingenieure wichtige Informationsintermediäre sind. Lässt sich ein gesundheitspolitisches Äquivalent des „Concurrent Engineering“ finden, sollte dies den Trade-off zwischen einer hohen Innovationsrate und hohen Anpassungskosten in der medizinischen Praxis erheblich entschärfen und so die Chancen für Effizienzgewinne durch ein hohes Innovationstempo im Gesundheitswesen verbessern. Die Erfahrungen aus dem industriellen Innovationsgeschehen legen es nahe, Anreize zu lateraler Kommunikation und regelmäßiger Job-Rotation der Gesundheitsmanager über die in vielen Ländern noch allzu starren organisatorischen Grenzen zwischen Leistungsanbietern, Versicherern und Regulatorischen Behörden zu setzen.

Das „Concurrent Engineering“ kann als eine Philosophie aufgefasst werden, derzufolge alle Aspekte des Lebenszyklus eines Produktes bereits in seiner Konzeption berücksichtigt werden. Das Produkt kann im Kontext der Gesund-

²⁰ Fuchs (1974), Kapitel 3, beschreibt diesen Übergang, während Somers und Somers (1961), S. 55 ff. den historischen Wandel in der Arbeitsteilung zwischen Ärzten und Krankenhäusern, ihren „Workshops“, in den USA beschreiben.

heitspolitik eine medizinische Behandlungsleitlinie sein. Deren Entstehung sollte von einem interdisziplinären Team geplant werden, das auf das Ziel verpflichtet ist, die Qualität dieses Produktes zu erhöhen und die Kosten sowie den Zeitaufwand seiner Entwicklung zu verringern. Um die Gesamtzeit zu reduzieren, sollte eine neue Leitlinie von Anfang an sehr sorgfältig definiert und an die Anforderungen ihrer praktischen Umsetzung angepasst werden. Dazu muss den Leitbildentwicklern gewissermaßen die Chance genommen werden, die besonderen Anforderungen und Restriktionen in der medizinischen Praxis zu ignorieren. Denn eine funktionale Trennung zwischen den Entwicklern medizinischer Leitlinien und den Anwendern in der medizinischen Praxis würde zu ähnlichen Problemen führen, wie sie der Ansatz des „Concurrent Engineering“ im verarbeitenden Gewerbe zum Teil mit großem Erfolg überwunden hat. Insbesondere wenn sich die technologischen Möglichkeiten der medizinischen Praxis rasch ändern und wenn die Zeit zwischen der Entwicklung und Anwendung neuen medizinischen Wissens verkürzt werden muss, ist eine frühzeitige Einbeziehung der Prioritäten und Restriktionen der medizinischen Praxis in der Entwicklung von Behandlungsleitlinien unerlässlich.

Ein wichtiges Element des „Concurrent Engineering“ ist die Entwicklung von vielseitig verwendbaren Bausteinen, die in eine Reihe unterschiedlicher Endprodukte oder Leitlinien eingesetzt werden können. Dies setzt ein gewisses Maß an Standardisierung und Kompatibilität voraus. Gleichzeitig kann so die Adoption beschleunigt werden, weil die Informationsanforderungen zur Anpassung neuer Behandlungsleitlinien reduziert werden, wenn auf bekannte Bausteine – wie zum Beispiel bekannte diagnostische Tests oder Operationsverfahren – zurückgegriffen werden kann. Auf diese Weise lässt sich nicht nur die Diffusion neuen Wissens beschleunigen, sondern auch die Entwicklung neuer medizinischer Behandlungsleitlinien vereinfachen.

Neben den Regeln, die auf eine Berücksichtigung der Umsetzungsrestriktionen, Qualitätsanforderungen, Produktionskosten medizinischer Leistungen und die Sicherheitsanforderungen abzielen, ist die Zusammensetzung und das Management interdisziplinärer Projektteams für den Erfolg medizinischer Behandlungsleitlinien wichtig. Hierzu sind verschiedene Ansätze denkbar. Eine Möglichkeit, den Informationstransfer zwischen verschiedenen Bereichen des Gesundheitswesens zu organisieren, besteht darin, Verbindungsspezialisten – mit Facharztbildung – als Informationsintermediäre zu beschäftigen. Eine andere Möglichkeit besteht in der Job-Rotation. Wenn ein Arzt vorübergehend auch einmal in der Forschungsabteilung eines Pharmaunternehmens oder in einer Krankenversicherung arbeiten müsste, würde dies automatisch den Austausch von Informationen erleichtern.

Eddy (1990c) beschreibt, wie Behandlungsleitlinien in der Vergangenheit entstanden sind. Im Allgemeinen wurde die vorherrschende medizinische Praxis zur Norm erhoben, so dass Verbesserungen allenfalls langfristig auf dem Wege vieler kleiner Veränderungsschritte eine Chance hatten. Medizinischer Fortschritt war im Wesentlichen das ungeplante Ergebnis eines evolutiven Prozesses. Die relevante Forschung und die Entwicklung von Leitlinien waren dezentral organisiert, die Festlegung von Standards erfolgte ex post – erst nachdem eine medizinische Technologie oder Behandlungsstrategie in der Praxis Verbreitung gefunden hatte und meist ohne systematische Berücksichtigung einer übergeordneten Zielsetzung. Großen Einfluss hatten traditionelle lokale Hierarchien, die sich in den geographisch abgegrenzten medizinischen „Schulen“ niederschlugen.²¹

Heute ist diese Herangehensweise nicht mehr akzeptabel. Die Rate, mit der neue medizinische Methoden eingeführt werden, hat erheblich zugenommen. Die Vielfalt und Komplexität medizinischer Entscheidungsprobleme hat

dadurch ebenfalls zugenommen. Dem einzelnen Arzt ist es längst nicht mehr möglich, einen vollständigen Überblick über die relevante Evidenz, alle möglichen Ergebnisse alternativer Behandlungsoptionen und deren spezifische Vor- und Nachteile für seinen Patienten zu haben. In dieser Situation muss die Entscheidungsfindung arbeitsteilig organisiert sein. Vor allem müssen die Leitlinien ex ante – also vor der breiten Adoption einer Technologie – festgelegt werden, um angesichts des gestiegenen Innovationstempos eine effiziente Diffusion und Anwendung neuen medizinischen Wissens und neuer technologischer Möglichkeiten zu gewährleisten. Wer aber soll die Behandlungsleitlinien in einem modernen Gesundheitswesen entwickeln? Wer hat einen Anreiz und ein besonderes Interesse, sich zu beteiligen: Ärzteorganisationen, Krankenhäuser, Pharma-Unternehmen, Krankenversicherungen oder Patientenverbände? Und wo sind die nötigen Informationen zu finden?

Eine Kernfrage ist, ob einheitliche Leitlinien auf der Ebene des gesamten Gesundheitssystems in einem Land nötig sind oder ob die Entwicklung lokalen Institutionen, zum Beispiel den einzelnen Fachabteilungen eines Krankenhauses, überlassen werden kann. Heimpel (2003) favorisiert als ehemals leitender Klinikarzt die Dezentralisierung, weil durch sie die besonderen Probleme des Klinikalltags besser berücksichtigt und eine höhere Compliance der Ärzte erreicht werden könne. Sackett et al. (2000), S. 181, dagegen sehen in der parallelen lokalen Entwicklung von Leitlinien nicht nur eine Ressourcenverschwendung, sondern sogar ein vergebliches Unterfangen, weil die zur Entwicklung hochwertiger Leitlinien erforderlichen Ressourcen, Erfahrungen und Kenntnisse fehlen würden.

Aus gesundheitsökonomischer Sicht ist eine Zentralisierung der Leitlinienentwicklung vorteilhaft, wenn es starke Skalenerträge dabei gibt und wenn die

²¹ Siehe Berger et al. (1997) für weitere Einzelheiten.

Ärzte vor Ort keine ausreichenden Anreize haben, selbst Leitlinien zu entwickeln. Andererseits sind es die Ärzte, die besser über die Präferenzen ihrer Patienten Bescheid wissen und jene praktischen Erfahrungen mit neuem medizinischen Wissen und neuen Technologien sammeln, die als wichtige Experimente zur Überprüfung der kontextbezogenen Effektivität und Verträglichkeit gelten können. Dezentrale Informationen aus der medizinischen Praxis sollten bei der Entwicklung neuer oder bei der Weiterentwicklung existierender Leitlinien nicht ignoriert werden. Die dezentrale Information und Erfahrung ist umso wichtiger, je stärker spezialisiert die betroffenen Ärzte sind, weil dann die Leitlinie auch sehr spezifisch formuliert sein muss. Zudem können komplementäre Investitionen zum Beispiel in neue diagnostische Großgeräte, über die niedergelassene Fachärzte oder leitende Klinikärzte entscheiden, für eine *erfolgreiche* Diffusion und Durchsetzung von Behandlungsleitlinien von großer Bedeutung sein. Die Beteiligung dieser Investoren an der Entscheidung über die Leitlinieninhalte kann daher wichtig sein. Die jeweiligen Verbände der betroffenen Fachärzte sollten ohnehin in jedem Fall an der Entwicklung von Leitlinien beteiligt sein.

Die Rolle von Interessengruppen. Es ist grundsätzlich sinnvoll, verschiedene Interessengruppen an eine zentrale Entwicklung medizinischer Behandlungsleitlinien zu beteiligen. Dabei ist jedoch das jeweilige Partikularinteresse zu beachten, um der Gefahr eines „regulatory capture“ durch einzelne Interessengruppen von vornherein zu begegnen. Aus Sicht der Ärzte kann die Verbreitung von Leitlinien dazu führen, dass ihr Berufsstand an faktischer Bedeutung und sozialem Prestige verliert und praktizierende Ärzte als Gruppe – möglicherweise zugunsten der Entwickler von Leitlinien – auf einen Teil ihres Einkommens verzichten müssen. Denn die Entwickler von Leitlinien werden mit deren zunehmender Verbreitung wichtiger und könnten durch ihren Wissensvorsprung sogar eine Art von Monopolstellung erwerben. Schon um die

Compliance mit neuen Leitlinien zu gewährleisten, müssen praktizierende Ärzte an ihrer Entwicklung beteiligt sein, worauf Heimpel (2003) und die von ihm zitierten empirischen Studien hinweisen. Eine ausführliche Analyse findet sich in Hasenbein et al. (2003).

Patientenverbände stellen insofern eine verzerrte Interessengruppe dar, als sie nur Menschen attrahieren, die bereits erkrankt sind, nicht jene, die in Zukunft erkranken *könnten*. Die bereits Erkrankten werden wie echte Lobbyisten nur ihr Partikularinteresse vertreten – eine möglichst unbegrenzte Behandlung ihrer eigenen Krankheit – und nicht das gesamtgesellschaftliche Optimum – eine rational begründete Verteilung der Ressourcen auf alle ex ante möglichen Krankheiten. Gleichwohl ist auch die Mitwirkung von Patientenverbänden an der Leitlinienentwicklung sinnvoll, wenn sich dadurch die Compliance der Patienten bei einer medizinischen Behandlung im Einklang mit der Leitlinie verbessert.

Die Krankenversicherungen ihrerseits werden ein großes Interesse an der Beteiligung in der Entwicklung von Leitlinien haben, da sie als wichtige Determinante der medizinischen Praxis ja auch einen wesentlichen Einfluss auf die Höhe der Kosten medizinischer Behandlungsepisoden haben. Aus volkswirtschaftlicher Sicht achten die Krankenversicherungen aber zu einseitig auf die Kosten und zu wenig auf die Nutzen einer neuen Behandlung, insbesondere wenn diese in Form von Externalitäten anderen Menschen als den eigenen Versicherten zufließen. Eine weitere Verzerrung in einem wettbewerblich organisierten Versicherungsmarkt ergibt sich aus dem stets möglichen Wechsel eines Versicherten zur Konkurrenz, was der abgebenden Krankenversicherung gewissermaßen den finanzierten Ertrag raubt, wenn eine Behandlungsepisode erfolgreich abgeschlossen ist. Krankenversicherungen werden daher grundsätzlich ein größeres Interesse an Leitlinien haben, die kurzfristige Heilerfolge zum Ziel haben, und wenig Interesse an Leitlinien, die auf eine

langfristige, wenn auch langsame Verbesserung des Gesundheitszustandes ihrer Versicherten haben.

Demnach hat keine einzelne Gruppe wirklich die notwendige Neutralität und Unabhängigkeit, die für eine Orientierung der Leitlinienentwicklung an einer sozialen Kosten-Effektivitätsanalyse erforderlich ist. Daher sollte es in jedem Land ein unabhängiges Institut geben, das bei der Entwicklung medizinischer Leitlinien eine Führungsrolle übernehmen muss. Ein Beispiel ist das britische *National Institute for Clinical Excellence* (NICE). Dabei kann jedoch ein typisches Bürokratieproblem auftreten: Das Management eines solchen Instituts könnte sich wie ein Budgetmaximierer verhalten, so dass wiederum das soziale Optimum verfehlt wird. Es werden dann möglicherweise zu viele – häufig wechselnde – oder zu komplizierte Behandlungsleitlinien entwickelt, so dass die Anpassungskosten in der medizinischen Praxis zu groß sind.

Organisationsprinzipien und Anreizstrukturen. Die vertikale Integration der Entwicklung medizinischer Behandlungsleitlinien mit der Produktion medizinischer Leistungen in Krankenhäusern und Ärztenetzwerken stellt einen denkbaren Ansatz dar, die Anreizprobleme zu mildern und den Wissenstransfer zu beschleunigen. Im Grunde ist dies nichts anderes als der Ansatz des Managed Care, auf den ich unten näher eingehe. Der Theorie von Teece (1988) folgend ist eine vertikale Integration schon deshalb geboten, weil Verträge zur Entwicklung neuer medizinischer Behandlungsleitlinien notwendigerweise *unvollständige* Verträge sein müssen. Denn die Ergebnisse können nicht im Voraus festgelegt werden, weil ja gerade *neu* entdecktes medizinisches Wissen in die Behandlungsleitlinien einfließen soll.

Während eine solche vertikale Integration zwar die Diffusionsbedingungen für neues medizinisches Wissen verbessern kann, löst sie noch nicht das Problem, Anreize für die Anbieter neuen medizinischen Wissens zu setzen. Sie birgt

sogar die Gefahr, alle Informationsquellen auszuschließen, die außerhalb der vertikal integrierten Organisation angesiedelt sind. Zudem gilt, dass eine zentralisierte Leitbildentwicklung wichtige Anpassungsschritte, die eine grundlegende Reorganisation erfordern, versäumen und die Bedürfnisse der Anwender in der medizinischen Praxis vernachlässigen könnte. Andererseits ist die Möglichkeit, Anreize durch die Zuweisung geistiger Eigentumsrechte zu setzen, im Gesundheitswesen weitgehend ausgeschlossen. Natürlich gibt es wichtige Ausnahmen, insbesondere die Entwicklung neuer Pharmaka und neuer Produkte der Medizintechnik, aber im Allgemeinen ist es weder möglich noch wünschenswert, neues medizinisches Wissen zu patentieren, zu Monopolpreisen anzubieten und auf diese Weise zu rationieren.

Obwohl die Praxis der Medizin überwiegend in den Bereich fällt, den Dasgupta und David (1994) als technologischen Bereich klassifiziert haben, reicht es heute nicht mehr aus, wenn allein die Universitätskliniken an Patienten forschen. Denn Universitätskliniken behandelten nicht nur die besonders schwer Erkrankten, sondern vor allem auch die von seltenen Krankheiten betroffenen Patienten. Die Forschung der Universitätskliniken ist daher nicht immer mit der Nachfrage nach neuem Wissen in anderen Bereichen der medizinischen Leistungserstellung kongruent. Vernachlässigt wird zum Beispiel die Forschung zur korrekten Dosierung von Pharmaka, zum Beispiel an Kindern oder alten Menschen. Der medizinische Innovationsprozess muss ein offener Prozess sein, zu dem eine Vielzahl von Spezialisten mit komplementären Fähigkeiten beitragen. Die Steuerungsaufgabe des Staates besteht darin, Bedingungen zu schaffen, unter denen die Spezialisten aus verschiedenen Teilgebieten der Medizin produktiv zusammenarbeiten können. Transaktionskosten, Informationsasymmetrien und die Unmöglichkeit vollständiger Verträge sind einige der Faktoren, die die Zusammenarbeit von angewandten Wissenschaftlern bei der Entwicklung komplexer Technologien und bei Erschließung neuer Wissens-

gebiete behindern können. Soweit geistige Eigentumsrechte an neuem medizinischem Wissen ausgeschlossen sind, muss der Innovationsprozess in der Medizin den Prinzipien eines offenen Innovationssystems folgen, wie es in der reinen Wissenschaft seit langem etabliert ist.

Insofern kommt das von Dasgupta und David (1994) beschriebene System der offenen Wissenschaft („Open Science“) den Anforderungen an die Informationsinfrastruktur in einem modernen Gesundheitswesen recht nahe. Im Unterschied zu Dasgupta und David (1994) kann jedoch die Produktion des relevanten medizinischen Wissens nicht ausschließlich an die Wissenschaft delegiert werden, sondern muss aus der medizinischen Praxis heraus getragen werden. Prioritätsregeln und Entlohnungsstrukturen müssen als Teil der Informationsinfrastruktur entsprechend angepasst werden.

Grundlegend ist dabei die Unterscheidung zwischen kodifiziertem und implizitem medizinischem Wissen („Codified“ versus „Tacit Knowledge“). „Codified Knowledge“ ist alles, was sich zu Papier bringen oder in elektronischen Medien speichern und übertragen lässt. Der Begriff des „Tacit Knowledge“ geht auf Polanyi (1966) zurück und bezeichnet die besonderen Fähigkeiten und Techniken, die durch wissenschaftliche Erfahrung erworben werden und zur Produktion und Nutzung kodifizierten Wissens erforderlich sind. Diese beiden Arten medizinischen Wissens sind komplementär, aber die Grenze zwischen ihnen ist endogen. Sie hängt – wie bei Dasgupta und David (1994) – einerseits von den relativen Zugangs- und Übertragungskosten für die Nutzer dieses Wissens ab. Andererseits hängen die relativen Anteile von kodifiziertem und implizitem Wissen auch von der Höhe der Entlohnung und der Kosten einer Kodifizierung ab. Je höher – *ceteris paribus* – die Entlohnung desto eher wird neues medizinisches Wissen *öffentlich* bereitgestellt und desto weniger werden die beteiligten Ärzte und Institutionen versuchen, ihr *privates*

Wissen zu maximieren und andere, zum Beispiel potentielle Wettbewerber, davon auszuschließen.

Weil die Anbieter eines öffentlichen Gutes sich nur einen kleinen Teil des Wertes für die Nutzer aneignen können, sind die Anreize zur Bereitstellung öffentlicher Güter im Allgemeinen viel zu gering. Zur Bewältigung dieses Problems im Gesundheitswesen sind grundsätzlich mehrere Lösungen denkbar: *Erstens* die Bereitstellung durch die Regierung, *zweitens* die Schaffung geistiger Eigentumsrechte und *drittens* die Subvention der Wissensproduktion durch autonome Forschungseinrichtungen. Da viele Regierungen die medizinische Forschung in Universitäten und Großforschungseinrichtungen finanzieren, sind sie direkt an der Bereitstellung neuen medizinischen Wissens beteiligt. Für die pharmazeutische Industrie ihrerseits sind geistige Eigentumsrechte an ihren Erfindungen der wichtigste Innovationsanreiz. Staatliche Subventionen werden in Deutschland unter anderem für die Entwicklung von Leitlinien gewährt.

Jede dieser Lösungsansätze des Anreizproblems hat seine eigenen Nachteile. Patente zum Beispiel schaffen neue Verzerrungen, weil der Preis eines patentgeschützten Gutes im Allgemeinen über den Grenzkosten liegt. Dies schränkt die Nutzung neuen Wissens ein. Ein besserer Anreizmechanismus ist daher im wissenschaftlichen System zu sehen. Der Entdecker einer wichtigen Erkenntnis oder Erfindung wird in diesem System für Originalität und zeitliche Priorität belohnt. Ähnlich wie in der Wissenschaft sollten auch praktizierende Mediziner und andere Mitarbeiter des Gesundheitswesens für Entdeckungen und die Bereitstellung von Informationen belohnt werden, die dann in die öffentliche Infrastruktur eingespeist werden. Hierzu müsste einerseits die Beförderungspraxis im Gesundheitswesen verändert werden, so dass sie der Bereitschaft zur Weitergabe von Erfahrungswissen aus der eigenen Praxis in geeigneter Weise entlohnt, oder es müssten, wie es auch in der Wissenschaft üblich ist, Preise vergeben werden, über die unabhängige Experten entscheiden.

Einen entsprechenden Vorschlag machen Wennberg et al. (2002) für das amerikanische Medicare-System. Dessen Finanzierungsstrukturen müssten so reformiert werden, dass sie Qualität und Effizienz belohnen. Dazu schlagen Wennberg et al. (2002) vor, Zentren für medizinische Exzellenz einzurichten. Um daran teilzunehmen, müssten Krankenhäuser, Ärztenetzwerke oder anderen Organisationen, die regionale Zusammenschlüsse repräsentieren, sich verpflichten, die jeweils beste medizinische Praxis umzusetzen. Diese Zentren sollen – Wennberg et al. (2002) folgend – für ihre Kapazität selbst verantwortlich gemacht werden, und zwar auf der Grundlage bevölkerungsbasierter Ansätze der Ressourcenplanung. Weitere Maßnahmen sollen darauf abzielen, die Sicherheit der Patienten zu verbessern, die Varianz in der Versorgung aufgrund individueller Präferenzen zu reduzieren, die Überversorgung aufgrund von Überkapazitäten zu reduzieren, die regelmäßige Überwachung der medizinischen Qualität einzuführen sowie ganz allgemein eine effiziente Ressourcenverwendung zu belohnen und die Implementierung dieser Maßnahmen zu beschleunigen.

5. Der ambulante Sektor: anbieterinduzierte Nachfrage und ärztliche Vergütungssysteme

Die wichtigste Trennungslinie zwischen den unterschiedlichen epidemiologischen Milieus, die für die Organisation und Effizienz medizinischer Entscheidungen von Bedeutung sind, verläuft zwischen dem ambulanten und dem stationären Sektor eines Gesundheitssystems. Der stationäre Sektor eines modernen Gesundheitswesens ist gekennzeichnet durch einen intensiven Einsatz medizinischer Technologie und die arbeitsteilige Zusammenarbeit zahlreicher Pflege- und Hilfskräfte bei der Durchführung diagnostischer und therapeutischer Maßnahmen sowie bei der Versorgung, Betreuung, Beobachtung und Verwaltung der Patienten. Der ambulante Sektor dagegen ist das

ursprüngliche und ureigene Feld des Arztes. Tönies (1993), S. 2, spricht vom „Arzt an der ersten ärztlichen Linie“ und zitiert die Berufsbeschreibung der internationalen Leeuwenhorst-Teaching-Gruppe, die 1977 vom Europarat (1978) als offizielle *Berufsdefinition*²² anerkannt wurde:

„Ein Allgemeinarzt ist ein Inhaber eines ärztlichen Diploms mit spezifischer Ausbildung, die ihn befähigt, Einzelpersonen, Familien, Gemeinschaften und einer Patientenschaft persönliche, primäre und fortlaufende *medizinische Versorgung* zu gewähren. Er betreut die Patienten unabhängig von deren Alter, Geschlecht und Krankheit. Die Eigenheit seiner Aufgabe besteht in der Synthese dieser Funktionen. Er betreut die Patienten in seiner Praxis oder in ihrem eigenen Heim, manchmal auch in einer Klinik oder in einem Krankenhaus. Sein Ziel ist die Frühdiagnose. Bei seinen *Entscheidungen* über Gesundheit und Erkrankung berücksichtigt er physische, psychologische und soziale Faktoren. Dies findet seinen Ausdruck in der Betreuung seiner Patienten. Er muss in der Lage sein, alle vom Patienten gelieferten *Informationselemente* entgegenzunehmen und zu deuten. Er gibt eine erste Stellungnahme zu allen Problemen, die ihm als Arzt vorgelegt werden. Er gewährleistet die fortlaufende Betreuung seiner Patienten. Diese ständigen Kontakte bieten ihm die Möglichkeit, in einer auf den einzelnen Patienten zugeschnittenen Zeitfolge Informationen zu sammeln und ein Vertrauensverhältnis zu entwickeln, das er beruflich nutzen kann. Er übt seine Tätigkeit in Zusammenarbeit mit ärztlichen und nicht-ärztlichen Kollegen aus. Er vermag seine Aufgabe als Mitglied eines Gesundheitsteams wahrzunehmen, ohne deswegen seine persönliche Verantwortung für die Behandlung des ihm anvertrauten Patienten abzulegen. Er muss wissen, wie und wann er durch pflegerische Betreuung, Vorsorge und erzieherische Maß-

²² Vgl. hierzu Meyer (2002), der eine Reihe von Definitionen aus unterschiedlichen Ländern gegenüberstellt und auf ihre gesundheitspolitische Bedeutung hinweist.

nahmen eingreift, um die Gesundheit seiner Patienten und ihrer Familien zu bewahren. Er bekennt sich weiterhin zu der Verpflichtung, fachkundigen Rat in, seine Lebensgemeinschaft betreffenden, Gesundheitsfragen zu geben.“

5.1 Informationsökonomische Grundlagen

So vielfältig, wie die Aufgaben des Arztes sein können, so anspruchsvoll sind auch die Anforderungen an sein Kommunikations- und Informationsverhalten. Tönies (1993), S. 19 ff., beschreibt den Weg vom Symptom zur gesicherten Diagnose, den im Prinzip jeder Arzt täglich neu beschreiten muss. Im Gespräch mit seinen Patienten muss der Arzt die vorliegenden Gesundheitsstörungen mit Bezeichnungen benennen, die eine unmittelbare Verständigung über die Erkrankung und deren Symptome ermöglichen. Voraussetzung für eine *Diagnose*, die vorab festgelegten erkenntnistheoretischen Bedingungen genügt, ist die professionelle Bezeichnung einzelner Gesundheitsstörungen im Sinne einer Klassifizierung. Die Aggregation der Information zur Diagnose ist nicht nur Voraussetzung für die Zuteilung von Medikamenten, Heilmitteln, Kuren und anderen therapeutischen Ressourcen und für die Annahme heilungsfördernder Verhaltensweisen durch den Patienten, sondern auch für eine *Prognose* des Krankheitsverlaufs durch den Arzt. Die Prognose hängt dabei in der Regel von der Einhaltung einer oder mehrerer alternativer *Handlungsanleitungen* ab, durch die eine gegebene Diagnose das weitere Vorgehen in der Therapie, der diagnostischen Absicherung und der Vorsorge festlegt.

Die Bedeutung eines allgemein akzeptierten Klassifizierungssystems für die medizinische Diagnose kann kaum überschätzt werden. Nicht nur Fehler in seiner Anwendung, sondern auch Fehlerquellen innerhalb des Systems können weitreichende, wenn auch schwer zu messende Folgen haben. Das einführende Vorwort zur „Internationalen statistischen Klassifikation der Krankheiten und

verwandter Gesundheitsprobleme“²³ beginnt mit der Feststellung: „Eine Klassifikation von Krankheiten kann als eine Systematik von Krankheitsgruppen definiert werden, der Krankheitsbilder nach feststehenden Kriterien zugeordnet werden. Es gibt unterschiedliche Klassifikationsachsen, wobei sich die jeweils gewählte Achse nach dem Verwendungszweck der zu erstellenden Statistik richtet. Eine statistische Klassifikation der Krankheiten muss den gesamten Bereich der möglichen Krankheitsbilder mit einer überschaubaren Anzahl von Kategorien abdecken.“

In welchem Maße sich die Konstruktion der Kategorien an den tatsächlichen medizinischen Handlungsmöglichkeiten oder an sozialen Konventionen orientiert, ist einerseits für die Effizienz und andererseits für die Akzeptanz eines Klassifizierungssystems von zentraler Bedeutung. Rosenberg (2002) argumentiert, dass die soziale Konstruktion diagnostischer Kategorien zu einer Reihe von Dilemmata der medizinischen Praxis geführt hat. Er nennt insbesondere die Gefahr, dass starre Krankheitskategorien abweichendes Verhalten definieren und so zur Durchsetzung von Normen und politischer Macht missbraucht werden, dass der Arzt den besonderen Umständen des einzelnen Patienten weniger gerecht wird, dass „Protokrankheiten“ entstehen, die die weitere Entwicklung der Medizin und die Lebensqualität der Betroffenen negativ beeinflussen können, und dass der Bürokratisierung und Anonymisierung der medizinischen Praxis Vorschub geleistet wird.

Obwohl die Klassifizierung von Wissen schon immer eine wesentliche Grundlage wissenschaftlichen Arbeitens war, gibt es auch heute noch keine *allgemein* gültige Klassifizierungsmethode. Booch (1994) beschreibt die historischen

²³ Dies ist amtliche deutschsprachige Ausgabe der von der Weltgesundheitsorganisation 1992 herausgegebenen zehnten Revision der *International statistical Classification of Diseases and Related Health Problems* (ICD-10). Die deutsche Ausgabe wurde im Auftrag des Bundesministeriums für Gesundheit vom *Deutschen Institut für medizinische Dokumentation und Information* (1994) herausgebracht.

Ursprünge der Klassifizierung von Wissen, die eigentlich schon immer auf relativ einfachen Regeln beruhte. Der „klassische“ Ansatz geht auf Plato zurück und legt bestimmte beobachtbare Eigenschaften von Objekten als Kriterien für ihre Klassifizierung fest. Um eine schnittmengenfreie Zuordnung zu erreichen, sollte man dabei nach Eigenschaften suchen, die möglichst wenig miteinander interagieren, wie zum Beispiel die Größe, die Farbe, die Form und das Material eines Objektes. Bei der Klassifizierung von Lebewesen kann auch die Art ihrer Bewegung oder ihre Entwicklung als Unterscheidungsmerkmal herangezogen werden. In der Praxis lassen sich allerdings nicht immer Eigenschaften finden, die zu einer schnittmengenfreien Zuordnung im Sinne des Verwendungszweckes der Klassifizierung führen. Auch wenn bestimmte Eigenschaftsmerkmale hoch mit den Kategorien einer Klassifizierung korreliert sind, kann es zu viele Ausnahmen geben; zum Beispiel können nicht alle Vögel fliegen.

Um diese Schwierigkeiten zu vermeiden, sind in der Literatur über lernende Maschinen (oder: maschinelles Lernen) zwei neue Ansätze der Klassifizierung entwickelt worden. Beim „Conceptual Clustering“ (Michalski und Stepp 1983) wird die Zuordnung der Objekte aus der konzeptionellen Beschreibung einzelner Klassen abgeleitet. Weil es sich dabei um eine Zuordnung von Objekten aufgrund von Wahrscheinlichkeiten handelt, entspricht das „Conceptual Clustering“ dem Wesen der medizinischen Diagnose von Krankheiten. In einer entsprechenden Anwendung nutzen Cheng und Fu (1985) die Ergebnisse eines „Conceptual Clustering“, um Krankheiten im Sinne der traditionellen chinesischen Medizin zu diagnostizieren.

Manchmal ist eine konsistente Beschreibung konzeptioneller Cluster nicht möglich. Dann bietet sich als letzte Möglichkeit eine Klassifizierung nach Prototypen an, bei der jede Klasse von Objekten durch einen Prototyp dargestellt wird. Die Zuordnung erfolgt dann danach, welchem Prototyp ein zu klassifizierendes Objekt am ähnlichsten ist. Verschiedene Prototypen unter-

scheiden sich dabei vor allem dadurch, dass einzelne Eigenschaften in unterschiedlicher Weise zusammenwirken. Die Klassifizierung aufgrund der Interaktion von Eigenschaften ist wesentlich für die Theorie der Prototypen, wie sie unter anderem von Lakeoff und Johnson (1980) beschrieben wird. Wenn die Ursachen einer Krankheit unbekannt sind und eindeutige Unterscheidungsmerkmale nicht definiert werden können, ist die Klassifizierung nach Prototypen auch in der Medizin relevant.

Im Hinblick auf therapeutische Entscheidungen sollte zunächst alles, was gleich oder ähnlich behandelt wird, in einer Kategorie zusammengefasst werden. Zusätzlich können andere Kriterien zur Kategorienbildung sinnvoll sein – zum Beispiel dann, wenn eine kausale Therapie nicht möglich ist. In diesem Fall kann anstelle des diagnostischen Vollbilds einer Krankheit auch eine Klassifizierung anhand von Symptomen sinnvoll sein, zumal diese selbst behandlungsbedürftig sein können. Um die Fehlerwahrscheinlichkeit zu verringern, sollte ein diagnostisches Klassifikationsschema grundsätzlich auch Kategorien für nicht-behandelbare Gesundheitsstörungen bereithalten, damit der Arzt weniger leicht verführt wird, diese Möglichkeit im diagnostischen Prozess zu ignorieren.

Jede Klassifizierung einer Gesundheitsstörung hat natürlich auch finanzielle Folgen, die von der Kategorienbildung und der Fehleranfälligkeit des Systems beeinflusst werden; eine falsche oder zu unspezifische Behandlung von Gesundheitsstörungen kann unnötige Kosten verursachen. Dies gilt unabhängig von der Art des Vergütungssystems. In einem prospektiven Vergütungssystem nimmt die Klassifizierung von Diagnosen und Prozeduren ohnehin eine zentrale Stellung ein. Aufgrund der anderen Zielsetzung kann jedoch eine buchhalterische Klassifizierung zum Zwecke der Ermittlung von Fallpauschalen als Teil eines Finanzierungssystems, das optimale Anreize zur kosteneffektiven medizinischen Behandlung setzen soll, nicht Eins-zu-Eins aus einem Klassifizierungssystem abgeleitet werden, das für die optimale Behandlung von

Gesundheitsstörungen und gegebenenfalls für die Prognose ihres Verlaufes entwickelt wurde.

Das deutsche System zur Klassifizierung von stationär zu behandelnden Gesundheitsstörungen ist eine Adaption des australischen Systems der Zuordnung von Krankheiten zu „Diagnostically Related Groups“ (AR-DRGs). Australien hatte Anfang der 80er Jahre als erstes Land der Welt die Finanzierung von Krankenhausleistungen auf ein System mit Fallpauschalen umgestellt. Eine einführende Beschreibung der grundlegenden Anforderungen an Diagnose- und Prozedurenklassifikationen in einem DRG-System findet sich bei Stausberg (2002). Die Gestaltung der Klassen muss sich an einzelnen Anforderungen in den Bereichen Gruppierung, Kalkulation, Prüfung von Fehlbelegung und Qualitätssicherung orientieren. Diese Anforderungen können aus den AR-DRGs sowie aus den vorgeschriebenen Verfahren zur Fehlbelegungsprüfung und Qualitätssicherung abgeleitet werden. Die in Deutschland gesetzlich vorgeschriebenen Diagnoseklassen für den stationären Bereich, spezifiziert in ICD-10-SGB-V 2.0 und OPS-301 2.0/2.1, sind allerdings nach Stausbergs Einschätzung viel differenzierter, als es für eine Definition von DRGs eigentlich notwendig ist.

Jedes für einen gegebenen Zweck optimierte Klassifizierungssystem muss in der Anwendung letztlich unvollkommen bleiben. Tönies (1993), S. 21, argumentiert zu Recht, dass die Leitidee, jeder Allgemeinarzt möge als *Produkt seiner Diagnostik* „wiederholbar eindeutige Beurteilungen an eindeutige Orte des Klassifizierungssystems setzen“, praktisch undurchführbar ist. Als wichtigste Klassifizierungsprobleme nennt er *erstens* das Problem der Abgrenzung sicherer und unsicherer diagnostischer Merkmale (insbesondere die Frage des Symptomausschlusses und –einschlusses bei der Bestimmung eines diagnostischen Vollbilds), *zweitens* das Problem der Klassifizierung vor Ablauf eines Krankheitsprozesses oder ohne volle Kenntnis der Erkrankung

(insbesondere die Abgrenzung eines ergebnisoffenen Prozesses gegen statische Zustände oder das Endstadium einer Erkrankung) und *drittens* das Problem des richtigen Kriteriums zur Kategorienbildung (zum Beispiel anatomisch, ätiologisch – nach den Ursachen gegliedert – oder handlungsbezogen).

Lange Zeit galt es in der Medizin, wie Tönies (1993), S. 23, ausführt, als Ideal eines Klassifizierungssystems, dass die *Symptome*, die ja gedanklich noch vor den diagnostischen Hypothesen stehen, möglichst ignoriert und in den Kodierungsvorschriften vollständige Diagnosen bevorzugt werden sollten, selbst wenn diese auf unsicheren Schlussfolgerungen beruhten. Für Symptome wird daher in den meisten Klassifizierungssystemen, die historisch aus Todesursachenstatistiken entstanden sind, weniger Platz eingeräumt als für die Bezeichnungen „statischer“ Vollbilder von Erkrankungen. Erst in neuerer Zeit hat hier ein Umdenken eingesetzt; auch das historisch gewachsene Klassifizierungssystem der WHO (ICD – International statistical Classification of Diseases) ermöglicht heute die Kodierung von Symptomen in einem Ausmaß, wie sie oft für die Beschreibung von Beratungsergebnissen gebraucht wird.

Das Ende einer primärärztlichen Beratung stellt ja in vielen Fällen nur einen Ausschnitt einer Krankheitsentwicklung im Frühstadium und den Beginn eines medizinischen Erkenntnisprozesses dar, so dass noch nicht mit allzu viel gesicherter Information zu rechnen ist: Die Beschreibung der Gesundheitsstörung durch Symptome ist dann unvermeidlich und hat auch insofern ihre Berechtigung, als oft erste therapeutische Handlungen einzuleiten sind, bevor eine gesicherte Diagnose vorliegt. In diesem Zusammenhang weist Tönies (1993), S. 24, allerdings zu Recht auf die Gefahr hin, dass die Zulassung von Symptomkategorien „dem Klassifizieren auf niedrigster Ebene, ohne Bemühen um Integration zur Diagnose, Vorschub leisten könnte“.

Rosenberg (2002) beschreibt die historischen Wurzeln der Methodik der Differentialdiagnose im 19. und frühen 20. Jahrhundert und analysiert die Veränderungen, die zur Herausbildung diagnostischer Krankheitskategorien geführt haben. Er weist nach, dass die Vorstellung klar abgegrenzter Krankheiten, die unabhängig von ihrer Ausprägung im einzelnen Patienten existieren, ein *soziales* Konstrukt ist, das in erster Linie durch technologische Veränderungen entstanden ist. Das Konstrukt erleichtert die Ausnutzung von Skalenerträgen in der Anwendung medizinischen Wissens. Als Beispiel hierfür nennt Rosenberg (2002) die Keimtheorie ansteckender Krankheiten, die einer personenunabhängigen Kategorisierung von Krankheiten nach der Art ihrer Ursachen, anstelle einer individuellen Beschreibung von Symptomen, den Boden bereitete. Ebenfalls im 19. Jahrhundert begann die Entwicklung immer präziserer Instrumente zur Messung von Symptomen, zur Überprüfung diagnostischer Hypothesen und zur Beobachtung des Krankheitsverlaufs. Erst dadurch wurde es möglich, verschiedene Gesundheitszustände als unterschiedliche Phasen ein und derselben Krankheit zu erkennen.

Weitere technologische Veränderungen, die eine Nachfrage nach klar definierten diagnostischen Kategorien entstehen ließen, lagen in der Entwicklung der Epidemiologie, der statistischen Erfassung der Verbreitung einzelner Krankheiten, und des Krankenhauses von einer Pflegeeinrichtung zu einem Zentrum der Hochleistungsmedizin. Durch diese Transformation wuchsen den Krankenhäusern neue Aufgaben zu, die ihrerseits die Herausbildung und Standardisierung diagnostischer Kategorien begünstigten. Medizinisches Wissen wird auf der Grundlage dieser Kategorien formuliert und kommuniziert.

Die moderne Informationstechnologie bietet die Chance, die Weiterentwicklung von Klassifizierungssystemen dynamisch an neue diagnostische und therapeutische Möglichkeiten der Medizin anzupassen. Methodische Ansätze dazu finden sich in der bereits genannten Literatur über Maschinen-Lernen („machine

learning“).²⁴ Die Weiterentwicklung eines Klassifizierungssystems sollte dieser Literatur folgend einen Algorithmus verwenden, der anhand von Beispielen darauf eingestellt wird, empirische Beziehungen zwischen Attributen einzelner Beobachtungen und vorab festgelegten Klassifizierungsattributen zu erkennen und für die Entwicklung effizienter Klassifizierungsregeln zu nutzen. Wenn die Kategorien der Klassifizierung bereits feststehen, wird die Entwicklung der Klassifizierungsregeln als „Supervised Learning“ oder „Concept Acquisition“ bezeichnet.²⁵ Wenn die Kategorien dagegen erst während des Klassifizierungsprozesses gebildet und kontinuierlich modifiziert werden, spricht man von „Unsupervised Learning“ oder von „Concept Formation“.

Das „Supervised Learning“ steht in enger Beziehung zur Technik des „Data Mining“, bei der einzelne Beobachtungen als Trainingsstichproben analysiert werden, um ein Modell des gegebenen Datensatzes zu erzeugen. Jede Beobachtung gehört annahmegemäß zu einer vorab definierten Klasse, die durch ein klassifizierendes Attribut bestimmt wird. Das durch Training gefundene Klassifizierungsmodell kann dann genutzt werden, um weitere Stichproben zu analysieren und die Beobachtungen effizient zu klassifizieren. Um eine Trainingsstichprobe zu klassifizieren, muss das Klassifizierungsprogramm die Werte der einzelnen Attribute in der Stichprobe gegen den jeweiligen Entscheidungsbaum testen. Auf diese Weise wird ein Pfad festgelegt, der die Vorhersagen des Modells abbildet. Sind die Attribute ausreichend, sollte es stets möglich sein, einen Entscheidungsbaum zu konstruieren, der jede Beobachtung in der Trainingsstichprobe korrekt klassifiziert. Gibt es sogar mehrere Entscheidungsbäume, die diese Bedingung erfüllen, ist im Allgemeinen der *einfachste* korrekte Entscheidungsbaum zu bevorzugen. Denn

²⁴ Siehe zum Beispiel in Alexopoulos et al. (1999), Bourlas et al. (1999), Magoulas und Pretza (1999), Moustakis und Charissis (1999), Kamber et al. (1996) oder Quinlan (1986).

²⁵ Siehe zum Beispiel Michalski (1983).

bei diesem ist am ehesten anzunehmen, dass er eine kausale Struktur des Klassifizierungsproblems erfasst.

Ein relativ einfaches und effizientes Verfahren des „Supervised Learning“ ist als „Induktion von Entscheidungsbäumen“ bekannt, obwohl es zutreffender als „Induktion von Klassifizierungsbäumen“ bezeichnet würde. Andere Möglichkeiten sind durch neuronale Netzwerke, genetische Algorithmen oder ähnliche mathematische Verfahren gegeben (Quinlan 1993, S. 12 ff.). Bei der Induktion eines Entscheidungsbaumes wird dieser Baum dadurch entwickelt, dass die gegebenen Beobachtungen rekursiv aufgeteilt werden, bis jede Teilmenge eindeutig ist. Danach werden Zweige des Baumes mit zu großen statistischen Abweichungen entfernt. Schließlich wird der Teil des Entscheidungsbaumes als Klassifizierungssystem verwendet, der sich durch die geringste geschätzte Fehlerrate auszeichnet. Für die praktische informationstechnische Umsetzung dieser Methode kommt es vor allem darauf an, das Verfahren skalierbar zu machen und eine große Zahl von Attributen effizient zu verarbeiten. Ein Beispiel hierfür bieten Mehta et al. (1996).

Beim „Unsupervised Learning“ soll das Klassifizierungssystem gleichzeitig eine optimale Abgrenzung von Kategorien und effiziente Zuordnungsregeln für die zu klassifizierenden Objekte finden. Eine umfassende Analyse dieses Problems bieten Fisher et al. (1991). Den frühesten Lösungsansatz stellt die statistische Cluster-Analyse (siehe Everitt 1993) dar, die durch einen paarweisen Vergleich aller Beobachtungen Kategorien nach dem Kriterium der größten Ähnlichkeit bildet und einem sogenannten Dendogramm abbildet. Damit ist zwar die Bildung von Kategorien erfolgt, aber das eigentliche Klassifizierungsproblem bleibt ungelöst. Denn die Zuordnung neuer Beobachtungen zu einem bereits vollständigen System von Clustern muss auf andere Methoden zurückgreifen.

Erst das „Conceptual Clustering“ beginnt, die Bildung von Teilmengen mit Methoden des „Supervised Learning“ zu integrieren. Im Einklang mit der Bayesianischen Theorie medizinischer Entscheidungen stehen insbesondere solche Ansätze, bei denen die Kategorienbildung und die Klassifizierung mithilfe Bayesianischer Methoden erfolgt; ein Beispiel hierfür ist AUTOCLASS von Cheeseman et al. (1988). Dieses System sucht nach den Klassen mit der größten Wahrscheinlichkeit für einen gegebenen Datensatz und ordnet diesen die einzelnen Beobachtungen mit unterschiedlichen Wahrscheinlichkeiten zu, die im Gegensatz zu früheren Verfahren des maschinellen Lernens stets unter Eins liegen.

Der anspruchsvollste Lösungsansatz wird „Concept Formation“ genannt und stellt eine inkrementelle Form des „Unsupervised Learning“ dar. Ein wichtiges Beispiel, das Bayesianische Entscheidungsregeln implementiert, ist das COBWEB-System von Fisher (1987). Es steht im Einklang mit Langleys (1987) Forderung, dass sinnvolle Modelle nicht einfach unverbundene Kategorien bilden sollten, sondern eine Art konzeptioneller Hierarchie, deren einzelne Konzepte hinreichend flexibel sind, um im Rahmen eines inkrementellen Lernprozesses angepasst zu werden. Dies ist offensichtlich für medizinische Klassifizierungssysteme relevant, weil das Auftreten neuer Krankheitsvarianten oder die Einführung neuer Behandlungsmöglichkeiten immer wieder einen Teil der Kategorienabgrenzungen in Frage stellt, die zuvor überzeugende Klassifizierungsergebnisse ermöglicht haben. Wenn Anpassungen der Kategorienabgrenzung erforderlich sind, kommt es nicht immer ausschließlich auf eine möglichst hohe Genauigkeit an, sondern – wie Moustakis und Charissis (1999) erklären – oft auch auf die Verständlichkeit für den Nutzer, zum Beispiel einen

praktischen Arzt, der nicht allzu viel Zeit hat, die Abgrenzungskriterien zwischen diagnostischen Kategorien im Detail zu studieren.²⁶

In der medizinischen Praxis wird die Effizienz eines diagnostischen Klassifizierungssystems grundsätzlich durch den Trade-off zwischen dem Aufwand zur Erstellung der Diagnose und den Kosten mangelnder Genauigkeit bestimmt. Die optimale Genauigkeit eines Klassifizierungssystems hängt dabei nicht nur von der Leistungsfähigkeit der verfügbaren diagnostischen Technologie ab, sondern auch von den technologischen Behandlungsmöglichkeiten für einzelne Krankheiten. Die Bayesianische Testtheorie kann zunächst nur erklären, wie ein Testverfahren auszuwählen ist, wenn das Klassifizierungsschema bereits feststeht. Bei der Weiterentwicklung eines Klassifizierungssystems ist aber zu fragen, nach welchen Regeln es an neue Informationen über Symptome oder Behandlungsmöglichkeiten anzupassen ist.

Insbesondere ist zu fragen, ob die Anpassung iterativ erfolgen sollte, jeweils *nachdem* eine therapeutische Innovation auf den Markt gekommen ist, oder ob eine Art dynamischer Optimierung die Anpassungsrate bestimmen sollte? Im ersten Fall ist die Anpassung adaptiv. Im zweiten Fall eröffnet die zunehmende Verfeinerung der diagnostischen Klassifizierung neue Chancen in der medizinischen Forschung, weil unterschiedliche Ausprägungen von ähnlichen Krankheiten deutlicher erkennbar werden. Ein Beispiel hierfür ist die zur Zeit diskutierte Hypothese, dass einige Arten von Krebs bakterielle Ursachen haben könnten. Dies würde bedeuten, dass sie ganz anders behandelt werden müssten als Krebsarten, die auf eine intrinsische Veränderung der Zellteilungsfunktion zurückzuführen sind.

²⁶ Eine neuere Weiterentwicklung dieser Ideen und Methoden findet sich in Iba und Langley (2001). Eine innovative Anwendung des „Unsupervised Learning“ im Bereich der Gentechnik entwickeln Monti et al. (2003).

Dieses Beispiel macht nebenbei deutlich, dass eine feinere diagnostische Klassifizierung den geringsten Schaden durch Fehlbehandlung erwarten lässt. Andererseits ist die Wahrscheinlichkeit von Fehldiagnosen größer, je mehr diagnostische Kategorien ein Klassifizierungssystem zulässt. Die Bedeutung des Trade-offs zwischen dem Aufwand der Diagnose und den Kosten möglicher Fehlbehandlung wird besonders deutlich, wenn eine neue Krankheit auftritt. Die ersten Patienten der AIDS-Seuche wurden über Jahre falsch diagnostiziert und falsch behandelt, weil die AIDS-Krankheit zunächst nicht bekannt war. Je größer ein Klassifizierungssystem desto geringer sind die Chancen, dass eine neue Krankheit als solche erkannt wird und eine Anpassung des Klassifizierungssystems veranlasst. Das Erkennen von Krankheitsfällen, die in einem gegebenen Klassifizierungssystem nicht oder nur schwer zuzuordnen sind, kann nur in der medizinischen Praxis erfolgen. Es ist daher zu fragen, welche Anreize das medizinische System dem einzelnen Arzt setzt, einen Verdacht auf eine neue Krankheit zu melden. Wie lässt sich vermeiden, dass Ärzte kritiklos diagnostizieren und rasch zur Therapie schreiten, um nicht als entscheidungsunfähig zu gelten?

Es gibt zweifellos bedeutende Vorteile einer internationalen Koordination bei der Entwicklung von Klassifizierungssystemen unter dem Dach der Weltgesundheitsorganisation. Dies darf jedoch nicht dazu führen, dass die dynamische Weiterentwicklung und Anpassung der internationalen Klassifizierung von Krankheiten an neue diagnostische und therapeutische Möglichkeiten vernachlässigt wird. Es ist klar, dass jedes gegebene Klassifizierungssystem revidiert werden muss, wenn neue Krankheiten auftreten. Aber auch neue therapeutische Möglichkeiten können Grund für eine Revision sein, wenn sie eine differenzierte Behandlung von ähnlichen Krankheitsfällen ermöglichen, die bislang gleich behandelt und deshalb in einer Krankheitskategorie zusammengefasst wurden. Schließlich kann selbst bei einer

konstanten Menge therapeutischer Technologien eine Verfeinerung des Klassifizierungssystems angezeigt sein – zum Beispiel dann, wenn sich die Informationsverarbeitungskapazität der medizinischen Praxis erhöht. Denn bei einer signifikanten Verringerung des Diagnoseaufwands, verschiebt sich der Trade-off zwischen Diagnoseaufwand und den Kosten der Fehlbehandlung, so dass es aus sozialer Sicht lohnend wird, Krankheiten genauer zu diagnostizieren. Dabei muss eine Verbesserung der Informationsverarbeitungskapazität der medizinischen Praxis nicht immer auf die Einführung neuer diagnostischer Technologien zurückzuführen sein. Sie kann auch in organisatorischen Veränderungen begründet sein – zum Beispiel in einer beschleunigten Diffusion medizinischen Wissens, die Praxisvariationen abbauen hilft.

5.2 Angebotsinduzierte Nachfrage und Nachfrageüberhang

Aus gesundheitsökonomischer Sicht liegt eine grundlegende Ursache für Effizienzprobleme des ambulanten Sektors in der *asymmetrischen Verteilung von Information* über den Wert medizinischer Leistungen. Eine asymmetrische Informationsverteilung kann zum Beispiel in einer unterschiedlichen Kontaktfrequenz verschiedener Patientengruppen zum Ausdruck kommen. So ist bekannt, dass Männer sehr viel seltener Hilfe beim Arzt suchen als Frauen. Tönies (1993), S. 94, nennt hierfür unter anderem als Ursache: „Es gehört zum (irrigen) Selbstbild der Männer, keine Hilfe zu brauchen. Hilfe zu suchen, muss erlernt werden. Dazu müsste es als rollenkonform erlebt werden. Die Frauen sind mit den Kindern öfter beim Arzt, sie fassen dadurch Vertrauen zur medizinischen Betreuung.“

Andererseits gilt auch, dass die Berufstätigkeit vielen Männer wenig Freizeit lässt, so dass die Opportunitätskosten eines Arztbesuches vielfach höher sind als bei Frauen. Unabhängig vom Geschlecht wird der Zeitaufwand für viele Patien-

ten eine ökonomisch höher zu bewertende Form der Selbstbeteiligung an den Gesamtkosten des Arztbesuches sein als die vom Gesetz vorgeschriebene finanzielle Selbstbeteiligung. Eine unterschiedliche Kontaktfrequenz muss daher nicht ausschließlich auf Informationsdefizite zurückzuführen sein. Aufgrund geringer Zugangsbarrieren und einer finanziellen Selbstbeteiligung, die bei allen Versicherten in der Regel weit unter den sozialen Opportunitätskosten eines Arztbesuches liegt, ist natürlich auch eine zu hohe Kontaktfrequenz der Frauen denkbar. Grundsätzlich ist eine unterschiedliche Kontaktfrequenz verschiedener Patientengruppen mit Tönies (1993), S. 94, aus zwei Gründen als problematisch anzusehen: *Einerseits* können unversorgte Symptome mit Krankheitswert bei verspäteter Behandlung größeres Leiden und langfristig höhere Kosten verursachen (Ausbleiben der tertiären Prävention). *Andererseits* droht eine unmittelbare Verschwendung von Ressourcen, wenn zu viele Symptome ohne Krankheitswert zum Arzt gebracht werden.

Aber auch *nachdem* der Patient den Weg zum Arzt gefunden hat, kann asymmetrische Information Effizienzverluste verursachen. Denn der *Informationsvorsprung des Arztes* beeinträchtigt die Konsumentensouveränität, wenn die Wahl einer therapeutischen Strategie von den Ergebnissen des vom Arzt gesteuerten Diagnoseprozesses abhängt. Der Diagnoseprozess reduziert nicht nur die anfängliche Unsicherheit des Arztes über Krankheitsursache und Heilungsmöglichkeiten, sondern verschafft dem Arzt gleichzeitig einen Informationsvorsprung gegenüber dem Patienten. Nutzt er diesen zu seinem eigenen finanziellen Vorteil aus, entsteht eine *anbieterinduzierte Nachfrage*, eine Ausweitung medizinischer Leistungen über den von einem vollständig informierten Patienten gewünschten Umfang hinaus. Diese Möglichkeit besteht unabhängig davon, ob der Diagnoseprozess zu einer gesicherten Diagnose führt oder ob auch die Information des Arztes im Hinblick auf die therapeutischen Handlungsmöglichkeiten unvollständig bleibt. Somit sind medizinische

Praxisvariationen aufgrund unvollständiger Information des behandelnden Arztes ohne Weiteres mit einer angebotsinduzierten Nachfrage vereinbar: auch ein unvollständig informierter Arzt, der einen Teil seiner a priori-Einschätzungen aus lokalen Quellen bezieht, kann gegenüber seinen Patienten einen Informationsvorsprung haben.

Die *Hypothese angebotsinduzierter Nachfrage* ist für die Gesundheitspolitik von zentraler Bedeutung, geht es doch um die grundsätzliche Frage, ob eine effiziente Ressourcenallokation im Gesundheitswesen eher durch Anreize auf der Nachfrageseite oder durch die Regulierung des Angebotsverhaltens der Ärzte erreicht werden kann. Mit McGuire (2000) können wir angebotsinduzierte Nachfrage nach medizinischen Leistungen als eine Nachfrageverschiebung definieren, die der Arzt gegen seine eigene Auffassung der optimalen Behandlungsintensität veranlasst. Diese Hypothese ist vor allem für Ärzte relevant, die eine Einzelleistungsvergütung erhalten. Die Evidenz deutet darauf hin, dass Ärzte in bestimmten Umständen tatsächlich dazu bereit sind, Kompromisse zu Lasten des Patienten zu machen, um ein höheres Einkommen zu erreichen.

Allerdings muss die extreme Version dieser Hypothese, die sogenannte *Zieleinkommenshypothese*, verworfen werden. Nach der Zieleinkommenshypothese ist der Grenznutzen eines Arztes aus zusätzlicher Freizeit und aus dem Berufsethos, unnötige Behandlungen zu vermeiden, gleich Null, sobald das Einkommen des Arztes unter ein bestimmtes Anspruchsniveau fällt, das als Zieleinkommen bezeichnet wird. Diese Verhaltensannahme garantiert im Rahmen eines entsprechenden Modells eine anbieterinduzierte Nachfrage bei zunehmender Ärztedichte ab einem bestimmten (bereits relativ hohen) Ausgangsniveau. Empirisch wurde ein positiver Zusammenhang zwischen der Ärztedichte und der Menge medizinischer Leistungen jedoch auch bei relativ niedriger Ärztedichte nachgewiesen.

Zudem konnte ein kausaler Einfluss der Ärzte auf die Menge medizinischer Leistungen, die ein durchschnittlicher Patient nachfragt, nicht eindeutig nachgewiesen werden. So konnten eine Reihe empirischer Untersuchungen zwar nachweisen, dass die Rate der Inanspruchnahme einzelner medizinischer Leistungen in einer Region zunahm, wenn die Ärztedichte zunahm, beziehungsweise dass die Rate der Nutzung einzelner medizinischer Leistungen höher war, wenn die Ärztedichte höher war. Allerdings ist auch ein umgekehrter Kausalzusammenhang denkbar: So könnte es sein, dass Regionen, in denen aufgrund der Altersstruktur oder aufgrund der Krankheitsanfälligkeit der Bevölkerung die Nachfrage nach ärztlichen Leistungen überdurchschnittlich hoch ist, eine größere Zahl von Ärzten attrahieren, wenn diese sich bei ihrer Niederlassungsentscheidung an der durchschnittlichen Auslastung der bestehenden Praxen orientieren.

In einer amerikanischen Untersuchung konnte zum Beispiel gezeigt werden, dass der Anteil von Geburten, die auf Anraten des jeweiligen Arztes mit Kaiserschnitt erfolgten, in jenen Bundesstaaten weit überdurchschnittlich zunahm, in denen der – zweifellos exogene, von ärztlichen Entscheidungen unabhängige – Geburtenrückgang der siebziger und achtziger Jahre besonders ausgeprägt war (Gruber und Owings 1996). Offenbar haben die Gynäkologen ihr Einkommen zu stabilisieren versucht, indem sie bei einer rückläufigen Menge die Erträge der durchschnittlichen Geburt erhöht haben. Die Gynäkologie stellt ein nahezu ideales natürliches Experiment zur Hypothese der angebotsinduzierten Nachfrage dar, weil es keinen plausiblen medizinischen Grund gibt, warum die relative Häufigkeit von Kaiserschnitten zwischen Bundesstaaten variieren sollte. Wie in Schaubild 15 kann daher die Nachfrage N_0 , die ohne die ärztliche Beeinflussung gewählt würde, als vollkommen preisunelastisch gelten. P_0 stellt hier den regulierten Produzentenpreis, die Vergütung der ärztlichen Leistungen M , dar. Wird das Ausgangsgleichgewicht

(M_0) durch den Markteintritt neuer Ärzte gestört, entfallen auf den einzelnen Arzt weniger Patienten. Alle Ärzte versuchen dann dem Modell zufolge, die Auslastung ihrer Praxis durch die Ausweitung der durchschnittlichen medizinischen Leistungen für den einzelnen Patienten sicherzustellen.

Auf aggregierter Ebene gibt es, wie Breyer und Zweifel (2003) berichten, sowohl für die USA als auch für Deutschland einen positiven statistischen Zusammenhang zwischen der regionalen Ärztedichte und den Pro-Kopf-Ausgaben für ärztliche Leistungen. Aber für die deutsche Situation während der siebziger und achtziger Jahre kann die Hypothese nicht widerlegt werden, dass ein Nachfrageüberhang abgebaut wurde. Ein Nachfrageüberhang liegt vor, wenn der regulierte Preis medizinischer Leistungen – P_{k0} in Schaubild 16 – zu gering ist, um die etablierten Ärzte zu einem markträumenden Angebot medizinischer Leistungen zu bewegen. Unabhängig von der Angebotskurve A_0 wird zum regulierten Preis P_{k0} die Menge N_1 nachgefragt. In einer solchen Situation sind unterschiedliche Anpassungsprozesse in der kurzen und in der langen Frist zu erwarten. Kurzfristig kann sich der Nachfrageüberhang auf dem „offiziellen Markt“ sogar noch verschärfen, wenn die etablierten Ärzte – wie beispielsweise im polnischen Gesundheitswesen der 90er Jahre – mit einem Teil ihrer Leistungen auf den Schwarzmarkt ausweichen, wo sich höhere Preise erzielen lassen. Längerfristig wird der Nachfrageüberhang durch den Markteintritt junger Ärzte abgebaut.

5.3 Ärztliche Vergütungssysteme

Vergütungssysteme sind eines der wichtigsten Instrumente zur Steuerung des Anbieterverhaltens im Gesundheitswesen, wobei die Vergütung ärztlicher Leistungen im Mittelpunkt der Analyse stehen muss. Ein optimales Vergütungssystem muss sich an einer klaren gesundheitsökonomischen Zielsetzung orientieren. Lautet das Ziel zum Beispiel Kosteneffektivität in der medizini-

schen Leistungserstellung (im Sinne mikroökonomischer Produktionseffizienz), dann ergeben sich daraus drei ökonomisch begründete Beurteilungskriterien für Vergütungssysteme – nämlich *erstens*: Ist die Produktion medizinischer Einzelleistungen effizient? Mit anderen Worten: Werden sie durch die kostenminimierende Kombination der Produktionsfaktoren hergestellt? *Zweitens*: Ist die Behandlung einzelner Fälle effizient? Mit anderen Worten: Werden die Einzelleistungen im Rahmen der medizinischen Substitutionsmöglichkeiten so kombiniert, dass der Gesamtpreis der Behandlung bei gegebenem Heilerfolg möglichst gering ist? Und *drittens*: Wird der optimale Heilerfolg erreicht? Mit anderen Worten: Steckt der Arzt in die Behandlung des einzelnen Patienten genügend Ressourcen, ohne sie zu verschwenden? Nach Breyer und Zweifel (2003) sind in erster Linie die folgenden vier Vergütungssysteme zu unterscheiden:

(i) *Kostenerstattung nach Faktoreinsatzmengen*. Die Erstattungsfunktion lautet $E = h^t(t_1 + t_2) + h^x(x_1 + x_2)$ und setzt sich additiv aus der Vergütung der eingesetzten ärztlichen Arbeitszeit, t_1 und t_2 , und des eingesetzten Verbrauchsmaterials, x_1 und x_2 , zusammen. Die Summe der Faktormengen wird dabei jeweils mit den Kostenerstattungssätzen, h^t und h^x , multipliziert. Der Arzt, der sein Einkommen maximieren will, hat dadurch einen Anreiz, Einzelleistungen ineffizient und insgesamt zu viele oder zu wenige Einzelleistungen zu produzieren, je nachdem ob die Kostenerstattung die Faktorpreise über- oder unterschreitet.

Schaubild 17 verdeutlicht dies, wobei ich annehme: Die neuen Faktorpreise sind im Punkt A aufkommensneutral. Die gestrichelte Linie stellt das Faktorpreisverhältnis $-w^t/w^x$ dar. Die durchgezogene Linie stellt das effektive Faktorpreisverhältnis $(-w^t + \Delta h^t)/(w^x + \Delta h^x)$ nach Berücksichtigung der Kostenerstattungs-

sätze h^t und h^x dar, wobei $\Delta h^t = |w^t - h^t|$ und $\Delta h^x = |w^x - h^x|$ mit $0 < h^t < w^t$ und $0 < w^x < h^x$.

(ii) *Einzelleistungsvergütung*. Die Erstattungsfunktion lautet $E = p_1 M_1 + p_2 M_2 = p_1 f_1(t_1, x_1) + p_2 f_2(t_2, x_2)$ und setzt sich additiv aus den mit ihren Preisen bewerteten Einzelleistungen, M_1 und M_2 , zusammen, die gemäß ihren Produktionsfunktionen, f_1 und f_2 , erstellt werden. Zwar wird der Arzt nun seine Einzelleistungen effizient produzieren. Aber nur wenn die relativen Vergütungssätze *genau* den relativen Grenzproduktivitäten bei der Behandlung einer Krankheit entsprechen, wird ein einkommensmaximierender Arzt auch die kostenminimierende Kombination der Einzelleistungen wählen.

Schaubild 18 verdeutlicht dies, wobei ich annehme: Die neuen Vergütungssätze sind im Punkt A aufkommensneutral. Die gestrichelte Linie stellt das Verhältnis der Herstellungskosten der Einzelleistungen $-p_1/p_2$ dar. Die durchgezogene Linie stellt das effektive Kostenverhältnis nach Berücksichtigung der relativen Vergütungssätze für die Einzelleistungen $(-p_1 + \Delta c_1)/(p_2 + \Delta c_2)$ dar, wobei $\Delta c_1 = |p_1 - c_1|$ und $\Delta c_2 = |p_2 - c_2|$ mit $0 < p_1 < c_1$ und $0 < c_2 < p_2$.

(iii) Bei einem *Periodenfixum* wird der Arzt zwar seine Einzelleistungen effizient produzieren, einen Anreiz, den Heilerfolg zu optimieren, hat er nicht.

(iv) Die Vergütung durch *Kopf-Pauschale (Capitation)* gewährleistet Produktionseffizienz, Behandlungseffizienz und ermöglicht dem Zahler auch die Steuerung des vom Arzt angestrebten Heilerfolgs – durch Wahl der optimalen Vergütungshöhe pro eingeschriebenem Patient, sofern die Zahl der eingeschriebenen Versicherten positiv auf Heilerfolge reagiert. Die Kopf-Pauschale ist der Prototyp eines *prospektiven Vergütungssystems*, das die Vergütungshöhe vor der Leistungserstellung festlegt und so die Kostenverantwortung dem Leistungserbringer zuweist. Im Gegensatz dazu ist jedes Vergütungssystem, das

den tatsächlichen Ressourcenverbrauch im Nachhinein berücksichtigt und den Leistungserbringer von dem Kostenrisiko teilweise oder vollständig entbindet, *retrospektiv*. Ein in jeder Hinsicht ideales Vergütungssystem gibt es nicht. Auf Anreizprobleme, die auftreten können, wenn Ärzte ausschließlich durch Kopf-Pauschalen für jeden eingeschriebenen Patienten entlohnt werden, gehe ich in Abschnitt 6.3 gesondert ein.

6. Der stationäre Sektor: Skalenerträge und Spezialisierung im Krankenhaus

Die *volkswirtschaftliche Funktion* von Krankenhäusern liegt in der Versorgung stationärer Patienten, also sowohl in der Pflege und Anwendung therapeutischer Maßnahmen, als auch in der Diagnose, wenn dafür besonders teure Geräte und Eingriffe am Körper des Patienten erforderlich sind. Nicht nur erfordern moderne bildgebende Verfahren in der Diagnose oft so hohe Investitionen, dass sie für niedergelassene Ärzte von vornherein ausscheiden. Auch die vielfältigen Möglichkeiten moderner operativer Verfahren erfordern oft einen stationären Aufenthalt, der vor allem einer intensiven Nachsorge und Beobachtung des Patienten dient. Hinzu kommt die Behandlung Schwerstkranker, die zum Beispiel eine Operation oder andere aufwändige Therapieverfahren benötigen, mit der eine ambulante Arztpraxis überfordert wäre. Diese Ausrichtung des stationären Sektors hat eine ganze Reihe von Implikationen für die Gesundheitspolitik. An erster Stelle ist die besonders hohe Preisinflation bei Krankenhausleistungen zu nennen. Diese darf aber natürlich bei zunehmendem Technologieeinsatz nicht isoliert betrachtet werden – bei gleichzeitigen Qualitätssteigerungen in der medizinischen Behandlung könnte die Kosteneffektivität eines Krankenhauses sogar insgesamt zunehmen.

Historisch betrachtet hat die Gesundheitspolitik den langfristigen Wandel der Krankenhäuser von Pflegeanstalten hin zu Zentren des Technologieeinsatzes

unterstützt. Auch in der Zukunft muss die Gesundheitspolitik die Entwicklung des stationären Sektors gestaltend begleiten. Ursprünglich waren Krankenhäuser kirchliche Pflegeanstalten für Langzeitkranke, in denen Ärzte nur als Gäste Zutritt hatten. Heute dagegen sind fast alle Krankenhäuser *Zentren des spezialisierten Technologieeinsatzes*, in denen meist Ärzte das Sagen haben. Sie entscheiden über den Einsatz teurer Diagnosegeräte, etwa die bildgebenden Verfahren, sowie über die Ausstattung und Nutzung der Einrichtungen für chirurgische Operationen und die Intensivmedizin.

Eine Ausnahmestellung haben die psychiatrischen Krankenhäuser; bei ihnen spielt der Technologieeinsatz keine so große Rolle. Sie existieren in erster Linie, weil es Menschen gibt, die aufgrund einer psychischen Krankheit sich selbst oder die Gemeinschaft gefährden würden, wenn sie außerhalb einer Anstalt (und ohne intensive ärztliche Beobachtung und Betreuung) selbstständig leben müssten. Bei der Ausstattung von Krankenhäusern der Akutmedizin muss die Gesundheitspolitik beachten, dass diese in einer Region zudem eine *Optionsnachfrage* befriedigen müssen und dazu ausreichend Betten für Notfälle, Katastrophen, Epidemien und Kriegsfälle vorhalten.

6.1 Alternative Formen der Krankenhausfinanzierung

Die *Finanzierung* von Krankenhäusern erfolgt nach ganz unterschiedlichen Prinzipien. *Eine* Möglichkeit besteht darin, jedem Krankenhaus ein fixes Budget zuzuweisen. Dies wird zum Beispiel bei Militärkrankenhäusern praktiziert, in den USA auch bei psychiatrischen Krankenhäusern. Eine *zweite* Möglichkeit besteht darin, Tagessätze für jeden behandelten Patienten zu zahlen, unabhängig von der Diagnose oder der Schwere des Krankheitsfalls. Dies wurde bis Ende 2002 in Deutschland praktiziert. Eine *dritte* Möglichkeit der Krankenhausfinanzierung besteht darin, die Kosten der Patientenpflege und -behandlung gegen Einzelnachweise des Ressourceneinsatzes zu erstatten. Wie

die Diskussion um ärztliche Vergütungsformen gezeigt hat, ist dies jedoch im Allgemeinen nicht sinnvoll.

Die *vierte* Möglichkeit besteht darin, Fallpauschalen zu bezahlen, deren Höhe von der Diagnose abhängig ist. Dieses Verfahren ist in Australien und den USA seit langem Standard und wird ab 2003 auch in Deutschland eingeführt. Damit *Fallpauschalen* effizient sind, müssen kostenhomogene Diagnoseklassen definiert werden, die als „Diagnostically Related Groups“ (DRGs) bezeichnet werden. Diese weichen in vielen Fällen von der rein medizinisch motivierten Klassifikation von Diagnosen ab, die Ärzte sich für den Zweck der optimalen Therapiewahl geschaffen haben. Abschnitt 6.3 stellt einige der organisatorischen Veränderungen dar, die Beobachter als Folge der Fallpauschalenfinanzierung für die deutschen Krankenhäuser erwarten.

Eine ausschließliche Finanzierung von Krankenhäusern auf der Basis von Fallpauschalen kann jedoch die Arbeitsteilung zwischen den Krankenhäusern behindern. Zum Beispiel stellt sich dann die Frage, wer für die Forschung der Universitätskliniken bezahlt. In der Vergangenheit ist die Forschung der Universitätskliniken zu einem erheblichen Teil durch Überschüsse aus der Patientenversorgung finanziert worden.²⁷ Eine weitere Problematik liegt darin, dass Fallpauschalen die Anreize zur Technologiedoption verändern und den Einsatz medizinisch vorteilhafter Innovation attraktiv machen können, wenn damit höhere Kosten pro Behandlungsfall verbunden sind. Ein Beispiel hierzu wird in Schmid et al. (2004) dargestellt.

6.2 Möglichkeiten und Grenzen des Wettbewerbs

Die zunehmende *Spezialisierung der Medizin* mit immer wieder neuen diagnostischen und therapeutischen Möglichkeiten hat dazu geführt, dass

²⁷ Eine ähnliche Frage stellt sich in Bezug auf die Ausbildung von Fachärzten und Krankenschwestern. Siehe hierzu Meltzer et al. (2000).

Krankenhäuser Größenvorteile realisieren können, die aus einer größeren Vielfalt der verfügbaren Technologien und der durch Fachärzte vertretenen Teildisziplinen resultieren. Der intensive Technologieeinsatz in modernen Krankenhäusern hat deshalb eine Reihe von organisatorischen Implikationen: Um Skalenerträge zu realisieren, sollte nicht nur die Größe von Krankenhäusern generell zunehmen, sondern auch Größenunterschiede sollten zunehmen. Denn die Arbeitsteilung zwischen Krankenhäusern kann die Vorteile der medizinischen Spezialisierung noch besser nutzen, wenn die Universitätskliniken sogar über die kostenminimierende Größe von Krankenhäusern hinaus wachsen, um die schwersten Fälle aus kleineren Krankenhäusern übernehmen zu können. Daher werden die aufwendigsten und modernsten Technologien in der Regel in Universitätskliniken konzentriert. Andere Krankenhäuser, insbesondere Kreiskrankenhäuser, sind mit Technologien weniger gut ausgestattet. Gleichwohl müssen auch sie vielfältig spezialisierte Fachärzte unter Vertrag haben, um jederzeit auf die Anforderungen unterschiedlicher Krankheitsfälle angemessen reagieren zu können.

Ex post ist es für die Effizienz bei der Nutzung spezialisierten Humankapitals und der Geräteausstattung eines Krankenhauses entscheidend, dass die Art der eingewiesenen Fälle den Erwartungen entspricht, unter denen das Krankenhaus konzipiert wurde. Deshalb besteht eine wichtige Möglichkeit zur Effizienzsteigerung bei der Nutzung vorhandener Krankenhauskapazitäten darin, dass man Patienten aus kleineren Krankenhäusern in die Universitätskliniken überweist, wenn die aufgetretenen Krankheiten von unerwarteter Komplexität sind und den Einsatz spezialisierter Fachärzte erfordern.

Die *Produktionsfunktion* eines Krankenhauses kann im Sinne der Theorie der Produktdifferenzierung von Dixit und Stiglitz wie folgt geschrieben werden:

$$Y = L^{1-\alpha} \int_0^A x_i^\alpha di \quad \text{mit } 0 < \alpha < 1.$$

Darin bezeichnet L die Anzahl der Pflegekräfte und x_i die Anzahl der mit allen nötigen Geräten ausgestatteten Fachärzte der i -ten Teildisziplin, wobei hier zugunsten der mathematischen Eleganz ein Kontinuum medizinischer Teildisziplinen in dem Intervall $[0, A]$ angenommen wird. Durch Romer (1990) wissen wir, dass im Gleichgewicht, das durch Gleichheit des Grenzerlöses und der Grenzkosten für jede Teildisziplin charakterisiert ist, $A = ((2 - \alpha)/2h)H$ gelten muss, wobei h die Fixkosten (für die Ausbildung des Personals und Anschaffung der Geräte) jeder Teildisziplin und H der Humankapitalstock ist. H kann als das gesamte verfügbare medizinische Wissen interpretiert werden, das hier der einzige in der Bereitstellung medizinischer Teildisziplinen benötigte Produktionsfaktor ist. Das Krankenhaus arbeitet dann nach der Produktionsfunktion $Y = bL^{1-\alpha}A^{1-\alpha}H^\alpha$, worin b eine positive Konstante ist. Setzt man für A den Gleichgewichtswert ein, weist die Produktionsfunktion des Krankenhauses

steigende Skalenerträge in Bezug auf L und H auf: $Y = b \left(\frac{2-\alpha}{2h} \right)^{1-\alpha} L^{1-\alpha} H$. Die

Summe der partiellen Produktionselastizitäten, $1+1-\alpha$, ist größer als Eins. Offensichtlich steigt der Output eines Krankenhauses mit dem Umfang des gesamten medizinischen Wissens und sinkt mit den Fixkosten der Einrichtung und Unterhaltung medizinischer Teildisziplinen. Die Fixkosten gehen in die Formel ein, weil die Entwickler neuer medizinischer Geräte, die durch Patente geschützt sind, für ihre Forschungsausgaben und Fachärzte für ihre langwierige Ausbildungsanstrengung entlohnt werden müssen. Solange diese Inputs für medizinische Teildisziplinen nicht nach ihren Grenzkosten entlohnt werden, bleiben jedoch der Output und die Gesamtgröße eines Krankenhauses suboptimal.

Neben den Vorteilen der Arbeitsteilung, der größeren Spezialisierung und Effizienz der Mitarbeiter, die ein größeres Krankenhaus ermöglicht, gibt es weitere Gründe für steigende Skalenerträge. So kann ein größeres Krankenhaus

Fluktuationen bei der Zahl der eingewiesenen Patienten besser ausgleichen, indem es die Patienten so auf die verschiedenen Stationen verteilt, dass fast immer deren optimale Auslastung erreicht wird. Die Evidenz deutet daraufhin, dass eine Krankenhausstation nicht mehr als 40 Betten haben sollte, um effizient arbeiten zu können. Für Krankenhäuser insgesamt zeigt die amerikanische Evidenz zudem, dass eine Gesamtgröße von weniger als 100 Betten zu klein ist, um das volle Leistungsspektrum eines allgemeinen Krankenhauses abzudecken. Weder die Fachärzte der einzelnen Teildisziplinen, noch die teuren diagnostischen Spezialgeräte, z.B. die bildgebenden Verfahren, können dann ausgelastet werden. *Abnehmende Skalenerträge* dagegen scheinen bei Krankenhäusern mit mehr als 500 Betten einzusetzen. Die Pflegekosten pro Patient und Tag steigen dann deutlich an. In der obigen Formel müsste b deshalb eine abnehmende Funktion der Bettenzahl B sein: $b = b(B) > 0 \quad \forall B$ mit $\partial b / \partial B < 0$ und $\partial^2 b / \partial B^2 < 0$.

Dass Universitätskrankenhäuser die optimale Größe eines allgemeinen Krankenhauses deutlich übersteigen, ist auch deshalb sinnvoll, weil sie neben der Patientenversorgung den Auftrag haben, die Lehre, die Ausbildung von Krankenschwestern und Fachärzten, sowie die Forschung im gesamten Spektrum der medizinischen Teildisziplinen zu gewährleisten. Neuerdings gibt es in Deutschland auch gesetzliche Regelungen, die im Rahmen von Disease-Management-Programmen Mindestmengen bei der Einrichtung von Früherkennungsprogrammen und Behandlungszentren für bestimmte Krankheitsbilder vorschreiben. Börchers et al. (2003) beschreiben diese Entwicklung am Beispiel der Brustkrebszentren. Sie berichten von amerikanischen und britischen Studien, in denen sich ein signifikanter Fünf-Jahres-Überlebensvorteil bei Frauen zeigte, die in einem Zentrum mit mehr als 150 Brustkrebsoperationen pro Jahr operiert wurden. Sie führen die im langfristigen Vergleich geringe Mortalitätsreduktion und das im europäischen Mittelfeld vergleichsweise hohe

Mortalitätsniveau in Deutschland auf eine zu geringe Spezialisierung und Zentralisierung der Behandlung des Mammakarzinoms zurück. Ein weiteres Beispiel für die Notwendigkeit der Konzentration medizinischer Ressourcen sind die Zentren für erbliche Krebskrankheiten, die sich mit Unterstützung der Deutschen Krebshilfe gebildet haben. Propping und Aretz (2004), S. S13, legen dar, dass nur derartige Zentren den besonderen Anforderungen an die klinische und genetische Diagnostik bei erblichen Krebskrankheiten gerecht werden und die schwierige Überwachung und Betreuung der Betroffenen leisten können.

Ein Beispiel für ein modernes Großkrankenhaus ist das in den 70er Jahren gebaute Klinikum der Rheinisch-Westfälischen-Technischen-Hochschule in Aachen, das in Ackerknecht und Murken (1992), S. 175 f., beschrieben wird. Damit dieses Klinikum den gesamten Fächerkanon der theoretischen und klinischen Medizin unter einem Dach vereinigen konnte, einschließlich der diagnostischen und therapeutischen Dienstleistungen wie Labormedizin, Röntgenologie, Nuklearmedizin, Physiotherapie, medizinische Bibliotheken und Unterrichtsräume, entstand ein Gebäude für insgesamt 28 Kliniken, 18 Institute, 1500 Betten, 3800 Vollzeitarbeitsplätzen und 2400 Ausbildungsplätze für Studenten und Pflegekräfte sowie anderen medizinischen Dienstleistungsberufen.

Es ist eine Folge des raschen medizinischen Fortschritts und der Einführung zahlreicher neuer Großtechnologien, dass die Aufteilung der Medizin in Teildisziplinen weit vorangeschritten ist. Einige dieser Entwicklungen finden ihren Niederschlag in zentralen Einrichtungen, wie radiologische Abteilungen und labormedizinische Zentren. Die Entwicklung der Anästhesiologie zum Beispiel hat die moderne Intensivmedizin möglich gemacht und dadurch die Technologieintensität von Krankenhäusern wesentlich erhöht. Daneben ist der Einsatz der Technologie vor allem durch die erhebliche Erweiterung der diagnostischen Möglichkeiten intensiviert geworden, die heute Isotopen-

medizin, Computertomografie, Echokardiografie, Endoskopie, Endokrinologie, Lungenfunktionstests und vieles anderes mehr umfassen. Allerdings können Probleme der Koordination zwischen den Teildisziplinen eine wichtige Quelle *abnehmender* Skalenerträge sein.

In Deutschland wurde der Ausbau großer, mit modernen Technologien ausgestatteter Krankenhäuser durch öffentliche Finanzierung seit den 60er Jahren stark vorangetrieben. Dadurch sind allerdings inzwischen *Überkapazitäten* entstanden, die von einigen Krankenhäusern systematisch im europäischen Ausland und in arabischen Ländern vermarktet werden. So wirbt man heute gezielt Patienten aus Mitgliedsländern der Europäischen Union an, die in ihrem Heimatland längere Zeit auf notwendige Operationen warten müssten. Dies gilt insbesondere für die skandinavischen Länder und das Vereinigte Königreich, wo ein nationaler Gesundheitsdienst zu großer Ressourcenknappheit und zur Rationierung vieler medizinischer Leistungen geführt hat.

6.3 Ausblick

Organisatorische Veränderungen. Durch die Einführung eines prospektiven Vergütungssystems im deutschen Krankenhaussektor wird es zu starken strukturellen Anpassungen kommen, in deren Folge mittelfristig mit einem Bettenabbau um 20 bis 30 Prozent zu rechnen ist. Klimpe (2002) und Vera (2003) beschreiben den organisatorischen Wandel zum modularen Großkrankenhaus, den die Einführung des DRG-Systems voraussichtlich in Deutschland auslösen wird. Im Mittelpunkt der Modularisierung steht dabei auch weiterhin die medizinische Leistungserstellung. Dieser Kernprozess des Krankenhauses lässt sich unterteilen in die vier Schritte Aufnahme, Diagnose, Therapie und Entlassung. Parallel dazu laufen die beiden unterstützenden Prozesse Pflege und Verwaltung ab. Für einen möglichst effizienten Ablauf dieser Prozesse ist eine weitgehende Standardisierung der Arbeitsabläufe

vorteilhaft. Hierzu wird die Einführung von „Clinical Pathways“ (Patientenpfade) empfohlen, die alle Leistungen beschreiben, die im Laufe eines Behandlungsprozesses zu erbringen sind. Die Grundlage hierfür sollten Behandlungsleitlinien sein, die auf der evidenzbasierten Medizin beruhen und explizite Vorgaben zur Behandlungsqualität machen. Vera (2003) argumentiert, dass erst „Clinical Pathways“, die eine friktionslose Versorgung von Patienten über Abteilungsgrenzen hinweg ermöglichen, die Voraussetzung für ein integriertes Prozess-, Qualitäts- und Kostenmanagement schaffen. Nicht mehr die isolierte diagnostische oder therapeutische Einzelleistung, sondern jeweils der gesamte Behandlungsprozess einer Krankheit wird im Mittelpunkt stehen, wenn Deutschlands Krankenhäuser beginnen, ein modernes Produktmanagement zu entwickeln.

Gleichzeitig ist eine stärkere Spezialisierung auf ausgewählte Patientengruppen zu erwarten. Nicht nur ist zu erwarten, dass einzelne Krankenhäuser nach Wettbewerbsvorteilen durch eine Spezialisierung auf ausgewählte DRGs suchen, in denen sie andere Krankenhäusern bei den Kosten unterbieten können. Eine zunehmende Spezialisierung auf Teilbereiche des Gesamtspektrums stationärer Leistungen ist schon deshalb unvermeidlich, weil im Zuge der weiteren Reform der deutschen Krankenhausfinanzierung für alle Arten von Operationen gesetzlich festgelegte Mindestmengen zur Voraussetzung der Erstattungsfähigkeit durch die Krankenkassen werden sollen. Aber auch innerhalb einzelner Diagnoseklassen kann es zu einer Spezialisierung auf Patientengruppen kommen, bei denen vorab erkennbar ist, dass sie aufgrund ihrer Fallschwere („case mix“) kostengünstiger zu behandeln sind. Weil Patienten mit hoher Fallschwere in der Behandlung oft besonders teuer sind, können sie ein Verlustgeschäft darstellen, wenn das Krankenhaus nur die für die Behandlung eines *durchschnittlichen* Falles als kostendeckend ermittelten Vergütungssätze erhält.

Ellis (1998) hat dazu aus einem theoretischen Modell drei aus einzelwirtschaftlicher Sicht rationale Verhaltensweisen von Krankenhäusern im Wettbewerb abgeleitet: Erstens, ein Überangebot von Leistungen für Patienten mit geringer Fallschwere (*Creaming*); zweitens, ein unzureichendes Angebot für Patienten mit hoher Fallschwere (*Skimping*); und drittens, das explizite Vermeiden der Behandlung von Patienten mit hoher Fallschwere (*Dumping*).²⁸ Die Relevanz dieser Verhaltenshypothesen hat eine Reihe von Studien für den amerikanischen Krankenhausesektor nach Einführung eines prospektiven Vergütungssystems durch die einflussreiche staatliche Medicare-Organisation, die die Krankenhauskosten der über 65jährigen erstattet, empirisch belegt: Unprofitable Patienten werden in den USA überwiegend in staatlichen Krankenhäusern behandelt. Zudem wurde in privaten Krankenhäusern beobachtet, dass die durchschnittliche Verweildauer bei Patienten mit geringer Fallschwere verlängert, bei Patienten mit hoher Fallschwere dagegen verkürzt wurde, was mit den Hypothesen des *Creaming* und *Skimping* im Einklang steht.

Ganz im Sinne des *Creaming* sieht Stausberg (2002), S. 298, eine unerwünschte Auswirkung jedes DRG-Systems in der Mengensteigerung durch eine Indikationsausweitung, die sich durch externe Überprüfungen der Notwendigkeit von Krankenhausaufnahmen begrenzen lassen. Dagegen entfällt die Prüfung überflüssiger Behandlungstage, da die Erlöse in einem DRG-System im Prinzip unabhängig von der Liegedauer sind. Aus Sicht eines gegen die Kosten versicherten Patienten ist vor allem ein Verlust an Behandlungsqualität zu befürchten, der zum Beispiel durch eine aus betriebswirtschaftlichem Kalkül zu starke Verkürzung der Liegedauer entstehen kann. Stausberg (2002) sieht außerdem die Gefahr, dass eine „zwar leistungsorientierte, aber nicht leistungsgerechte Definition der DRGs“ zu übermäßigem Sparen bei der

²⁸ Letzteres kann zum Beispiel dadurch erreicht werden, dass ein Krankenhaus bestimmte diagnostische Geräte nicht anschafft, die nur im Falle besonders aufwändig zu behandelnder Patienten gebraucht würden.

Qualität einzelner Leistungen führen könnte. Diese Gefahr erscheint jedoch nicht allzu groß, denn zur Vorbeugung gegen eine „Billigmedizin“ enthält § 135a SGB V präzise Anforderungen an eine externe Qualitätssicherung und verlangt zudem die Einführung und Weiterentwicklung eines internen Qualitätsmanagements (vgl. Mohr 2002).

7. Integrierte Versorgung: Gatekeeper und medizinische Leitlinien

7.1 Managed Care

Wie die Theorie medizinischer Entscheidungen gezeigt hat, können Anreiz- und Informationsprobleme einer effizienten Allokation medizinischer Ressourcen in einem innovationsfreundlichen Gesundheitssystem im Wege stehen. Bei einem raschen technologischen Wandel in der Medizin ist das öffentliche Gut des Wissens um die bestmögliche medizinische Praxis von besonderer Bedeutung. Der einzelne Arzt hat jedoch unzureichende Anreize, sich durch Fortbildung stets auf dem neuesten Stand zu halten. Im Ergebnis ist die Diffusion neuen medizinischen Wissens und neuer medizinischer Technologien eingeschränkt.

Es ist daher zu fragen, durch welche *organisatorischen* Innovationen die Anreizsituation der Ärzte verbessert und die Diffusion neuer Informationen über die relative Häufigkeit von Krankheiten und die Möglichkeiten medizinischer Technologien beschleunigt werden kann. Organisatorische Innovationen mit dieser Zielsetzung werden unter dem Begriff „Managed Care“ zusammengefasst. Verschiedene Varianten des Managed Care werden heute vor allem in den USA und in der Schweiz praktiziert. Ein gemeinsames Kennzeichen ist dabei, dass in der Regel die freie Arztwahl des Patienten eingeschränkt wird.

In den USA stehen Managed Care-Organisationen in einem intensiven Wettbewerb um Versicherte und haben somit einen starken finanziellen Anreiz,

nach effizienten Organisationsformen zu suchen und dabei auch immer wieder neue organisatorische Experimente zu wagen. Aufgrund der geringen Regulierung des Marktes für medizinische Leistungen haben sie eine Vielzahl von Methoden entwickelt, um die kostentreibenden Auswirkungen der Unsicherheit medizinischer Entscheidungen zu begrenzen. Die Effizienzvorteile der einzelnen Methoden sind leicht zu verstehen, wenn man die einschlägige ökonomische Theorie über die Auswirkungen von Unsicherheit und asymmetrischer Information im Gesundheitswesen kennt.

Im einzelnen verwenden Managed Care-Organisationen vor allem folgende Methoden: *Gatekeeper* – ein Arzt wird dafür verantwortlich gemacht, die Gesamtkosten der Behandlung eines im zugewiesenen Patienten unter Kontrolle zu halten. Er wird also verantwortlich gemacht für alle Laboruntersuchungen, Überweisungen, Krankenhausbehandlungen und den gesamten Medikamentenkonsum, der mit einer von ihm diagnostizierten Krankheit im Zusammenhang steht. *Capitation* – der Arzt hält einen großen Teil seines Einkommens in Form eines Budgets für jeden Patienten, der bei ihm eingeschrieben ist, und hat deshalb einen Anreiz, möglichst effizient mit den Ressourcen umzugehen (prospektive Vergütung). *Rückforderungen* bei Budgetüberschreitung – viele Managed Care-Organisationen machen die Ärzte finanziell mitverantwortlich für die Überschreitung vorgegebener Budgets. Hierzu wird in der Regel ein Teil der Vergütung in einen Sonderfonds eingezahlt, der erst am Ende einer Periode an die Ärzte verteilt wird, wenn sie innerhalb des vorgegebenen Budgets geblieben sind. Haben sie es überschritten, wird der Fonds an die Health Maintenance Organisation zurückgezahlt.

Weil die Managed Care-Organisationen Anreize haben, ihre Kosten mit diesen und anderen Methoden klein zu halten, wird vielfach vermutet, dass sie einen wesentlichen Beitrag auch zur gesamtwirtschaftlichen Effizienz eines Gesundheitssystems leisten. Das folgt jedoch nicht zwingend. Denn die Theorie

zeigt, dass die Managed Care-Organisationen über sehr unterschiedliche Möglichkeiten verfügen, ihre Kosten zu reduzieren, von denen längst nicht alle aus volkswirtschaftlicher Sicht effizient sind. Die Evidenz deutet vielmehr darauf hin, dass die Kostenersparnisse der Managed Care-Organisationen zumindest teilweise auf *positive Risikoselektion* und erfolgreiche *Kostenüberwälzung* („cost shifting“) auf andere Kostenträger oder Drittzahler zurückzuführen sind.

Glied (2000) bietet einen Überblick über die theoretische und empirische Literatur zur Entwicklung des Managed Care in den USA, wo diese Form der Krankenversicherung seit Mitte der achtziger Jahre zur dominanten Form geworden ist. Ende 1995 waren bereits mehr als 90 Millionen Privatversicherte Mitglied in einer Health Maintenance Organisation, Preferred Provider Organisation oder einer Zwischenform. Darüber hinaus hatten fast alle konventionellen Krankenversicherer begonnen, Elemente des Managed Care in ihre Versicherungsverträge aufzunehmen.

Die zentrale Frage ist: Wie effizient ist Managed Care im Vergleich zu anderen Arten der Krankenversicherung? Die ökonomische Theorie gibt hierauf keine eindeutige Antwort. Zwar wird die Ausbreitung von Managed Care als Reaktion auf spezifische Formen des Marktversagens im Gesundheitswesen angesehen, aber zu jedem einzelnen Marktversagen gibt es jeweils unterschiedliche Lösungsansätze. Ramsey und Pauly (1997) haben gezeigt, dass es aus theoretischer Sicht keinen Grund gibt anzunehmen, dass Managed Care alle Formen des Marktversagens besser bewältigt als andere Organisationsansätze. Auch die empirische Literatur hat eine höhere Effizienz von Managed Care-Organisationen im Vergleich zu konventionellen Versicherungsverträgen nicht eindeutig nachweisen können.

Die Vielfalt der vertraglichen Detaillösungen in Managed Care macht es nicht nur schwierig, Managed Care allgemeingültig zu definieren, sondern macht es auch schwierig, aus empirischen Studien Schlussfolgerungen über die Effizienz einzelner Elemente zu ziehen. Während ein konventioneller Versicherungsvertrag im Wesentlichen durch drei Elemente definiert ist – die Versicherungsprämie, die Menge der Leistungen und die Selbstbeteiligungssätze des Versicherten –, unterscheiden sich Managed Care-Organisationen zusätzlich dadurch, wie sie die Ersteller medizinischer Leistungen auswählen, bezahlen und überwachen.

Viele Managed Care-Organisationen bieten eine breitere Palette von Leistungen an als konventionelle Krankenversicherungsgesellschaften. Insbesondere bieten Managed Care-Organisationen mehr vorbeugende Gesundheitsleistungen an. In der Literatur wird argumentiert, dass dies ein wichtiges Element ist, um gute Risiken – relativ gesunde Versicherte – zu attrahieren. Die finanzielle Selbstbeteiligung der Versicherten an ihren Krankheitskosten ist bei vielen Managed Care-Organisationen deutlich geringer als in der konventionellen Krankenversicherung. Zudem spielt die finanzielle Selbstbeteiligung bei Managed Care-Organisationen eine andere Rolle. Sie zielt nicht in erster Linie auf eine Einschränkung der nachgefragten Leistungen, sondern auf die Nutzung von Leistungen der mit einer Managed Care Organisation unter Vertrag stehenden Leistungsanbieter. Den Versicherten ist es vielfach freigestellt, Leistungen von Anbietern außerhalb des Netzwerkes in Anspruch zu nehmen, aber die Versicherten müssen dann eine höhere Selbstbeteiligung tragen.

Die Kontrolle ihrer Kosten erreichen die Managed Care Organisationen nicht nur dadurch, dass sie kostengünstige Leistungsanbieter auswählen und unter Vertrag nehmen, sondern auch durch besondere Methoden der Vergütung, die häufig stärkere Anreize zu einer kosteneffektiven Behandlung setzt als die konventionelle Krankenversicherung. Zusätzlich zu periodischen Festbeträgen,

Einzelleistungsvergütung und Kopf-Pauschalen nutzen die Managed Care Organisationen erfolgsabhängige Zahlungen und andere Anreizmechanismen, die für einzelne medizinische Leistungen unterschiedlich gestaltet sein können. Kopf-Pauschalen wälzen einen Teil der finanziellen Risiken der Krankheit auf die Leistungsanbieter ab. Ellis und McGuire (1993) haben Kopf-Pauschalen daher als eine Art der anbieterseitigen Selbstbeteiligung charakterisiert. Die Anbieter medizinischer Leistungen sind oft besser in der Lage, finanzielle Risiken zu tragen als die Versicherten. Sie sind meist nicht nur weniger risikoscheu, sondern verfügen auch über mehr Informationen über Nutzen und Risiken und können daher besser effiziente Allokationsentscheidungen treffen. Andererseits können Kopf-Pauschalen dazu führen, dass schlecht informierte Patienten notwendige medizinische Leistungen nicht erhalten oder dass die Leistungsanbieter versuchen, besonders kostenträchtige Patienten abzuwimmeln.

Glied (2000) diskutiert, welchen Beitrag Managed Care-Organisationen zur Bewältigung der verschiedenen Formen von Marktversagen machen, die den Markt für Krankenversicherungsverträge charakterisieren. Dazu ist zunächst festzustellen, dass Managed Care-Organisationen eine stärkere Rationierung medizinischer Leistungen praktizieren und so möglicherweise günstigere Risiken attrahieren können als konventionelle Versicherungen. Theoretisch spricht dies dafür, dass Managed Care-Organisationen das Problem der asymmetrischen Information zwischen Versicherten und Versicherern und die daraus resultierende *adverse Selektion* zumindest teilweise – durch „positive“ Risikoselektion – überwinden können. Die Evidenz hierfür ist jedoch unklar. Aus gesamtwirtschaftlicher Sicht ist zudem festzuhalten, dass eine erfolgreiche Selektion guter Risiken dazu führen würde, dass die schlechteren Risiken sich stärker in der konventionellen Krankenversicherung konzentrieren und deren

Kosten in die Höhe treiben. Die Effizienz des Krankenversicherungsmarktes insgesamt kann sich dadurch verringern.

Moral Hazard wird im Wesentlichen durch anbieterseitige Selbstbeteiligung bekämpft. Das (2002) argumentiert, dass Kopfpauschalen erfolgreich die finanziellen Anreize zur ärztlichen Induktion von Nachfrage beseitigt hätten. Die rasche Ausbreitung des Managed Care seit Ende der 80er Jahre könne den zwischen 1989 und 1996 zu beobachtenden Rückgang der relativen Häufigkeit von Kaiserschnitten bei Geburten in den USA erklären, während die These von der angebotsinduzierten Nachfrage bei konventionellen Krankenversicherungen gültig bleibe. In seiner empirischen Untersuchung findet Das (2002) eine starke *negative* Korrelation zwischen dem regionalen Wachstum von Managed Care und der regionalen Rate von Kaiserschnittgeburten.

Aus theoretischer Sicht hängt die optimale Kombination anbieterseitiger und nachfrageseitiger Anreizmechanismen hängt von der Verteilung der Entscheidungsbefugnisse, der relativen Fähigkeit von Leistungsanbietern und Konsumenten, Risiken zu tragen, und von den relativen Verwaltungskosten ab. Wenn es Managed Care über eine Verbesserung der statischen Allokationseffizienz hinaus gelingt, die Intensität des Technologieeinsatzes zu reduzieren, kann dies die Kosten medizinischer Behandlungen auch für die Versicherten konventioneller Krankenversicherungsorganisationen senken. Dies ist ein langfristiger volkswirtschaftlicher Vorteil, der jedoch bislang nicht ausreichend empirisch belegt ist. Eine beschreibende Darstellung, wie Managed Care die Einführung und Verbreitung medizinischer Technologien in den USA beeinflusst, findet sich in Cohen (1997).

Ein weiterer wichtiger Unterschied bei der Bewältigung von Marktversagen liegt in den besseren Anreizen und in der größeren Fähigkeit von Managed Care-Organisationen, Informationen über die Qualität medizinischer Leistungs-

anbieter zu sammeln, auszuwerten und den Versicherten zur Verfügung zu stellen. Patienten haben im Allgemeinen große Schwierigkeiten, die Qualität einzelner Leistungsanbieter zu beurteilen. Managed Care-Organisationen haben einen besonderen Anreiz, diese Informationen zu sammeln, weil sie durch ihre Auswahl der Leistungsanbieter und durch ihren Einfluss auf medizinische Behandlungsleitlinien ihren eigenen Erfolg im Hinblick auf die Qualität der Leistungen verbessern können. Zudem haben Managed Care-Organisationen durch die Integration der Leistungserstellung einen Vorteil, Informationen über Behandlungsepisoden insgesamt und über die Ergebnisse der medizinischen Behandlungen zu sammeln und zu vergleichen.

Schell und Lauterbach (2002) beschreiben die von amerikanischen Managed Care-Organisationen verwendeten Instrumente eines umfassenden Qualitätsmanagements. Danach verwenden mehr als 90 Prozent dieser Organisationen zur Evaluation ihrer medizinischer Leistungen die vom *National Committee for Quality Assurance* (NCQA) entwickelten HEDIS-Indikatoren, die die Qualität eines Anbieters anhand einer ganzen Reihe relevanter Kriterien wie Effektivität, Verfügbarkeit, Stabilität messen. Zusätzlich werden standardisierte Indikatoren zur Messung der Patientenzufriedenheit verwendet. Schell und Lauterbach (2002) argumentieren, dass der hohe Wettbewerbsdruck den amerikanischen Managed Care Organisationen einen ausreichenden Anreiz gibt, sich freiwillig einem Qualitätsvergleich zu stellen und die Ergebnisse öffentlich zu machen. Als Ergänzung könne eine Zertifizierung von Managed Care-Organisation den Konsumenten helfen, das Preis-Leistungsverhältnis unterschiedlicher Anbieter zu beurteilen.

Im Gegensatz zu konventionellen Versicherungsgesellschaften verfügen die Managed Care-Organisationen über einen klaren Anreiz und über die nötigen Instrumente, um durch die Zufriedenheit ihrer Kunden einen Markennamen für eine hohe Qualität der durch ihre Vertragsärzte angebotenen Leistungen zu

entwickeln. Bei konventionellen Krankenversicherungen bestimmt im Grunde der Versicherte durch seine freie Arztwahl selbst über die Qualität der empfangenen Leistungen. Allerdings besteht bei Managed Care immer die Gefahr, dass Leistungsanbieter, die nach der Qualität ihrer medizinischen Ergebnisse bewertet werden, einen Anreiz erhalten, „schwierige“ Patienten von vornherein abzuwimmeln, um eine schlechte Beurteilung der Ergebnisse zu vermeiden. Bei „schwierigen“ Patienten ist zu erwarten, dass die Ergebnisse bei gleichem Aufwand hinter den Ergebnissen zurückbleiben, die bei einem weniger schwierigen Patienten erzielt werden können.

7.2 Entwicklung und Anwendung medizinischer Leitlinien

Grundsätzlich positiv ist zu werten, dass viele Managed Care-Organisationen *medizinische Leitlinien* verwenden, an denen sich die Vertragsärzte bei ihren diagnostischen Entscheidungen und therapeutischen Handlungen orientieren müssen. Für die Einführung von Disease Management-Programmen, die aus Leitlinien abgeleitet sind, ist zudem die Vernetzung der Versorgungsstrukturen, die von Managed Care-Organisationen bereitgehalten werden, von Vorteil. Lauterbach (1997) sieht jedoch auch erhebliche Konfliktpotentiale zwischen den Prinzipien des Managed Care und der Einführung von „Disease Management-Programmen“. Vor allem ist es bei einer Fluktuation von bis zu 30 Prozent der Versicherten pro Jahr aus Sicht einer Managed Care-Organisation wenig lohnend, „Disease Management-Programme“ für chronische Krankheiten zu finanzieren, denn die damit erreichbaren längerfristigen Verbesserungen des Gesundheitszustandes der Kranken drohen bei deren Abwanderung einem Wettbewerber zu zufließen.

Wenn Leitlinien immer wieder auf den neuesten Stand des medizinischen Wissens gebracht werden, können sie, wie oben dargelegt, wesentlich zu einer effizienten Diffusion neuen Wissens und neuer medizinischer Technologien

beitragen. Die vertikale Integration von Versicherungs- und Versorgungsfunktion im Managed Care kann – indem sie medizinische Leitlinien ins Zentrum stellt – die Diffusion neuen Wissens und das Lernen aus praktischen Erfahrungen mit Leitlinien grundsätzlich erleichtern. Andererseits besteht auch die Gefahr, dass eine vertikale Organisation zu konservativ ist und zu wenig mit neuen Technologien experimentiert. Dies könnte durch eine zu starke Betonung kurzfristiger Kostenersparnisse verursacht sein, bei der die mittel- und langfristigen Nutzen einer besseren Kranken-Behandlung nicht ausreichend berücksichtigt werden.

Aber auch wenn eine Managed Care-Organisation ihren eigenen Vertragsärzten die richtigen Anreize setzt, sich am Informationsaustausch zur Entwicklung neuer Behandlungsleitlinien zu beteiligen, bleibt aus gesamtwirtschaftlicher Sicht zu fragen, welche Anreize sie im Wettbewerb haben, über organisatorische Grenzen hinweg Informationen auszutauschen. Ein Pool neuen medizinischen Wissens, aus dem alle Managed Care-Organisationen und alle Anbieter medizinischer Leistungen schöpfen können, stellt ja die optimale Nutzung eines in der Anwendung nicht-rivalisierenden Gutes dar. Harhoff et al. (2003) entwickeln – in einem etwas anderen Kontext – ein spieltheoretisches Modell, das Bedingungen definiert, unter denen Wettbewerber freiwillig Information preisgeben, die zu einer Innovation führen kann, von der alle Wettbewerber in einem Wirtschaftszweig profitieren. Aus diesem Ansatz lassen sich nun für unseren Zweck Hypothesen ableiten, unter denen auch Managed Care-Organisationen im Wettbewerb um Patienten bereit sein könnten, Informationen an einen externen Entwickler medizinischer Leitlinien zu geben, selbst wenn konkurrierende Managed Care-Organisationen davon direkt oder indirekt profitieren können. Der externe Leitlinien-Entwickler könnte zum Beispiel eine medizinische Fachgesellschaft oder das kommende deutsche *Institut für Qualität und Wirtschaftlichkeit in der Medizin* (IQWiM) sein.

Der externe Leitbildentwickler hätte dabei das exklusive Recht, einzelne Informationen in eine neu zu entwickelnde Behandlungsleitlinie aufzunehmen oder auszulassen. Harhoff et al. (2003) betonen, dass eine freiwillige Informationsweitergabe in ihrem Modell bereits dann zu erwarten ist, wenn mindestens ein Nutzer hinreichend von der Weitergabe seiner Informationen profitiert. Der Gesundheitspolitik ist die Ermutigung und eine nicht richtungsvorgebende Unterstützung der Entwicklung von Behandlungsleitlinien auf der Grundlage freiwillig zur Verfügung gestellter Informationen aus Wissenschaft und medizinischer Praxis als die derzeit beste Strategie zu empfehlen.

Eine ähnliche Schlussfolgerung lässt sich aus der Analyse von Dasgupta und David (1994) zur Rolle des Wettbewerbs im wissenschaftlichen System ableiten: Wettbewerb zwischen Managed Care-Organisationen muss demnach nicht unvereinbar mit einem freien Zugang zu allen relevanten Informationen aus der Praxis des Managed Care sein, die zur Schaffung des öffentlichen Gutes medizinisches Wissen beitragen. Um den Wettbewerb mit einem Teilen dieser Informationen zu vereinbaren, ist eine *Kultur* zu entwickeln, die einen kooperativen Wettbewerb stabilisiert und reguliert. Dasgupta und David (1994) beschreiben die Mechanismen dieser Kultur, die mit dem Begriff der „Republic of Science“ belegt ist und dem Ideal der naturwissenschaftlichen Forschung entspricht. Die Teilnahme an diesem offenen System der Wissensschaffung ist der Analyse zufolge für alle Beteiligten vorteilhaft, weil sie ohne die Nutzung gepoolter Informationen kaum Chancen hätten, Priorität für eigene Ideen, Entdeckungen und Erfindungen zu etablieren und so ihren Anteil an deren Ertrag zu ernten.

Im Gesundheitswesen schlagen sich die Ergebnisse kooperativer Wissensgenerierung in neuen Behandlungsleitlinien nieder, die gewissermaßen die Software für die Erstellung medizinischer Leistungen darstellen. Aus diesen Überlegungen schließe ich, dass auch im Gesundheitswesen eine der

Wissenschaft entsprechende Kultur zu entwickeln ist, die die Weitergabe von Informationen und Erfahrungen aus der medizinischen Praxis belohnt, wenn diese für die Entwicklung von Behandlungsleitlinien nützlich sind. Die Karrierechancen von Ärzten und anderen an der Erstellung medizinischer Leistungen Beteiligter müssen den sozialen Wert freiwillig gegebener Informationen honorieren und die Vernetzung des ganzen medizinischen Systems durch Job-Rotation und laterale Kommunikation über die traditionellen Grenzen zwischen Versicherungen, Leistungsanbietern und Regulierungsbehörden hinweg fördern.

Kritisch anzumerken ist, dass eine zu starke Standardisierung der medizinischen Entscheidungsfindung durch Behandlungsleitlinien auch Kosten im Sinne von Effizienzverlusten zur Folge haben kann. Solche Kosten treten zum Beispiel dann auf, wenn Diagnose und Therapie nicht mehr optimal an die besonderen Umstände des einzelnen Patienten angepasst werden. Auf ein ähnliches Problem hatte ich bei der Diskussion über Fallpauschalen in der Vergütung von Krankenhausleistungen hingewiesen, wo Anreize zur suboptimalen Diagnose von Krankheiten entstehen können. Wie Morris und Shin (2002) zeigen, kann sich die Wohlfahrt im Rahmen entsprechender Modelle verschlechtern, wenn die Bereitstellung öffentlicher Information, an der sich eine Vielzahl komplementärer Allokationsentscheidungen orientiert, die privaten Bemühungen um eigene Informationsquellen erlahmen lässt, obwohl die öffentliche Information unpräzise ist. Der Grund für dieses Ergebnis ist, dass allein die öffentliche Information geeignet ist, die komplementären Allokationsentscheidungen zu koordinieren.

8. Schlussbemerkung

Die Produktionsbedingungen medizinischer Leistungen sind durch die revolutionären Erfolge der medizinischen Forschung und die rasche Aus-

weitung der technologischen Möglichkeiten seit dem frühen 20. Jahrhundert einem ständigen Wandel unterworfen. Ein fortgesetzter – aber in seinen Auswirkungen diskontinuierlicher – Wandel wird auch die Zukunft des Gesundheitswesens prägen. Neben dem Einsatz von Arbeit, Kapital und anderen tangiblen Ressourcen, bei denen private Eigentumsrechte die Allokation erleichtern, erfordert die Produktion medizinischer Leistungen in zunehmendem Maße Inputs, die den Charakter eines öffentlichen Gutes haben. Öffentliche Güter sind dadurch definiert, dass mehrere Verwendungszwecke nicht miteinander rivalisieren und das Ausschlussprinzip nicht anwendbar ist. Der wichtigste Input dieser Art ist das medizinische Wissen, das breit gefasst neben dem Wissen um die Ursachen und Behandlungsmöglichkeiten einzelner Krankheiten auch deren Epidemiologie umfasst. Einer ständigen Aktualisierung bedarf insbesondere das Wissen um die relative Häufigkeit unterschiedlich zu behandelnder Gesundheitsstörungen und um die relativen Nutzen und Kosten therapeutischer Technologien. Durch den raschen technologischen Wandel in der Medizin nimmt die Bedeutung dieses öffentlichen Gutes stetig zu.

Eine immer stärkere Spezialisierung der Ärzte ist aufgrund des enormen Wachstums medizinischen Wissens und immer vielfältigerer technologischer Möglichkeiten unvermeidlich und im Grundsatz auch sinnvoll. Gleichzeitig vervielfältigen sich jedoch auch die Möglichkeiten zu Fehldiagnosen. So haben zum Beispiel „self-referrals“ zum falschen Facharzt in der Regel einen besonders hohen Anteil von Fehldiagnosen zur Folge. Eine Stärkung der Allgemeinärzte als Gatekeeper des Gesundheitswesens kann hier eine heilsame Wirkung entfalten; ihre Aufgabe muss es sein, die Patienten sorgfältig zu filtern, bevor sie zu den unterschiedlichen Spezialisten mit ihren kostenintensiven Diagnose- und Behandlungsformen gelangen.

Medizinische Praxisvariationen können als die quantitativ bedeutsamste Form des Marktversagens in der Erstellung medizinischer Leistungen gelten. Die

regionalen Praxisvariationen werden durch einen quantitativen und qualitativen Mangel an wichtigen öffentlichen Gütern, insbesondere bei der Bewertung neuer medizinischer Technologien, und durch die noch unzureichende Entwicklung einer modernen Informationsinfrastruktur zur raschen Diffusion neuen medizinischen Wissens verstärkt.

Der Wohlfahrtsverlust aus den regionalen Praxisvariationen wird zudem durch den technologischen Wandel vergrößert, weil die durchschnittliche Anwendungsrate einer neuen Technologie verzerrt ist, wenn der Diffusionsprozess verzögert ist. Hierzu trägt zusätzlich bei, dass ein rascher technologischer Wandel die Varianz der Anwendungsrate neuer Technologien innerhalb einer Region vergrößert. Auch innerhalb einer Region kann sich deshalb erst allmählich – nach Ablauf eines längeren Bayesianischen Lernprozesses – eine einheitliche Anwendungsrate herausbilden, wenn eine neue medizinische Technologie eingeführt worden ist.

Die Opportunitätskosten einer Steuerung durch Anreize auf der Angebotsseite haben sich infolge technologischer Innovationen, insbesondere durch die phantastischen Möglichkeiten der modernen Informationstechnologie, deutlich verringert. Die im verarbeitenden Gewerbe bewährte Methode des „Concurrent Engineering“ kann für die Weiterentwicklung des Innovationsprozesses im Gesundheitswesen eine nützliche Metapher sein. Organisatorischer und technologischer Wandel sind in der Medizin zwei Seiten ein und derselben Medaille. Obwohl es zweifellos Wechselwirkungen gibt, wie sie die Literatur zum Beispiel für den Einfluss der organisatorischen Form von Krankenversicherungsschutz auf die technologischen Innovationsanreize diskutiert, muss die Gesundheitspolitik doch davon ausgehen, dass der technologische Wandel grundlegender ist. Er ist bei langfristiger Betrachtung unabhängig von der Gestaltung einzelner nationaler Gesundheitssysteme, weil er global getrieben wird. Die deutsche Gesundheitspolitik muss sich überlegen, wie sie

dazu beitragen kann, die Chancen des medizinisch-technischen Fortschritts für die Menschen optimal zu nutzen.

Literaturverzeichnis

- Ackerknecht, E. und A. Murken (1992). *Geschichte der Medizin*. Stuttgart: Enke.
- Aghion, P. und P. Howitt (1998). *Endogenous Growth Theory*. Cambridge MA: MIT Press.
- Albert, U.-S., M. Koller, W. Lorenz, K.-D. Schulz (2003). Report über die Entwicklung der Leitlinie Brustkrebs-Früherkennung in Deutschland: methodisches Vorgehen, Erkenntnisse und Implikationen. *Gesundheitsökonomie und Qualitätsmanagement* 8: 39–51.
- Alexopoulos, E., G. D. Dounias und K. Vemmos (1999). Medical Diagnosis of Stroke using Inductive Machine Learning. In: *Proceedings of the Workshop on Machine Learning in Medical Applications, Advanced Course on Artificial Intelligence (ACAI 99)*. Chania, Greece. <http://www.iit.demokritos.gr/skel/eetn/acai99/Workshops.htm>.
- Antioch, K. M. und M. K. Walsh (2004). The Risk-adjusted Vision beyond Casemix (DRG) Funding in Australia, International Lessons in High Complexity and Capitation. *European Journal of Health Economics* 5: 95–109.
- Baetjer, H. Jr. (1998). *Software as Capital: An Economic Perspective on Software Engineering*. Los Alamitos CA: IEEE Computer Society Press.
- Balas, E. A. und S. A. Boren (2000). *Managing Clinical Knowledge for Health Care Improvement. Yearbook of Medical Informatics*. Bethesda: National Library of Medicine.
- Becker, G. S. und K. M. Murphy (1992). The Division of Labor, Coordination Costs, and Knowledge. *Quarterly Journal of Economics* 107 (4): 1137–1160.
- (2000). *Social Economics*. Cambridge MA: Harvard University Press.
- Berger, M., B. Richter und J. Mühlhauser (1997). Evidence-based Medicine: Eine Medizin auf rationaler Grundlage. *Der Internist* 38: 344–351.
- Bernstein, J. (1997). Test-indication Curves. *Medical Decision Making* 17: 103–106.
- Booch, G. (1994). *Object-oriented Analysis and Design*. Menlo Park: Addison-Wesley.
- Börchers, K., B. Driller, A. Lang und G. Aleff (2003). Brustkrebs in Deutschland: Ohne Qualitätsmanagement geht es nicht. *Gesundheitsökonomie und Qualitätsmanagement* 8: 332–333.
- Bourlas, P., E. Giakoumakis und G. Papkonstantinou (1999). A Knowledge Acquisition and Management System for ECG Diagnosis. In: *Proceedings of the Workshop on Machine Learning in Medical Applications, Advanced Course on Artificial Intelligence (ACAI 99)*. Chania, Greece. <http://www.iit.demokritos.gr/skel/eetn/acai99/Workshops.htm>.
- Breyer, F. und P. Zweifel (2003). *Gesundheitsökonomie*, 4. Auflage. Berlin, Heidelberg: Springer.

- Cheeseman, P., J. Kelly, M. Self, J. Stutz, W. Taylor und D. Freeman (1988). AutoClass: A Bayesian classification system. *Proceedings of the Fifth International Conference on Machine Learning*. Ann Arbor MI: Morgan Kaufmann.
- Cheng, Y. und K. Fu (1985). Conceptual Clustering in Knowledge Organization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7: 592–598.
- Cohen, A. B. (1997). Managed Care und die Einführung und Verbreitung medizinischer Technologien. In: M. Arnold, K. W. Lauterbach und K.-J. Preuß (Hrsg.), *Managed Care: Ursachen, Prinzipien, Formen und Effekte*. Stuttgart: Schattauer.
- Cutler, D. M., R. S. Huckman und M. B. Landrum (2004). The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. NBER Working Paper No. 10489. Cambridge MA.
- Daniel, B. L. und T. M. Daniel (1993). Graphic Representation of Numerically Calculated Predictive Values. *Medical Decision Making* 13: 355–358.
- Daniels, M. und S. A. Schroeder (1977). Variations among Physicians in Use of Laboratory Tests II: Relation to Clinical Productivity and Outcomes of Care. *Medical Care* 15 (6): 482–487.
- Das, M. (2002). Is there Evidence Against the Induced Demand Hypothesis? Explaining the Large Reduction in Cesarean Rates. Columbia University Department of Economics Discussion Paper Series No. 0102-40.
- Dasgupta, P. und P. A. David (1994). Toward a New Economics of Science. *Research Policy* 23: 487–521.
- Detmer, D. E. (2003). Building the National Health Information Infrastructure for Personal Health, Health Care Services, Public Health, and Research. *BMC Medical Informatics and Decision Making* 3: 1. <http://www.biomedcentral.com/1472-6947/3/1>.
- Deutsches Institut für medizinische Dokumentation und Information (1994). Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision. Bern, Göttingen: Hans Huber.
- DiMasi, J. A. (2003). The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics* 22: 151–185.
- Eddy, D. M. (1990a). Anatomy of a Decision. *JAMA – Journal of the American Medical Association* 263 (3): 441–443.
- (1990b). Practice Policies – What Are They? *JAMA* 263 (6): 877–880.
- (1990c). Practice Policies: Where Do They Come From? *JAMA* 263 (9): 1265–1275.
- (1990d). Practice Policies – Guidelines for Methods. *JAMA* 263 (13): 1839–1841.
- (1990e). Guidelines for Policy Statements: The Explicit Approach. *JAMA* 263 (16): 2239–2243.

- (1990f). Comparing Benefits and Harms: The Balance Sheet. *JAMA* 263 (18): 2493–2505.
- (1990g). Designing a Practice Policy. *JAMA* 263 (22): 3077–3084.
- (1990h). Resolving Conflicts in Practice Policies. *JAMA* 264 (3): 389–391.
- (1990i). What Do We Do About Costs? *JAMA* 263 (9): 1161–1170.
- Eisenberg, J. E. (2002). Globalize the Evidence, Localize the Decision: Evidence-Based Medicine and International Diversity. *Health Affairs* 21 (3): 166–168.
- Ellis, R. P. (1998). Creaming, Skimping, and Dumping: Provider Competition on the Intensive and Extensive Margins. *Journal of Health Economics* 17(5): 537–555.
- Erdmann, E. (2003). Die Positivliste – Ein Beitrag zum Qualitätsmanagement? *Gesundheitsökonomie und Qualitätsmanagement* 8: 1.
- Estes, G. W. und R. M. Mandel (1995). Sexuell übertragbare Krankheiten. In: G. Sachse (Hrsg.), *Medizinische Entscheidungen: vom Symptom zur Diagnose*. Berlin, Wiesbaden.
- Europarat, Ministerkomitee, Resolution (77) 30 (1978): Der Allgemeinpraktiker, seine Ausbildung und Wege zur Weckung seines Selbstbewusstseins. *Internationale Allgemeinmedizin und Hochschule* 9: 771–776.
- Everitt, B. (1993). Cluster Analysis. London: Heinemann.
- Evidence-Based Medicine Working Group (1992). Evidence-based medicine: A New Approach to Teaching the Practice of Medicine. *JAMA* 268: 2420–2425.
- Fagan, T. J. (1975). Nomogram for Bayes's Theorem. *The New England Journal of Medicine* 293: 257.
- Fisher, D. H., M. J. Pazzani und P. Langley (1991). Concept Formation: Knowledge and Experience in Unsupervised Learning. San Mateo CA: Morgan Kaufmann.
- Fisher, D. H. (1987). Knowledge Acquisition Via Incremental Conceptual Clustering. *Machine Learning* 2: 139–172.
- Folland, S., A. Goodman und M. Stano (1993). The Economics of Health and Health Care. New York: Macmillan.
- Fuchs, V. (1974). Who Shall Live? Health, Economics, and Social Choice. New York: Basic Books.
- Gerdtham, U.-G. und B. Jönsson (2000). International Comparison of Health Expenditure. In: A. Culyer und J. Newhouse (Hrsg.), *Handbook of Health Economics* Vol. 1A. Amsterdam: Elsevier.
- Getzen, T. (1997). Health Economics: Fundamentals and Flow of Funds. New York: John Wiley & Sons.
- Glasziou, P. (1991). Threshold Analysis via the Bayes Nomogram. *Medical Decision Making* 11: 61–62.
- Glied, S. (2000). Managed Care. In: A. Culyer und J. Newhouse (Hrsg.), *Handbook of Health Economics* Vol. 1A. Amsterdam: Elsevier.

- Goddeeris, J. H. (1980). *Insurance, Technology, and Medical Expenditures: A Study of the Interactions*. University of Wisconsin – Madison.
- Greene, H. L., W. P. Johnson und M. J. Maricic (1995). *Medizinische Entscheidungen: vom Symptom zur Diagnose*. Deutsche Ausgabe herausgegeben von G. Sachse. Berlin, Wiesbaden: Ullstein Mosby.
- Gruber, J., und M. Owings (1996). Physician Financial Incentives and Cesarean Section Delivery. *Rand Journal of Economics* 23: 99–123.
- Harhoff, D., J. Henkel und E. von Hippel (2003). Profiting from Voluntary Information Spillovers: How Users Benefit by Freely Revealing their Innovations. *Research Policy* 32: 1753–1769.
- Hasenbein, U., C.-W. Wallesch und J. Rübiger (2003). Ärztliche Compliance mit Leitlinien. *Gesundheitsökonomie und Qualitätsmanagement* 8: 363–375.
- Heimpel, H. (2003). Zur Bedeutung nationaler und lokaler Leitlinien für ärztliche Entscheidungen. *Medizinische Klinik* 98: 226–230.
- Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems*. Cambridge MA: The MIT Press.
- Iba, W., und P. Langley (2001). Unsupervised Learning in Probabilistic Concept Hierarchies. In: G. Paliouras, V. Karkaletsis und C. D. Spyropoulos (Hrsg.), *Proceedings of the Workshop on Machine Learning in Medical Applications, Advanced Course on Artificial Intelligence (ACAI 99)*. Berlin, Heidelberg: Springer.
- Institute of Medicine (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington DC: National Academy Press.
- Kamber, M., L. Winstone, W. Gong, S. Cheng und J. Han (1996). *Generalization and Decision Tree Induction: Efficient Classification in Data Mining*. Mimeo: Simon Fraser University.
- Klimpe, D. (2002). Das modulare Großkrankenhaus – Ein Weg in die Zukunft mit einem DRG-optimierten Betriebskonzept. *Betriebswirtschaftliche Forschung und Praxis* 2: 118–129.
- Knoper, S. R. (1995). Husten. In: G. Sachse (Hrsg.), *Medizinische Entscheidungen: vom Symptom zur Diagnose*. Berlin, Wiesbaden.
- Lakeoff, G. und M. Johnson (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Langley, P. (1987). Machine Learning and Concept Formation. *Machine Learning* 2: 99–102
- Lauterbach, K. W. (1997). Zum Verhältnis von Disease Management und Managed Care. In: M. Arnold, K. W. Lauterbach und K.-J. Preuß (Hrsg.), *Managed Care: Ursachen, Prinzipien, Formen und Effekte*. Stuttgart: Schattauer.
- Lerner, J. und J. Tirole (2002). Some Simple Economics of Open Source. *The Journal of Industrial Economics* 50 (2): 197–233.
- Linde, K. (1999). Gibt es gesicherte Therapien in der Homöopathie? *Der Internist* 40: 1271–1274.

- Lohmann, H. und O. Bornemeier (2002). DRGs als Chance für ein wettbewerbsorientiertes Gesundheitssystem. *Gesundheitsökonomie und Qualitätsmanagement* 7: 304–309.
- Magoulas, G. D., und A. Prentza (1999). Machine Learning in Medical Applications. In: G. Paliouras, V. Karkaletsis und C. D. Spyropoulos (Hrsg.), *Proceedings of the Workshop on Machine Learning in Medical Applications, Advanced Course on Artificial Intelligence (ACAI 99)*. Berlin, Heidelberg: Springer.
- McGuire, T. (2000). Physician Agency. In A. J. Culyer und J.-P. Newhouse (Hrsg.), *Handbook of Health Economics*. Amsterdam.
- Mehta, M., R. Agrawal und J. Rissanen (1996). SLIQ: A Fast Scalable Classifier for Data Mining. In: *Proceedings of the International Conference on Extending Database Technology*. Avignon, France, March 1996.
- Meltzer, D., F. L. Hiltz und D. Bates (2000). Implications of Managed Care for Teaching Hospitals. In: Cutler, D. M. (Hrsg.), *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*. Chicago: University of Chicago Press.
- Meyer, R. L. (2002). Definitionen der Allgemeinmedizin im Laufe der Zeit. *Schweizerische Ärztezeitung* 83 (4): 2681–2690.
- Michalski, R. S. (1983). A Theory and Methodology of Learning from Examples. In: R. S. Michalski, J. G. Carbonell und T. M. Mitchell (Hrsg.), *Machine Learning: An Artificial Intelligence Approach*. Los Altos CA: Morgan Kaufmann.
- Michalski, R. und R. Stepp (1983). Learning from Observation: Conceptual Clustering. In: R. S. Michalski, J. G. Carbonell und T. M. Mitchell (Hrsg.), *Machine Learning: An Artificial Intelligence Approach*. Los Altos CA: Morgan Kaufmann.
- Michel, E., M. Stiletto, W. Gründler und M. Walka (2002). Diagnosencodierung und Erlös im AR-DRG-System. *Gesundheitsökonomie und Qualitätsmanagement* 7: 292–296.
- Mohr, V. D. (2002). Versorgungsqualität im DRG-System – Die Rolle des externen Qualitätsvergleichs. *Gesundheitsökonomie und Qualitätsmanagement* 7: 310–315.
- Monti, S., P. Tamayo, I. Mesirov und T. Golub (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning* 52: 91–118.
- Morris, S. und H. S. Shin (2002). Social Value of Public Information. *American Economic Review* 92 (5): 1521–1534.
- Moustakis, V. und G. Charissis (1999). Machine Learning and Medical Decision Making. In: *Proceedings of the Workshop on Machine Learning in Medical Applications, Advanced Course on Artificial Intelligence (ACAI 99)*. Chania, Greece. <http://www.iit.demokritos.gr/skel/eetn/acai99/Workshops.htm>.

- National Academy of Sciences (1999). *To Err Is Human*. Washington, D.C.
- Neumann, P. J. , K. Claxton und M. C. Weinstein (2000). The FDA's Regulation of Health Economics Information. *Health Affairs* September/Oktober: 129–137.
- Newhouse, J.-P. (1981). The Erosion of the Medical Market Place. In: R. Scheffler (Hrsg.), *Advances in Health Economics and Health Services Research*, Vol. 2. Westport CT: JAI Press.
- Nyman, J. und R. Maude-Griffin (2001). The Welfare Economics of Moral Hazard. *International Journal of Health Care Finance and Economics* 1: 23–42.
- OECD (2004). *Towards High-Performing Health Systems*. Paris.
- Pauker, S. G. und J. P. Kassirer (1980). The Threshold Approach to Clinical Decision Making. *The New England Journal of Medicine* 302: 1109–1117.
- (1987). Decision Analysis. *The New England Journal of Medicine* 316: 250–258.
- Pauly, M. (1968). The Economics of Moral Hazard: Comment. *American Economic Review* 58: 531–536.
- Phelps, C. (1992). Diffusion of Information in Medical Care. *Journal of Economic Perspectives* 6 (3): 23–42.
- (2000). Information Diffusion and Best Practice Adoption. In A. Culyer und J. Newhouse (Hrsg.), *Handbook of Health Economics*. Amsterdam.
- Phelps, C. und C. Mooney (1992). Priority Setting for Medical Technology and Medical Practice Assessment: Correction and Update. *Medical Care* 31 (8, August).
- (1993). Variations in Medical Practice Use: Causes and Consequences. In R. Arnould, R. Rich und W. White (Hrsg.), *Competitive Approaches to Health Care Reform*. Washington D.C.
- Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge.
- Prasad, B. (1996). *Concurrent Engineering Fundamentals Vol. I: Integrated Product and Process Organization*. Upper Saddle River NJ: Prentice Hall.
- (1997). *Concurrent Engineering Fundamentals Vol. II: Integrated Product Development*. Upper Saddle River NJ: Prentice Hall.
- Propping, P. und S. Aretz (2004). Die genetische Revolution: Bedeutung für Therapie und Prävention. *Der Internist* 45 (Supp. 1): S6–S14.
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning* 1: 81–106.
- (1993). *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann.
- Reng, C.-M., H.-J. Friedrich, A. Timmer und J. Schölmerich (2003). Fachinformationen für Ärztinnen und Ärzte in Deutschland: Akzeptanz, Qualität und Verfügbarkeit von Fachinformationen unter besonderer Berücksichtigung der elektronischen Informationsmedien. *Medizinische Klinik* 98: 648–655.

- Richter, K. und S. Lange (1997). Methoden der Diagnoseevaluierung. *Der Internist* 38: 325–336.
- Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy* 98 (5): S71–S102.
- Roos, N. P. (1984). Hysterectomy: Variations in Rates across Small Areas and across Physicians' Practices. *American Journal of Public Health* 74 (4): 327–335.
- Rosenberg, C. E. (2002). The Tyranny of Diagnosis: Specific Entities and Individual Experience. *The Milbank Quarterly* 80 (2): 237–260.
- Rothschild, M. und J. Stiglitz (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics* 90 (4): 630–649.
- Rychlik, R. (1999). Gesundheitsökonomie und Krankenhausmanagement. Stuttgart: Kohlhammer.
- Sackett, D. L., S. E. Straus, W. S. Richardson, W. Rosenberg und R. B. Haynes (2000). Evidence-Based Medicine: How to Practice and Teach EBM. Churchill Livingstone.
- Silvey, D. (1975). Statistical Inference. London: Chapman and Hall.
- Schell, H. und K. W. Lauterbach (2002). Evaluation, Benchmarking, Qualitätsmanagement und Zertifizierung als Instrument für eine evidenzbasierte Gesundheitspolitik. In: K.-J. Preuß, J. Rübiger und J. H. Sommer (Hrsg.), *Managed Care: Evaluation und Performance-Measurement integrierter Versorgungsmodelle*. Stuttgart: Schattauer.
- Schmid, A., O. Moerer, A. Neilson, H. Burchadi und H. Schneider (2004). Die Einführung von neuen Therapien in der Intensivbehandlung: klinische Vorteile bei zusätzlichen Kosten. *Gesundheitsökonomie und Qualitätsmanagement* 9: 93–96.
- Schnurrer, J. U. und J. C. Frölich (2003). Zur Häufigkeit und Vermeidbarkeit von tödlichen unerwünschten Arzneimittelwirkungen. *Der Internist* 44: 889–895.
- Smythe, J. G. (2002). Reputation, Public Information, and Physician Adoption of an Innovation. *European Journal of Health Economics* 3: 103–110.
- Somers, H. M. und A. R. Somers (1961). Doctors, Patients, and Health Insurance. Garden City, NY: Doubleday Anchor Books.
- Sox, H., M. Blatt, M. Higgins und K. Marton (1988). Medical Decision Making. Stoneham: Butterworth Publishers.
- Stausberg, J. (2002). Gestaltung von Diagnosen- und Prozedurenklassifikationen in einem DRG-System. *Gesundheitsökonomie und Qualitätsmanagement* 7: 297–303.
- Tassey, G. (1992). Technology Infrastructure and Competitive Position. Dordrecht, Kluwer.
- Teece, D. (1988). Technological Change and the Nature of the Firm. In: G. Dosi (Hrsg.) *Technical Change and Economic Theory*. London: Pinter Publishers.

- Tönies, H. (1993). *Entscheidungen in der Allgemeinpraxis – Die Medizin der Symptome*. Wien: Springer.
- Vayda, E. (1973). Comparison of Surgical Rates in Canada, England, and Wales. *New England Journal of Medicine* 289: 1224–1228.
- Vera, A (2003). Neue Organisationsstrukturen in deutschen Krankenhäusern nach der DRG-Einführung. *Gesundheitsökonomie und Qualitätsmanagement* 8: 25–31.
- Weiss, S. M. und C. A. Kulikowski (1991). *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann.
- Wendt, J. K. (1995). Akuter und Chronischer Kopfschmerz. In: G. Sachse (Hrsg.), *Medizinische Entscheidungen: vom Symptom zur Diagnose*. Berlin, Wiesbaden.
- Wennberg, J. E. (1996). On the Appropriateness of Small-Area Analysis for Cost Containment. *Health Affairs* 15: 164–167.
- (2003). The More Things Change...: The Federal Government's Role in the Evaluative Sciences. *Health Affairs – Web Exclusives*. 25 June 2003.
- , E. S. Fisher und J. S. Skinner (2002). Geography and the Debate over Medicare Reform. *Health Affairs – Web Exclusives*. 13 February 2002.
- Wennberg, D. E. und J. E. Wennberg (2003). Addressing Variations: Is There Hope for the Future? *Health Affairs – Web Exclusives*. 10 December 2003.
- Wennberg, J. E., J. L. Freeman und W. J. Culp (1987). Are Hospital Services Rationed in New Haven or Over-Utilised in Boston? *Lancet* 1: 1185–1189.
- Zweig, M. H. und G. Campbell (1993). Receiver-Operating Characteristics (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry* 39: 561–577.

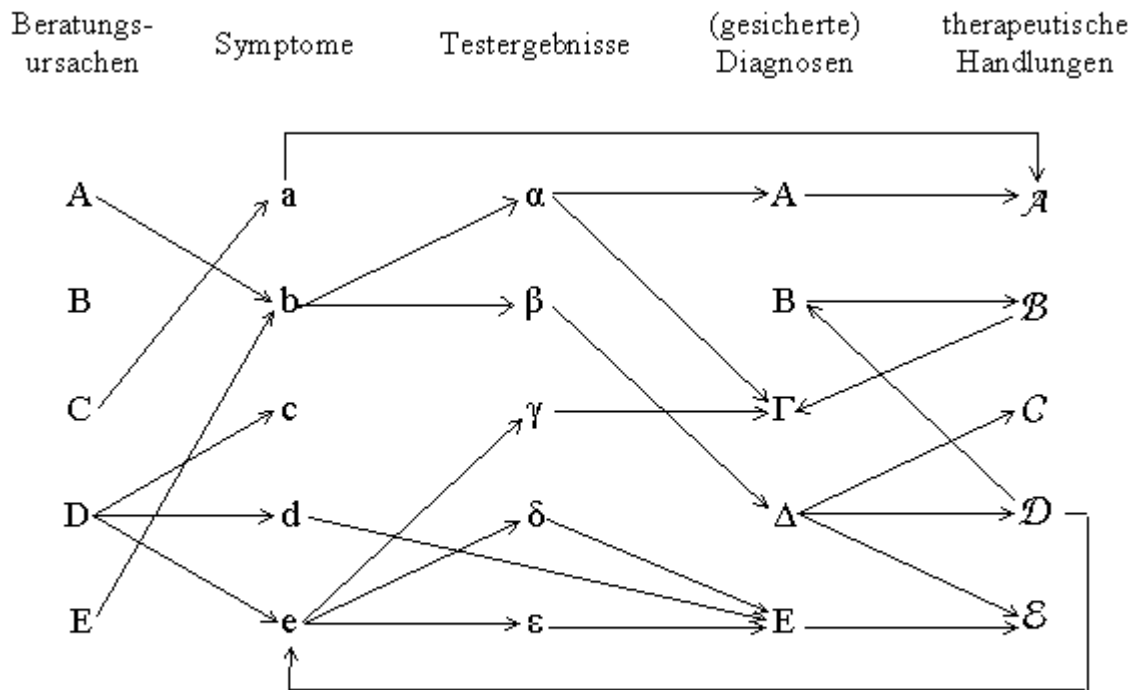
Schaubild 1: Diagnose als Zuordnungsproblem

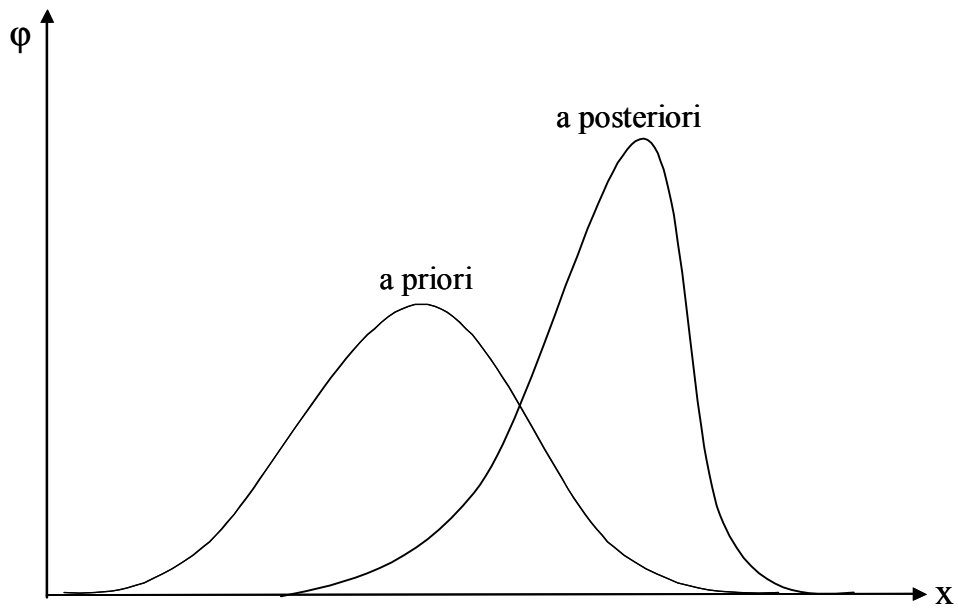
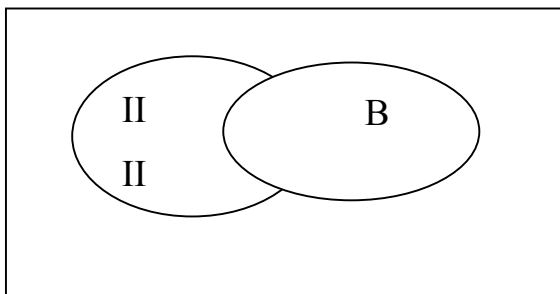
Schaubild 2a: A priori und a posteriori Verteilung**Schaubild 2b:** Vennsches Mengendiagramm

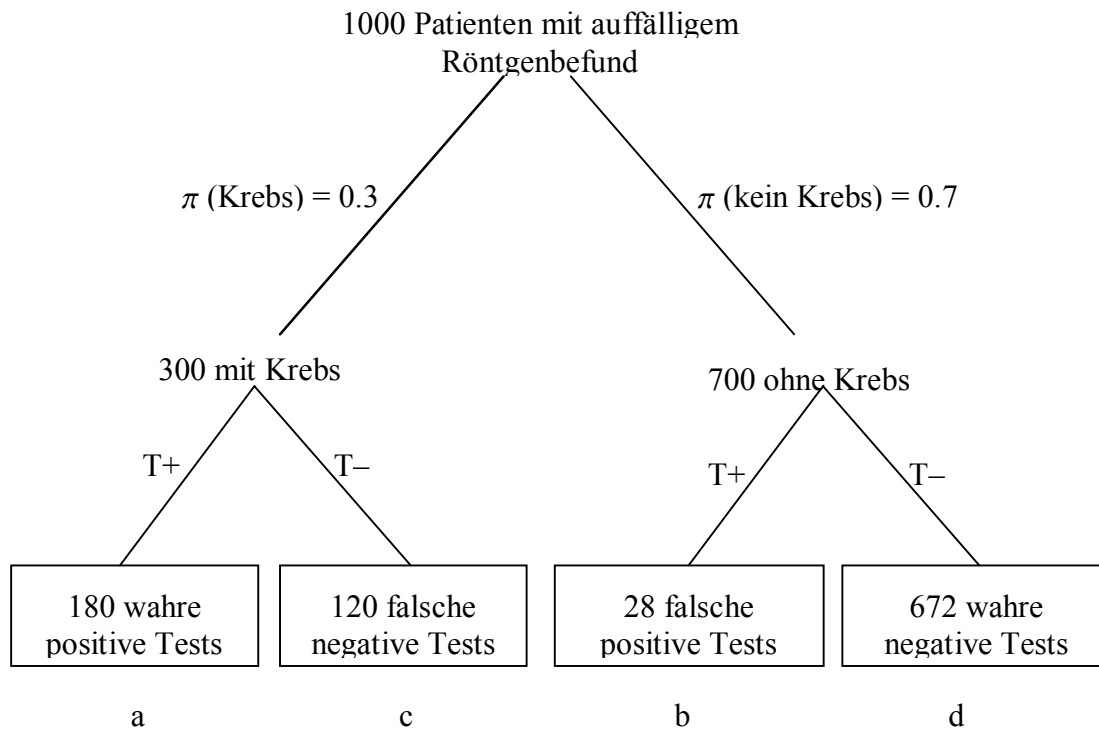
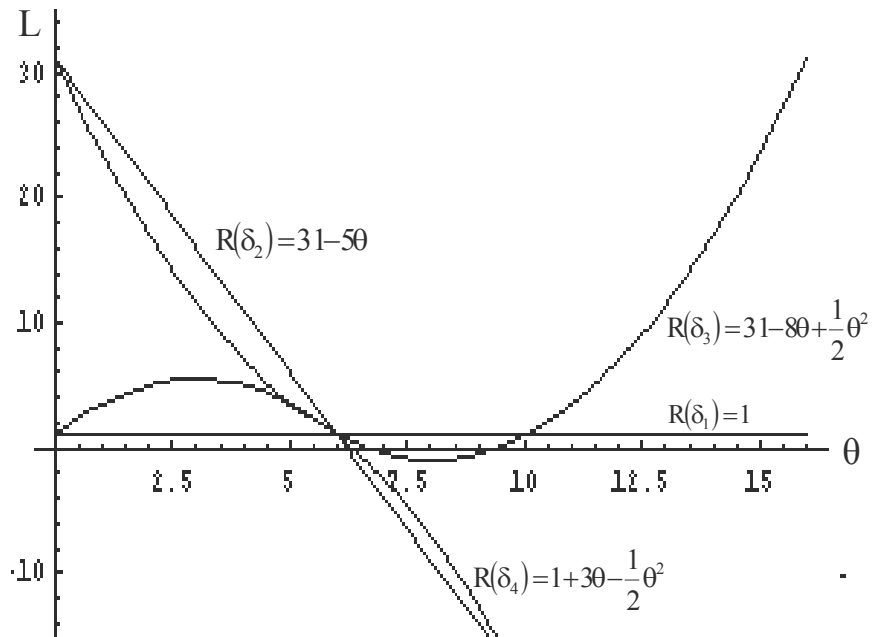
Schaubild 3: Baumdarstellung des Bayes Theorems

Schaubild 4: Risikofunktionen alternativer Entscheidungsfunktionen

Quelle: Diese Darstellung beruht auf den in Silvey (1975), S. 163 ff., gegebenen Beispielen.

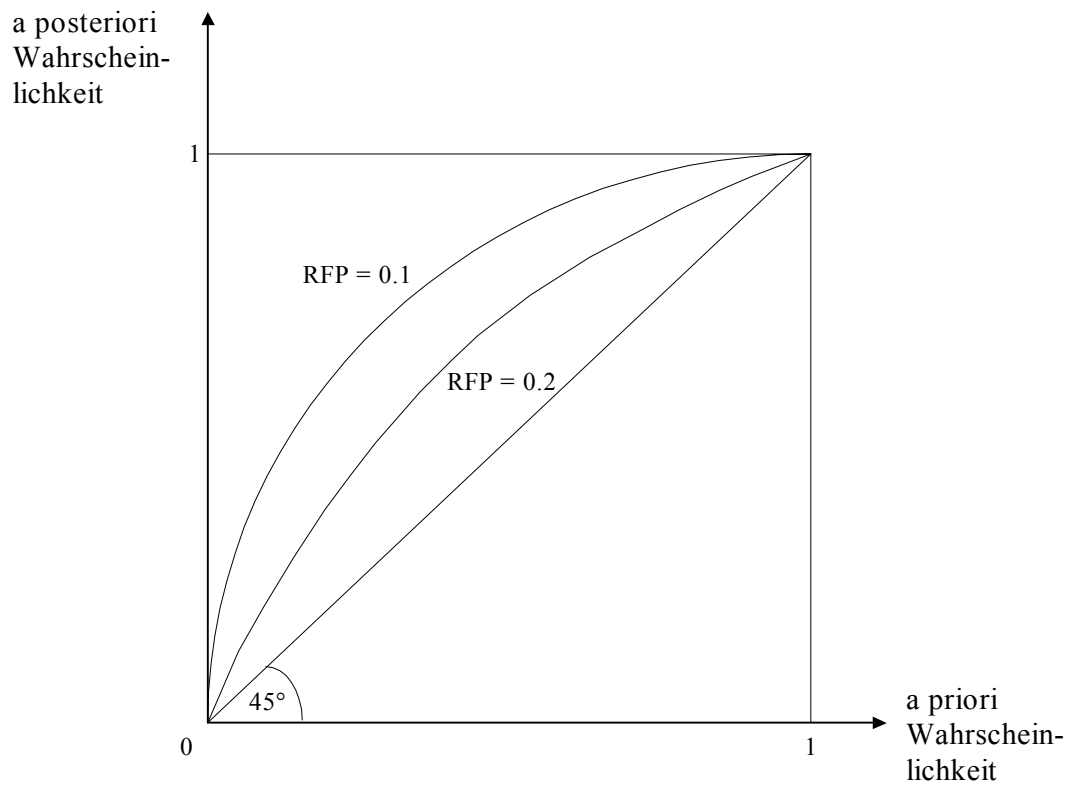
Schaubild 5: Diagnostischer Test mit positivem Ergebnis

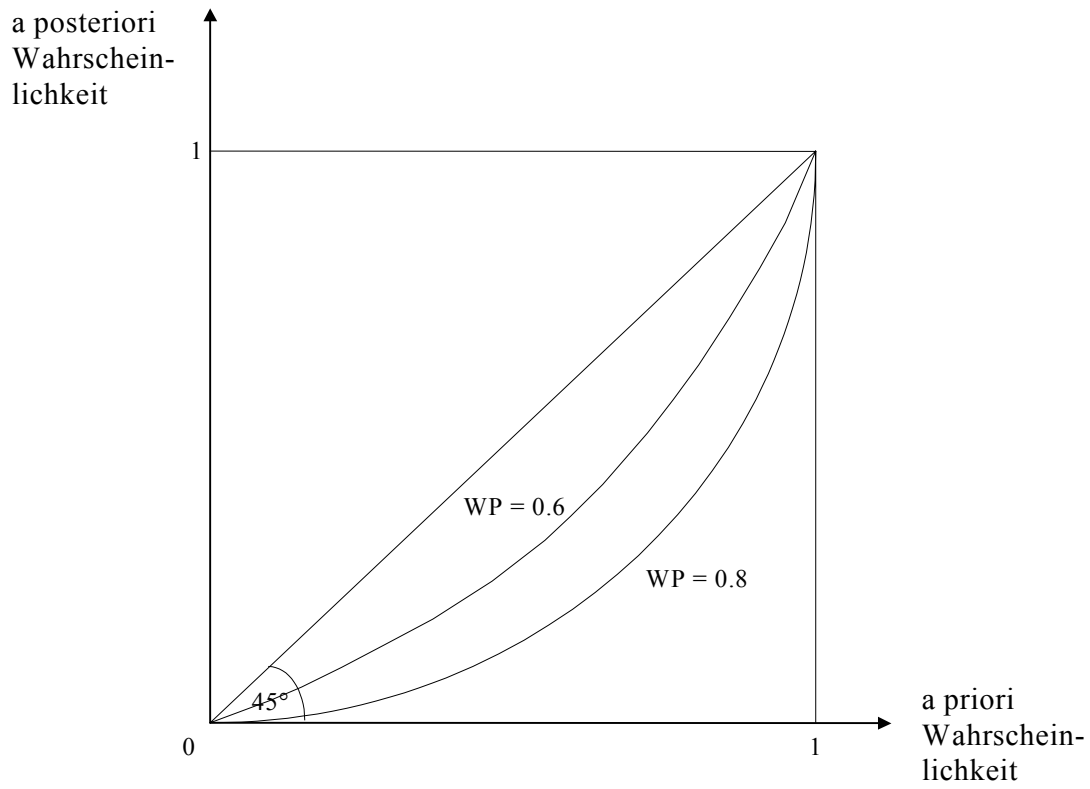
Schaubild 6: Diagnostischer Test mit negativem Ergebnis

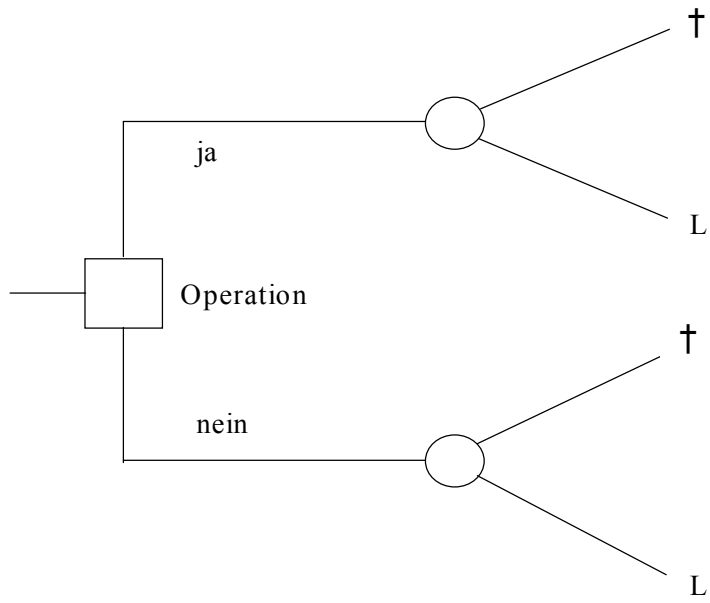
Schaubild 7: Medizinischer Entscheidungsbaum

Schaubild 8: Schwellenwahrscheinlichkeit für eine Behandlung und das Kosten-Nutzen-Verhältnis

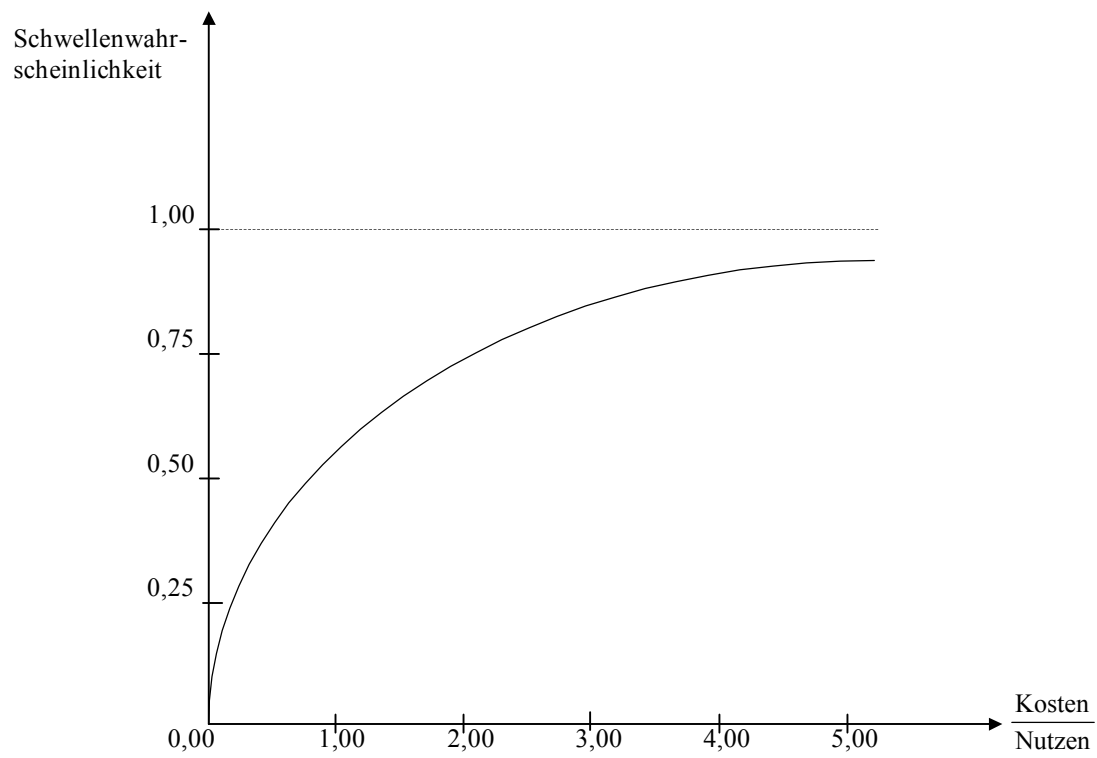
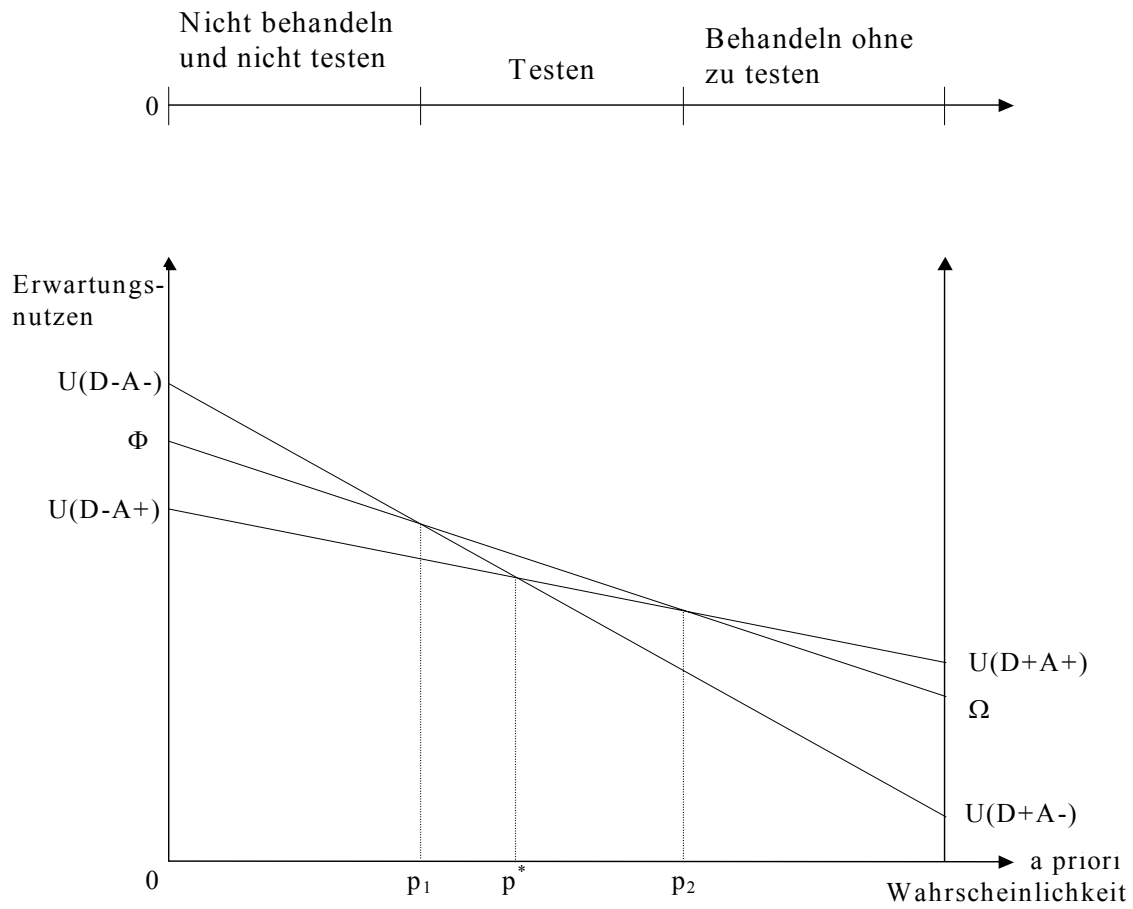


Schaubild 9: Entscheidung über die Durchführung eines diagnostischen Tests

$$\Phi = \text{RFP} \cdot U(\text{D-A}+) + (1-\text{RFP}) \cdot U(\text{D-A-}) + U(\text{T})$$

$$\Omega = \text{SEN} \cdot U(\text{D+A}+) + (1-\text{SEN}) \cdot U(\text{D+A-}) + U(\text{T})$$

Schaubild 10: Die Abhängigkeit der Durchführung eines Tests von seiner Sensitivität und der Rate falscher positiver Ergebnisse

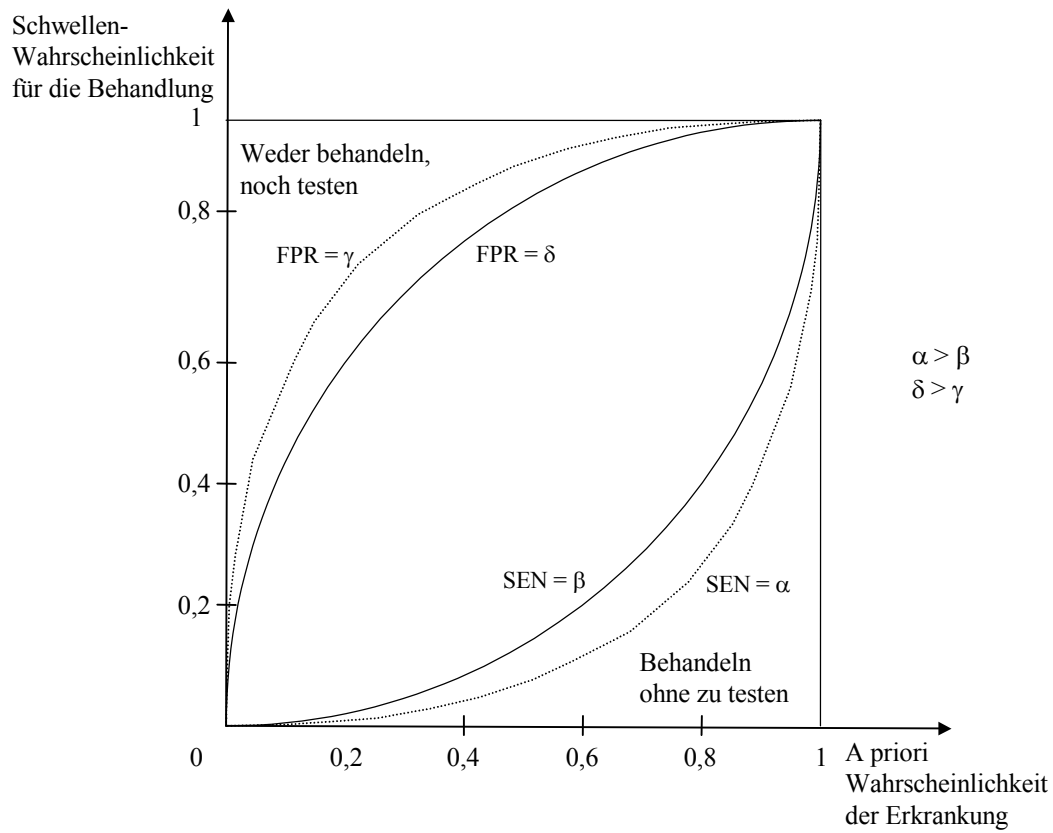
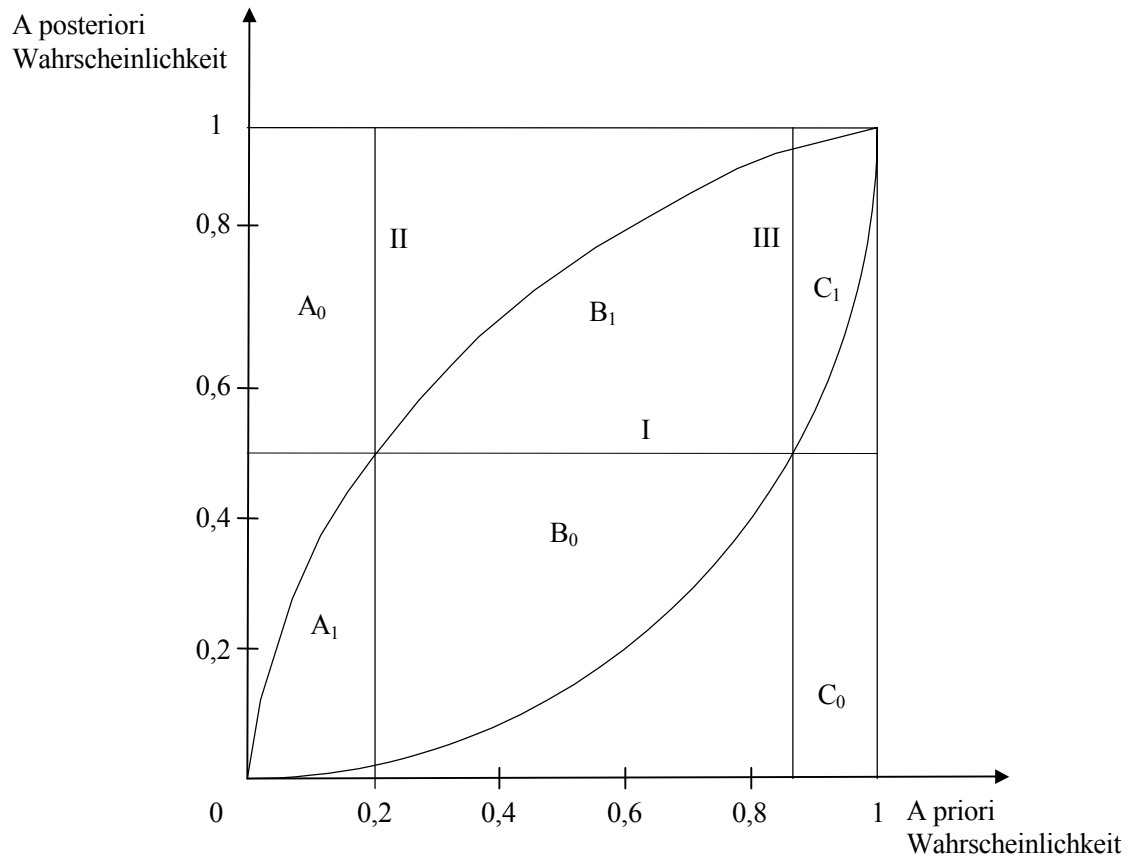


Schaubild 11: Die Interdependenz von Test- und Behandlungsentscheidung

Quelle: In Anlehnung an Bernstein (1997).

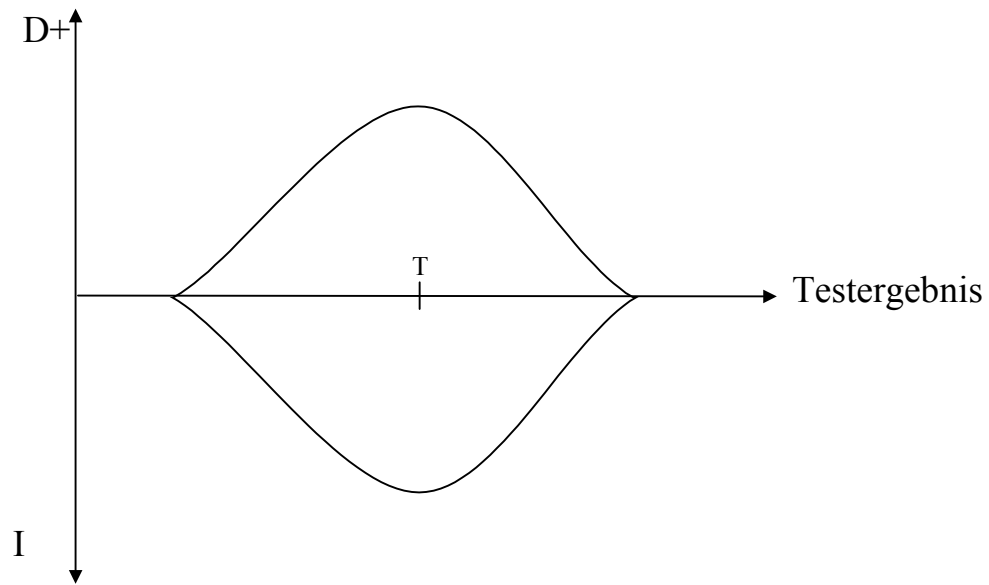
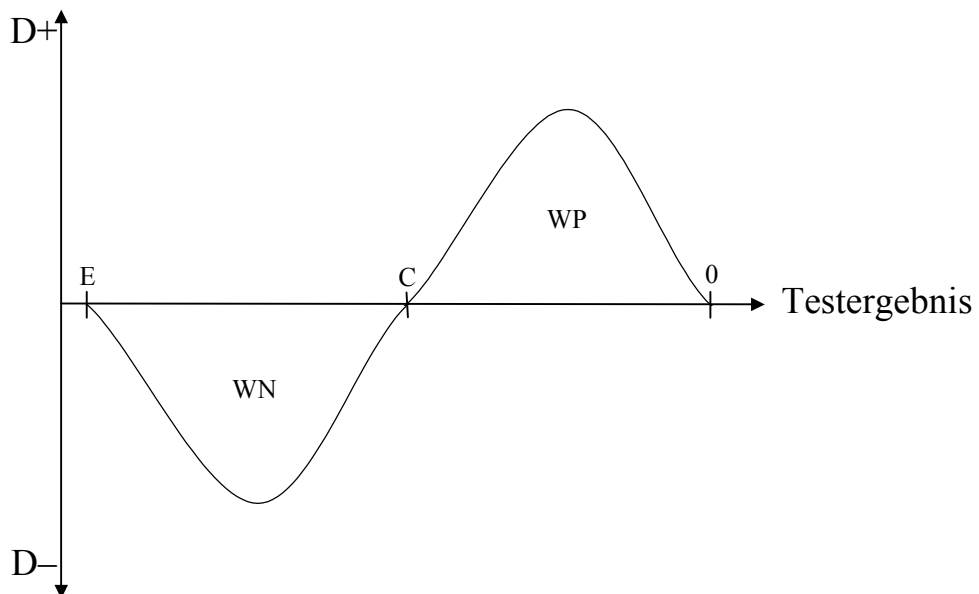
Schaubild 12a: Uninformativer Test**Schaubild 12b:** Perfekter Test

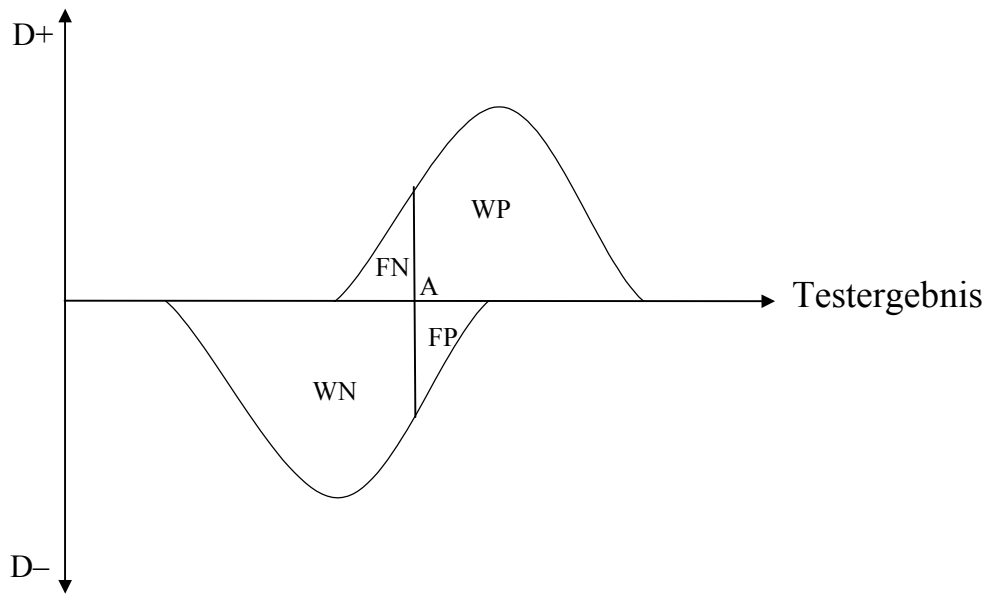
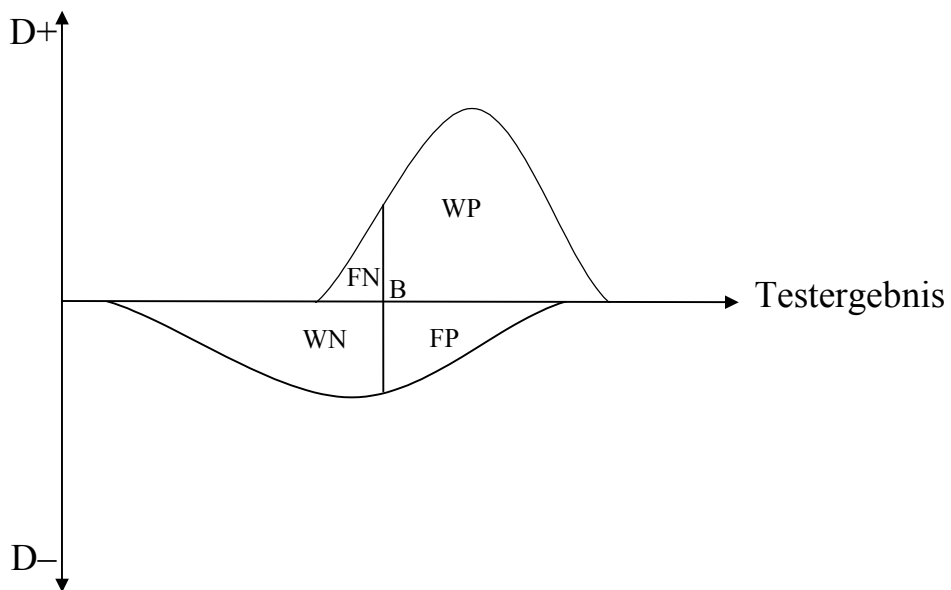
Schaubild 12c: Typischer Test für eine Reihenuntersuchung**Schaubild 12d:** Typischer Test zu Bestätigung einer Hypothese

Schaubild 12e: Receiver Operating Characteristics (ROC) für zwei unterschiedliche medizinische Testverfahren

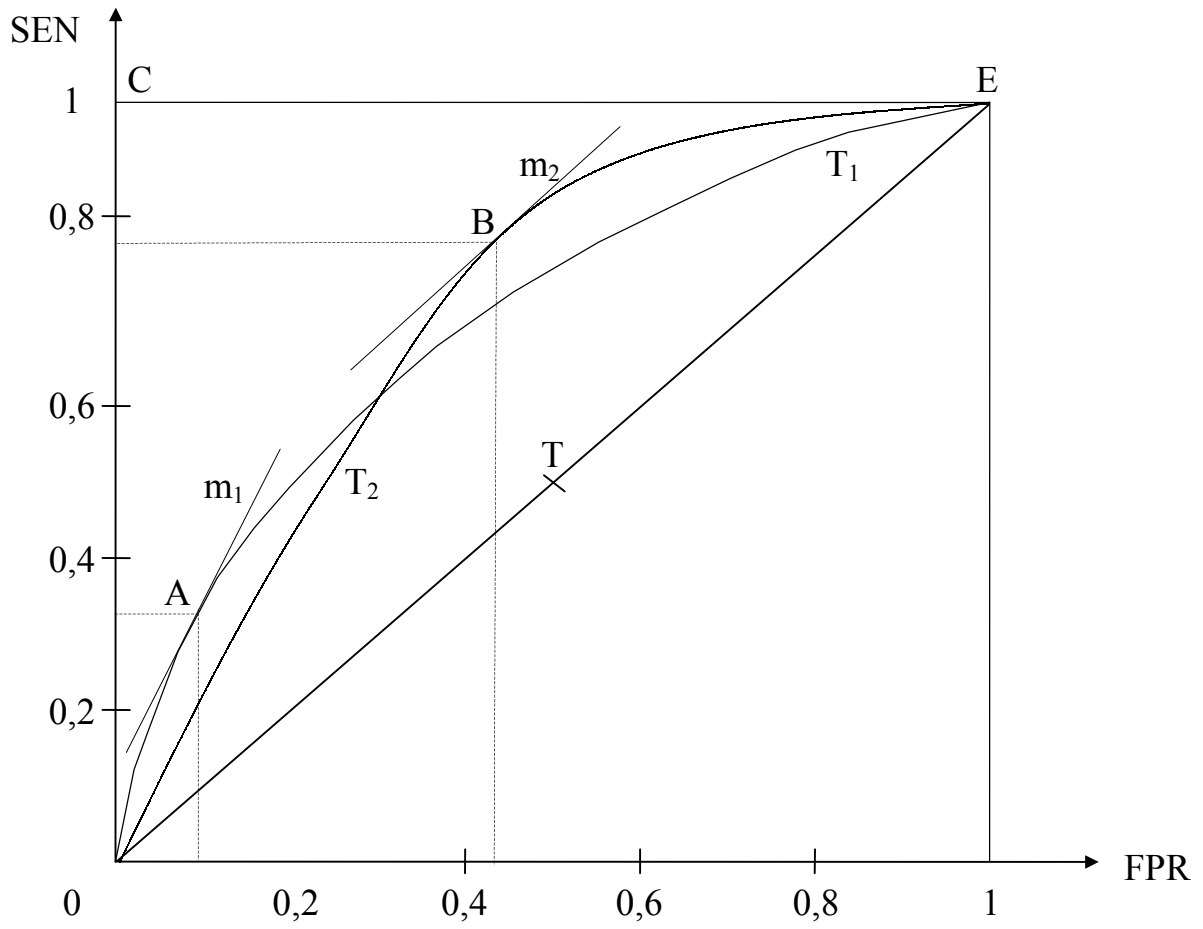
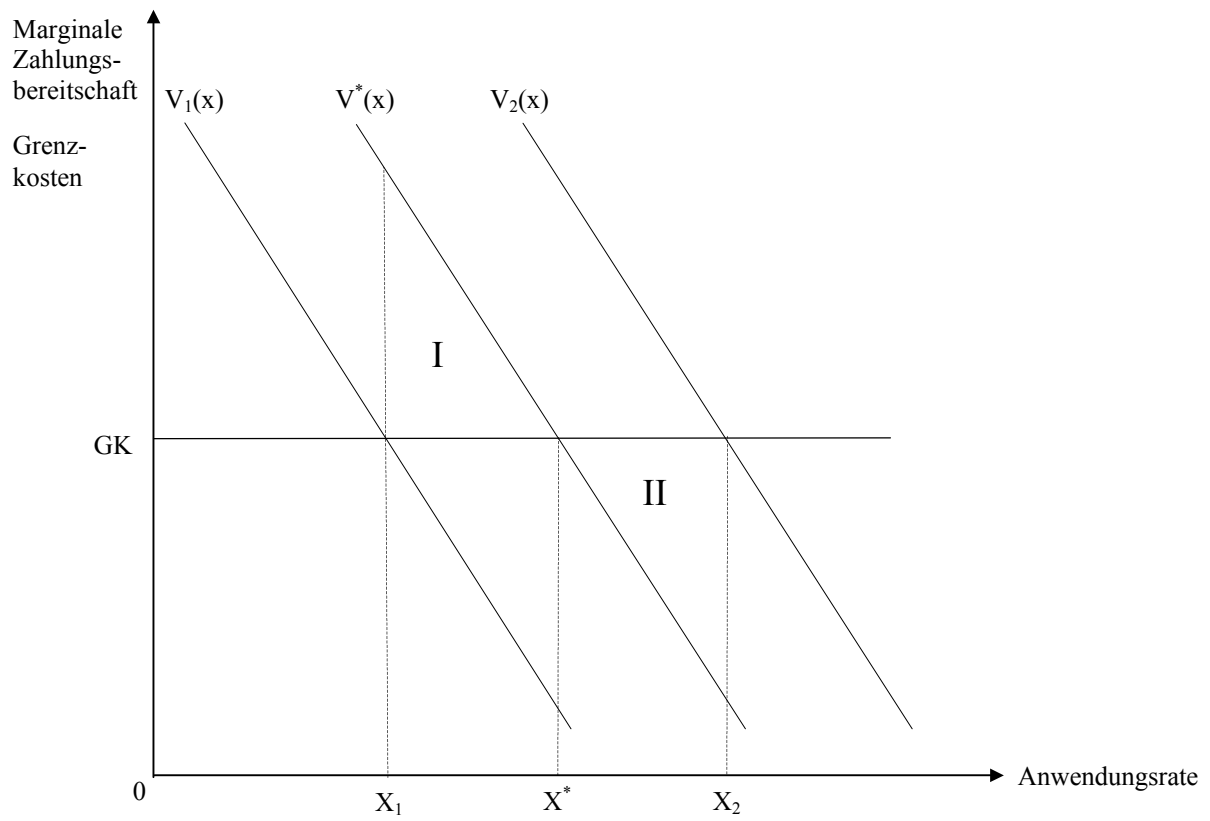
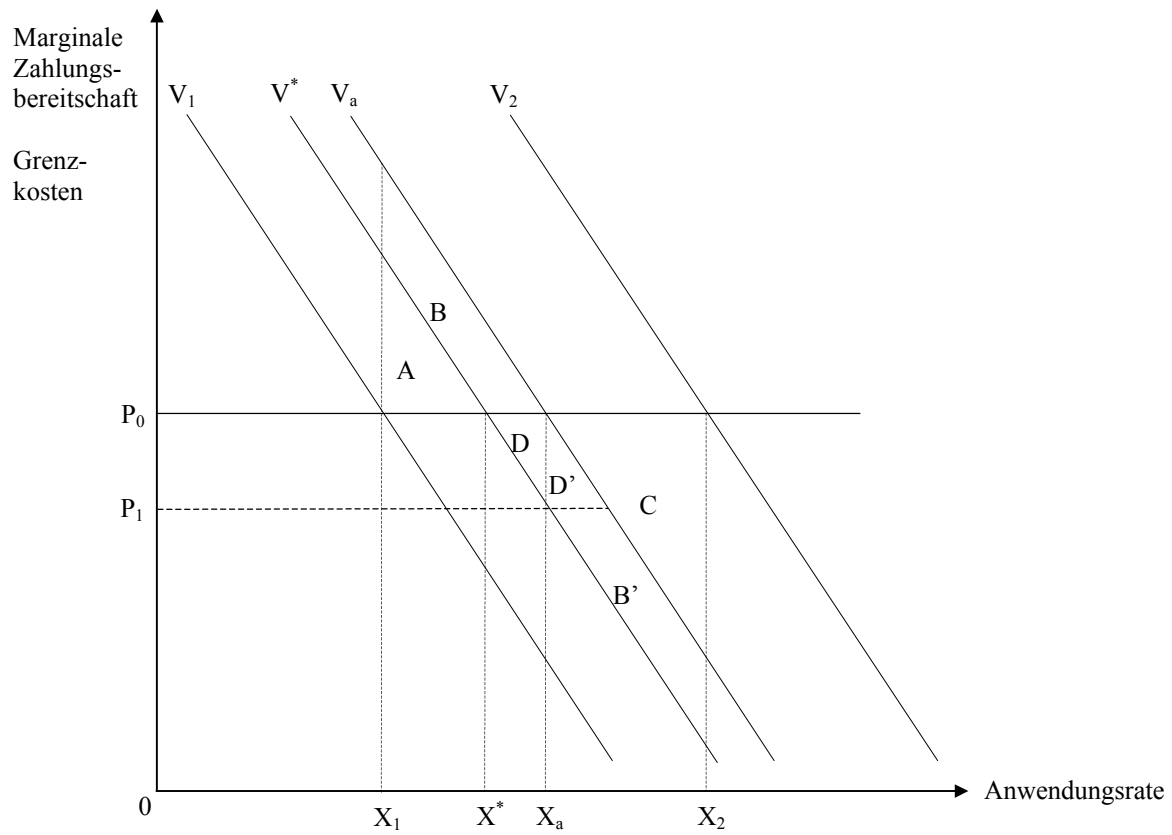


Schaubild 13a: Effizienzverluste durch medizinische Praxisvariationen

Quelle: In Anlehnung an Phelps und Mooney (1993), S. 163.

Schaubild 13b: Medizinische Praxisvariationen und Moral Hazard

Quelle: In Anlehnung an Phelps und Mooney (1993), S. 165.

Schaubild 13c: Medizinische Praxisvariationen bei unzureichender Diffusion einer neuen Medizintechnologie

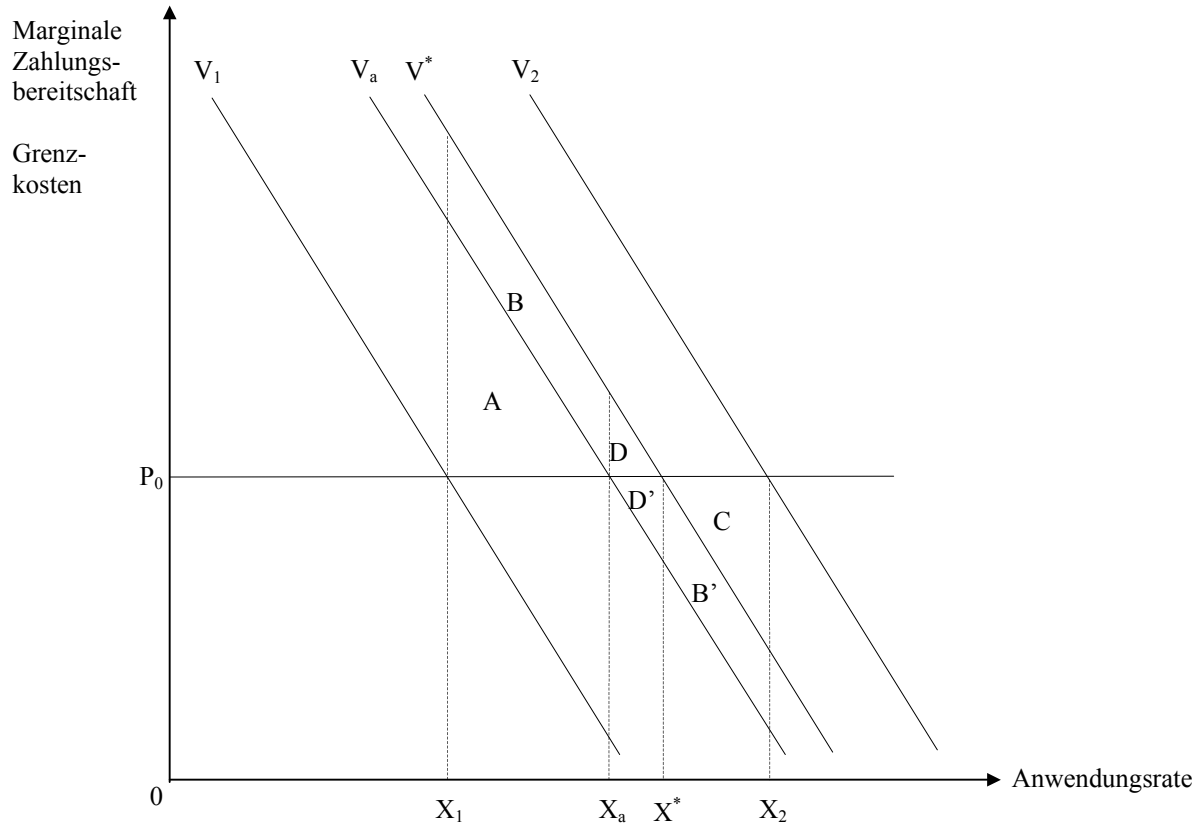


Schaubild 13d: Medizinische Praxisvariationen und Monopolpreisbildung

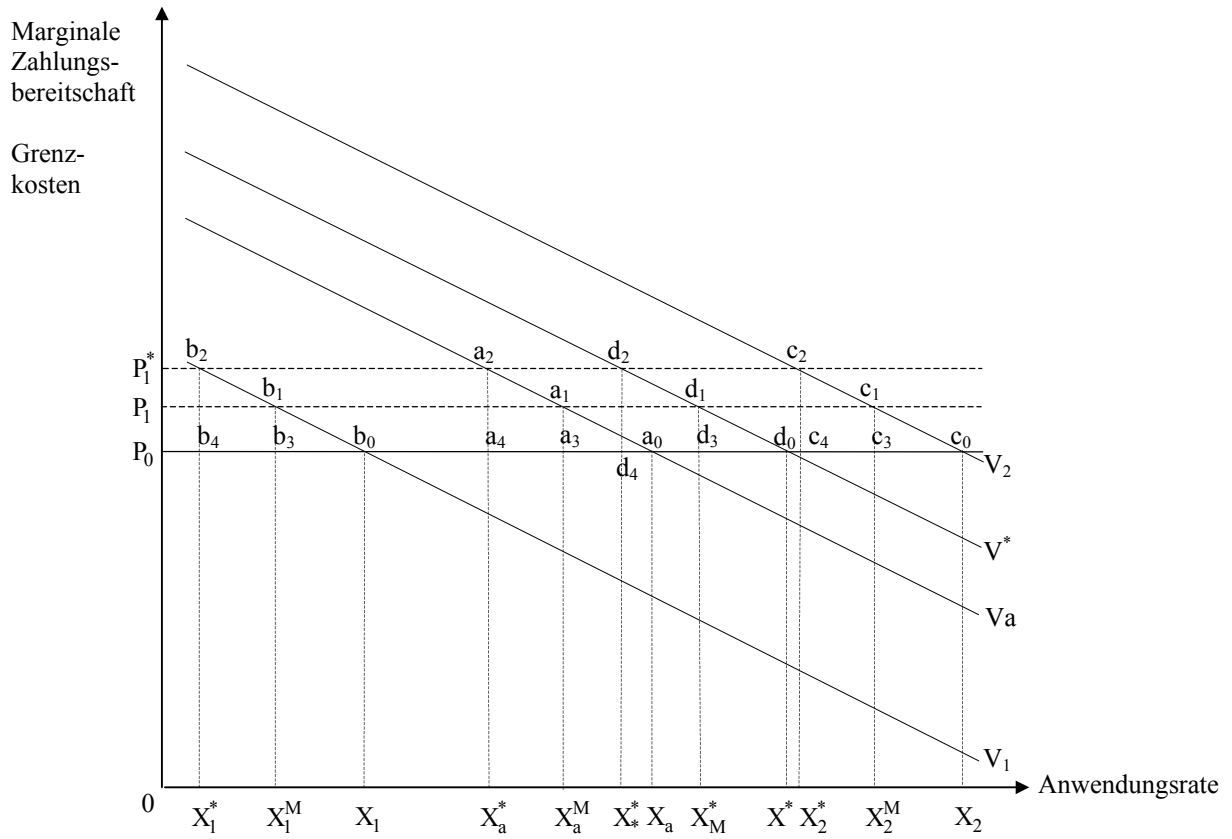


Schaubild 14: Angebot und Nachfrage neuer medizinischer Information

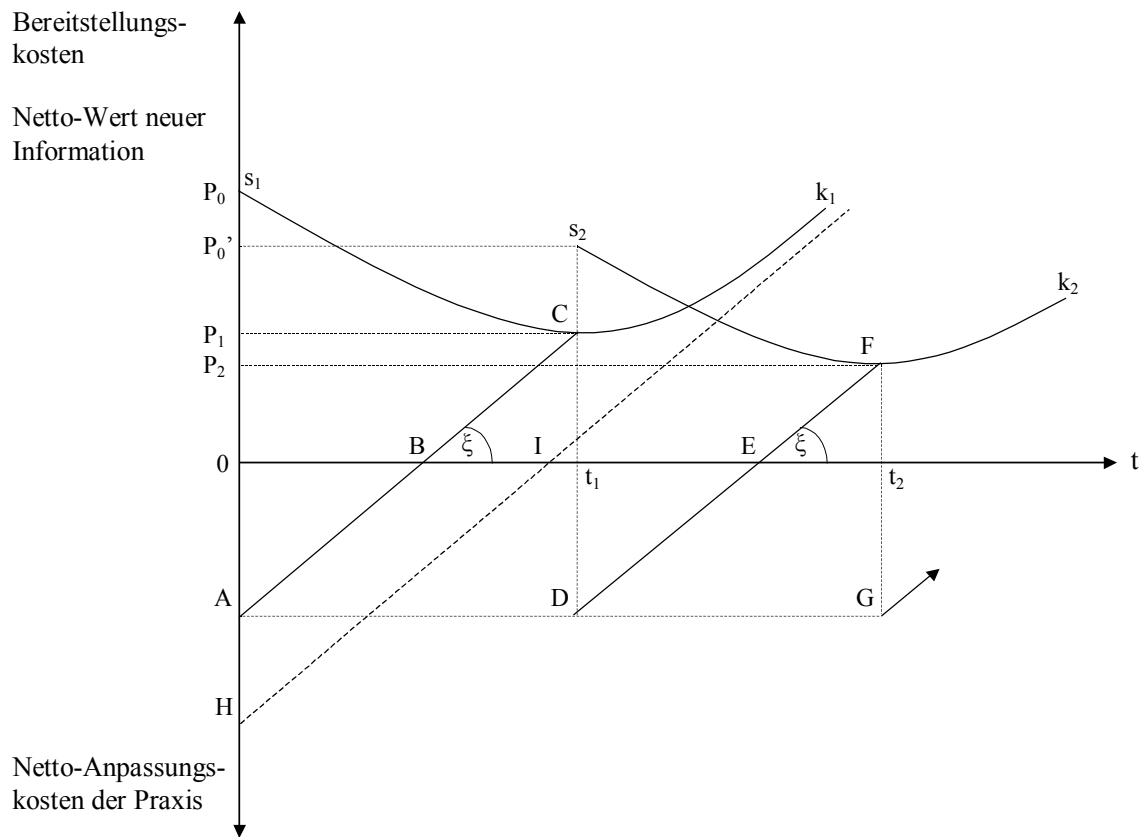


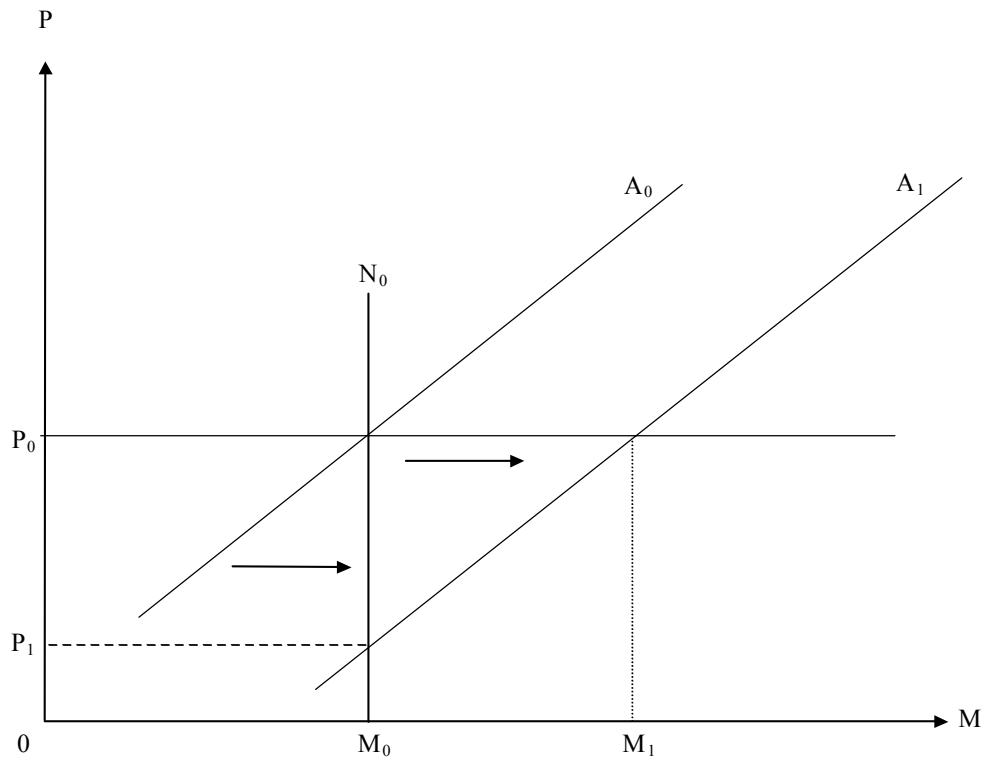
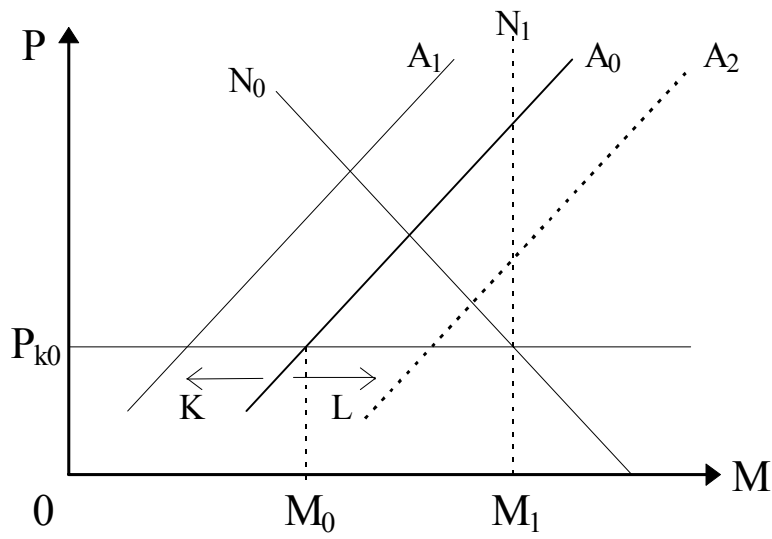
Schaubild 15: Anbieterinduzierte Nachfrage

Schaubild 16: Nachfrageüberhang

K: Kurzfristiges Ausweichen der etablierten Ärzte mit einem Teil ihrer Leistungen auf den „Schwarzmarkt“.

L: Langfristiger Markteintritt neuer Ärzte.

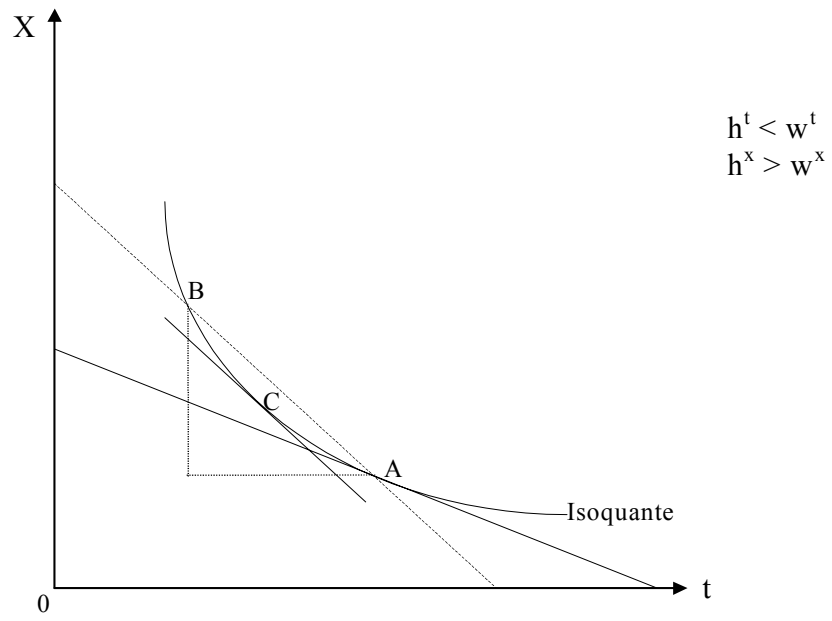
Schaubild 17: Kostenerstattung

Schaubild 18: Einzelleistungsvergütung