

UDO SCHNEIDER

**THEORIE UND EMPIRIE
DER ARZT-PATIENT-
BEZIEHUNG**



UDO SCHNEIDER

THEORIE UND EMPIRIE DER ARZT-PATIENT- BEZIEHUNG

Auf den Märkten für medizinische Leistungen und in der Beziehung zwischen Arzt und Patient existieren Informationsasymmetrien in einer Vielzahl, wie sie in kaum einem anderen Bereich einer Ökonomie anzutreffen sind. In der Arbeit wird die beidseitige Informationsasymmetrie zwischen Arzt und Patient untersucht. Zum einen geschieht dies hinsichtlich des Einflusses von Änderungen der Vertragsparameter auf die medizinische Leistung und das Verhalten des Patienten. Zum anderen betrifft dies die Wahl der Versicherungs- und Vergütungsparameter. Ergänzend wird die Arzt-Patient-Beziehung mittels einer empirischen Analyse untersucht. In deren Mittelpunkt stehen die Determinanten der Gesundheitszufriedenheit des Patienten und dabei insbesondere die Inanspruchnahme medizinischer Leistungen.

Udo Schneider wurde 1971 in Worms geboren und studierte von 1990 bis 1996 Volkswirtschaftslehre an der Universität Mannheim. Von 1997 bis 2001 arbeitete er als wissenschaftlicher Mitarbeiter am Lehrstuhl für Allgemeine Volkswirtschaftslehre, insbesondere Finanzwissenschaft an der Universität Greifswald. Zur Zeit ist er als wissenschaftlicher Assistent am Lehrstuhl für Volkswirtschaftslehre III, insbesondere Finanzwissenschaft der Universität Bayreuth tätig.

Theorie und Empirie der Arzt-Patient-Beziehung

ALLOKATION IM MARKTWIRTSCHAFTLICHEN SYSTEM

Herausgegeben von
Heinz König, Hans-Heinrich Nachtkamp,
Ulrich Schlieper, Eberhard Wille

Band 47



PETER LANG

Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien

UDO SCHNEIDER

THEORIE UND EMPIRIE DER ARZT-PATIENT- BEZIEHUNG

Zur Anwendung
der Principal-Agent-Theorie
auf die Gesundheitsnachfrage



PETER LANG
Europäischer Verlag der Wissenschaften

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Schneider, Udo:

**Theorie und Empirie der Arzt-Patient-Beziehung : zur
Anwendung der Principal-Agent-Theorie auf die
Gesundheitsnachfrage / Udo Schneider. - Frankfurt am Main ;
Berlin ; Bern ; Bruxelles ; New York ; Oxford ; Wien : Lang, 2002
(Allokation im marktwirtschaftlichen System ; Bd. 47)
Zugl.: Greifswald, Univ., Diss., 2002
ISBN 3-631-39759-3**

Open Access: The online version of this publication is published on www.peterlang.com and www.econstor.eu under the international Creative Commons License CC-BY 4.0. Learn more on how you can use and share this work: <http://creativecommons.org/licenses/by/4.0>.



This book is available Open Access thanks to the kind support of ZBW – Leibniz-Informationzentrum Wirtschaft.

**Gedruckt auf alterungsbeständigem,
säurefreiem Papier.**

9

ISSN 0939-7728

ISBN 3-631-39759-3

ISBN 978-3-631-75606-5 (eBook)

© Peter Lang GmbH

Europäischer Verlag der Wissenschaften

Frankfurt am Main 2002

Alle Rechte vorbehalten.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Printed in Germany 1 2 3 4 5 7

www.peterlang.de

Vorwort

Das deutsche Gesundheitswesen und dessen Finanzierungsprobleme stehen seit Jahren im Zentrum der politischen Diskussion. Trotz zahlreicher Reformansätze ist es in der Vergangenheit jedoch nicht gelungen, die Probleme in diesem Bereich zu lösen. Die Fehlentwicklungen im Gesundheitswesen sind in vielerlei Hinsicht auf Steuerungsmängel und falsche Anreizstrukturen zurückzuführen. Die vorliegende Arbeit greift Anreizstrukturen in der Beziehung zwischen Arzt und Patient auf und analysiert diese auf Grundlage eines informationsökonomischen Modells.

Die Arbeit entstand während meiner Tätigkeit am Lehrstuhl für Allgemeine Volkswirtschaftslehre, insbesondere Finanzwissenschaft der Ernst-Moritz-Arndt-Universität Greifswald und wurde im Wintersemester 2001/2002 von der Rechts- und Staatswissenschaftlichen Fakultät als Dissertation angenommen.

Mein besonderer Dank gilt zunächst Herrn Prof. Dr. Volker Ulrich, der während der gesamten Zeit meinem Forschungsvorhaben aufgeschlossen gegenüberstand und wichtige Hinweise zu dessen Umsetzung lieferte. Dem Zweitgutachter der Arbeit, Herrn Prof. Dr. Armin Rohde, danke ich besonders für seine Unterstützung während der Entstehung der Arbeit. Für wertvolle Hinweise, Anregungen und kritische Anmerkungen danke ich Herrn Dr. Stefan Heng, Herrn Dipl.-Volkswirt Paul Marschall, Frau Dipl.-Kffr. Brit Svenja Albers sowie den Teilnehmern des Forschungsseminars des Bereichs Wirtschaftswissenschaften der Universität Greifswald.

Greifswald, im März 2002

Udo Schneider

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Fehlentwicklungen und Anreize im Gesundheitswesen	1
1.2	Aufbau der Arbeit.....	4
2	Die Arzt-Patient-Beziehung im deutschen Gesundheitswesen	7
2.1	Das System der Gesundheitssicherung in Deutschland	7
2.2	Die finanzielle Situation des Gesundheitswesens	10
2.3	Folgerungen für die Beziehung zwischen Arzt und Patient.....	20
3	Informationsökonomische Grundlagen	23
3.1	Problematik	23
3.2	Adverse Selektion.....	24
3.3	Moral Hazard.....	27
3.4	Die Grundstruktur von Principal-Agent Problemen	32
4	Die Anwendung der Principal-Agent-Theorie auf das Gesundheitswesen.....	43
4.1	Die These der anbieterinduzierten Nachfrage.....	43
4.2	Die Arzt-Patient-Beziehung als Principal-Agent-Modell	47
4.2.1	Die Übertragung eines Standardansatzes	47
4.2.2	Modellkritik	55
4.3	Die Agency-Theorie im Rahmen der Informationsbeziehungen im Gesundheitswesen	56
4.3.1	Informationsbeziehungen und Anreize	56
4.3.2	Anforderungen des Gesundheitswesens an Agency-Modelle.....	59
4.3.3	Interdependenzen und Anreize	64
4.4	Kritische Würdigung der Agency-Theorie im Gesundheitswesen	67
5	Forschungsansätze über das Verhältnis Arzt - Patient - Versicherung	69
5.1	Übersicht	69
5.2	Optimale Vergütungs- und Versicherungssysteme.....	72
5.3	Die Rolle des Arztes als doppelter Agent	79
5.4	Optimale Krankenversicherungsverträge und die Entlohnung ärztlicher Leistungen	85

5.5	Die Vergütung ärztlicher Leistungen bei unbeobachtbarem Aufwand des Arztes und des Patienten	91
6	Double Moral Hazard zwischen Arzt und Patient	99
6.1	Zusammenhang zwischen ärztlicher Leistung und Compliance des Patienten	99
6.2	Kontakte zwischen Arzt und Patient	101
6.3	Simultane versus sequentielle Entscheidungen	102
6.4	Modellstufen des Double Moral Hazard-Problems	106
6.5	Das theoretische Modell der Behandlung	107
6.5.1	Grundstruktur	107
6.5.2	Kooperative Lösung	110
6.5.3	Nichtkooperative Lösung	112
6.5.4	Vergleich beider Lösungen	114
6.6	Das Verhältnis zwischen Compliance des Patienten und ärztlicher Behandlungsleistung	115
6.6.1	Substitutions- versus Komplementaritätsbeziehung	115
6.6.2	Resultate ohne Selbstbeteiligung	117
6.6.3	Zur Wirkung einer Selbstbeteiligung	122
6.7	Die Vertragsstufe	127
6.7.1	First-best-Lösung	130
6.7.2	Lösung bei asymmetrischer Information	132
6.8	Kritische Würdigung des Double Moral Hazard-Problems	135
6.A	Anhang zu Kapitel 6	137
7	Empirische Analyse der Arzt-Patient-Beziehung	139
7.1	Umsetzung theoretischer Befunde	139
7.2	Überblick über empirische Studien	143
7.3	Ökonometrische Untersuchung anhand von Querschnittsdaten	148
7.3.1	Einführung in Modelle diskreter abhängiger Variablen	148
7.3.2	Modelle für geordnete Kategorien	157
7.3.3	Daten	165
7.3.4	Schätzergebnisse	172
7.4	Ökonometrische Untersuchung von Paneldaten	181
7.4.1	Einführung in die Panel-Analyse	181

7.4.2	Panel-Modelle für binäre Variablen	184
7.4.3	Schätzverfahren für Panel-Modelle mit ordinalen abhängigen Variablen	186
7.4.4	Daten.....	189
7.4.5	Schätzergebnisse.....	193
7.5	Kritische Würdigung	200
8	Zusammenfassung und Ausblick.....	202
	Literaturverzeichnis	205

Abbildungsverzeichnis

Abbildung 2.1:	Die Hauptbeteiligten im deutschen Gesundheitswesen	9
Abbildung 2.2:	Gesundheitsausgaben 1998 nach Ausgabenträgern in Mrd. DM.....	12
Abbildung 2.3:	Gesundheitsausgaben 1998 nach Leistungsarten in Mrd. DM.....	14
Abbildung 3.1:	Zeitliche Struktur des Problems adverser Selektion	25
Abbildung 3.2:	Zeitliche Struktur bei Moral Hazard.....	28
Abbildung 3.3:	Abgrenzungen des Moral Hazard.....	29
Abbildung 3.4:	Probleme der Lösung des First-order-approach	36
Abbildung 3.5:	Gewinn- und Risikoaufteilung bei Risikoaversion	40
Abbildung 4.1:	Anbieterinduzierte Nachfrage als Reaktion auf eine Angebotszunahme	45
Abbildung 4.2:	Zusammenhang zwischen ärztlicher Leistung und Gesundheitszustand	49
Abbildung 4.3:	Optimale Honorierung bei gegebenen Werten von θ	54
Abbildung 4.4:	Honorierung bei verändertem Gesundheitszustand.....	54
Abbildung 4.5:	Akteure im Gesundheitswesen	58
Abbildung 4.6:	Informationsbeziehungen im ambulanten Sektor.....	59
Abbildung 4.7:	Ein dreiteiliges Modell des Gesundheitsprozesses.....	66
Abbildung 5.1:	Spielstufen des Modells von Ma und McGuire.....	87
Abbildung 6.1:	Simultane Entscheidungen unter unvollständiger Information	104
Abbildung 6.2:	Sequentielle Entscheidungen ohne Unsicherheit	105
Abbildung 6.3:	Stufen des Modells	107
Abbildung 6.4:	Unabhängigkeit von Compliance und ärztlicher Leistung...	118
Abbildung 6.5:	Compliance und medizinische Leistung bei strategischen Komplementen.....	119
Abbildung 6.6:	Gleichgerichtete Effekte bei strategischen Substituten.....	120
Abbildung 6.7:	Gegenläufige Effekte bei strategischen Substituten.....	121
Abbildung 6.8:	Unabhängigkeit und Selbstbeteiligung.....	123

Abbildung 6.9: Strategische Komplemente und Selbstbeteiligung.....124
 Abbildung 6.10: Strategische Substitute und erhöhte Genesungswahrscheinlichkeit.....125
 Abbildung 6.11: Strategische Substitute und verminderte Genesungswahrscheinlichkeit.....126
 Abbildung 7.1: Bestimmungsfaktoren des Gesundheitszustands.....140
 Abbildung 7.2: Lineares Wahrscheinlichkeitsmodell für eine einzelne erklärende Variable151
 Abbildung 7.3: Zusammenhang von latenter und beobachteter Variable153

Tabellenverzeichnis

Tabelle 2.1 Struktur der Einnahmen und Ausgaben in der gesetzlichen Krankenversicherung in Mrd. DM16
 Tabelle 2.2: Leistungsausgaben für die Mitglieder der gesetzlichen Krankenversicherung.....18
 Tabelle 7.1: Variablenbeschreibung der Querschnittsanalyse166
 Tabelle 7.2: Erwarteter Einfluß der erklärenden Variablen (Querschnittsdatensatz)170
 Tabelle 7.3: Deskriptive Statistik des Querschnittsdatensatzes171
 Tabelle 7.4: Ergebnisse der Querschnittsschätzung für die Gesamtstichprobe und die Nichterwerbstätigen.....175
 Tabelle 7.5: Ergebnisse der Querschnittsschätzung für Erwerbstätige180
 Tabelle 7.6: Variablenbeschreibung des Paneldatensatzes192
 Tabelle 7.7: Deskriptive Statistik des Paneldatensatzes.....193
 Tabelle 7.8: Schätzergebnisse für die Gesamtstichprobe und die Nichterwerbstätigen (Panel-Analyse)196
 Tabelle 7.9: Schätzergebnisse Erwerbstätige (Panel-Analyse)199

1 Einleitung

1.1 Fehlentwicklungen und Anreize im Gesundheitswesen

Das deutsche Gesundheitswesen und dabei insbesondere die gesetzliche Krankenversicherung (GKV) steht nach knapp drei Jahrzehnten an Reformbemühungen weiterhin im Zentrum des wissenschaftlichen und politischen Interesses. Während in den siebziger und achtziger Jahren noch vornehmlich die Kostendämpfung auf der Ausgabenseite im Mittelpunkt stand, wird seit den neunziger Jahren der Einnahmenseite und den Anreizstrukturen im Gesundheitswesen größere Beachtung beigemessen. Ein Grund für diese Gewichtsverlagerung dürfte die relative Erfolglosigkeit der zahlreichen gesetzlichen Eingriffe auf der Ausgabenseite der GKV sein. Diese konnten den mittelfristigen Wachstumstrend der Gesundheitsausgaben und die Steigerung der Beitragssätze nur vorübergehend abbremsen, aber niemals stoppen oder gar umkehren. So stieg das Defizit trotz zahlreicher Eingriffe nahezu fortwährend an und lag im ersten Halbjahr des Jahres 2001 bei 5 Mrd. DM (vgl. Bundesministerium für Gesundheit 2001).

Ob eine Erhöhung der finanziellen Mittel im Bereich des Gesundheitswesens die vorherrschenden strukturellen Probleme dort zu lösen vermag, bleibt fraglich, solange keine grundsätzlichen Reformen auf der Einnahmenseite stattfinden. Durch zunehmende strukturelle Arbeitslosigkeit, ungünstige Konjunkturerwicklung und ein sich abschwächendes Wirtschaftswachstum sind die Defizite in der GKV nicht nur ausgabenbedingt, sondern beruhen z.T. auch auf einem nur geringen Anstieg der beitragspflichtigen Einnahmen (vgl. Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (SVRKAiG) 1998, Ziffer 423). Eine solche Betrachtung darf jedoch nicht den weiterhin hohen Ausgabendruck verneinen, der teilweise auf medizinischem Fortschritt und erhöhter Lebenserwartung, teilweise aber auch auf verbleibende Steuerungsmängel zurückzuführen ist (vgl. SVRKAiG 1998, Ziffer 423).

Die in den letzten Jahren ausgeschöpften beachtlichen Wirtschaftlichkeitsreserven bleiben vor diesem Hintergrund ohne begleitende strukturelle Reformen im Gesundheitswesen wirkungslos bzw. der Erfolg ist zumeist, wie in der Vergangenheit zu beobachten war, nur kurzfristig (vgl. SVRKAiG 1988, Ziffer 428). Viele der Reformen blieben auch deshalb hinter den hohen Erwartungen zurück, da es bei den betroffenen Akteuren des Gesundheitswesens zu Verhaltensänderungen und Gegenreaktionen kam. Reformen, deren Ansatzpunkte der einzelne Patient und der behandelnde Arzt sind, müssen bei einer solchen Konstellation auch deren Reaktionen auf sich ändernde Rahmenbedingungen in das

Kalkül mit einbeziehen. Hierzu gehören insbesondere die Berücksichtigung der Interdependenzen der Entscheidungen der Beteiligten und deren Auswirkungen auf die Ausgestaltungen der Verträge im Gesundheitswesen, d.h. der Vereinbarungen über den Versicherungsschutz und über die Vergütung der ärztlichen Leistungen.

In allen Betrachtungen bildet dabei der Patient ein zentrales Element des Gesundheitswesens. Abstrahiert man von der reinen medizinischen Versorgung, so erscheint es geeigneter, vom „Nutzer des gesundheitlichen Versorgungssystems“ (Deutscher Bundestag 2001, Ziffer 301) zu sprechen. Dieser läßt sich je nach Interessenlage in Bürger, Versicherter, Kunde, Konsument oder Patient unterscheiden. Bürger besitzen ein hohes Interesse an gesundheitsfördernden Lebensbedingungen und den notwendigen Versorgungsstrukturen. Ein Versicherter weist hauptsächlich finanzielle Motive hinsichtlich des Versicherungsschutzes auf, insbesondere für möglichst geringe Beitrags- oder Prämienzahlungen in Verbindung mit umfassenden Leistungen. Als Kunden bezeichnet man einen Nutzer dann, wenn es um aktive Nachfrage nach medizinischen Produkten oder Gesundheitsleistungen, Marktcompetenz oder auch um den Einfluß auf das medizinische Angebot geht. Der Begriff Konsument stellt direkt auf die in Anspruch genommene Leistung ab. Am häufigsten wird jedoch der Ausdruck ‚Patient‘ verwendet.¹ Dabei geht man davon aus, daß dieser sich aufgrund einer akuten oder chronischen Erkrankung für die Angebote der medizinischen Versorgung interessiert. Aus ökonomischer Sicht sind vor allem die Interaktionen und Interdependenzen eines Patienten bzw. Nutzers mit den Leistungserbringern und den Kostenträgern, den Sozial- und Privatversicherern, von hervorgehobenem Interesse. Die Beziehung zwischen Arzt und Patient stellt in diesem Zusammenhang die Kernbeziehung des Gesundheitswesens dar. Dabei besitzt der Arzt einen Wissensvorsprung gegenüber dem Patienten. Dieser erstreckt sich sowohl auf Diagnose- als auch auf Therapiemöglichkeiten. In dieser Situation, in der der Patient die Erbringung der medizinischen Leistung an den Arzt delegiert, herrscht eine asymmetrische Informationsverteilung nicht nur zugunsten des Arztes, sondern auch zugunsten des Patienten. Dieser Umstand rührt daher, daß der Arzt das behandlungsbegleitende Verhalten des Patienten i.d.R. weder beobachten kann noch genau einzuschätzen vermag.

Die Existenz einer wechselseitigen asymmetrischen Informationsbeziehung stellt den Kern der vorliegenden Untersuchung dar. Sie baut auf den informationsökonomischen Aspekten der Kontrakttheorie auf. Ausgangspunkt ist dabei die Betrachtung des Verhältnisses zwischen Arzt und Patient als Principal-

¹ Im weiteren Verlauf der Arbeit wird grundsätzlich der Begriff ‚Patient‘ benutzt, vor allem, da in der Untersuchung i.d.R. direkt auf den Behandlungsprozeß abgestellt wird.

Agent-Beziehung. Diese befaßt sich mit Situationen der Delegation von Aufgaben und Leistungen aufgrund von Spezialisierungsvorteilen und den damit verbundenen Problemen asymmetrischer Information. Dabei können sowohl vorvertragliche Informationsasymmetrien (adverse Selektion) als auch nachvertragliche Asymmetrien (Moral Hazard) Gegenstand einer Untersuchung sein. Der Focus der Arbeit liegt auf dem Verhalten der Entscheidungsträger im Gesundheitswesen nach Abschluß eines Vertrages, sei es eine Versicherungs- oder eine Vergütungsvereinbarung. Ausgehend von abgeschlossenen Verträgen sollen die Auswirkung der Vertragsgestaltung auf die resultierenden Anreize und die individuellen Handlungen untersucht werden, wobei nach Abschluß des Vertrages zwischen den beteiligten Akteuren Informationsasymmetrien vorliegen. Anders ausgedrückt wird die Fragestellung untersucht, wie nachvertragliche Aktionen durch die Wahl der Vertragsparameter beeinflusst werden können. Aus diesem Grund werden Probleme adverser Selektion in diesem Kontext nicht näher diskutiert, sondern es erfolgt eine Konzentration auf die existierenden Informationsasymmetrien nach Vertragsabschluß. Da diese nicht nur einseitig auftreten, sondern wechselseitig zwischen Arzt und Patient vorliegen, kann in diesem Zusammenhang von einem Double Moral Hazard-Problem gesprochen werden.

Im Rahmen der Analyse der Arzt-Patient-Beziehung mit Hilfe der Principal-Agent-Theorie zeigt sich, daß die Besonderheiten des Gesundheitswesens eine simple Anwendung dieses Konzeptes nicht erlauben. Dies betrifft hauptsächlich drei Kritikpunkte: Erstens treten in der Beziehung zwischen Arzt und Patient sogenannte „ergänzende Sachwalter“ (Zweifel 1994, S. 20) auf, bspw. der Gesetzgeber, die Verbände des Gesundheitswesens oder vor allem die Versicherer. Zweitens betrifft dies das Zusammenwirken von medizinischer Leistung des Arztes und dem gesundheitsrelevanten Verhalten des Patienten. Drittens besitzt der Patient ex ante kein Wissen über den Einfluß der medizinischen Leistung auf die Gesundheit. Er kann weiterhin auch ex post nicht beurteilen, ob sich eine Verbesserung seines Gesundheitszustands ursächlich auf die in Anspruch genommene medizinische Leistung zurückzuführen läßt oder ob andere Faktoren wie seine Selbstheilungskräfte oder Umweltzustände dafür verantwortlich sind (vgl. McGuire et al. 1988, S. 155). Der Übergang von den Standardmodellen der Principal-Agent-Theorie hin zu Ansätzen, die den Anforderungen des Gesundheitswesens gerecht werden, erweist sich somit als schwierig (vgl. McGuire et al. 1988, S. 155). Erstens liegen die bereits angesprochenen Aspekte vor, in denen die Principal-Agent-Theorie im Gesundheitswesen von einer herkömmlichen abweicht. Zweitens ist es möglich, daß die Nutzenfunktionen bzw. deren Argumente in der Beziehung zwischen Arzt und Patient nicht mehr separat betrachtet werden können, sondern daß zwischen ihnen Interdependenzen bestehen (können).

Der Zusammenhang zwischen ärztlicher Behandlungsleistung und gesundheitsrelevantem Verhalten des Patienten kann mit Hilfe eines zweistufigen Modellansatzes untersucht werden. Zum einen geschieht dies durch die Analyse der Behandlungsstufe, auf der sowohl der Arztbesuch des Patienten als auch die Handlungen der beiden Akteure untersucht werden, zum anderen durch die Betrachtung der Vertragsstufe, auf der die Bedingungen des Versicherungs- und Vergütungsvertrages festgelegt werden. Von besonderer Relevanz ist dabei weiterhin die beschriebene gegenseitige Beeinflussung der gewählten Aktionen und der Einfluß von Vertragsmodifikationen auf diese Aktionen, bspw. durch die Einführung einer Selbstbeteiligung. Dadurch soll überprüft werden, ob eine höhere Beteiligung des Patienten an den Behandlungsausgaben Auswirkungen auf das Verhalten des Patienten und die medizinische Leistung des Arztes besitzt.

Neben den Ergebnissen einer theoretischen Untersuchung sind für die Betrachtung der Situation zwischen Arzt und Patient im Gesundheitswesen auch empirische Befunde relevant. Eine direkte Umsetzung der theoretischen Modellstrukturen der Principal-Agent-Theorie und der Double Moral Hazard-Problematik scheidet jedoch im allgemeinen bereits an den zur Verfügung stehenden Daten. Aus diesem Grund kommt eine Analyse der Determinanten der Gesundheitszufriedenheit in Betracht, da auch in der theoretischen Analyse der Gesundheitszustand u.a. von den Aktionen des Patienten und des Arztes abhängt. Mit ihrer Hilfe soll gezeigt werden, inwiefern die medizinische Inanspruchnahme und das gesundheitsrelevante Verhalten des Patienten zusammen mit prädisponierenden, sozioökonomischen und gesundheitsrelevanten Variablen sowie Variablen des Wohnumfeldes Einfluß auf die subjektive Gesundheitszufriedenheit ausüben.

1.2 Aufbau der Arbeit

Im Rahmen der skizzierten Problemstellung der vorliegenden Arbeit werden im *zweiten Kapitel* Grundlagen der Beziehung zwischen Arzt und Patient im deutschen Gesundheitswesen erläutert. Dazu gehört neben einem Überblick über das System der Gesundheitssicherung auch eine Darstellung der heutigen finanziellen Situation, wobei die Gesundheitsausgaben einmal nach den Trägern und dann nach den Leistungsarten aufgeschlüsselt werden. Im Anschluß daran findet eine nähere Betrachtung der Situation der Gesetzlichen Krankenversicherung im Zeitablauf statt, wobei eine Darstellung der Struktur der Einnahmen und Ausgaben sowie der Leistungsausgaben nach den Mitgliedern erfolgt. Den Abschluß dieses Kapitels bildet die Erörterung der Fehlanreize und Schwachstellen der Beziehung zwischen Arzt und Patient, die als Ursache der Defizite in der medizinischen Versorgung oder der Organisationsmängel gesehen werden können.

Das *dritte Kapitel* widmet sich der Präsentation der informationsökonomischen Grundlagen. Zunächst werden die grundsätzliche Problematik erläutert, die wesentlichen Kennzeichen der unterschiedlichen Bereiche der adversen Selektion und des Moral Hazard vorgestellt und die Annahmen der Informationsökonomik geschildert. Der folgende Abschnitt behandelt die Informationsasymmetrie nach Abschluß eines Vertrages. Dazu wird auf unterschiedliche Begriffsabgrenzungen des Moral Hazard näher eingegangen. Der nächste Teil behandelt die Grundstruktur von Principal-Agent-Problemen, wie sie in der Literatur als Holmström/Mirrlees-Ansatz bekannt ist.² Dabei werden auch die mit diesem Ansatz bestehenden Schwierigkeiten geschildert und eine graphische Darstellung des Ergebnisses des Vertrages zwischen Principal und Agenten vorgestellt. Ein Überblick über die weiteren Ansätze und Entwicklungen im Rahmen der Principal-Agent-Theorie schließt diesen Bereich ab.

Das *vierte Kapitel* behandelt die Möglichkeit, den vorgestellten Ansatz der Principal-Agent-Theorie auf die Situation zwischen Arzt und Patient zu übertragen und die Besonderheiten des Gesundheitswesens hinsichtlich der verfügbaren Information der beteiligten Akteure zu analysieren. Dazu wird zuerst die Theorie der anbieterinduzierten Nachfrage vorgestellt, die davon ausgeht, daß der Arzt seinen aus dem Informationsvorsprung resultierenden diskretionären Handlungsspielraum zur Beeinflussung der Nachfrage nach den von ihm erbrachten medizinischen Leistungen ausnutzen kann. Anschließend wird der Standardansatz der Agency-Theorie auf das Verhältnis zwischen Arzt und Patient übertragen und kritisch hinterfragt. Sodann werden die Hauptbeteiligten des Gesundheitswesens sowie ihre Beziehung zueinander unter informationsökonomischen Gesichtspunkten betrachtet. Dies betrifft die bestehenden Anreize, die Anforderungen, die sich aus diesem für die modelltheoretische Analyse ergeben sowie die Interdependenzen, die zwischen den einzelnen Akteuren bestehen.

Das *fünfte Kapitel* gibt einen Überblick über theoretische Forschungsansätze des Verhältnisses zwischen Arzt, Patient und Versicherung. Dazu werden zunächst in einer kurzen Übersicht bestehende Modelle vorgestellt, beginnend mit den Auswirkungen asymmetrischer Information zwischen den Vertragsparteien im Gesundheitswesen, dargestellt anhand von Untersuchungen, die sich speziell an ein Versicherungssystem, sei es die amerikanische Health Maintenance Organization oder den britischen National Health Service, anlehnen, sowie Ansätzen, die direkt den Arzt als Agenten des Patienten betrachten. Daran schließt die Darstellung von vier ausgewählten Modellansätzen an, deren zentrale Elemente für die folgende theoretische Untersuchung von großer Bedeutung sind.

² Zu einer Übersicht über den Holmström/Mirrlees-Ansatz siehe S. 34.

Diese ist im *sechsten Kapitel* dargestellt und behandelt die beidseitige Informationsasymmetrie zwischen Arzt und Patient (Double Moral Hazard). Zuerst wird der grundsätzliche Zusammenhang zwischen ärztlicher Leistung und der Compliance des Patienten geschildert, bevor auf die Möglichkeiten der Einordnung des Kontaktes zwischen Arzt und Patient eingegangen wird. Das sich daran anschließende Unterkapitel behandelt, wie sich die Reihenfolge der von Arzt und Patient gewählten Handlungen auf das Behandlungsergebnis auswirkt. Anschließend wird die Grundstruktur des Modells für den Fall der Behandlung erläutert. Dazu gehört die Analyse der Situation bei vollständiger und asymmetrischer Information. Es folgt die Darstellung des Verhältnisses zwischen Compliance und medizinischer Leistung, den sich daraus ergebenden Handlungen und die Betrachtung des Einflusses einer Selbstbeteiligung für den Patienten. Im nächsten Abschnitt erfolgt die Analyse der Vertragsstufe, insbesondere im Hinblick auf den Einfluß der asymmetrischen Information auf die Vertragsgestaltung.

Kapitel 7 widmet sich der empirischen Analyse der Arzt-Patient-Beziehung. Zuerst werden, auf Basis der theoretischen Ergebnisse die grundlegenden Voraussetzungen für eine geeignete empirische Untersuchung festgelegt und testbare Hypothesen abgeleitet. Daran schließt sich ein Überblick über empirische Studien an, die Anhaltspunkte für die eigene Vorgehensweise bieten. Für die ökonometrische Untersuchung bieten sich Ansätze an, die neben Individualdaten auch zeitliche Aspekte mit in die Analyse einbeziehen. Zunächst werden jedoch als Ausgangspunkt Modelle für Querschnittsdatensätze untersucht. Dabei werden die angewendeten Methoden, die verwendeten Variablen und die Ergebnisse der Querschnittsschätzung präsentiert. Der nächste Abschnitt befaßt sich mit der Panel-Analyse, die auch die Zeitstruktur berücksichtigt. Hierbei werden nach einer Einführung die Schätztechniken erläutert, bevor auch hier auf die Datenauswahl eingegangen wird. Danach folgt die Vorstellung der Ergebnisse der Panel-Analyse sowie damit einhergehend ein Vergleich mit den Schätzergebnissen des Querschnittsdatensatzes.

Im *letzten Kapitel* sind die Ergebnisse der Arbeit zusammengefaßt und es erfolgt ein Ausblick auf die Erweiterungsmöglichkeiten sowohl der theoretischen Modellierung der Arzt-Patient-Beziehung im Rahmen der Principal-Agent-Theorie als auch deren Umsetzung in eine empirische Analyse.

2 Die Arzt-Patient-Beziehung im deutschen Gesundheitswesen

2.1 Das System der Gesundheitssicherung in Deutschland

Das deutsche Gesundheitswesen ist ein komplexes Geflecht der verschiedenen Beziehungen zwischen Leistungserbringern, den Nachfragern bzw. Konsumenten, den Finanziers der Leistungen sowie den Produzenten von Arznei-, Heil- und Hilfsmitteln. Vor diesem Hintergrund erscheint es für einen ersten Ansatz der Darstellung der Arzt-Patient-Beziehung angebracht, das deutsche Gesundheitssystem und dessen Hauptbeteiligte näher zu betrachten. Dabei stehen vor allem die Rolle des Patienten und die Unterschiede im Versicherungssystem im Mittelpunkt des Interesses. Im weiteren Verlauf werden die gegenwärtige finanzielle Situation und die zeitliche Entwicklung der Ausgaben und Einnahmen dargestellt. Ausgangspunkt der Überlegungen sind hierbei die Ausgaben für das Jahr 1998, unterteilt nach Trägern und Arten. Darüber hinaus sind die Struktur der Einnahmen und der Ausgaben seit 1970 für die gesetzliche Krankenversicherung als größtem Ausgabenträger sowie die Leistungsausgaben nach Mitgliedern im Zeitablauf von Interesse. Aus der Organisationsstruktur und der finanziellen Situation des deutschen Gesundheitswesens können dann Folgerungen für die Beziehung zwischen Patient und behandelndem Arzt abgeleitet werden, die an den Fehlanreizen ansetzen.

Zwar stellt das im Zentrum der Untersuchung stehende Verhältnis zwischen Arzt und Patient die Kernbeziehung im Gesundheitswesen dar, jedoch sind im Gesundheitswesen zahlreiche andere Akteure wie der Gesetzgeber, die Krankenkassen und die Versicherungen und weitere Interessengruppen vertreten. Die Organisationsstruktur des deutschen Gesundheitswesens ist über die Jahre hinweg gewachsen. Dabei ist ein komplexes System entstanden, das einer Vielzahl von Regulierungen unterliegt und in dem Informationsasymmetrien in den verschiedensten Ausprägungen anzutreffen sind. In Deutschland sind 91 % der Bevölkerung durch die etwa 400 gesetzlichen Krankenkassen abgesichert (vgl. hierzu Sachverständigenrat für die gesamtwirtschaftliche Entwicklung 2000, Ziffer 468). Nur etwa 8,9 % sind ausschließlich bei einer der etwa 50 bundesweiten privaten Krankenversicherungen (PKV) versichert, ca. 0,2 % besitzen keinen Krankenversicherungsschutz. Obwohl in Deutschland keine allgemeine Versicherungspflicht existiert, besitzt die gesetzliche Krankenversicherung (GKV) aufgrund staatlicher Regulierung eine überragende Bedeutung im Gesundheitswesen.³ Abhängig Beschäftigte (mit Ausnahme der Beamten) mit einem Arbeitseinkommen bis zur Beitragsbemessungsgrenze sowie Empfänger

³ Zu einer Übersicht über die gesetzliche Krankenversicherung siehe Statistisches Bundesamt 1998, S. 359-366.

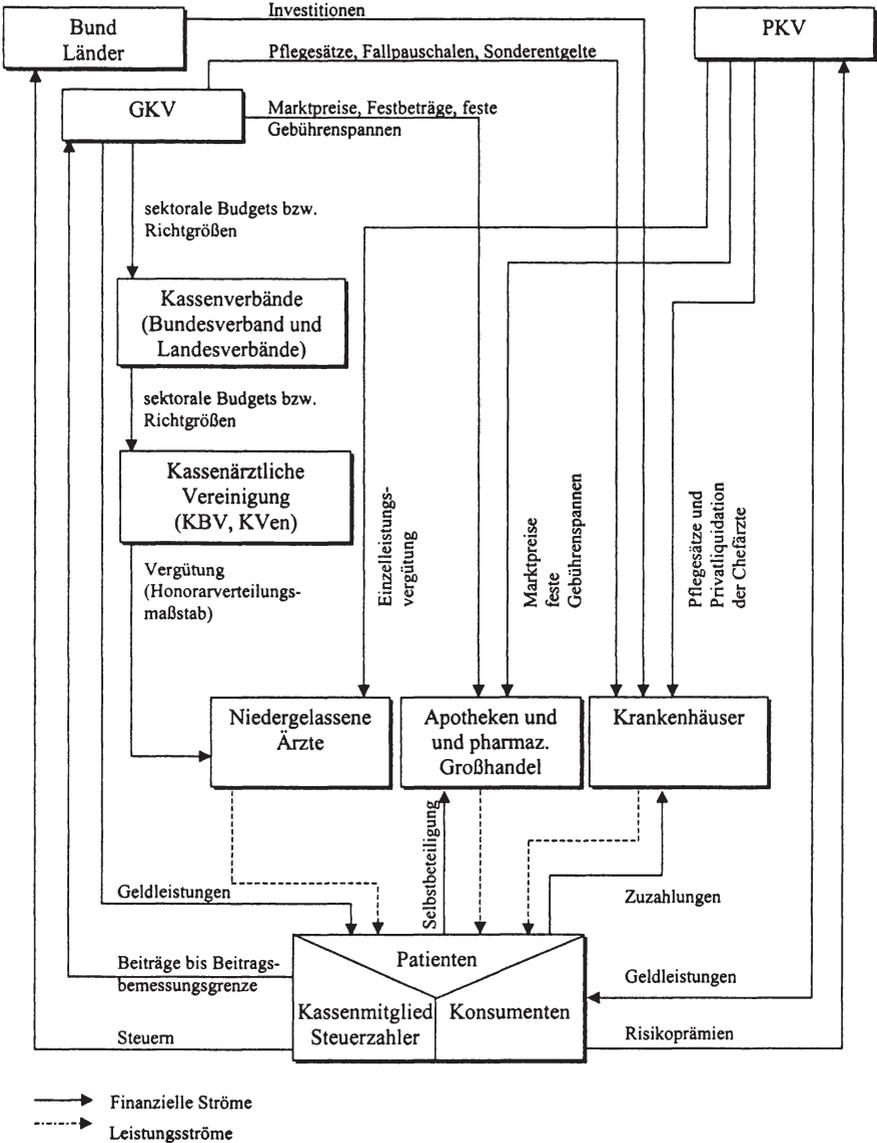
von Lohnersatzleistungen und Bezieher von Sozialhilfe sind über eine gesetzliche Krankenkasse abgesichert.⁴ Hinzu kommen die nicht erwerbstätigen Familienangehörigen, die ebenfalls in den gesetzlichen Versicherungsschutz mit einbezogen sind.

Abbildung 2.1 zeigt eine Übersicht über die Hauptbeteiligten dieses Systems. Die durchgezogenen Pfeile symbolisieren finanzielle Ströme zwischen den Beteiligten, die gestrichelten Pfeile Leistungsströme. Im Zentrum des Interesses steht die Beziehung zwischen den Leistungserbringern (niedergelassene Ärzte, Krankenhäuser, Apotheken und pharmazeutischer Großhandel) und den Patienten.

Direkte Kontakte zwischen Arzt und Patient treten dabei im Fall niedergelassener Ärzte und im Krankenhaus auf, während bei Apotheken und pharmazeutischem Großhandel der Arzt nur mittelbar in Form von Verschreibungen in Erscheinung tritt. Zuzahlungen oder Selbstbeteiligungen existieren im deutschen Gesundheitswesen für die GKV lediglich im Arzneimittelbereich, bei der stationären Versorgung und bei Heil- und Hilfsmitteln. Innerhalb der PKV sind Zuzahlungen und Selbstbeteiligungen vom individuell vereinbarten Versicherungsvertrag abhängig. Dieser richtet sich nach dem gewählten Versicherungsumfang, dem Geschlecht, dem Eintrittsalter und persönlichen Gesundheitsrisiken (vgl. hierzu Schneider et al. 1995, S. 205). Wie in Abbildung 2.1 zu sehen, agiert eine einzelne Person in unterschiedlichen Rollen im Gesundheitswesen. Im Falle der Inanspruchnahme medizinischer Leistungen spricht man vom Patienten. Im Bereich der PKV ist es der Konsument, der gegen risikoäquivalente Prämien versichert ist und im Versicherungsfall Geldleistungen erhält. Bei gesetzlicher Pflichtversicherung ist das Individuum Kassenmitglied, das gegen Beiträge die medizinischen Leistungen im wesentlichen als Sachleistungen erhält. Weiterhin zahlt die zuständige Krankenkasse auch Geldleistungen direkt an das Mitglied, bspw. das Krankengeld. Zusätzlich ist das Individuum auch Steuerzahler. Diese Zahlungen fließen in die Haushalte von Bund und Ländern wobei aus den allgemeinen Steuereinnahmen die Investitionen in den Krankenhausesektor finanziert werden.

⁴ Die Beitragsbemessungsgrenze ist zur Zeit auf 75 % der in der Rentenversicherung gültigen Grenze festgelegt und beträgt in Deutschland seit dem 1.1.2001 6525 DM pro Monat (vgl. Bundesministerium für Gesundheit 2001, S. 12). Diese lohnabhängigen Beiträge stehen jedoch schon seit einiger Zeit in der Diskussion (vgl. hierzu o.V. 2000, S. 4 oder Barbier 2000, S. 17).

Abbildung 2.1: Die Hauptbeteiligten im deutschen Gesundheitswesen



Quelle: Eigene Darstellung nach Hoffmeyer 1994, S. 432.

Neben den Unterschieden im Versicherungssystem existieren auch Unterschiede im Bereich der Vergütung der ärztlichen Leistungen zwischen gesetzlicher und privater Krankenversicherung. Im Rahmen der GKV verhandeln die Verbände der Leistungserbringer mit den Verbänden der Krankenkassen über die Gesamtvergütung.⁵ Dabei existieren zur Zeit in nahezu allen Bereichen sektorale Budgets bzw. Richtgrößen, die das Wachstum der Ausgaben an den Anstieg der beitragspflichtigen Einnahmen koppeln.⁶ In einer zweiten Stufe kommt der Honorarverteilungsmaßstab zum tragen, nach dem die Kassenärztliche Vereinigung die Gesamtvergütung auf die Kassenärzte verteilt (vgl. Schneider et al. 1995, S. 211).⁷ Für den Krankenhaussektor besteht ein System aus Pflegesätzen, Fallpauschalen und Sonderentgelten (vgl. Ulrich und Wille 1997), im Arzneimittelbereich regeln ein System aus Markt- und Festpreisen sowie feste Gebührenspannen die Vergütung. In der PKV existiert im ambulanten Sektor weitgehend noch die Einzelleistungsvergütung, die auf der Gebührenordnung für Ärzte (GOÄ) beruht. Im stationären Sektor erfolgt die Vergütung auf der Basis von Pflegesätzen und der Privatliquidation von Chefarzten. Für den Arzneimittelbereich gelten Marktpreise sowie feste Gebührenspannen.

2.2 Die finanzielle Situation des Gesundheitswesens

Der Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (SVRKAiG) stellte bereits 1994 fest, daß „[d]ie Gesundheitsausgaben [...] oft im Mittelpunkt der gesundheitspolitischen Diskussion [stehen]. Ihre Entwicklung führt zu Fragen der Effizienz des Gesundheitswesens und zu Steuerungsnotwendigkeiten.“ (SVRKAiG 1994, Ziffer 48). In Deutschland wurden im Jahre 1998 nach Berechnungen des Statistischen Bundesamtes ca. 412,7 Mrd. DM für Gesundheitsleistungen ausgegeben, was 10,9 % des Bruttoinlandsproduktes entspricht (vgl. hierzu und zu den Angaben Zifonun und Cordes 2001, S. 198ff.).⁸ In der Zeit von 1992 bis 1998 sind die Ausgaben von 320,6 Mrd. DM nominal

⁵ Auf Seiten der Leistungserbringer sind dies die Kassenärztliche Bundesvereinigung (KBV) bzw. die Kassenärztliche Vereinigung und auf Seiten der Kassen deren Bundes- oder Landesverbände.

⁶ Die bisher bestehenden Budgets im Arzneimittelmarkt werden aufgrund des Arzneimittelbudget-Ablösungsgesetzes – ABAG abgeschafft.

⁷ Die Verteilung der Gesamtvergütung ist eine Selbstverwaltungsaufgabe der Vertragsärzte. Die Kassenärztliche Vereinigung legt dabei die Bedingungen fest. Der Honorarverteilungsmaßstab wirkt flankierend zum ‚Einheitlichen Bewertungsmaßstab‘ (EBM), indem er die Zahlungen der Krankenkassen und die Vergütungen der Ärzte ins Gleichgewicht bringt (vgl. Sydow 1999, S. 42f.).

⁸ Zu den genannten 412,7 Mrd. DM kommen noch 116,7 Mrd. DM in Form von Einkommensleistungen hinzu, so daß sich die Gesamtausgaben auf 529,4 Mrd. DM belaufen (vgl. Zifonun und Cordes 2001, S. 197).

um 92,1 Mrd. DM angestiegen, was einem Zuwachs von 28,7 % entspricht. Dabei hat sich der Anstieg gegen Ende der betrachteten Zeitperiode jedoch verlangsamt. Zwischen 1996 und 1998 lag er lediglich bei 1,7 %, während er zwischen 1992 und 1994 bei 10,1 % und zwischen 1994 und 1996 bei 15 % lag.⁹ Im gleichen Zeitraum stieg der Preisindex für die Lebenshaltung aller privaten Haushalte um 13,9 %. Für Güter für Gesundheits- und Körperpflege stiegen die Preise um 18,0 % und für die Waren der Gesundheitspflege sogar um 44,2 % (vgl. Zifonun und Cordes 2001, S. 198).

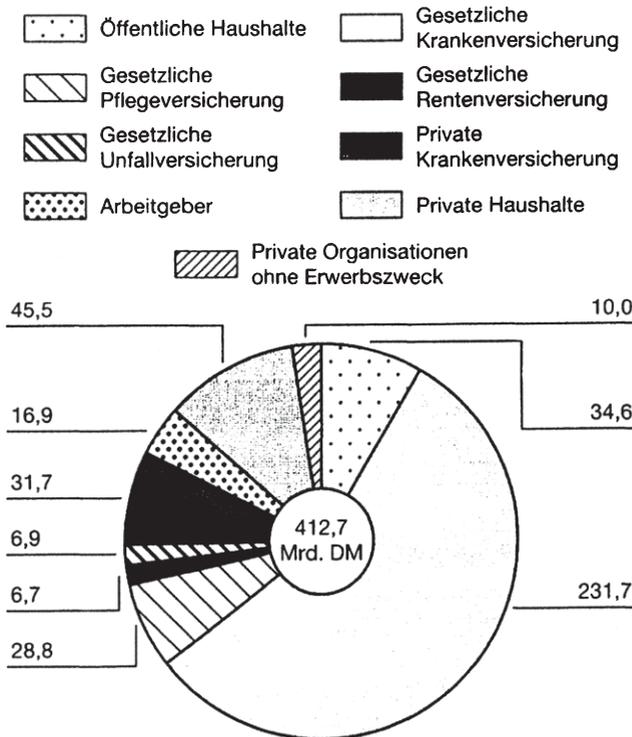
Spaltet man die Gesundheitsausgaben nach den Ausgabenträgern auf, so ergibt sich, daß die Gesetzliche Krankenversicherung mit Ausgaben von 231,7 Mrd. DM den Großteil (56,1 %) übernimmt (siehe hierzu Abbildung 2.2). Dabei stellen die Leistungsausgaben den Hauptposten dar. Es folgen die privaten Haushalte mit 45,5 Mrd. DM (11,0 %). Die öffentlichen Haushalte gaben 34,6 Mrd. DM (8,4 %) für die Gesundheit aus. Darunter fallen die Haushalte von Bund, Ländern und Gemeinden sowie die von der Bundesanstalt für Arbeit erbrachten Leistungen. Die Schwerpunkte bei den Ausgaben liegen bei der Krankenhausfinanzierung. Bei der privaten Krankenversicherung entfiel der Großteil der 31,7 Mrd. DM (7,7 %) auf die Behandlungsleistungen. Neben diesen wurden 28,8 Mrd. DM (7,0 %) durch die gesetzliche Pflegeversicherung getragen, die hauptsächlich für betreuende Maßnahmen in Form von Sachleistungen und Pflegegeldern anfielen. Rund 16,9 Mrd. DM (4,0 %) wurden durch die Arbeitgeber getragen. Diese Ausgaben dienten zur Erhaltung und Wiederherstellung der Gesundheit, wobei die Entgeltfortzahlungen den größten Teil ausmachten. Die verbleibenden 52,3 Mrd. DM entfielen auf private Organisationen ohne Erwerbszweck, die gesetzliche Unfallversicherung und die gesetzliche Rentenversicherung.. Diese Darstellung verdeutlicht, daß neben den Ausgaben von Versicherungen und Krankenkassen auch noch andere öffentliche oder private Träger existieren.

Vergleicht man die Angaben aus Abbildung 2.2 mit den Daten des Jahres 1996, so fällt auf, daß vor allem die Ausgabend der gesetzlichen Pflegeversicherung um 43,2 % gestiegen sind (vgl. hierzu Statistisches Bundesamt 2001, S. 3). Der Grund hierfür liegt in der Einführung der zweiten Stufe der Pflegeversicherung zum 1. Juli 1996, ab dem die Leistungen für die stationäre Pflege in die Pflegeversicherung übernommen wurden. Aus gleichem Grund sind die

⁹ Der Grund für den unterdurchschnittlichen Anstieg im Zeitraum von 1996 bis 1998 liegt in der Einführung der Pflegeversicherung im ambulanten Bereich zum 1. April 1995 und im stationären Bereich zum 1. Juli 1996 (vgl. Zifonun und Cordes 2001, S. 198). Die Ausgaben der gesetzlichen Pflegeversicherung wuchsen im Zeitraum zwischen 1996 und 1998 um 43,2 % (vgl. o.V. 2001a, S. 19).

Ausgaben der öffentlichen Haushalte um 23 % zurückgegangen, was besonders auf die Ausgaben der Sozialhilfe für die Pflege zutrifft. Der Ausgabenanstieg in der privaten Krankenversicherung von 10,4 % wurde durch die private Pflegepflichtversicherung beeinflusst. Durch die Einführung des Wachstums- und Beschäftigungsförderungsgesetzes zum 1. Januar 1997 kam es zu einer Abnahme der Ausgaben in der gesetzlichen Rentenversicherung von 27,8 %, da die Zahl der medizinischen Rehabilitationsleistungen zurückging. Insgesamt kann man festhalten, daß innerhalb der Gesundheitsausgaben eine Verschiebung der Ausgaben für Pflege hin zur Pflegeversicherung stattfand.

Abbildung 2.2: Gesundheitsausgaben 1998 nach Ausgabenträgern in Mrd. DM



Quelle: Zifonun und Cordes 2001, S. 198.

Neben der Betrachtung der Träger der Gesundheitsausgaben ist auch die Aufteilung der Ausgaben auf die verschiedenen Leistungsarten von Interesse, da

dadurch das Gewicht der einzelnen Bereiche der Gesundheitsversorgung ersichtlich ist. In Abbildung 2.3 ist zu erkennen, daß der größte Bereich durch die Gruppe ‚Waren‘ mit 107,5 Mrd. DM abgedeckt wird. Dahinter verbergen sich Arzneimittel, Hilfsmittel, Zahnersatz und Implantate, Blutprodukte und sonstiger medizinischer Bedarf. In dieser Gruppe besitzen die Arzneimittel mit 55,9 Mrd. DM das größte Gewicht (vgl. Zifonun und Cordes 2001, S. 199). Der zweithöchste Posten sind ärztliche Leistungen mit Ausgaben von 106,7 Mrd. DM. Diese setzen sich aus Grund- und Sonderleistungen, Laborleistungen und strahlendiagnostischen Leistungen zusammen. Auf pflegerische und therapeutische Leistungen entfielen Ausgaben in Höhe von 88,1 Mrd. DM. Verwaltungsleistungen machen mit 21,5 Mrd. DM immerhin noch 5,2 % der Ausgaben aus.

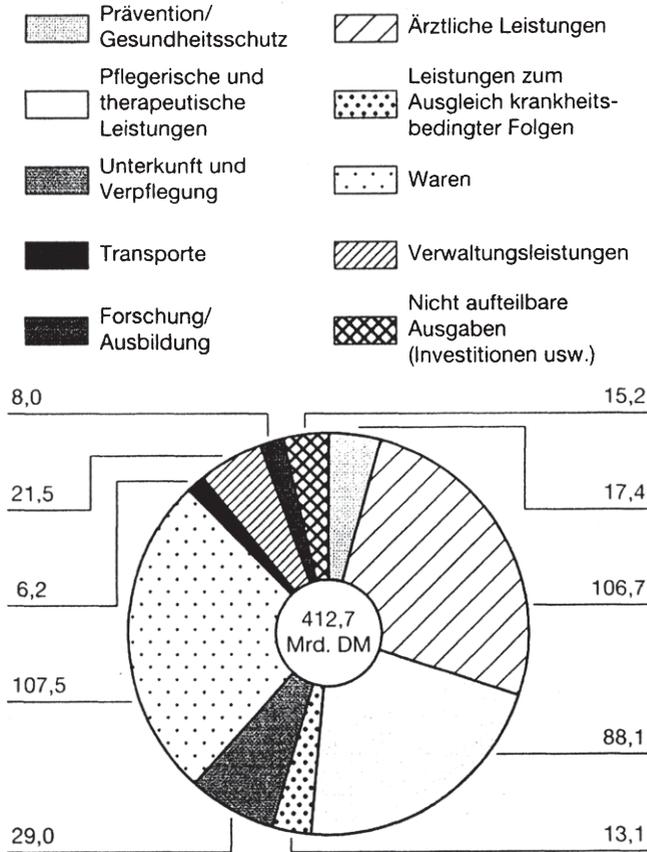
Betrachtet man die Entwicklung der Einnahmen und Ausgaben im größten Bereich des Gesundheitswesens, der gesetzlichen Krankenversicherung, so stellt man fest, daß die Gesamtausgaben von 25,18 Mrd. DM im Jahre 1970 (alte Bundesländer) auf über 255 Mrd. DM im Jahre 1999 (Gesamtdeutschland) gestiegen sind (siehe Tabelle 2.1).¹⁰ Davon entfallen auf die Leistungsausgaben ca. 241 Mrd. DM (94,5 %). Den größten Ausgabenbereich bildet dabei die stationäre Versorgung mit 85,54 Mrd. DM (35,5 %), gefolgt von der ärztlichen Behandlung mit 46,44 Mrd. DM (19,3 %). Im Jahre 1970 lag der Anteil der Krankenhausbehandlung noch bei 25,2 % und der der ärztlichen Behandlung bei 22,9 %. Allein dadurch ist die Verschiebung innerhalb der Ausgaben der GKV hin zur stationären Versorgung zu erkennen. Im gleichen Zeitraum stieg der Anteil der Verwaltungskosten an den Gesamtausgaben von 4,9 % im Jahre 1970 auf 5,5 % im Jahre 1999 an.

Anzumerken ist, daß die einzelnen Leistungsbereiche in unterschiedlichem Maße zur Kostenentwicklung in der GKV beitragen (vgl. zu diesem Punkt SVRKAiG 1994, Ziffer 65ff.). Zwar hat auch im Bereich Arznei-, Heil und Hilfsmittel absolut betrachtet eine Ausgabensteigerung stattgefunden, mit 15 % der Ausgaben für Gesamtdeutschland im Jahre 1999 hat sich jedoch der Anteil dieser Gruppe an den gesamten Leistungsausgaben reduziert. Für die alten Bundesländer betrug dieser Anteil 1970 noch 17,7 %. Ähnliches gilt auch für die

¹⁰ Die Unterschiede in den Ausgaben der GKV im Jahre 1998 zwischen Abbildung 2.2 und Tabelle 2.1 beruhen einerseits auf Änderungen der Berechnung in der neuen Gesundheitsausgabenrechnung der Gesundheitsberichterstattung des Bundes sowie andererseits auf den Unterschieden zwischen Statistischem Bundesamt und Bundesgesundheitsministerium. So weist das Statistische Bundesamt für 1996 Ausgaben in der GKV in Höhe von 244,3 Mrd. DM aus, während es 1998 lediglich 231,7 Mrd. DM waren (vgl. Statistisches Bundesamt 1999 und Zifonun und Cordes 2001).

zahnärztliche Behandlung und den Zahnersatz, bei denen auch eine deutliche Ausgabensteigerung festzustellen war, deren Anteil allerdings im Zeitablauf leicht rückläufig ist (von 10,6 auf 8,9 %).

Abbildung 2.3: Gesundheitsausgaben 1998 nach Leistungsarten in Mrd. DM



Quelle: Zifonun und Cordes 2001, S. 200.

Aus Tabelle 2.2 ist die Entwicklung der Leistungsausgaben für die Mitglieder in der GKV in den Jahren 1970 bis 1999 zu entnehmen. Dabei wurde zwischen Mitgliedern und deren Familienangehörigen in der GKV und Rentnern und Familienangehörigen unterschieden, die in der Krankenversicherung der Rentner

(KVdR) versichert sind. Zunächst ist festzuhalten, daß für alle Gruppen eine deutliche Ausgabensteigerung vorliegt. Gaben die Mitglieder 1970 in den alten Bundesländern noch 17,3 Mrd. DM für die Leistungen aus, so waren es 1998 bereits 111,8 Mrd. DM. Im gleichen Zeitraum stiegen die Ausgaben für die Gruppe der Rentner von 6,6 auf 84,0 Mrd. DM. In den neuen Länder erhöhten sich die Leistungsausgaben für die Gruppe der Mitglieder in dem Zeitraum von 1991 bis 1998 um 6,6 Mrd. DM von 13,1 auf 19,7 Mrd. DM, für die Rentner von 8,8 auf 19,5 Mrd. DM.

Die Ausgaben pro Kopf stiegen im Zeitraum von 1970 bis 1998 in den alten Ländern für die Mitglieder nominal von 763 auf 3872 DM je Mitglied und von 821 auf 7101 DM je Rentner. Die durchschnittlichen Gesamtausgaben in der GKV pro Kopf stiegen von 822 DM auf 5100 DM an. Während also jeder Rentner 1970 noch das 1,07fache der Ausgaben je Mitglied hatte, so war es 1998 bereits das 1,83fache. In den neuen Ländern gab ein Rentner durchschnittlich das 1,93fache eines Mitglieds aus und lag damit noch über dem Verhältnis in den alten Ländern. Die Ausgaben je Mitglied verfünffachten sich von 1970 bis 1998, die Ausgaben je Rentner haben sich im gleichen Zeitraum sogar auf über das achtfache erhöht.

Als Ursache für die steigenden Ausgaben benennt der Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen verschiedene Gründe (vgl. SVRKAiG 1994, Ziffer 64). Zu berücksichtigen ist bspw. die demographische Entwicklung, insbesondere der Anstieg des Anteils älterer Menschen an der Gesamtbevölkerung. Weiterhin wirkt der medizinisch-technische Fortschritt in Form von verbesserter, teurerer Diagnose, neuen Arzneimitteln und kostenintensiver Apparatechnik ausgabensteigernd. Zusätzlich bewirkt der Preisstruktureffekt, daß die Preise im personalintensiven Gesundheitssektor schneller als in anderen Bereichen der Volkswirtschaft ansteigen. Darüber hinaus bestehen Steuerungsdefizite, Überkapazitäten und Unwirtschaftlichkeiten. Schließlich kam es im Zeitablauf zu einer Zunahme der Behandlungen von kostenintensiven Verschleißerkrankungen und zu einer verstärkten Inanspruchnahme durch ein allgemein gestiegenes Gesundheitsbewußtsein.¹¹

¹¹ Der Effekt der stärkeren Inanspruchnahme institutioneller Pflege wurde mit der Einführung der gesetzlichen Pflegeversicherung 1995/96 auf diese verlagert.

Tabelle 2.1: Struktur der Einnahmen und Ausgaben in der gesetzlichen Krankenversicherung in Mrd. DM

Jahr	Einnahmen		Ausgaben ¹⁾									Einnahmen ./ Ausgaben ⁴⁾
	ins- gesamt	darunter	ins- gesamt	Leistungsausgaben ²⁾					Netto- Verwal- tungs- kosten ³⁾	Vermögens- aufwen- dungen und sonstige Aufwen- dungen		
		Beiträge		zusammen	darunter: ausgewählte Leistungsausgaben							
					ärztliche Behand- lung	zahn- ärztliche Behand- lung, Zahnersatz	Arznei- Verband, Heil- und Hilfs- mittel aus Apotheken	Kranken- haus- behand- lung			Kranken- geld	
Früheres Bundesgebiet												
1970	26,12	24,98	25,18	23,85	5,46	2,54	4,22	6,01	2,47	1,23	0,10	0,94
1971	31,28	29,94	31,14	29,59	6,81	3,23	4,97	7,65	2,96	1,48	0,08	0,14
1972	36,21	34,63	36,40	34,56	7,58	3,77	5,75	9,36	3,44	1,73	0,11	-0,19
1973	44,46	42,11	43,37	41,01	8,60	4,53	6,75	11,70	3,90	2,06	0,30	1,09
1974	51,11	48,43	51,81	49,19	9,93	5,48	7,88	15,25	4,26	2,44	0,19	-0,70
1975	60,74	58,16	60,99	58,17	11,26	8,31	8,90	17,53	4,66	2,72	0,10	-0,25
1976	70,20	67,52	66,56	63,59	11,92	9,61	9,64	19,26	4,73	2,87	0,11	3,64
1977	73,50	70,24	69,82	66,61	12,49	10,01	9,85	20,46	4,91	3,02	0,19	3,68
1978	76,43	73,19	74,79	71,46	13,19	10,72	10,65	21,87	5,31	3,19	0,14	1,65
1979	80,83	77,10	81,06	77,44	14,12	11,69	11,37	23,25	5,94	3,47	0,16	-0,24
1980	88,45	83,53	89,83	85,96	15,36	12,87	12,57	25,47	6,65	3,75	0,13	-1,38
1981	96,49	90,99	96,39	92,20	16,49	14,05	13,63	27,32	6,44	4,06	0,13	0,10
1982	101,71	96,52	97,22	92,68	16,92	13,06	13,78	29,60	5,90	4,46	0,09	4,48
1983	103,48	98,54	100,69	95,90	17,76	12,94	14,45	30,97	5,78	4,70	0,10	2,79
1984	105,81	100,73	108,68	103,56	18,92	13,90	15,54	32,38	6,31	4,98	0,13	-2,87
1985	111,83	106,91	114,11	108,70	19,66	14,32	16,60	34,02	6,38	5,26	0,14	-2,28
1986	118,51	113,99	119,87	114,06	20,30	14,06	17,63	36,53	6,88	5,67	0,14	-1,36
1987	125,07	120,87	125,00	118,93	20,97	13,65	18,89	38,19	7,40	5,97	0,10	0,08

1988	132,64	128,19	134,38	128,06	21,65	17,34	20,44	39,49	7,79	6,20	0,11	-1,74
1989	139,68	134,89	129,93	123,24	22,65	12,55	20,22	40,81	7,81	6,57	0,11	9,76
1990	147,75	141,86	141,65	134,24	24,37	13,01	21,84	44,60	8,83	7,28	0,14	6,10
1991	154,22	147,72	159,81	151,63	26,74	14,74	24,49	49,12	10,24	7,93	0,25	-5,59
1992	167,78	161,34	176,88	167,85	28,93	17,00	27,08	53,94	11,02	8,68	0,35	-9,10
1993	184,42	178,00	175,36	166,09	29,89	15,32	21,81	56,95	10,90	8,77	0,50	9,06
1994	191,16	184,51	188,28	178,46	31,97	16,57	22,89	61,36	11,64	9,28	0,54	2,88
1995	196,62	188,94	201,59	190,29	32,69	17,42	25,05	66,53	15,24	9,67	1,63	-4,96
1996	202,24	195,24	207,58	196,39	33,67	18,98	26,77	66,97	14,94	10,35	0,84	-5,34
1997	205,04	198,74	203,34	192,13	34,24	19,17	25,81	68,42	11,79	10,30	0,91	1,70
1998	208,64	202,58	207,47	195,77	34,79	17,60	27,28	70,45	11,34	10,91	0,79	1,17
1999	214,82	208,62	213,24	200,83	35,62	17,80	29,51	70,46	11,67	11,53	0,89	0,37 ⁵⁾
Neue Bundesländer und Berlin-Ost⁶⁾												
1991	26,00	25,44	23,23	21,93	3,33	2,77	3,99	7,97	1,01	1,29	0,01	2,77
1992	33,36	32,53	33,62	31,71	4,45	5,07	5,46	10,31	1,63	1,88	0,03	-0,26
1993	37,77	36,76	36,42	34,04	5,17	3,70	5,67	11,55	2,11	2,34	0,04	1,35
1994	41,18	40,09	41,32	38,77	5,86	4,03	6,28	13,42	2,82	2,45	0,10	-0,14
1995	38,75	37,65	40,95	38,53	5,78	3,79	6,36	13,16	3,16	2,33	0,09	-2,19
1996	41,00	39,46	42,61	40,03	5,67	4,04	6,67	13,90	3,25	2,45	0,13	-1,61
1997	41,70	40,64	41,71	39,22	5,84	4,13	6,24	14,38	2,66	2,33	0,17	-0,01
1998	41,22	40,49	41,84	39,16	5,85	3,61	6,33	14,78	2,44	2,42	0,26	-0,63
1999	41,79	40,74	42,81	40,14	5,82	3,64	6,72	15,09	2,31	2,50	0,17	0,18 ⁵⁾

¹⁾ Leistungsausgaben, Verwaltungskosten, Vermögensaufwendungen und sonstige Aufwendungen ohne Risikostrukturausgleich (RSA).

²⁾ Für alle Versicherten: Mitglieder (einschließlich Rentner) und deren Familienangehörige.

³⁾ Verwaltungs- und Verfahrenskosten.

⁴⁾ Abzüglich des RSA-Saldo.

⁵⁾ Unter Berücksichtigung des RSA-West-Ost Transfers in Höhe von 1,2 Mrd. DM.

⁶⁾ Ab 1995 Zuordnung von Berlin-Ost zu den alten Bundesländern.

Quelle: Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2000, S. 288

Tabelle 2.2: Leistungsausgaben für die Mitglieder der gesetzlichen Krankenversicherung

Jahr ¹⁾	Mitglieder			Rentner			Nachrichtlich:		
	insgesamt	je Mitglied ²⁾		insgesamt	je Rentner ³⁾		Gesamtausgaben ⁴⁾		
		insgesamt	je Mitglied ⁵⁾		insgesamt	je Mitglied ⁵⁾			
	Mrd DM	DM	1970=100	Mrd DM	DM	1970=100	Mrd DM	DM	1970=100
Früheres Bundesgebiet									
1970	17,3	763	100	6,6	821	100	25,2	822	100
1971	21,3	916	120,0	8,3	1 007	122,7	31,1	989	120,4
1972	24,5	1 039	136,1	10,0	1 171	142,6	36,4	1 131	137,7
1973	28,6	1 183	155,0	12,4	1 374	167,3	43,4	1 306	158,9
1974	33,6	1 392	182,4	15,6	1 664	202,6	51,8	1 547	188,3
1975	39,7	1 663	217,9	18,5	1 920	233,8	61,0	1 821	221,6
1976	42,9	1 808	237,0	20,7	2 099	255,6	66,6	1 982	241,2
1977	44,5	1 865	244,4	22,1	2 202	268,1	69,8	2 064	251,2
1978	47,3	1 953	256,0	24,2	2 377	289,5	74,8	2 175	264,8
1979	51,0	2 074	271,8	26,4	2 580	314,2	81,1	2 327	283,2
1980	56,5	2 252	295,1	29,5	2 859	348,1	89,8	2 538	308,9
1981	59,6	2 347	307,6	32,6	3 160	384,8	96,4	2 700	328,6
1982	58,7	2 310	302,7	34,0	3 265	397,6	97,2	2 714	330,4
1983	59,1	2 333	305,8	36,8	3 514	427,9	100,7	2 812	342,3
1984	63,2	2 485	325,7	40,4	3 821	465,3	108,7	3 019	367,5
1985	65,4	2 557	335,1	43,3	4 074	496,2	114,1	3 151	383,6
1986	67,9	2 632	345,0	46,1	4 332	527,6	119,9	3 289	400,3
1987	70,1	2 697	353,5	48,8	4 554	554,6	125,0	3 404	414,3
1988	75,3	2 873	376,6	52,7	4 888	595,3	134,4	3 632	442,1

1989	72,4	2 752	360,6	50,8	4 659	567,4	129,9	3 490	424,8
1990	79,1	2 936	384,8	55,2	5 011	610,2	141,7	3 734	454,4
1991	88,6	3 212	421,0	63,0	5 667	690,1	159,8	4 129	502,5
1992	97,6	3 485	456,7	70,3	6 250	761,1	176,9	4 507	548,6
1993	96,0	3 417	447,8	70,1	6 168	751,1	175,4	4 444	540,9
1994	102,2	3 637	476,6	76,3	6 661	811,2	188,3	4 760	579,4
1995	110,4	3 837	502,8	79,9	6 821	830,6	201,6	4 979	606,0
1996	114,4	3 996	523,7	82,0	6 970	848,8	207,6	5 102	621,0
1997	110,4	3 813	499,8	81,7	6 924	843,3	203,3	4 990	607,3
1998	111,8	3 872	507,5	84,0	7 101	864,8	207,5	5 100	620,7
1999	...	3 950	517,7	213,2
Neue Bundesländer und Berlin -Ost									
1991	13,1	1 536	 	8,8	2 905	 	23,2	2 009	
1992	19,0	2 224	 	12,7	4 163	 	33,6	2 900	
1993	19,7	2 406	 	14,3	4 618	 	36,4	3 222	
1994	21,7	2 768	 	17,1	5 311	 	41,3	3 737	
1995	21,2	3 014	 	17,3	5 456	 	40,9	4 010	
1996	21,3	3 112	 	18,7	5 690	 	42,6	4 202	
1997	20,3	3 002	 	19,0	5 663	 	41,7	4 132	
1998	19,7	2 980	 	19,5	5 752	 	41,8	4 189	
1999	...	3 081	 	 	42,8	...	

¹⁾ Bis 1998 Rechnungsergebnisse, für 1999: vorläufige Werte aus den Ergebnissen für das 1. – 4. Quartal.

²⁾ Für Mitglieder (ohne Rentner) und deren Familienangehörige.

³⁾ Für Rentner und deren Familienangehörige.

⁴⁾ Leistungsausgaben, Verwaltungskosten, Vermögensaufwendungen und sonstige Aufwendungen.

⁵⁾ Einschließlich Rentner.

Quelle: Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2000, S. 289

2.3 Folgerungen für die Beziehung zwischen Arzt und Patient

In den letzten zwanzig Jahren wurde mit über 200 Einzelgesetzen versucht, die Beitragssätze, die sich aus dem Verhältnis der Einnahmen und der Ausgaben ergeben, zu stabilisieren (vgl. dazu Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2000, Ziffern 470-471). Dabei waren die Erfolge dieser Kostendämpfungspolitik im allgemeinen nur von kurzer Dauer, da sie nicht auf den Kern des Problems, eine Strukturreform des staatlichen Gesundheitswesens, abzielten. Denn trotz eines hohen Versicherungsgrades, einer hohen Arzt- und Krankenhausedichte, der umfassenden Versorgung mit Heil- und Hilfsmitteln und hohen Forschungsausgaben ist das deutsche Gesundheitswesen im internationalen Vergleich nicht in der Spitzengruppe.¹²

Neben den Defiziten in der medizinischen Versorgung ist das deutsche Gesundheitssystem auch durch falsche Anreizstrukturen, unzureichende Qualitätssicherung und Organisationsmängel gekennzeichnet (vgl. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2000, Ziffern 472 und 473). Dabei sind auf der Arzt-Patient-Ebene fünf ursächliche Fehlanreize und Schwachstellen zu finden:

- Die Entscheidung, ob eine Befindlichkeitsstörung als Krankheit anzusehen ist und sich hieraus ein Versicherungsfall ergibt, liegt im Ermessen des einzelnen.¹³ Je freier die Arztwahl ausgestaltet ist und je geringer die durch den Nachfrager zu tragenden Kosten sind, desto häufiger wird dieser auch bei einer nur geringfügigen Beeinträchtigung des Gesundheitszustands den Arzt aufsuchen. Ein derartiges Verhalten ist für den einzelnen rational, da die Beiträge zur GKV Zwangsabgaben darstellen, denen er weder durch eine gesunde Lebensweise, noch durch eine geringere Nachfrage ausweichen kann. Im Durchschnitt liegt die Anzahl der Arzt-Patienten-Kontakte in Deutschland bei ca. zwölf pro Jahr und damit doppelt so hoch wie in den Vereinigten Staaten oder Großbritannien, im Vergleich mit Skandinavischen Ländern sogar viermal so hoch.

¹² In einem Ranking von 191 Ländern der Weltgesundheitsorganisation (WHO) nimmt Deutschland hinsichtlich des medizinischen Leistungsstandards den 14. Platz und hinsichtlich der Kosten den 25. Platz ein. In einer OECD-Studie wird dem deutschen Gesundheitssystem eine durchschnittliche Gesundheitsleistung bei überdurchschnittlichem Ressourceneinsatz bescheinigt (vgl. Sachverständigenrat für die Begutachtung der gesamtwirtschaftlichen Entwicklung 2000, Ziffer 471).

¹³ So gaben 18 % der ambulanten Patienten an, daß ihre Befindlichkeitsstörungen, die sie zum Besuch eines Arztes veranlaßten, als geringfügig einzustufen sind. Die behandelnden Ärzte sahen sogar 30 % der Gesundheitsstörungen als geringfügig an (vgl. Sachverständigenrat zur Beurteilung der gesamtwirtschaftlichen Entwicklung 2000, Ziffer 485).

- Zweitens bestimmt der Arzt, sofern der Patient ihn erst einmal konsultiert hat, nicht nur die Diagnose und Therapie, sondern aufgrund seiner medizinischen Kenntnisse über sein Angebot hinaus auch die Nachfrage des Patienten nach medizinischen Leistungen. Diese Tendenz zu einer angebotsseitigen Nachfrageausweitung wird dabei durch Abrechnungsverfahren auf der Basis der Vergütung von Einzelleistungen noch verstärkt.
- Drittens wird ein solches exzessives Angebotsverhalten durch eine nur begrenzt mögliche Kontrolle der erbrachten Leistungen sowie der Behandlungsqualität und des -ergebnisses erleichtert.
- Viertens wird ein exzessives Nachfrageverhalten der Patienten, das durchaus als individuell rational angesehen werden kann, durch die für ihn mangelnde Kostentransparenz hinsichtlich der Höhe und Verteilung der entstandenen Kosten begünstigt.¹⁴ In der Regel besitzt der in der GKV versicherte Patient keine Kenntnis über die tatsächlichen Kosten seiner Inanspruchnahme, da das Sachleistungsprinzip vorherrschend ist.
- Fünftens führt die uneingeschränkt freie Arztwahl zu einer Vielzahl von irrelevanten Mehrfachuntersuchungen und u.U. zu beträchtlichen Kostensteigerungen oder auch zu nicht abgestimmten Parallelbehandlungen. Dabei ist Deutschland das einzige Industrieland, in dem eine freie Wahl auch der Fachärzte möglich ist. Es gibt keine vorgelagerte Gatekeeper-Institution (vgl. Breyer und Zweifel 1999, S. 15), bei der bspw. der Hausarzt als Lotse tätig ist und neben der eigenen Behandlung auch die Funktion übernehmen würde, bei einem seinen Kompetenzbereich überschreitenden Krankheitsverdacht oder –befund den Patienten an einen Facharzt weiterzuleiten.¹⁵

Aus diesen bestehenden Organisationsmängeln und fehlenden Anreizstrukturen folgert der Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen, daß es einer Förderung der Eigenverantwortung im Gesundheitswesen bedarf (vgl. SVRKAiG 1994, Ziffer 606). Daneben sollte die zukünftige Steuerung neben der Verbandsebene und der Lenkung durch den Staat auch stärker auf Individualebene stattfinden, so zum Beispiel zwischen Versichertem und Krankenversicherung, Patient und Leistungserbringer oder Krankenversicherer und Leistungserbringer. Dies kann durch preisliche Instrumente, Mengenvorgaben

¹⁴ Zur Überkonsumption von versicherten Gütern und Leistungen aufgrund von Informationsmängeln (Moral Hazard) siehe Abschnitt 3.3.

¹⁵ In der politischen Diskussion wurden solche Ansätze bereits als Primärarztmodell diskutiert (vgl. o.V. 1999, S. 19). Die Lostenfunktion des Hausarztes soll nach Ansicht der Bundesgesundheitsministerin in Zukunft gestärkt werden (vgl. o.V. 2001a, S. 20).

oder Informationen erfolgen (vgl. SVRKAiG 1994, Ziffern 538-601). Besonders die Informationsverteilung zwischen den Beteiligten ist durch eine Schiefelage zugunsten des Arztes geprägt. Er verfügt über die notwendigen Informationen über die zu behandelnde Krankheit und das zur Auswahl der Therapie benötigte Wissen, während der Patient und der Versicherer diese Informationen nicht besitzen. Dies wirkt sich neben der eigentlichen Behandlungssituation vor allem beim Vertragsabschluß über die Honorierung und Leistungserbringung aus. Es ist daher von vorrangigem Interesse, welche Auswirkung die Informationsverteilung auf die Entscheidungen der Beteiligten des Gesundheitswesens ausübt und wie diese Entscheidungen durch die Vertragsgestaltung beeinflußt werden können.

3 Informationsökonomische Grundlagen

3.1 Problematik

Bei jedem Abschluß eines Vertrages, sei er explizit oder implizit, spielt die den Vertragspartnern vorhandene Information und deren strategische Interaktion eine wichtige Rolle. Verträge werden abgeschlossen, wenn sich die involvierten Parteien Tauschgewinne erhoffen (vgl. Schweizer 1996, S. 230). Bedingt durch Informationsasymmetrien kann es allerdings durch Fehlanreize zu einem Abweichen von der Pareto-optimalen Allokation kommen (vgl. Müller 1997, S. 1 und Schweizer 1999, S. 1).¹⁶ Im Gegensatz zur Theorie des allgemeinen Gleichgewichts, die zwar Situationen mit Unsicherheit, aber ausschließlich symmetrischer Informationen behandelt, untersucht die Informationsökonomik Situationen, in denen eine asymmetrische Informationsverteilung vorliegt (vgl. Salanié 1997, S. 1f.).¹⁷

Informationsasymmetrien kommen in nahezu allen Beziehungen innerhalb einer Ökonomie vor. An dieser Stelle der Untersuchung wird zum einen die Informationsverteilung vor dem eigentlichen Vertragsabschluß, zum anderen der Zustand nach dem Vertragsabschluß näher beleuchtet. Den ersten Fall bezeichnet man dabei als adverse Selektion. Hierbei verfügt eine Partei vor Vertragsabschluß über relevante Informationen, die der anderen Partei unbekannt sind. Der zweite Fall, Moral Hazard, ist dadurch gekennzeichnet, daß eine Vertragspartei Handlungen unternimmt, die den Wert des Vertrages für die andere Partei beeinflussen und die diese nicht beobachten und steuern kann (vgl. Kreps 1990, S. 577). Daran schließt sich die Darstellung eines weiteren Bereichs der Informationsökonomik, der Principal-Agent-Theorie, an. Hier werden im weitesten Sinne Informationsasymmetrien und die Vertragsgestaltung in Situationen der Delegation von Aufgaben analysiert. In diesem Zusammenhang ist es prinzipiell möglich, beide Situationen, adverse Selektion und Moral Hazard, zu betrachten. Die Darstellung in diesem Abschnitt beziehen sich lediglich auf die Principal-Agent-Theorie im Fall des Moral Hazard.

Zunächst ist es notwendig, einige grundlegende Begriffe, Annahmen und Vorgehensweisen der Informationsökonomik und der Vertragstheorie kurz zu beleuchten (vgl. hierzu Salanié 1997, S. 2f.). Zum einen werden im größten Teil

¹⁶ Stiglitz (2000) gibt einen Überblick über den Beitrag der Informationsökonomik zu der Entwicklung der Wirtschaftswissenschaften.

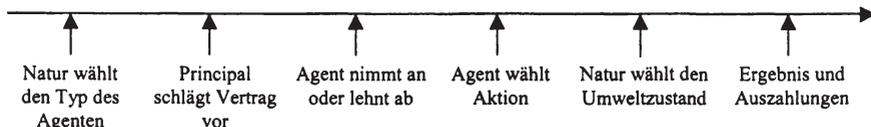
¹⁷ Unvollständige, aber symmetrische Information würde keiner Vertragsseite einen Vorsprung verschaffen, den diese ausnutzen könnte. Dies hat zur Folge, daß Unsicherheit alleine die kompetitive Allokation nicht ändert (vgl. Arrow 1970, S. 121ff.).

der Untersuchungen Partialmodelle verwendet, die Märkte mit einem oder wenigen Gütern vom Rest einer Ökonomie trennen. Weiterhin beschreiben diese Modelle Interaktionen einer kleinen Zahl von Agenten, die über unterschiedliche Informationen verfügen. Vorliegende Restriktionen, seien es institutionell gegebene oder stillschweigend akzeptierte, werden dabei in einem Vertrag zusammengefaßt. Dieser kann sowohl schriftlich fixiert sein (explizit), als auch implizit, d.h. auf einem System von Verhaltensnormen basierend. Dabei ist es vor allem wichtig, die Anreize dieser Verträge auf die Handlungen der einzelnen Akteure zu berücksichtigen. Explizite Verträge werden dabei von einer dritten Partei, bspw. einem Gericht oder einem Vermittler, garantiert. Daneben gewährleistet auch der Wunsch des Agenten, Reputation aufzubauen oder zu erhalten, die Einhaltung eines Vertrages. Dagegen resultiert bei einem impliziten Vertrag das Gleichgewicht aus den Interaktionen der betroffenen Parteien. Darüber hinaus basieren Modelle der Vertragstheorie auf der nichtkooperativen Spieltheorie, obwohl der zugrunde liegende Verhandlungsprozeß sich oft durch eine einfache Principal-Agent-Beziehung charakterisieren läßt.

3.2 *Adverse Selektion*

Betrachtet man die Situation vor Abschluß eines Vertrages, so existiert hierbei eine asymmetrische Informationsverteilung derart, daß eine Vertragspartei die Qualität einer Sache oder einer Person kennt, während die andere Partei nur Informationen über die durchschnittliche Qualität besitzt. Mit anderen Worten kann ein Auftraggeber (Principal) zwar das Verhalten des Handlungsgehilfen (Agent) überprüfen, deren optimale Entscheidung oder die Kosten dieser Entscheidung hängen aber vom Typ des Agenten ab, bspw. von bestimmten Produktions-Charakteristika, über die nur der Agent informiert ist (vgl. hierzu Macho-Stadler und Pérez-Castrillo 1997, S. 11). Besteht die asymmetrische Informationssituation über persönliche Charakteristika des Agenten, so besitzt der Principal lediglich Kenntnis über die verschiedenen Typen von Agenten, kann diese aber nicht unterscheiden. Vereinfacht läßt sich die Entscheidungssituation mit Hilfe eines Zeitstrahls darstellen (siehe Abbildung 3.1). Hierbei wählt ein neutraler Spieler, der als Natur bezeichnet wird, den Typ des Agenten, d.h. es werden dessen Charakteristika festgelegt. Im zweiten Schritt bietet der Principal dem Agenten einen Vertrag an. Dieser kann den Vertrag akzeptieren oder ablehnen. Lehnt er ihn ab, so endet die Beziehung. Nimmt er dagegen an, so wählt der Agent seine Aktion. Das Ergebnis dieser Aktion muß nun nicht immer deterministisch von ihr abhängen, sondern kann ebenso von Umweltbedingungen beeinflußt werden, die wiederum die Natur festlegt. In einem letzten Schritt wird das Ergebnis beobachtet und die Pay-offs werden realisiert.

Abbildung 3.1: Zeitliche Struktur des Problems adverser Selektion



Quelle: Vgl. Macho-Stadler und Pérez-Castrillo 1997, S. 11.

Bei adverser Selektion verdrängt schlechte Qualität die gute vom Markt (vgl. Akerlof 1970, S. 489). Grundsätzlich geht man davon aus, daß in einem Markt ein bestimmtes Gut in unterschiedlicher Qualität angeboten wird. Wenn nun bei einer Transaktion, d.h. einem Kauf, nur eine Vertragsseite (Verkäufer) im voraus Informationen über die Qualität besitzt, muß die andere Vertragspartei (Käufer) damit rechnen, nur eine negative Auswahl aus dem Angebot zu erhalten (vgl. Kreps 1990, S. 626). Sie ist dann nur bereit, einen relativ geringen Preis für das Produkt zu bezahlen. Am Beispiel des Automobilmarktes ist ersichtlich, daß durch die asymmetrische Information zwischen Verkäufer und Käufer Gebrauchtwagen schlechter Qualität diejenigen guter Qualität vom Markt verdrängen, da beide zum gleichen Preis gehandelt werden. Die Übertragung dieses Argumentes auf die Krankenversicherung ergibt, daß Personen, die älter als 65 Jahre sind und ein höheres Krankheitsrisiko darstellen, nicht versichert werden. Der Grund dafür, daß selbst höhere Prämien das Risiko nicht auffangen können, liegt darin, daß „[...] as the price level rises, the people who insure themselves will be those who are increasingly certain that they will need the insurance“ (Akerlof 1970, S. 492). Höhere Prämien führen immer dann zu einem höheren Anteil schlechter Risiken, wenn bessere Risiken nicht bereit sind, diese hohen Prämien zu zahlen. Der durchschnittliche Gesundheitszustand der Versicherten verschlechtert sich deshalb.¹⁸

Für den Fall zweier Risikogruppen läßt sich der Zusammenhang vereinfacht beschreiben: „By their very being, low-risk individuals are worse off than they would be in the absence of high-risk individuals“ (Rothschild und Stiglitz, 1976, S. 629). Im Gegensatz dazu werden Patienten mit hohem Risiko durch die An-

¹⁸ Einen ähnlichen Effekt kann man für die Ärzteseite annehmen. Falls die für die GKV ausgehandelten Honorare sinken, werden hochqualifizierte Ärzte ihre Lizenz zurückgeben und nur noch private Patienten behandeln. Dies beruht auf ihrer Überlegung, daß sie aufgrund ihrer Qualifikation außerhalb der GKV bessere Verdienstmöglichkeiten besitzen als bei den niedrigen Tarifen. Durch eine Absenkung der Honorare sinkt also die Zahl der hochqualifizierten Ärzte in der GKV, auch wenn dieser Trend durch ebenfalls sinkende Honorare im privaten Bereich zum Teil konterkariert wird (vgl. Neudeck 1989, S. 102f.)

wesenheit guter Risiken nicht beeinflusst. Der Umstand, daß gute Risiken keinen umfassenden Versicherungsschutz zu fairen Konditionen finden, stellt einen negativen externen Effekt der schlechten Risiken auf die guten dar (vgl. Strassl 1988, S. 125 und Breyer und Zweifel 1999, S. 161ff.).

Versicherungen können in einer solchen Situation Pooling-Verträge oder separierende Verträge anbieten. Pooling Verträge liegen vor, wenn für alle Versicherten ein Durchschnittsvertrag angeboten wird. Bei separierenden Verträgen liegt ein Gleichgewicht vor, bei dem die Versicherungsnachfrager sich selbst in den Vertrag ihrer jeweiligen Gruppe durch Selbstselektion einordnen sollen (Screening-Verträge). Dann werden die schlechten Risiken eine Vollversicherung zu einer höheren Prämie nachfragen, während eine Teilsicherung nur von den guten Risiken nachgefragt wird (vgl. Kortendieck 1993, S. 212f.). Es läßt sich zeigen, daß in einem kompetitiven Markt ein Pooling-Gleichgewicht nicht existiert, während separierende Gleichgewichte zwar möglich sind, teilweise jedoch nicht die Eigenschaft der Stabilität aufweisen.¹⁹

In diesem Zusammenhang wird oft die Frage gestellt, ob in einer staatlichen Regulierung des Versicherungsmarktes eine Lösung des Problems der adversen Selektion liegt und ob ein Eingriff die Existenz eines Gleichgewichtes sichert. Eine Möglichkeit liegt in einer staatlichen Monopolversicherung, bei der durch Ausschaltung des Wettbewerbes ein separierendes Gleichgewicht nicht durch die Abwanderung guter Risiken zu anderen Versicherungen gestört werden kann. Dieser Eingriff erscheint dann probat, wenn ansonsten kein Markt für die betreffende Versicherungsleistung zustande kommen würde (vgl. Strassl 1988, S. 179). Allerdings gibt Strassl zu bedenken, ob ein solcher drastischer staatlicher Eingriff mit dem Subsidiaritätsprinzip noch verträglich ist und ob nicht auch ein privater, regulierter Markt eine vergleichbare Second-best-Lösung hervorbringt.²⁰

Eine weitere Möglichkeit ist eine Versicherungspflicht und ein gleichzeitiges Angebot eines Durchschnittsvertrages (vgl. Kortendieck 1993, S. 208). Die Grundüberlegung besteht darin, daß schlechte Risiken grundsätzlich eine höhere Prämie zahlen müßten, da sie von Natur aus anfälliger für Krankheiten sind (vgl. Breyer und Zweifel, 1999, S. 164f.). Da dies oft als verteilungspolitisch

¹⁹ Die Probleme der Stabilität eines Gleichgewichtes sollen in diesem Zusammenhang nicht näher erläutert werden. Siehe hierzu Rothschild und Stiglitz (1976).

²⁰ Das Prinzip der Subsidiarität kann als ein wichtiges sozialpolitisches Zuordnungsinstrument angesehen werden. In diesem Zusammenhang sollen die höheren Einheiten „die Rahmenbedingungen und Voraussetzungen schaffen, damit die (...) untergeordneten Einheiten die ihnen derart übertragenen Aufgaben zu lösen in der Lage sind“ (SVRKAiG 1994, Ziffer 39-41).

unerwünscht angesehen wird, erreicht man durch den Kontrahierungszwang, daß schlechte Risiken nicht diskriminiert, gleichzeitig aber auch die guten Risiken versichert werden und diese somit die schlechten subventionieren (vgl. Kortendieck 1993, S. 209f.).²¹ Vor diesem Hintergrund läßt sich folgende Aussage formulieren: „Ein Ausgleich unterschiedlicher Startchancen läßt sich nicht durch die Außerkräftsetzung des Marktmechanismus und durch staatliches Angebot von Gesundheitsleistungen bewerkstelligen, wohl aber durch gesetzliche Versicherungspflicht, verbunden mit Kontrahierungszwang und dem Verbot, die Versicherungsprämie vom Gesundheitszustand abhängig zu machen“ (Breyer und Zweifel 1999, S. 164). Das Erreichen einer Pareto-optimalen Situation durch staatliche Eingriffe ist allerdings nicht gewährleistet, da der Gesetzgeber den gleichen Informationsbeschränkungen wie ein Versicherer unterliegt (vgl. Strassl 1988, S. 178).

3.3 *Moral Hazard*

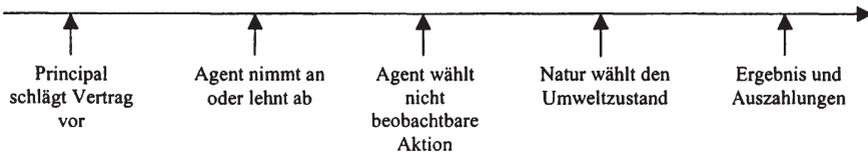
Die Situation nach Vertragsabschluß weist ebenfalls Informationsprobleme auf, die eine Pareto-optimale Allokation verhindern.²² „We consider the problem of moral hazard, where one party to a transaction may undertake certain actions that (a) affect the other party's valuation of the transaction but that (b) the second party cannot monitor/enforce perfectly“ (Kreps 1990, S. 577). Wichtig dabei ist die Beeinflussung des Vertragswertes eines Vertragspartners, ohne daß dieser auf Handlungen Einfluß nehmen könnte bzw. die Handlungen beobachten oder überprüfen könnte (vgl. Theilen 1996, S. 9). Kennzeichnend für dieses Problem ist wiederum die Existenz asymmetrischer Information zwischen zwei Vertragsparteien, in diesem Fall nach Vertragsabschluß. Der Wert des Vertrages hängt dabei vom resultierenden Output und damit bspw. von den Handlungen des Versicherungsnehmers ab. Die häufig anzutreffenden deutschen Begriffe moralisches Risiko oder moralische Versuchung sind insofern irreführend, da Moral Hazard keine moralisch verwerfliche Handlung darstellt, sondern lediglich auf das rationale ökonomische Verhalten des Individuums abstellt (vgl. Pauly 1968, S. 535). Die Entscheidungssituation kann auch hier anhand eines Zeitstrahls dargestellt werden (vgl. Abbildung 3.2).

²¹ Für die GKV läßt sich eine zweifache Umverteilung feststellen: Zum einen durch die Unabhängigkeit des Beitrages vom individuellen Krankheitsrisiko zugunsten der Krankheitsanfälligen, zum anderen durch den festen Prozentsatz des Einkommens zu den Einkommenschwachen (vgl. Breyer und Zweifel 1999, S. 165). Weiterhin findet eine Umverteilung von den jungen Versicherten zu den alten aufgrund der Multimorbidität und von Ledigen zu Familien durch die beitragsfreie Mitversicherung von Familienmitgliedern statt.

²² Vgl. zu einer Übersicht zur Moral Hazard-Problematik Kotowitz (1987) und Arnott und Stiglitz (1988).

Zuerst schlägt der Principal dem Agenten einen Vertrag vor, den der Agent annimmt oder ablehnt (take it or leave it). Anschließend wählt der Agent seine Handlung, die nicht durch den Principal beobachtet werden kann. Die Natur wählt dann den Umweltzustand, der zusammen mit der Aktion des Agenten das Ergebnis und die Auszahlungen determiniert. Eine alternative Formulierung des Problems ist, daß der Agent, bevor er seine Aktion wählt, den Naturzustand beobachtet, der Principal hingegen nicht. Die Unsicherheit ist für beide Akteure gleich, wenn der Vertrag aufgesetzt wird. Bevor der Agent seine Aktion wählt, erhält er aber zusätzliche Informationen. Die erste Zeitstruktur repräsentiert den Versicherungssektor, während die zweite Situation z.B. bei der Delegation von Aufgaben über räumliche Entfernungen Anwendung findet, bei der regionale Besonderheiten nur dem Agenten bekannt sind (vgl. Macho-Stadler und Pérez-Castrillo 1997, S. 10).

Abbildung 3.2: Zeitliche Struktur bei Moral Hazard



Quelle: Vgl. Macho-Stadler und Pérez-Castrillo 1997, S. 9.

Beispiele für Moral Hazard-Situationen sind die Beziehung zwischen Arbeitnehmer und Arbeitgeber, Landarbeiter und Landbesitzer, Arzt und Patient sowie Versicherungsbeziehungen. Der Großteil der Literatur bezieht das Phänomen auf letzteren Fall und begründet das Vorgehen damit, daß die anderen Beziehungen ebenfalls den Charakter einer Versicherungsleistung hätten.²³ Der Handlungsspielraum des Versicherungsnehmers beruht dabei auf der Unbestimmtheit des Vertrages, so daß der Versicherer das Risiko nicht näher bestimmen kann (vgl. Arrow, 1970, S. 142). Einerseits sieht er sich einem unvermeidlichen Risiko gegenüber, das es grundsätzlich zu versichern gibt, andererseits dem Risiko aus den Entscheidungen des Versicherten, d.h. einem subjektiven Risiko, bspw. durch mangelnde Vorsorge (vgl. Hackmann 1990, S. 71f.). Eine speziell auf den Versicherungsmarkt zugeschnittene Definition lautet: „*Moral hazard problems arise, when there is imperfect information concerning the actions of these who purchase insurance, because those actions cannot be*

²³ Das Verhältnis zwischen Landbesitzer und Pächter kann insofern als Versicherung aufgefaßt werden, da auch für diese Beziehung die optimale Risikoaufteilung gesucht wird (vgl. Stiglitz 1983, S. 6).

perfectly monitored and the insurance contract cannot specify all of the actions which the insured is to undertake. Under those circumstances, the provision of insurance implies that individuals do not bear fully the consequences of their actions.“ (Stiglitz 1983, S. 5) Dies bedeutet zum einen, daß das Verhalten des Versicherten sein Risiko beeinflußt und zum anderen, daß die Aufwendungen und die Sorgfalt, die zur Vermeidung von Schadensfällen getroffen werden, nicht im Vertrag festgelegt und vom Versicherer geltend gemacht werden können (vgl. Winter 1992, S. 61). Die Folgen des Auftretens von Moral Hazard gleichen denjenigen eines externen Effektes. Dafür ist es notwendig, daß der Versicherte einen Anreiz besitzt, mehr (medizinische) Leistungen in Anspruch zu nehmen, als es der Vertrag ex ante vorsieht (vgl. Kortendieck 1993, S. 194).

Oft wird als zusätzliche Annahme für Moral Hazard aufgeführt, daß der Zusammenhang zwischen dem Ergebnis und den Handlungen der einen Vertragsseite aufgrund der Existenz eines exogenen Zufallsprozesses bzw. sogenannter verschiedener Zustände der Welt für die andere Vertragsseite nicht klar erkennbar ist (vgl. Nell 1993, S. 103). Dies erweist sich als relevant, wenn die Handlungen nicht beobachtbar sind, sondern nur das Ergebnis (siehe hierzu Abschnitt 3.4). Eine Klassifikation der Formen des Moral Hazard kann in Abhängigkeit der Anreize, der Wirkung und des zeitlichen Auftretens vorgenommen werden. Problematisch bei einer systematischen Übersicht sind vor allem die unterschiedlichen Ansatzpunkte und die Vielzahl von Definitionen (siehe hierzu Nell 1993, S. 102ff.). Eine Übersicht über die hier dargestellten Varianten ist Abbildung 3.3 zu entnehmen.

Abbildung 3.3: Abgrenzungen des Moral Hazard



Eine in der Literatur weniger gebräuchliche Unterscheidung der Klassifikation von Moral Hazard-Problemen besteht in der Einteilung nach der Anreizquelle, also in *internes und externes Moral Hazard* (vgl. Nell 1993, S. 102ff.). Als

internes Moral Hazard bezeichnet man versicherungsinduzierte Verhaltensänderungen, d.h. Reaktionen, die der Patient nur aus dem Verhältnis zu seiner Krankenversicherung heraus unternimmt. Daneben ergeben sich durch den Einfluß der Ärzte auf die Nachfrage des Patienten nach medizinischen Leistungen Rückwirkungen auf den Versicherungsvertrag (externes Moral Hazard).²⁴

Einen weiteren Ansatzpunkt für die Strukturierung der Moral Hazard-Probleme bieten die Auswirkungen der Reaktionen des Versicherungsnehmers auf den Versicherungsvertrag (vgl. Pauly 1970, S. 414f.). *Moral Hazard* ist dann *risikoerhöhend*, wenn durch die Versicherung die Aktivitäten zur Schadensvermeidung dergestalt beeinflußt werden, daß das Engagement des Versicherungsnehmers sinkt.²⁵ Beispiele hierfür sind eine riskantere Lebensweise, wie das Ausüben gefährlicher Sportarten wie Bergsteigen oder Motorradrennen, oder die Vernachlässigung von Vorsorgeuntersuchungen. *Mengenerhöhendes Moral Hazard* liegt vor, wenn das Ausmaß des Schadens durch das Verhalten des Versicherungsnehmers beeinflußt wird. Beispiele hierfür sind das Motorradfahren ohne Sturzhelm oder der Einfluß des Patienten auf den Heilungsprozeß, wenn bspw. eine Atemwegserkrankung verschleppt wird (vgl. Kortendieck 1993, S. 194). Die dritte Variante ist der Fall eines *preiserhöhenden Moral Hazards*, bei der die höhere Nachfrage nach Gesundheitsleistungen die Preise der Leistungsanbieter ansteigen läßt. Dies ist allerdings nur möglich, sofern die Preise nicht einer staatlichen Festsetzung unterliegen. Eine ähnliche Wirkung kann auch bei externem Moral Hazard auftreten (vgl. Kortendieck 1993, S. 194).

Eine andere Abgrenzung im Falle der Krankenversicherung besteht in der Unterscheidung in zeitlicher Hinsicht.²⁶ Betreibt der Patient infolge einer Krankenversicherung vor Schadenseintritt keinen optimalen Aufwand zur Schadensvermeidung, liegt ex ante Moral Hazard vor. Die Schadensverminderungsaktivität des Versicherungsnehmers ist dabei um so niedriger, je höher der Deckungsgrad der Versicherung ist (vgl. Nell 1993, S. 223). Insbesondere bei einer Vollversicherung liegt ex ante Moral Hazard in der stärksten Ausprägung vor. Im Fall des ex post Moral Hazard besteht für den Patienten der Anreiz, die Intensität und den Umfang der Behandlung zu beeinflussen. Dies bedeutet, daß ein Patient im Krankheitsfall aufgrund des Versicherungsschutzes kein gesteig-

²⁴ Dieser Fall wird in der Literatur auch unter dem Begriff anbieterinduzierte Nachfrage behandelt (vgl. hierzu Kapitel 4.1)

²⁵ Die Schadensvermeidung bzw. der Konsum von Schadensvermeidungsgütern kann keinen unmittelbaren Nutzen für den Versicherten entfalten, sondern wird den Erwartungsnutzen durch die Senkung der Schadenswahrscheinlichkeit beeinflussen (vgl. Strassl 1988, S. 37).

²⁶ Siehe hierzu u.a. Breyer und Zweifel 1997, S. 186ff., Nell 1993, S. 102ff., Pauly 1986, S. 640 oder auch Strassl 1988, S. 39.

gertes Interesse an einer raschen Gesundung an den Tag legt mit der Konsequenz, daß der Arzt ihn eventuell erneut krank schreibt oder die Nachfrage nach medizinischen Leistung ansteigt.

Risikoerhöhendes und ex ante Moral Hazard erscheinen auf den ersten Blick identisch zu sein. Schließt man jedoch die Verschlimmerung einer Krankheit oder die zusätzliche Erkrankung, die durch mangelnde Vermeidungsaktivitäten hinzutritt, von der Betrachtung aus, so bedeutet dies, daß ein risikoerhöhendes Moral Hazard ex post nicht auftreten kann. Ein mengenerhöhendes Moral Hazard kann dagegen sowohl ex ante stattfinden als auch ex post, was den Fall der exzessiven Nachfrage widerspiegelt (vgl. Kortendieck 1993, S. 194).

Ein wichtiger Aspekt, der in der Diskussion um Moral Hazard oft vernachlässigt wird, besteht in der Gesamtwirkung einer Versicherung auf den Patienten, die über den Konsum von Gesundheitsleistungen hinausgeht. Die wichtigsten Effekte des Moral Hazard-Phänomens bestehen in einem größeren Erwartungsschaden und einer höheren Nachfrage nach Gesundheitsleistungen (vgl. Kortendieck 1993, S. 194). Die Einstufung dieser Wirkungen als negativ vernachlässigt einen wichtigen Aspekt, da „es der Sinn einer Krankenversicherung sein [soll], einen Teil der Risiken des Versicherten zu übernehmen, so daß dieser infolge der Wohlfahrtssteigerung und des Einkommenseffektes eine größere Nachfrage insgesamt entfalten kann“ (Kortendieck 1993, S. 192).²⁷ Anzumerken bleibt jedoch, daß eine solche Sichtweise möglicherweise die Finanzierungsprobleme im Gesundheitswesen verharmlost oder sie zumindest bewußt in Kauf nimmt.

Ob eine staatliche Regulierung die negativen Auswirkungen des Moral Hazard eindämmen kann, ist nicht eindeutig. Möglichkeiten der Beeinflussung des Umfangs des Moral Hazard bieten sich nur, falls das Verhalten des Versicherten beobachtet werden kann. Dies betrifft zum Beispiel den Konsum risikobeeinflussender Güter. Bei der Entscheidung, welche Güter von einem Eingriff betroffen sein sollten, maßt sich der Staat ein Wissen über die Nutzenwirkungen und externen Effekte des Konsums der Versicherten an, das er nicht besitzt.²⁸ Im Sinne der Versicherung wäre eine Besteuerung des Konsums

²⁷ Im Falle der Eigenvorsorge durch den Versicherungsnehmer liegt der Aufwand in Höhe des Maximalrisikos vor, während die Versicherung nur die Absicherung des Durchschnittsrisikos in ihre Prämie aufnimmt (vgl. Kortendieck 1993, S. 192).

²⁸ Ob solche Eingriffe grundsätzlich durch den Staat vorgenommen werden sollten, hängt davon ab, ob der Staat oder die Versicherung die niedrigeren Kosten bei der Beobachtung des Versichertenverhaltens aufweist. Problematisch ist vor allem, daß staatliche Eingriffe auch das Verhalten und die angebotenen Verträge der Versicherungen beeinflussen können (vgl. Bruce und Wong 1996, S. 77ff.).

risikoerhöhender Güter und eine Subvention von Schadensvermeidungsgütern optimal. Dies könnte jedoch auch durch die individuelle Vertragsgestaltung mit Hilfe von Bonussystemen und Haftungsausschlußregelungen erreicht werden.

3.4 Die Grundstruktur von Principal-Agent Problemen

Hierarchische Beziehungen bzw. die Delegation von Aufgaben finden sich in fast allen Bereichen einer Ökonomie. Individuen führen Tätigkeiten nicht selbst aus, sondern übergeben diese an Dritte, die sich auf die Ausführung dieser Aufgaben spezialisiert haben. Die Ursachen für die Delegation basieren auf einer begrenzten Informationsverarbeitungskapazität, beschränktem Wissen oder knapper Zeit und resultieren letztlich aus den komparativen Vorteilen einzelner Individuen in bezug auf bestimmte Tätigkeiten (vgl. Pfaff und Zweifel 1998, S. 2). Die ökonomischen Konsequenzen einer ungleichen Informationsverteilung zwischen den Vertragspartnern bei divergierenden Zielen bestehen darin, daß eine Vertragsseite ihren diskretionären Handlungsspielraum zu ihren Gunsten ausnutzen kann.²⁹

Die mikroökonomische Theorie behandelt diese Schwierigkeiten im Rahmen der Principal-Agent-Theorie.³⁰ Ein Auftraggeber (Principal) delegiert eine Aktion an seinen Agenten. Sein Nutzen hängt von der Entscheidung des Agenten ab. In vielen Fällen kann er diese Handlung jedoch nicht beobachten, sondern nur das Ergebnis der Aktion. Es besteht dabei eine Form des Moral Hazard, die als einseitig beschrieben werden kann (vgl. Dutta und Radner 1994, S. 870).³¹ Bedingt durch die Unwissenheit des Principals bezüglich der Aktionen, ergeben sich durch den Handlungsspielraum des Agenten vielschichtige Zusammenhänge. Insbesondere bei den Anreizstrukturen und den damit verbundenen Entlohn

²⁹ Müller (1997) behandelt die Grundlagen des Moral Hazard in der Principal-Agent-Theorie und geht dabei auch auf zahlreiche Erweiterungen des Grundmodells wie wiederholte Delegation oder dynamische Principal-Agent-Probleme ein.

³⁰ Im Kontext von Moral Hazard-Phänomenen erscheint es sinnvoll, gleichzeitig über Principal-Agent-Strukturen zu sprechen (siehe hierzu u.a. Dutta und Radner 1994). Die Trennung beider Gebiete ist allerdings insofern geboten, da die Principal-Agent Theorie nur eine von vielen möglichen theoretischen Spielformen des Umgangs mit der Moral Hazard-Problematik darstellt. Einen Überblick über die Theorie und die Umsetzung verschiedener Problembereiche geben Bamberg und Spremann (1987), Hart und Holmström (1987), MacDonald (1984), Rees (1985) und Rees (1985a).

³¹ Im Gegensatz dazu bezeichnet man ein mehrseitiges Moral Hazard mit mehreren Agenten jedoch ohne Principal als „partnership“ (Dutta und Radner 1994, S. 870). In der Principal-Agent-Theorie wird zwar immer nur das Moral Hazard des Agenten betrachtet, in der Realität ist ein solches Verhalten des Principals jedoch nicht auszuschließen.

nungssystemen ergeben sich komplexe Lösungen der Delegationsproblematik (vgl. Levinthal 1988, S. 155f.).

Aufgrund der bestehenden Informationsasymmetrie kann eine Pareto-optimale Situation nicht erreicht werden. Ein Ziel der Theorie besteht deshalb darin, die Struktur einer Beziehung so zu organisieren, daß der Wohlfahrtsverlust (agency cost bzw. loss) minimal ist (vgl. Pratt und Zeckhauser 1985, S. 3). Die Theorie läßt sich in einen positiven und einen normativen Zweig unterscheiden (vgl. Fischer 1995, S. 320 und Eisenhardt 1989, S. 59ff.). Die positive Theorie findet ihren Schwerpunkt in der Beschreibung und Erklärung von Auftragsbeziehungen. Die normative Theorie versucht, die durch die Annahmen erzielten Ergebnisse in Empfehlungen bezüglich der Gestaltung von Beziehungen abzuleiten. Dabei wird untersucht, unter welchen Bedingungen und vertraglichen Vereinbarungen eine Lösung Pareto-optimal sein kann.

Für die theoretische Ausgestaltung der Agency-Beziehungen sind eine Vielzahl von Varianten möglich. In der Literatur findet sich eine Spezialisierung auf Fälle, in denen die Handlungen des Agenten nicht direkt vom Principal zu beobachten sind. Der Output wird dabei zwar durch die Handlung beeinflusst, jedoch nicht vollständig determiniert (vgl. Arrow 1985, S. 37). Das bedeutet, daß der Principal nicht vom Output auf die Handlungen seines Agenten schließen kann.

Die Folgen lassen sich am Beispiel des Landbesitzers und des Pächters illustrieren. Die Ernte, die der Pächter für den Besitzer einfährt, hängt von seinen Anstrengungen beim Anbau und der Pflege, aber auch vom Einfluß des Wetters ab. Für das Ergebnis der Ernte sind somit zwei Faktoren verantwortlich, deren unterschiedliche Ausprägungen die Interpretation des Aufwands des Agenten erschweren. Zum einen besteht die Möglichkeit, daß der Agent hart gearbeitet hat, es durch einen verregneten Sommer allerdings zu einer Mißernte kommt. Umgekehrt hat der Agent möglicherweise nur minimale Anstrengungen unternommen, bedingt durch hervorragendes Wetter ist die Ernte jedoch überdurchschnittlich gut. Wäre die Bezahlung des Agenten allein vom Umfang der Ernte abhängig, so würde er im ersten Fall ungerechtfertigt bestraft, im zweiten würde sein Müßiggang noch belohnt werden. Es muß also eine Entlohnungsform gefunden werden, die den Agenten gerecht bezahlt und seine Anstrengungen für den Principal fördert.

Grundsätzlich lassen sich zwei Klassen von Modellen unterscheiden Hidden Action und Hidden Information (vgl. Arrow 1985, S. 38ff.).³² Letzteres spiegelt den Fall wider, in dem der Agent dem Principal Informationen vorenthält. Mit Hilfe dieses Wissen trifft er sodann seine Entscheidung. Der Principal dagegen kann nicht überprüfen, ob der Agent die Information in seinem Interesse verwendet oder ob er sich eigennützig verhalten hat. Hidden Action hingegen spiegelt die bereits angesprochene Situation wider, in welcher der Principal die Handlung des Agenten nicht genau beobachten kann.

Die Umsetzung der Agency-Überlegungen in ein theoretisches Modell erfolgt durch ein Erwartungsnutzenmodell.³³ Hierbei maximiert der Principal seinen Erwartungsnutzen unter Beschränkungen, die sich aus dem Verhalten des Agenten ergeben. Die Unsicherheit bezüglich des durch die Handlung des Agenten beeinflussten Outputs, kann sowohl diskret mit expliziten Wahrscheinlichkeiten bestimmter Ergebnisse als auch stetig modelliert werden. Aus der Vielzahl der Ansätze wird an dieser Stelle der Holström-Mirrlees-Ansatz vorgestellt (vgl. Holmström 1979 und Mirrlees 1976).³⁴ Zwei Gründe sprechen für dieses Vorgehen: Zum einen lieferten diese Arbeiten einen großen Beitrag zur Entwicklung der Principal-Agent-Modelle, zum anderen orientiert sich das in Abschnitt 4.2 vorgestellte Modell formal sehr eng an diesen Ausführungen.

In der Principal-Agent-Beziehung unternimmt der Agent Handlungen $a \in A \subseteq \mathbb{R}$, wobei A die Menge aller möglichen Aktionen innerhalb der reellen Zahlen beschreibt. Diese sind durch den Principal nicht direkt zu beobachten. Er nimmt nur das monetäre Ergebnis x dieser Handlung wahr, das zusätzlich durch eine Zufallsvariable θ beeinflusst wird. Diese Variable läßt sich als den realisierten Zustand der Welt interpretieren. Das Ergebnis kann somit als Funktion der Handlungen und der Zufallsvariable geschrieben werden als $x = x(a, \theta)$.

³² Die beiden Fälle werden auch als Moral Hazard und adverse Selektion bezeichnet. Diese Begriffsdefinition ist jedoch nur bedingt anwendbar, da Agency-Modelle nur eine bestimmte Ausprägung der beiden Probleme darstellen (vgl. Arrow 1985, S. 38). Darüber hinaus unterscheiden Picot et al. (vgl. 1997, S. 85ff.) zusätzlich in Hidden Characteristics und Hidden Intention. Falls der Principal unveränderliche Eigenschaften des Agenten oder seiner Leistung erst ex post, d.h. nach Vertragsabschluß, erfährt, spricht man von Hidden Characteristics. Hidden Intention liegt vor, wenn der Principal opportunistisches Verhalten des Agenten erkennen, jedoch nicht verhindern kann.

³³ Zu einer Einführung in die Probleme der Entscheidung unter Unsicherheit siehe bspw. Henderson und Quandt (1983) und Varian (1996).

³⁴ Einen ersten Ansatz zu dem Problem lieferte Ross (1973). Allerdings vernachlässigte er den negativen Nutzen, der sich aus dem Aufwand für den Agenten ergibt.

Die Nutzenfunktion des Principals sei nur über sein Vermögen w definiert als $G(w)$, die des Agenten über dessen Vermögen w und seine Handlungen a als $H(w, a)$. Zur Vereinfachung wird angenommen, daß die Nutzenfunktion des Agenten additiv-separabel in Vermögen und Aktionen ist: $H(w, a) = U(w) - V(a)$ mit $V'' > 0$. Zusätzlich gilt, daß durch eine marginale Erhöhung der Aktion des Agenten das Ergebnis nicht schlechter wird ($x_a \geq 0$). Die Handlung des Agenten kann als produktiver Inputfaktor beschrieben werden, der darüber hinaus negative Auswirkungen auf dessen Nutzen ausübt, so daß deshalb die Zielvorstellungen des Principals und des Agenten differieren (vgl. Holmström 1979, S. 76). Der Nutzen aus w steigt für beide Akteure mit zunehmendem Vermögen, und für beide wird angenommen, daß sie nicht risikofreudig sind, d.h. $U'' \leq 0$ und $G'' < 0$.

Da der Principal lediglich den Output x beobachten kann, wird er die Entlohnungsfunktion $s(x)$ nur von diesem abhängig machen. Der verbleibende Anteil des Principals beträgt folglich $r(x) = x - s(x)$. Bezüglich der Verteilungsfunktion von θ wird angenommen, daß diese beiden Parteien bekannt ist.³⁵ Der Agent wählt seine Handlung, bevor die Ausprägung von θ bekannt ist. Der Maximierungsansatz lautet:

$$\max_{s(x), a} E = \left\{ G \left(x - s(x) \right) \right\} \tag{3.1}$$

$$\text{u.d.N.} \quad E \left\{ G \left(s(x), a \right) \right\} \geq \bar{H} \tag{3.2}$$

$$a \in \underset{a \in A}{\operatorname{argmax}} E \left\{ H \left(s(x), a' \right) \right\}. \tag{3.3}$$

Die Nebenbedingung (3.2) stellt sicher, daß der Agent bei Annahme des Vertrages einen Mindestnutzen von \bar{H} garantiert bekommt. Diese stellt die Teilnahmebedingung (participation constraint) dar. Die Anreizbedingung (incentive compatibility constraint) des Agenten ist durch (3.3) gegeben.³⁶ Er soll seine Handlungen so wählen, daß sie für ihn optimal sind. Falls der Principal die Aktion a beobachten könnte, würde ein sogenannter ‚forcing contract‘ sicherstellen, daß der Agent die notwendige Handlung wählt, auch wenn die Honorierung

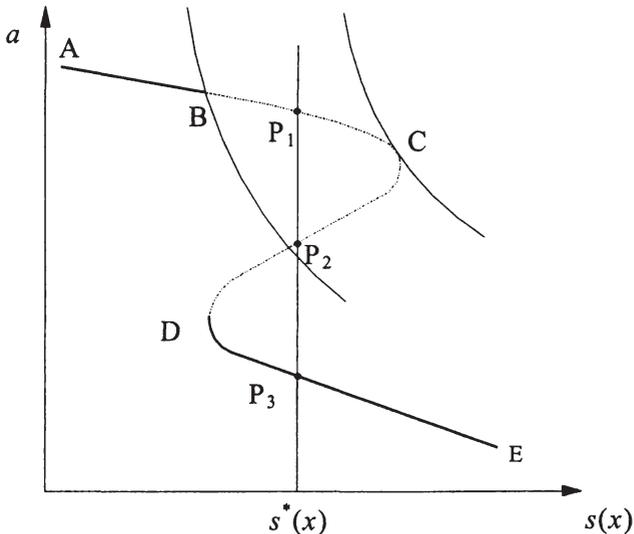
³⁵ Diese Annahme bezüglich der Informationsstruktur des Problems dient lediglich der Vereinfachung.

³⁶ Der Ausdruck „argmax“ beschreibt die Menge der Argumente, welche die nachfolgende Funktion maximiert (vgl. Holmström 1979, S. 76).

ohne Beachtung der Anreizkompatibilität gewählt wird (vgl. Holmström 1979, S. 76). Eine solche Situation stellt die First-best-Lösung dar. Unter den zuvor getroffenen Annahmen ergibt sich, daß das Resultat des obigen Maximierungsansatzes nur second-best sein kann.

Zur Herleitung der Lagrange-Funktion wird die Gleichung (3.3) durch die zugehörige Bedingung erster Ordnung für die Aktion des Agenten ersetzt. Dieses Vorgehen wird in der Literatur als First-order-approach bezeichnet. Allerdings muß dazu erfüllt sein, daß die Gleichung ein eindeutiges, stabiles Optimum besitzt. Dies läßt sich anhand der Abbildung 3.4 erläutern.

Abbildung 3.4: Probleme der Lösung des First-order-approach



Quelle: Grossman und Hart 1983, S. 303.

Die Abszisse gibt die Honorierungsfunktion für den Agenten, $s(x)$ an, die Ordinate die ihr zugehörige Aktion. Punkte auf der Kurve ABCDE genügen der Bedingung erster Ordnung zu (3.3) und werden im weiteren als Aktionslinie bezeichnet. Allerdings sind nur die Punkte auf den Abschnitten AB und DE der Aktionslinie globale Maxima (durchgezogene Linie). Für eine bestimmte Zahlung $s^*(x)$ bedeutet dies, daß die optimale Handlung des Agenten im Punkt P_3 und nicht in den Punkten P_1 oder P_2 sein wird. Die Indifferenzkurven geben Kombinationen von a und $s(x)$ an, die dem Principal denselben Nutzen stiften. Die höchste erreichbare Indifferenzkurve tangiert die Aktionslinie in Punkt C.

Die Abschnitte AB und DE stellen die Menge der möglichen Lösungen aus Sicht des Principals dar, für ihn ergibt sich Punkt B als Lösung. Allerdings genügt dieser Punkt nicht den Bedingungen erster Ordnung. Wird der Nutzen des Principals über $s(x)$ und a unter Berücksichtigung der Bedingung erster Ordnung des Agenten maximiert, so erhält man Punkt C als Lösung und nicht Punkt B. Dieser Widerspruch kommt durch die Vernachlässigung der hinreichenden Bedingungen zu (3.3) zustande (vgl. Grossman und Hart 1983, S. 302ff.).

Unter gewissen Annahmen an die Verteilungsfunktion ist der First-order-approach allerdings dennoch durchführbar (vgl. Rogerson 1985, S. 1357ff. und Mirrlees (1999)).³⁷ Im folgenden wird daher zur Vereinfachung angenommen, daß die aufgezeigten Schwierigkeiten nicht auftreten.

Der in den Gleichungen (3.1) bis (3.3) dargestellte Standardansatz erfährt im weiteren Verlauf eine Modifikation (vgl. Holmström 1979, S. 77). Der Output x wird nun als Zufallsvariable mit der Verteilung $F(x | a)$ angesehen. Das bedeutet, daß ihre Lage von der Handlung des Agenten beeinflusst wird. Gegeben die Zufallsvariable θ stellt $F(x | a)$ die Verteilung von x durch die Beziehung $x = x(a, \theta)$ dar. Falls durch die Aktionen des Agenten der Output positiv beeinflusst wird, d.h. $x_a \geq 0$, so folgt daraus $F_a(x | a) \leq 0$. Ein höherer Wert von a verschiebt demnach die Funktion in Richtung besserer Ergebnisse, schlechtere werden damit unwahrscheinlicher, jedoch nicht ausgeschlossen. $f(x | a)$ gibt die zugehörige Dichtefunktion an, so daß sich hier für das Maximierungsproblem schreiben läßt:

$$\max_{s(x), a} \int G(x - s(x)) f(x | a) dx \tag{3.4}$$

$$\text{u.d.N.} \quad \int U(s(x)) f(x | a) dx - V(a) \geq H \tag{3.5}$$

$$\int U(s(x)) f_a(x | a) dx = V'(a). \tag{3.6}$$

³⁷ Bei diesen Annahmen handelt es sich um die ‚Monotone Likelihood Ratio Condition‘ (MLRC) und die ‚Concavity of the Distribution Function‘ (CDF) (vgl. Rogerson, 1985, S. 1361ff. und Dutta und Radner, 1994, S. 877f.). Jewitt (1988) argumentiert, daß diese Annahmen in einigen Fällen zu restriktiv sind und präsentiert Bedingungen, die den First-order-approach rechtfertigen.

Gleichung (3.6) gibt dabei die Ableitung von (3.3) nach der Aktion a an. Das zugehörige Lagrange-Funktional lautet:³⁸

$$\begin{aligned}
 L = & \int G(x - s(x)) f(x | a) dx \\
 & + \lambda \left[\int U(s(x)) f(x | a) dx - V(a) - \bar{H} \right] \\
 & + \mu \left[\int U(s(x)) f_a(x | a) dx - V'(a) \right].
 \end{aligned} \tag{3.7}$$

Nimmt man abschnittsweise Linearität der Funktion an, so ergibt sich durch punktweises Optimieren von (3.7) nach $s(x)$ und Umformung:

$$\frac{G'(x - s(x))}{U'(s(x))} = \lambda + \mu \frac{f_a(x | a)}{f(x | a)}. \tag{3.8}$$

Dieser Ausdruck bestimmt die Risikoaufteilung zwischen Principal und Agenten, wenn dessen Handlungen nicht beobachtbar sind. Pareto-Optimalität liegt dann vor, wenn die rechte Seite konstant ist. Dies kann allerdings nur für $\mu = 0$ der Fall sein (vgl. Holmström 1979, S. 78). Der Lagrange-Parameter μ bestimmt sich aus der Ableitung der Lagrange-Funktion nach a . Für $\mu > 0$ gibt der Ausdruck f_a/f die Abweichung von der Pareto-optimalen Situation an.³⁹ Je größer die Werte sind, die a annimmt, desto größer ist der Aufwand des Agenten. Daher wird die Wahrscheinlichkeit eines Output für kleine x im Durchschnitt sinken und für große x steigen (vgl. Mirrlees 1976, S. 123). Weiterhin ist f_a/f eine steigende Funktion im Output (vgl. Levinthal 1988, S. 163).⁴⁰ Es ergibt sich, daß f_a/f für kleine Werte von x negativ und für große Werte von x positiv ist. Das bedeutet, daß mit größerem Aufwand die Wahrscheinlichkeit eines höheren Outputs und damit die Entlohnung ansteigt.

³⁸ Gleichung (4.5) beschreibt dabei nicht eine Funktion, sondern ein Funktional. Im Gegensatz zu einer Funktion, die einer reellen Zahl eine andere zuweist, ordnet ein Funktional einer Funktion, in diesem Fall $p(\theta)$, eine reelle Zahl zu (vgl. Chiang 1992, S. 7).

³⁹ Der Fall $\mu > 0$ wird genau dann gegeben sein, wenn der Principal einen hohen Wert von a präferiert. Wenn er die Handlungen beobachten und kontrollieren kann, wäre die Risikoaufteilung fix (vgl. Mirrlees 1976, S. 124).

⁴⁰ Der Ausdruck $|f_a(x | a)|$ reflektiert den Effekt der gewählten Aktion des Agenten auf die Wahrscheinlichkeit eines bestimmten Ergebnisses. Dann kann $|f_a/f|$ auch als Ableitung der *Log-Likelihood-Funktion* nach dem unbekanntem Parameter a betrachtet werden (vgl. Holmström 1979, S. 79 und Levinthal 1988, S. 164f.).

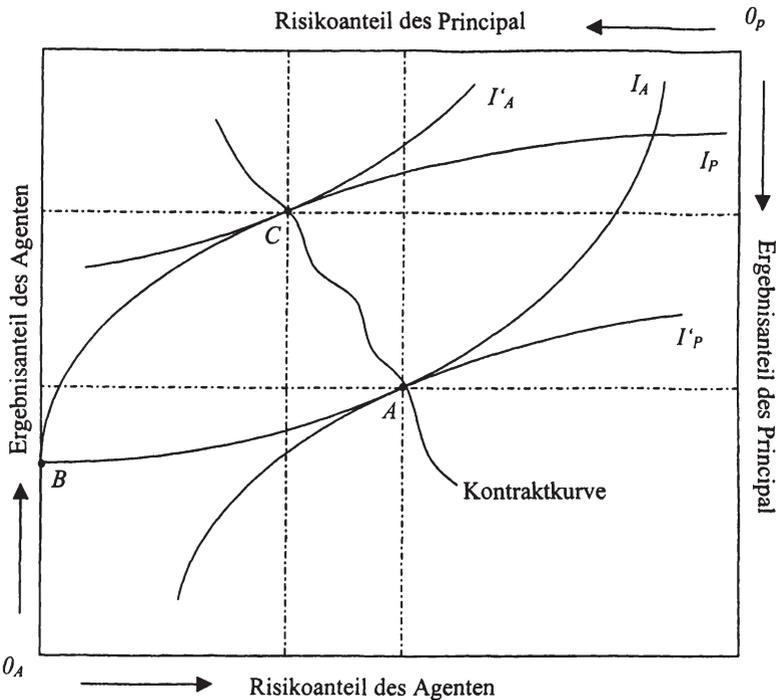
Es bleibt anzumerken, daß Gleichung (3.8) dann eine Pareto-optimale Situation beschreibt, falls der Principal die Aktion des Agenten genau beobachten kann. In diesem Fall besteht das Maximierungsproblem aus den Gleichungen (3.4) und (3.5) und das resultierende Grenznutzenverhältnis ist konstant.⁴¹ Ist der Principal darüber hinaus risikoneutral ($G' = \text{const}$), so ist der Lohn des Agenten unabhängig vom Output und damit eine fixe Zahlung, da der Principal das gesamte Risiko trägt (vgl. Levinthal 1988, S. 164). Weiterhin wird die First-best-Lösung erreicht, wenn der Agent risikoneutral ist, da kein Wohlfahrtsverlust dadurch entsteht, daß der Agent das Risiko eines unsicheren Outputs trägt (vgl. Sobel 1993, S. 267).

Eine graphische Darstellung des Zusammenhangs zwischen Gewinn- und Risikoaufteilung ist in Abbildung 3.5 zu erkennen (vgl. zu der folgenden Darstellung Douglas (1989)). Dabei wird angenommen, daß sowohl der Principal als auch der Agent risikoavers sind. Die Darstellung erfolgt mittels einer Edgeworth-Box, bei der der Ursprung des Agenten in der linken unteren Ecke abgetragen ist, der des Principal in der rechten oberen Ecke. Auf den Achsen sind der jeweilige Anteil am Ergebnis und der Risikoanteil des Principals bzw. des Agenten abgetragen. Die Höhe der Box ergibt sich aus der Summe der Anteile am Ergebnis, die Breite entspricht der Summe der Risiken. Ausgangspunkt ist das Entlohnungsschema B , bei dem der Erwartungsnutzen des Principals maximiert wird. Bei dieser Entlohnung erreicht der Principal die Indifferenzkurve I_P und der Agent die Kurve I_A .

Der Principal kann nun eine höhere Indifferenzkurve erreichen, indem er dem Agenten einen Vertrag A anbietet. Dieser ist durch einen höheren Anteil des Agenten am Ergebnis gekennzeichnet, wobei der Risikoanteil des Agenten ebenfalls ansteigt. Dadurch ergibt sich ein Pareto-optimaler Austausch zwischen Ergebnis- und Risikoanteil, da der Principal eine höhere Indifferenzkurve I'_P erreicht, während der Agent auf der alten Kurve verbleibt. Falls statt einer Festlegung des Vertrages durch den Principal nun Verhandlungen über den Vertrag erlaubt sind, liegen die resultierenden Anteile am Ergebnis und am Risiko auf der Kontraktkurve zwischen den Punkten A und C . Je größer dabei die Verhandlungsmacht des Agenten im Vergleich zum Principal ist, desto näher wird der endgültige Vertrag an Punkt C liegen, bei stärkerer Verhandlungsmacht des Principal näher an Punkt A .

⁴¹ Es liegt eine Kooperation zwischen Principal und Agent vor, um eine Bezahlung zu finden, die die gewichtete Summe der Nutzen maximiert (vgl. Ross 1973, S. 135).

Abbildung 3.5: Gewinn- und Risikoaufteilung bei Risikoaversion



Quelle: Vgl. Douglas 1989, S. 44.

Falls Unterschiede in der Risikoaversion vorliegen, so daß bspw. der Agent im Vergleich zum Prinzipal risikoaverser ist, verlaufen die Indifferenzkurven des Principals flacher im Vergleich zur Abbildung 3.5. Ausgehend vom gleichen Vertrag B ergibt sich, daß die resultierende Kontraktkurve nun näher am Ursprung des Agenten liegt, und daß der Vertrag, der dem Agenten angeboten wird, sowohl einen geringeren Risiko- und Ergebnisanteil beinhaltet (vgl. Douglas 1989, S. 45). Liegt der Extremfall eines risikoneutralen Principals vor, und der Agent ist risikoavers, so würde die Kontraktkurve mit der linken Seite der Box zusammenfallen. Dies bedeutet, daß der Principal das gesamte Risiko übernehmen würde und der Ergebnisanteil des Agenten eine fixe Zahlung darstellt, unabhängig vom tatsächlichen Ergebnis. Der ursprüngliche Vertrag wäre daher Pareto-optimal und auf den Agenten würde kein Risiko entfallen. Falls hingegen der Principal risikoaverser ist, so würde der Agent einen größeren Anteil sowohl am Ergebnis als auch am Risiko akzeptieren. Bei Risikoneutralität

des Agenten trägt dieser das gesamte Risiko und der Principal würde einen fixen Anteil am Gewinn erhalten. Falls beide Parteien risikoneutral sind, wäre der ursprüngliche Vertrag Pareto-optimal.

Die abgeleiteten Ergebnisse der Principal-Agent Theorie signalisieren eine partiell aufwands- und erfolgsbezogene Entlohnung. Durch erhöhte Anstrengungen wird ein gutes Ergebnis wahrscheinlicher, der Agent erhält einen Zuschlag auf sein Honorar. Im umgekehrten Fall wird er einen Abzug erhalten. Damit soll sichergestellt werden, daß der Agent große Anstrengungen für den Principal unternimmt. Allerdings weichen diese komplizierten Regelungen stark von den in der Realität vorherrschenden ab (vgl. Arrow 1985, S. 48ff.). Die Theorie legt zwar die Gründe für eine Risikoaufteilung vor, kann aber nicht zur Ausarbeitung von Entlohnungsstrukturen herangezogen werden, da dies u.a. zu hohe Kosten verursachen würde.

Erweiterungen und Anwendungen des Grundmodells bestehen in vielfacher Hinsicht. Shavell (1979) geht insbesondere auf die Aspekte Risikoallokation und Anreizstrukturen ein und analysiert Pareto-optimale Vergütungen. Die Aufteilung von Eigentum und Kontrolle in großen Unternehmen wird von Fama (1980) untersucht. In einer Arbeit aus dem Jahr 1982 analysiert Gjesdal in einem Principal-Agent-Modell den Wert von Informationen. Harris und Raviv (1978) beschreiben Anreizverträge, wie sie bspw. im Bildungsbereich oder im Gesundheitswesen vorkommen. Einen Überblick über die Anwendung des Grundmodells auf die Sozialversicherung bei risikoreicher Produktion und Moral Hazard auf dem Versicherungsmarkt gibt Rees (1985a). Darüber hinaus betrachtet er Modelle mit adverser Selektion in der Principal-Agent-Theorie. Eine generelle Übersicht über Anreize in Principal-Agent-Beziehungen findet sich in Sappington (1991), der auch auf die Auswahl des Agenten und dynamische Aspekte eingeht.

Des weiteren wurde das dargestellte Standardmodell noch in vielfacher Hinsicht erweitert. Holmström (1982) untersucht Beziehungen mit mehr als einem Agenten (moral hazard in teams). In diesem Rahmen sind zwei Aspekte von besonderer Bedeutung: die Trittbrettfahrerproblematik und der Wettbewerb unter den Agenten. Die Situation, in der ein Principal zwei Agenten anheuert, die in ihren Tätigkeiten von einer Kooperation profitieren, wird von Macho-Stadler und Pérez-Castrillo (1993) behandelt. Dabei gelangen sie zu dem Ergebnis, daß weitreichendere Kooperation ein effizientes Ergebnis begünstigt. Demski und Sappington (1984) analysieren optimale Anreizverträge, falls der Principal mit zwei Agenten gleichzeitig in Verbindung steht. Wichtigstes Resultat ist dabei, daß die private Information für die Agenten nicht mehr von Vorteil ist, es sei

denn, sie sind risikoavers. Sappington (1983) beschreibt Beziehungen zwischen Principal und Agent, in denen Haftungsbeschränkungen eine Rolle spielen.

Bernheim und Whinston (1987) behandeln Situationen, in denen mehrere risiko-neutrale Principale simultan und unabhängig das Verhalten eines (gemeinsamen) Agenten beeinflussen (common agency). Im Gegensatz dazu hat die Arbeit von Holmström und Milgrom aus dem Jahr 1987 die intertemporalen Anreize einer Principal-Agent-Beziehung zum Gegenstand. Fudenberg et al. (1990) analysieren den Zusammenhang von langfristigen Agency-Beziehungen und kurzfristigen Verträgen. In einem Aufsatz aus dem Jahr 1991 betrachten Holmström und Milgrom Beziehungen, in denen der Principal mehrere Aufgaben an den Agenten oder die Agenten delegiert bzw. in denen eine Aufgabe mehrere verschiedene Dimensionen aufweist (multitask agency). Dewatripont et al. (2000) beschreiben neuere Entwicklungen im Rahmen der ‚multitask agency theory‘, wie bspw. die Substituierbarkeit von Aufwendungen, Konflikte zwischen den einzelnen Aufgaben oder auch implizite Anreize.

Man erkennt, daß sich in den letzten zwanzig Jahren der Anwendungsbereich der Principal-Agent-Theorie und in diesem Zusammenhang behandelte Problemstellungen beträchtlich erweitert haben. Im Gesundheitswesen hingegen finden diese Theorien bislang nur selten Anwendung, obwohl es gerade dort, bedingt durch die Existenz einer Versicherung, durch asymmetrische Informationsverteilung zu Steuerungs- und Anreizproblemen kommt, oder, wie es Cutler und Zeckhauser (2000, S. 576) formulieren: „Health insurance involves a fundamental tradeoff between risk spreading and appropriate incentives. Increasing the generosity of insurance spreads risk more broadly but also leads to increased losses because individuals choose more care (moral hazard) and providers supply more care (principal-agent problems).“

4 Die Anwendung der Principal-Agent-Theorie auf das Gesundheitswesen

4.1 Die These der anbieterinduzierten Nachfrage

Die bisherige Darstellung konzentrierte sich entweder auf die Beziehung zwischen Arzt und Patient innerhalb des deutschen Gesundheitswesens oder auf informationsökonomische Aspekte von vertraglichen Beziehungen. Dabei ergab sich, daß aus der bestehenden Informationsasymmetrie zwischen den Akteuren Handlungsspielräume resultieren, die ausgenutzt werden können, ohne daß ein anderer dies verhindern könnte. Von hervorgehobener Bedeutung sind dabei die Vorteile, über die der Arzt gegenüber dem Patienten verfügt und ob es ihm möglich ist, die Nachfrage des Patienten zu beeinflussen. Ein weiterer wichtiger Punkt ist die Frage, ob es dem Patienten möglich ist, die Leistungen des Arztes durch die individuelle Vertragsgestaltung zu steuern. Dazu wird ein Standardmodell der Principal-Agent-Theorie auf die Arzt-Patient-Beziehung übertragen und anschließend kritisch analysiert. Daran schließt sich die Einordnung der Agency-Theorie in das Gesundheitswesen unter besonderer Berücksichtigung der bestehenden Informationsbeziehungen an. Insbesondere spielen hierbei die Strukturen des Gesundheitssektors und die bestehenden Interdependenzen zwischen den Akteuren eine Rolle.

In der Beziehung zwischen Arzt und Patient entstehen dem Arzt durch sein Spezialwissen Vorteile, die er zur Beeinflussung der Nachfrage ausnutzen kann. In seiner Tätigkeit als Agent des Patienten erhält er durch seinen Informationsvorsprung einen Anreiz, die durch den Patienten gewählte Behandlung so zu steuern, daß sich für ihn der höchste Nettoertrag ergibt (vgl. Evans, 1974, S. 163).⁴² Dem Arzt ist es bspw. über die Häufigkeit der Kontakte, die Intensität oder die Kosten der Behandlung möglich, auf die Nachfrage des Patienten zu Einfluß nehmen (vgl. Kortendieck 1993, S. 234). Er besitzt dabei eine Doppelrolle: Als Anbieter der medizinischen Leistungen sowie als Berater bezüglich der Entscheidung des Patienten über die nachgefragte Leistung (vgl. Breyer und Zweifel 1999, S. 241).

Das Verhalten des Arztes, seinen Informationsvorsprung auszunutzen, kann vor allem im Hinblick auf eine gestiegene Arztdichte untersucht werden. In einem normalen Markt bewirkt die Ausweitung des Angebotes durch eine höhere Ärztezahl eine Rechtsverschiebung der Angebotskurve. Daraus resultiert bei

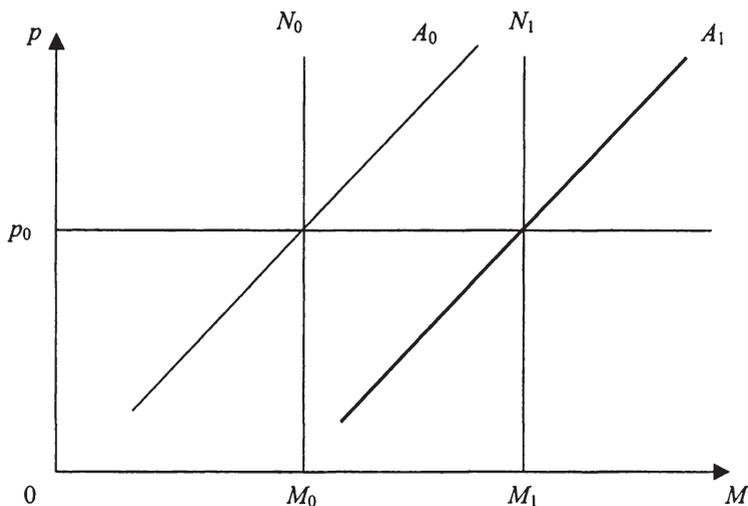
⁴² Shmanske (1996, S.197ff.) sieht im Marktversagen durch asymmetrische Information keinen Grund für staatliche Interventionen, sondern betrachtet dieses Versagen als Knappheit, die der Markt beheben sollte.

einer fallenden Nachfragekurve eine gestiegene Menge und ein gesunkener Preis. Die Veränderungen der Ausgaben für ärztliche Leistungen bleibt unklar und hängt von der Preiselastizität der Nachfrage und des Angebotes ab (vgl. Breyer und Zweifel 1999, S. 242f.). Problematisch für die Beziehung zwischen Arzt und Patient ist es, wenn der Arzt in die Entscheidungen, die er für den Patienten trifft, seine eigenen Interessen einfließen läßt. Falls bspw. die Ärzte als Reaktion auf eine gestiegene Arztdichte die Informationen, die sie dem Patienten bereitstellen, variieren, so entwickelt sich die anbieterdeterminierte Nachfrage zu einer angebots- bzw. anbieterinduzierten Nachfrage (vgl. Breyer und Zweifel 1999, S. 244).⁴³

Falls die Patienten vollversichert sind, ist ihre Nachfrage hingegen nicht mehr vom Preis abhängig. Die Folgen lassen sich mit Hilfe von Abbildung 4.1 illustrieren (vgl. hierzu Breyer und Zweifel 1999, S. 244f.). Bei staatlicher Preisfestsetzung für die medizinischen Leistungen (bspw. über eine Gebührenordnung) auf p_0 sei angenommen, daß in der Ausgangssituation bei gegebener Arztdichte das Leistungsangebot A_0 gerade so hoch ist, daß es zu einem Ausgleich mit der Primärnachfrage N_0 kommt. Die Angebotskurve verschiebt sich durch Hinzukommen weiterer Ärzte nach außen auf A_1 . Ohne Einfluß auf die Nachfrage bzw. bei alleiniger Verfolgung der Interessen des Patienten bliebe die Nachfragekurve stabil und Preises die Leistungsmenge aufgrund des regulierten bei $N_0 = M_0$ konstant. In diesem Fall sind die Anbieter rationiert und somit dazu gezwungen, eine geringere als die angestrebte Menge zu erbringen. Verhält sich der Arzt dagegen nicht als perfekter Sachwalter des Patienten, sondern verfolgt er bei der Beratung eigene Interessen, dann würde er über das medizinisch indizierte Maß hinaus zusätzliche Leistungen empfehlen, die nur marginal wirksam oder medizinisch wertlos sind, um zu erreichen, daß die gewünschte Angebotsmenge $M_1 = A_1(p_0)$ auch erreicht wird. In der Abbildung ist dies durch eine Rechtsverschiebung der Nachfragekurve bis zu dem Punkt, an dem die nachgefragte Menge zum Preisniveau p_0 gleich der angebotenen ist, gegeben. Dies bedeutet, daß der Arzt, sobald er Eigeninteressen verfolgt, z.B. Gewinn- oder Erwerbsmotive, die Nachfrage soweit ausdehnen wird, bis der Grenznutzen der Nachfrageausweitung seinen Grenzkosten entspricht (vgl. Schulenburg 1987, S. 6ff.).

⁴³ Ein Überblick über die These der anbieterinduzierten Nachfrage findet sich in McGuire (2000) und Folland et al. (2001). Zu einer formalen Umsetzung siehe Breyer (1984 und 1984a) und Dranove (1988).

Abbildung 4.1: Anbieterinduzierte Nachfrage als Reaktion auf eine Angebotszunahme



Quelle: Breyer und Zweifel 1999, S. 245.

Bei einer medizinischen Behandlung herrscht somit keine vollständige Konsumentensouveränität, da der Patient i.d.R. nur den Erstkontakt bestimmt, den weiteren Umfang der medizinischen Leistung jedoch der Arzt festlegt (vgl. Breyer 1984, S. 415). Durch das Vertrauen, das der Patient dem Arzt gegenüber entgegenbringen muß, wird die nachgefragte Menge letztendlich durch den Arzt determiniert. Der Schluß von der einzelnen Arzt-Patient-Beziehung auf den gesamten Markt läßt die Vermutung zu, daß der Patient nicht die ausschlaggebende Instanz ist. „Die Nachfragekurve, die das geplante Volumen der Inanspruchnahme ärztlicher Leistungen bei alternativen Preisen angibt, spiegelt danach in überwiegendem Maße Entscheidungen der Anbieter und nicht der Nachfrager wider“ (Breyer 1984, S. 15).⁴⁴

Dieses Problem kann alternativ als Auseinanderfallen individueller Aktionen aufgefaßt werden. Die Einheit von Nachfrager, Konsument und Finanzier, wie sie sich in den meisten Märkten präsentiert, ist im ambulanten Gesundheitswesen nicht mehr gegeben. Als Konsument der medizinischen Leistung tritt der

⁴⁴ Im Markt für medizinische Leistungen kann daher keine Dichotomisierung in Angebots- und Nachfrageseite vorgenommen werden (vgl. Evans 1974, S. 163f.).

Patient auf, die Krankenversicherung übernimmt die Finanzierung, während die Nachfrage maßgeblich vom Arzt abhängt (vgl. Wille und Ulrich 1991, S. 29).

Begünstigt wird die Angebotsinduzierung der Nachfrage vor allem durch ein Vollversicherungssystem und den medizinischen Fortschritt (vgl. Breyer und Zweifel 1999, S. 245f.). Bei einer Vollversicherung spielt die Zahlungsbereitschaft des Patienten keine Rolle, da sie nicht als Begrenzung für den Konsum dient. Die moderne Diagnosetechnik erlaubt es dem Arzt, umfangreiche Therapien zu verordnen, so daß der Patient keine Nutzenverschlechterung durch den Mehrkonsum erleidet.

Als Ziel des ärztlichen Handelns wird oft die Gewinnmaximierung angesehen. Daneben kann er auch versuchen, ein bestimmtes Zieleinkommen zu erreichen (vgl. Evans 1974, S. 167ff.). Der Arzt verfügt über eine zumindest grobe Zielvorstellung bezüglich seines Einkommens und dem damit verbundenen Arbeitsaufwand. Falls zwischen seinen Zielvorstellungen und dem Ist-Zustand eine zu starke Abweichung vorliegt, kommt es zu einem Anpassungsprozeß. Für den Fall, daß das Ziel nicht erreicht wurde, wird der Arzt eine Ausweitung der Menge anstreben, was er über sein Behandlungsverhalten oder seine Preispolitik erreichen kann. Werden die Preise nicht durch den einzelnen Arzt festgelegt, kann er seine Zielvorstellungen entweder nur für sein Einkommen oder seinen Arbeitsaufwand erreichen (vgl. Breyer 1984, S. 422f.). Kritisch bei der Annahme eines Zieleinkommens stellt sich die Höhe des Anspruchsniveaus dar und wie es durch ein Kosten-Nutzen-Kalkül ermittelt wird (vgl. Schulenburg 1987, S. 68).

Die Nachfrage nach medizinischen Leistungen und ihr Zusammenhang mit der Arztdichte kann allerdings nicht alleine auf die anbieterinduzierte Nachfrage zurückgeführt werden.⁴⁵ Das erste Argument folgt der Überlegung, daß die Nachfrage aufgrund staatlicher Regulierung rationiert ist, so daß ein permanenter Nachfrageüberhang besteht. Sind die Ärzte voll ausgelastet, wird durch neue Anbieter also bei einer damit höheren Arztdichte nur ein bestehender Überhang abgebaut.⁴⁶

Im zweiten Fall wird die Nachfragekurve durch eine höhere Qualität der Behandlung nach außen verschoben. Die höhere Arztdichte gewährleistet eine intensivere Diagnose und Therapie, da der einzelne Arzt nun mehr Zeit aufwenden

⁴⁵ Vgl. hierzu Breyer 1984a, S. 17ff., Beyer und Zweifel 1999, S. 237ff., Feldstein 1993, S. 87ff. oder insbesondere bezüglich des ärztlichen Verhaltens Enthoven (1981).

⁴⁶ Warteschlangen können sich allerdings auch durch eine Vollversicherung ergeben, da in diesem Fall die Preise keine Rationierungsfunktion haben.

kann. Der Patient wird bei gleichem Preis für die medizinische Leistung mit einer Nachfrageausweitung reagieren. Weiterhin nehmen durch die gestiegene Arztdichte die indirekten Kosten eines Besuches, dazu gehören Wege- und Zeitkosten, ab, wodurch ein Arztbesuch für den Patienten verbilligt wird.

Eine dritte Einschränkung liegt in einem möglicherweise umgekehrten Kausalzusammenhang. Die veränderte Arztdichte begründet nicht die stärkere Inanspruchnahme, sondern kann als Reaktion auf sie gesehen werden. Bei einer Neuansiedlung wählen die Ärzte bewußt Regionen, die einen Nachfrageüberhang aufweisen, z.B. durch höhere Krankheitsanfälligkeit, eine bessere finanzielle Absicherung des Krankheitsrisikos oder ein höheres Einkommen.

Die aufgelisteten Gründe werden von Vertretern der These der anbieterinduzierten Nachfrage nicht bestritten (vgl. Kortendieck 1993, S. 240ff.). Nach ihrer Ansicht spielt der technische Fortschritt als Shift-Parameter für eine Ausweitung des Angebotes und der Nachfrage eine wichtige Rolle. Newcomer verfügen gegenüber etablierten Ärzten über ein aktuelleres und größeres Wissen, aber auch über die leistungsfähigeren Geräte. Damit sich die teureren Investitionen in Ausstattung und Humankapital lohnen, sind Newcomer dazu bereit, ihre Patienten intensiver zu behandeln.

In wieweit die theoretische Erklärung einer durch die höhere Arztdichte determinierten gesteigerten Nachfrage empirisch als bestätigt gelten kann, ist fraglich, da die sich überlappenden Einflußfaktoren oftmals nicht voneinander isoliert werden können (vgl. Wille und Ulrich 1991, S. 29ff. und Kortendieck 1993, S. 237f.). Auch Ergebnisse, nach denen die ambulanten Ausgaben stärker von der Facharztdichte als von der Dichte der Allgemeinmediziner abhängen, sind nicht einheitlich. Deshalb die Gültigkeit der These in Zweifel zu ziehen, stellt jedoch eine vorschnelle Reaktion dar. Selbst wenn nur ein Teil der Nachfragesteigerung auf eine höhere Arztdichte zurückzuführen sind, so sind die damit verbundenen Implikationen für das Verhältnis von Patient, Versicherung und Arzt von Bedeutung.

4.2 Die Arzt-Patient-Beziehung als Principal-Agent-Modell

4.2.1 Die Übertragung eines Standardansatzes

Die hier präsentierte Modellierung der Arzt-Patient-Beziehung als Principal-Agent-Ansatz konzentriert sich auf die Situation nach der Entscheidung über

den Erstkontakt und damit auf die Behandlung.⁴⁷ Die Problematik der adversen Selektion bleibt im folgenden ausgeklammert. Die Konzentration auf einen Teil des Entscheidungsprozesses verneint nicht die ökonomische Relevanz vorangehender Prozesse. Die Betrachtung beschränkt sich auf sogenannte Principal-Agent-Probleme bei Hidden Action (vgl. Abschnitt 3.4).

Die Analyse der Beziehung zwischen Arzt und Patient erfolgt mit Hilfe des klassischen Ansatzes von Holmström und Mirrlees (vgl. Holmström, 1979 und Mirrlees, 1976). Entsprechend der Einteilung läßt sich dieser Ansatz der normativen Ausrichtung und der Klasse der Hidden Action-Modelle zuordnen. Die Abweichung von der optimalen Risikoaufteilung resultiert dabei aus der asymmetrischen Informationsbeziehung zwischen Arzt und Patient. Zweifel (1994) überträgt diesen Ansatz auf das Gesundheitssystem und leistet eine gesundheitsökonomische Interpretation der zugrunde liegenden Strukturen.

Im vorliegenden Modell wird die Honorierungsform gesucht, die der Patient wählen würde, damit der Arzt sich ihm gegenüber als perfekter Agent verhält. Der Nutzen des Patienten (Principal) hängt hierbei von dem monetär bewertetem Gesundheitszustand nach der Behandlung ab, abzüglich der Ausgaben für die Entlohnung des Arztes (Agent).⁴⁸ Der Principal maximiert in diesem Ansatz den aus diesem Nettowert resultierenden Erwartungsnutzen unter bestimmten Restriktionen, die sich aus dem Verhalten des Arztes ergeben.

Die Handlung des Arztes besteht in der Auswahl der medizinischen Leistung $a \in A \subseteq \mathbb{R}$, wobei A die Menge aller ärztlichen Leistungen darstellt. Die ärztliche Handlung kann durch den Patienten nicht eingeordnet bzw. bewertet werden, allerdings nimmt der Patient den resultierenden Gesundheitszustand θ wahr.⁴⁹ Dieses Ergebnis des medizinischen Leistungserstellungsprozesses wird dabei nicht alleine durch den Inputfaktor ‚ärztliche Leistung‘ determiniert. Der resultierende Gesundheitszustand stellt vielmehr eine Zufallsvariable mit der be-

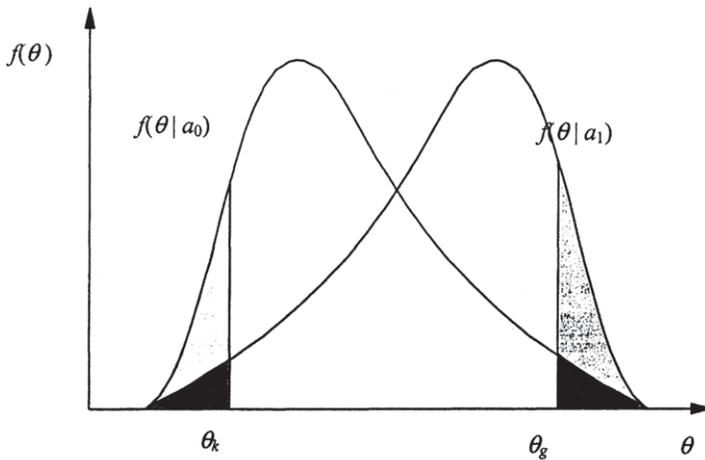
⁴⁷ Eine Würdigung der Agency-Theorie im Rahmen des Gesundheitswesens findet sich in Mooney und Ryan (1993).

⁴⁸ Die Annahme, daß der Gesundheitszustand in Geldeinheiten ausgedrückt werden kann, folgt nach aus der Herleitung der Zahlungsbereitschaft für marginale Verbesserungen des Gesundheitszustands (vgl. Zweifel 1994, S. 13 und Zweifel et al. 2001, S. 5). Diese Annahme kann als kritisch betrachtet werden, da die Bewertung von Leben und Gesundheit mit vielfältigen Problemen verbunden ist (vgl. hierzu Ried und Ulrich (1996) und Ried und Wille (1997).

⁴⁹ Sicherlich erkennt der Patient welcher Art die Handlungen des Arztes sind. Der Einfluß auf seinen Gesundheitszustand ist für ihn aber nur schwer abzuschätzen bzw. bleibt ihm verborgen. Weiterhin kann er nicht beurteilen, ob eine andere Entscheidung nicht angemessener wäre (vgl. Ried 1992, S. 165).

dingten Verteilungsfunktion $F(\theta|a)$ und der bedingten Dichte $f(\theta|a)$ dar. Die Lage der Verteilungsfunktion wird dabei durch die Handlungen des Arztes a beeinflusst. Ein höherer Wert von a verschiebt die Verteilungsfunktion in Richtung besserer Ergebnisse, schlechtere Resultate werden damit unwahrscheinlicher, können aber nicht ausgeschlossen werden.⁵⁰ Abbildung 4.2 zeigt den Verlauf zweier Dichtefunktionen in Abhängigkeit von der gewählten medizinischen Leistung. Hierbei gibt θ_k einen schlechten und θ_g einen guten Gesundheitszustand an. Die Lage der Dichtefunktion $f(\theta|a_0)$ ist durch die Wahl der medizinischen Leistung a_0 determiniert, die für einen geringen ärztlichen Behandlungsaufwand steht. Demgegenüber kennzeichnet die Dichte $f(\theta|a_1)$ einen hohen Einsatz an medizinischer Leistungsintensität a_1 . Höhere Anstrengungen des Arztes lassen einen schlechten Gesundheitszustand unwahrscheinlicher werden und vice versa.

Abbildung 4.2: Zusammenhang zwischen ärztlicher Leistung und Gesundheitszustand



Quelle: Zweifel 1994, S. 12.

Das zentrale Problem für die Entlohnung des Arztes besteht darin, daß auch ein guter Gesundheitszustand mit geringem Arbeitsaufwand des Arztes erreicht

⁵⁰ Die Verteilungsfunktion genügt daher der stochastischen Dominanz erster Ordnung (vgl. Holmström 1979, S. 77).

werden kann, wenn auch nur mit geringer Wahrscheinlichkeit. „Wenn nun der Patient mangels Fachwissen die Bemühungen des Arztes [...] nicht beobachten kann, so bleibt ihm nur der erreichte Gesundheitszustand θ als Maßstab für die Honorierung“ (Zweifel 1994, S. 12). Unter Umständen kann der Arzt bei einer erfolgsbezogenen Honorierung sein Einkommensziel mit nur geringen Aufwänden erreichen. Im umgekehrten Fall erscheint es möglich, daß trotz hoher Anstrengungen des Arztes der Gesundheitszustand des Patienten schlecht bleibt und der Arzt lediglich eine geringe Entlohnung erhält. Geht man dagegen auf ein fixes Honorar über, fördert man gute Leistungen nicht ausreichend, da der Arzt unabhängig von seiner geleisteten Arbeit immer die gleiche Entlohnung erhält.

Die Entlohnung des Arztes ist in diesem Ansatz eine Funktion des erreichten Gesundheitszustands, $p(\theta)$. Der verbleibende Nettowert $\theta - p(\theta)$, d. h. die Differenz zwischen dem monetär bewerteten Gesundheitszustand und dem monetären Behandlungsaufwand des Arztes, geht dann als Argument in die Nutzenfunktion, U^p des Patienten ein. Die Nutzenfunktion sei vom ‚von Neumann-Morgenstern-Typ‘, stetig differenzierbar und konkav, d. h. der Patient ist risikoavers. Er maximiert seinen erwarteten Nutzen über die Entlohnungsregel des Arztes:

$$\max_{p(\theta)} \int U^p(\theta - p(\theta)) f(\theta | a^*) d\theta. \quad (4.1)$$

Die Nebenbedingungen dieses Optimierungsproblems ergeben sich aus der Teilnahmebedingung für den Arzt und einer Bedingung für die Anreizkompatibilität der ärztlichen Handlung. Die Teilnahmebedingung des Arztes folgt aus seiner Nutzenfunktion. Diese ist vom ‚von Neumann-Morgenstern-Typ‘ und additiv-separabel. Sie besteht aus einer Entlohnungskomponente U^A und einer Aufwandskomponente V^A , wobei die gewählte medizinische Leistung durch die Aufwandskomponente negativ in die Nutzenfunktion eingeht. Die ärztliche Leistung a läßt sich somit als produktiver Inputfaktor beschreiben, der eine negative Auswirkung auf den Nutzen des Arztes ausübt, weshalb die Zielvorstellungen des Patienten und des Arztes bezüglich der medizinischen Leistung differieren (vgl. Holmström 1979, S. 76). Der Arzt ist nicht risikofreudig in bezug auf sein Einkommen und muß, damit er einen Vertrag mit dem Patienten eingeht, mindestens seinen Reservationsnutzen (\bar{U}^A) erzielen. Gleichung (4.2) stellt die Teilnahmebedingung des Arztes dar:

$$\int U^A(p(\theta)) f(\theta | a^*) d\theta - V^A(a) \geq \bar{U}^A. \quad (4.2)$$

Die medizinische Leistung wird dann durch den Arzt festgelegt. Da der Patient die Handlungen des Arztes nicht beobachten kann, kann dieser stets die optimale Höhe a^* der Behandlung wählen. Dies wird dann erreicht, wenn sein Nettoeinkommen maximal ist.⁵¹

$$a^* = \arg \max_a \left[\int_{\theta} U^A(p(\theta)) f(\theta | a) d\theta - V^A(a) \right]. \quad (4.3)$$

Aus Gleichung (4.3) läßt sich die Bedingung für die Anreizkompatibilität in Gleichung (4.4) durch partielles Differenzieren nach a herleiten: Gleichung (4.4) zeigt, daß der Erwartungswert des Grenzertrages einer medizinischen Leistung im Optimum den Grenzkosten der Bereitstellung entspricht.⁵²

$$\int_{\theta} U^A(p(\theta)) f_a(\theta | a^*) d\theta - V^{A'}(a^*) = 0. \quad (4.4)$$

Das Maximierungsproblem des Patienten besteht dann aus den Gleichungen (4.1), (4.2) und (4.4) (Zielfunktion, Teilnahmebedingung des Arztes und Bedingung für Anreizkompatibilität), woraus folgendes Lagrange-Funktional resultiert:

$$\begin{aligned} L = & \int_{\theta} U^P(\theta - p(\theta)) f(\theta | a^*) d\theta \\ & + \lambda \left[\int_{\theta} U^A(p(\theta)) f(\theta | a^*) d\theta - V^A(a^*) - \bar{U}^A \right] \\ & + \mu \left[\int_{\theta} U^A(p(\theta)) f_a(\theta | a^*) d\theta - V^{A'}(a^*) \right]. \end{aligned} \quad (4.5)$$

Die notwendige Bedingung für ein Nutzenmaximum des Patienten erhält man durch punktweises Differenzieren nach $p(\theta)$. Dies bedeutet, daß für alle Werte,

⁵¹ Allerdings muß dazu erfüllt sein, daß die Gleichung ein eindeutiges, stabiles Optimum besitzt (vgl. Grossman und Hart 1983, S. 302ff.). Unter gewissen Annahmen an die Verteilungsfunktion läßt sich der *First-order-approach*, die Ableitung von Gleichung (4.3) nach a , dennoch durchführen. Bei diesen handelt es sich um die Monotone Likelihood Ratio Condition (MLRC) und die Concavity of the Distribution Function (CDF) (vgl. Rogerson 1985, S. 1357ff. und Dutta und Radner 1994, S. 877f.).

⁵² Hierbei bezeichnet $f_a(\theta | a^*)$ die partielle Ableitung der Dichtefunktion nach der medizinischen Leistung.

welche die Funktion p an der Stelle θ annimmt, die folgende notwendige Bedingung erfüllt sein muß:

$$\frac{U^p'(\theta - p(\theta))}{U^A'(p(\theta))} = \lambda + \mu \frac{f_a(\theta | a^*)}{f(\theta | a^*)}. \quad (4.6)$$

Das Grenznutzenverhältnis auf der linken Seite gibt die optimale Aufteilung des Vorteils aus der Transaktion $p(\theta)$ zwischen dem Arzt und dem Patienten wieder (vgl. Zweifel 1994, S. 15). Ein hoher Wert des Bruches resultiert aus einem geringen Nettowert für den Patienten, gleichbedeutend mit einem höheren Anteil des Arztes. Nimmt man für beide Akteure Risikoaversion an, so sinkt das Grenznutzenverhältnis aufgrund der Konkavität der Nutzenfunktionen (U^p und U^A) mit zunehmendem Nettowert und vice versa.

Die rechte Seite gibt die objektiven Bestimmungsgründe der Aufteilung des Vorteils durch die anreizkompatible Honorierung $p^*(\theta)$ wieder. Ein Teil λ ist ein Fixbetrag, der den Arzt zur Annahme des Vertrages bewegen soll. Da λ der Lagrange-Multiplikator der Teilnahmebedingung (4.2) ist, gibt er die marginale Änderung des Wertes der Zielfunktion durch eine marginale Änderung der Nebenbedingung, d.h. \hat{U}^A , an (vgl. Takayama 1985, S. 139). Der Wert von λ ist dann groß, „wenn ein Nichtzustandekommen des Vertrages für den Patienten erhebliche gesundheitliche Konsequenzen hätte“ (Zweifel 1994, S. 15).

Für eine Honorierung $p(\theta)$ liegt nur dann Pareto-Optimalität im Sinne einer optimalen Risikoaufteilung vor, wenn die Grenzraten der Substitution des Einkommens des Patienten und des Arztes zwischen zwei beliebigen Zuständen gleich sind (vgl. Rees 1985, S. 7). Die rechte Seite der Gleichung (4.6) ist in diesem Fall konstant. Eine perfekte Risikoaufteilung gilt somit nur für den Fall $\mu = 0$ (vgl. Holmström 1979, S. 78).⁵³ Der Lagrange-Parameter μ bestimmt sich aus der Ableitung des Lagrange-Funktional nach der Leistung des Arztes a . Für $\mu > 0$ gibt der Ausdruck f_a/f die Abweichung gegenüber der Pareto-optimalen Situation an. Je größer die Werte für a sind, um so größer ist der Behandlungsaufwand des Arztes. Die Wahrscheinlichkeit für einen schlechten Gesundheitszustand sinkt in diesem Fall bzw. sie nimmt zu für einen guten Gesundheitszustand (vgl. Mirrlees 1976, S. 123). Weiterhin ist f_a/f eine steigende Funktion im

⁵³ Die Pareto-optimale Risikoaufteilung wird durch die Beachtung der Anreizeffekte verhindert. Diese Effekte berücksichtigen die Auswirkungen der Wahl einer Honorierungsfunktion $p(\theta)$, gegeben θ , auf die Entscheidung über eine medizinischen Leistung a durch den Arzt (vgl. Rees 1985, S. 21).

Ergebnis θ (vgl. Levinthal 1988, S. 163). Dies impliziert, daß der Quotient f_a/f für niedrige Werte des Gesundheitszustands negativ und für große Werte positiv ausfällt. Mit größerem Behandlungsaufwand nimmt die Wahrscheinlichkeit für einen besseren Gesundheitszustand zu und damit letztlich auch die Entlohnung des Arztes. Dieser Teil der Optimalitätsbedingung bezeichnet somit die Erfolgsbeteiligung des Arztes. Für einen schlechten Gesundheitszustand bedeutet dies, daß der Arzt zusätzliche Anstrengungen unternehmen müßte, damit ein negatives Ergebnis mit geringerer Wahrscheinlichkeit eintritt. In diesem Fall ist $f_a(\theta|a^*)$ negativ und der Arzt muß einen Abzug von seinem Honorar in Kauf nehmen.

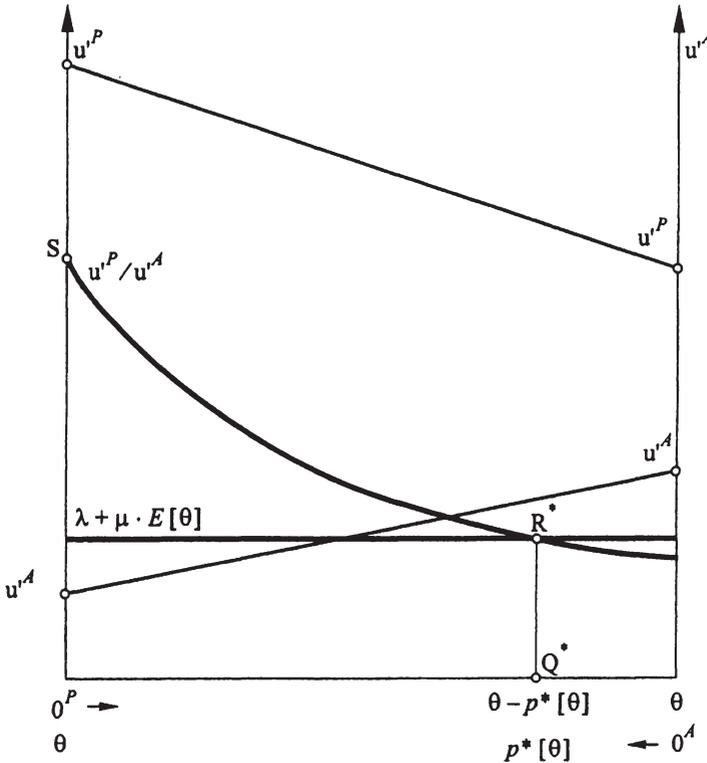
Eine graphische Darstellung der Herleitung der optimalen Vergütung bei gegebenen Werten von θ findet sich in Breyer und Zweifel (1999). In Abbildung 4.3 gibt die Distanz zwischen 0^P und θ ein bestimmtes Behandlungsergebnis wieder, falls man vom Ursprung des Patienten ausgeht. Ausgehend vom Arzt ist dies die Distanz zwischen 0^A und θ . Der Grenznutzen des Patienten (U^{rP}) ist aufgrund seiner Risikoaversion eine fallende Gerade, d.h. mit zunehmendem Anteil des Patienten am Behandlungsergebnis wird sein Grenznutzen geringer und damit sein Nutzen größer. Analog dazu sinkt der Grenznutzen des Arztes mit der Entfernung vom Ursprung. Das Grenznutzenverhältnis (U^{rP}/U^{rA}) ist eine absinkende Kurve, beginnend in Punkt S , und gibt die linke Seite von Gleichung (4.6) wieder, die mit steigender Vergütung des Arztes abnimmt.

Die rechte Seite von Gleichung (4.6) gibt die objektiven Bestimmungsgründe der Vergütung für bestimmte Werte von Q an. Der Ausdruck $\lambda + \mu E$ (mit $E = f_a/f$) hängt dabei nicht von der Vergütung des Arztes $p(\theta)$, sondern nur vom Ergebnis des Leistungserstellungsprozesses θ ab und läßt sich in der Abbildung als horizontale Gerade darstellen. Diese schneidet die Kurve für das Grenznutzenverhältnis von Patient und Arzt in Punkt R^* , der damit der Optimalitätsbedingung aus Gleichung (4.6) entspricht. Der daraus resultierende Punkt Q^* entspricht der optimalen Aufteilung des Behandlungsergebnisses zwischen Arzt und Patient. Letzterer erhält $\theta - p^*(\theta)$ vom Ergebnis, der Arzt der einen Anteil in Höhe von $p^*(\theta)$.

Für ein schlechteres Behandlungsergebnis $\tilde{\theta}$ folgt, daß sich die Entfernung zwischen 0^P und $\tilde{\theta}$ verkürzt und sich der Ursprung des Arztes verschiebt (siehe Abbildung 4.4). Damit ändert sich auch die Lage der Grenznutzenfunktion des Arztes, sie verschiebt sich nach oben (\tilde{U}^{rA}). Durch die Verschiebung der Grenznutzenfunktion des Arztes verläuft das Verhältnis der Grenznutzen nun unterhalb der ursprünglichen Lage, beginnend in Punkt S' . Aufgrund der Abhängigkeit des Ausdrucks $\lambda + \mu E$ vom erreichten Gesundheitszustand θ ,

verlagert sich die Gerade nach unten. Der Schnittpunkt mit der neuen Kurve für das Grenznutzenverhältnis (\tilde{R}^{**}) gibt die neue Aufteilung des erzielten Gesundheitszustands $\tilde{\theta}$ auf Arzt und Patient an. Man erkennt, daß der Anteil von Arzt und Patient am erzielten Ergebnis zurückgeht.

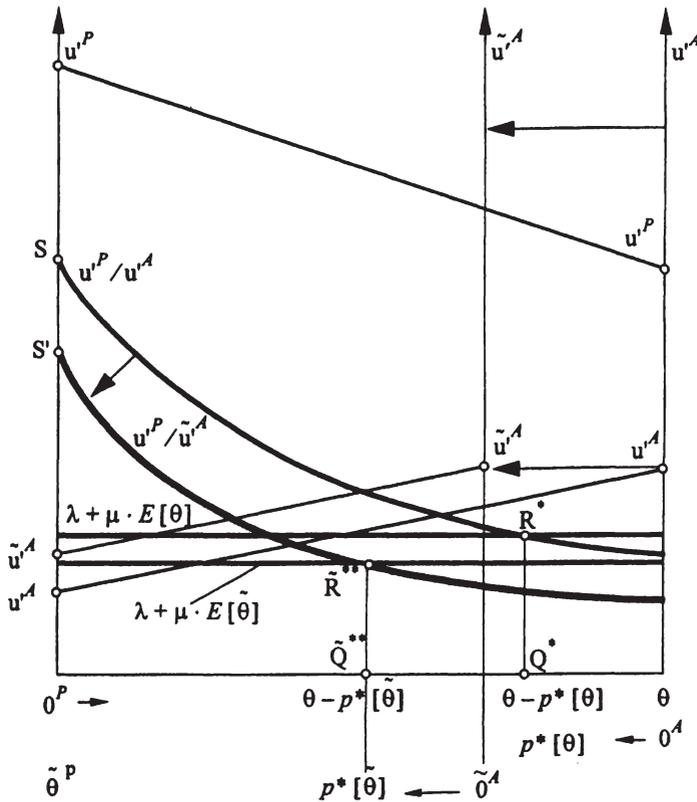
Abbildung 4.3: Optimale Honorierung bei gegebenen Werten von θ



Quelle: Breyer und Zweifel 1999, S. 286.

Die vorgestellten Ergebnisse signalisieren im Kern eine aufwandsbezogene und erfolgsorientierte Entlohnung. Durch erhöhte Anstrengungen wird ein gutes Behandlungsergebnis wahrscheinlicher und der Arzt erhält einen Honorarzuschlag. Im umgekehrten Fall wird er einen Abzug in Kauf nehmen müssen. Damit wird sichergestellt, daß der Arzt hinreichende Anstrengungen für den Patienten unternimmt.

Abbildung 4.4: Honorierung bei verändertem Gesundheitszustand



Quelle: Breyer und Zweifel 1999, S. 287.

4.2.2 Modellkritik

Einer Umsetzung der abgeleiteten Vergütungsform des Arztes stehen vor allem zwei Aspekte entgegen (vgl. hierzu Zweifel 1994, S. 17f.). Zum einen unterliegt der Vertragsabschluß zwischen Patient und Arzt häufig einer gewissen Dringlichkeit. Dies trifft insbesondere dann zu, wenn der Gesundheitszustand des Patienten vor Vertragsabschluß schlecht ist, so daß bei einer Nichtbehandlung ernste Konsequenzen drohen, im Extremfall sogar der Tod. In einer solchen Situation würde der Parameter λ einen hohen Wert aufweisen, mit der Folge, daß der Anteil des Patienten am erreichten Gesundheitszustand niedrig ausfällt und

damit der Nettowert sinkt. Weiterhin erscheint der Patient nicht in der Lage, die Angaben des Arztes bezüglich des Reservationsnutzens und damit der Höhe des Fixums zu überprüfen. Ein zweites Problem ergibt sich bei der Umsetzung der erfolgsabhängigen Vergütung. Dies betrifft sowohl den Multiplikator μ , der die marginalen Auswirkungen einer zu geringen Behandlungsleistung angibt, als auch den Quotienten f_{df} , der die Auswirkung auf den Gesundheitszustand erfaßt. Unter Umständen hat der Patient keine Informationen über beide Komponenten, so daß eine Vergütung nach dem Erfolg einer Behandlung kaum bzw. gar nicht möglich erscheint.

Aus diesen Problemen folgt, daß „[d]ie Chance für den Patienten als Auftraggeber, dem Arzt über die Wahl der Honorierungsfunktion die Anreize für eine optimale Behandlungsintensität geben zu können, [...] vergleichsweise klein [sind]. Insbesondere fehlen ihm dazu die Informationen über die Effektivität des Arztes, d. h. über seinen Beitrag zur Verbesserung des Gesundheitszustands“ (Zweifel 1994, S. 19). Der Versuch, die Honorierung des Arztes vom erreichten Gesundheitszustand abhängig zu machen, stößt auf weitere Schwierigkeiten. Zunächst ist die Ermittlung des Gesundheitszustands mit vielfältigen Meßproblemen behaftet. Hinzu kommt, daß die subjektive Sichtweise des Patienten bezüglich des Gesundheitszustands von der des Arztes divergieren kann. Weiterhin ist die Frage, ob wirklich der erreichte Gesundheitszustand als Maßstab für eine Honorierung der ärztlichen Leistung ausreichen würde. Ein optimaler Gesundheitszustand läßt sich bei einer leichten Krankheit einfacher erreichen als bei einer schweren. Für die Vergütung erscheint es bei Berücksichtigung dieses Aspektes geeigneter, die Verbesserung des Gesundheitszustands heran zuziehen. Es bleibt auch festzuhalten, daß die abgeleiteten Ergebnisse der Agency-Theorie bezüglich der Honorierung des Arztes ohne praktische Umsetzung bleiben (vgl. Arrow 1985, S. 49). Die Frage, ob und wie die ermittelten Vergütungsformen auf das Gesundheitswesen anwendbar sind, bleibt zum gegenwärtigen Zeitpunkt noch unbeantwortet, insbesondere vor dem Hintergrund der komplexen Beziehungen der Hauptakteure.

4.3 Die Agency-Theorie im Rahmen der Informationsbeziehungen im Gesundheitswesen

4.3.1 Informationsbeziehungen und Anreize

Analog zu anderen Bereichen der Ökonomie, besitzt auch das Gesundheitswesen ein hohes Maß an Komplexität, da die im Prozeß der Gesundheitsproduktion involvierten Teilnehmer in unterschiedlichem Maße über Informationen verfügen. Die Interdependenzen zwischen den verschiedenen Akteuren und die unterschiedlichen Interessen lassen es für den einzelnen Teilnehmer interessant

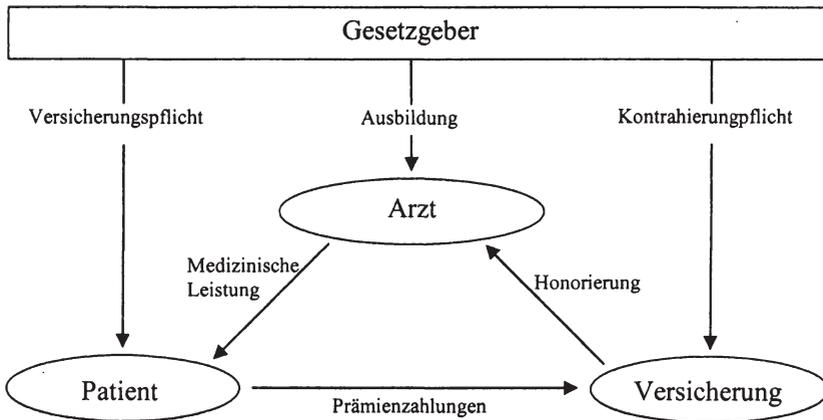
erscheinen, durch einen Informationsvorsprung gegenüber den anderen Teilnehmern die Erreichung individueller Ziele zu fördern (vgl. auch Cutler und Zeckhauser 2000, S. 588ff.).

Von grundsätzlichem Interesse erscheint zunächst, welche Akteure im Hinblick auf die Modellierung von Agency-Beziehungen zu unterscheiden sind. Zum einen existiert die klassische Arzt-Patient-Beziehung. Andere Beziehungen bestehen zwischen Patient und Versicherung sowie zwischen Versicherung und Arzt. Diese Beziehungen treten in Gesundheitssystemen nahezu simultan auf, weshalb eine Analyse, die sich nur auf das Verhältnis zwischen zwei Akteuren beschränkt und hier insbesondere auf das Verhältnis zwischen Arzt und Patient, wichtige Rückkopplungen und Interdependenzen vernachlässigt.

Abbildung 4.5 zeigt das Zusammenwirken der zentralen Akteure im Gesundheitswesen.⁵⁴ Der Gesetzgeber beeinflusst das Verhalten der anderen Akteure, indem er die Rahmenbedingungen des Gesundheitswesens festlegt. Er versucht beispielsweise, den Patienten zu schützen, indem er der Versicherung eine Kontrahierungspflicht vorschreibt oder er versucht, das Niveau der Ausbildung der Ärzte zu sichern, indem er Marktzutrittsbarrieren aufstellt. Zum einen kann es sich um einen numerus clausus oder Zugangstest vor Studienbeginn handeln, zum anderen stellt auch die Kassenzulassung eine Beschränkung des Marktzugangs dar. Gleichzeitig besteht für den Patienten die Pflicht zum Abschluß einer Versicherung, sofern sein monatliches Erwerbseinkommen nicht über der Versicherungspflichtgrenze liegt. Hinzu kommt, daß der Gesetzgeber im Falle einer Sozialversicherung auch den Umfang des Leistungskataloges festlegt, so daß die bisher beschriebenen Akteure in ihrer Handlungsfreiheit in vielfältiger Weise eingeschränkt sind. Der Patient fragt beim Arzt die Diagnose, Therapie sowie medizinische Leistungen in Form von Behandlungen und Verschreibungen nach und wird bei einer Versicherung gegen eine Prämie bzw. den Krankenversicherungsbeitrag gegen das Risiko der Ausgaben für medizinische Leistungen abgesichert. Dabei hängt es von der Art der Versicherung ab, ob und in welcher Höhe eine Eigenbeteiligung für den Versicherten anfällt. Die Vergütung des Arztes für erbrachte Leistungen liegt schließlich, abgesehen von durch den Patienten zu tragenden Selbstbeteiligungen, bei der Versicherung oder Krankenkasse. Deren Spitzenverbände verhandeln mit den Verbänden der Leistungserbringer die Vergütungsvereinbarungen.

⁵⁴ Der Einfluß von Interessengruppen und Verbänden wird an dieser Stelle nicht näher untersucht.

Abbildung 4.5: Akteure im Gesundheitswesen

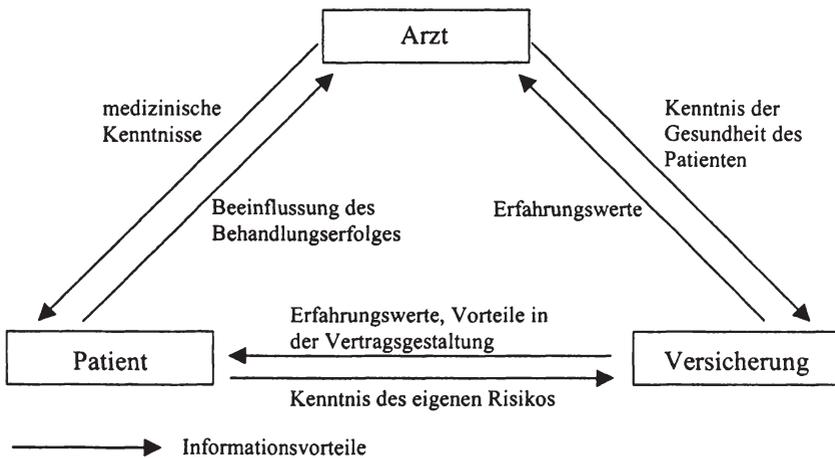


In diesem komplexen Umfeld des Gesundheitswesens entwickeln sich zwischen den einzelnen Marktteilnehmern Informationsbeziehungen, die für die Ausübung ihrer Tätigkeiten und das Ergebnis der Gesundheitsproduktion von großer Relevanz sind. Die Informationsstrukturen zwischen Arzt, Patient und Krankenversicherung lassen sich aus Abbildung 4.6 entnehmen.

Es besteht bei allen Beziehungen ein wechselseitiger Informationsvorsprung, der von einer der beiden Seiten zu ihrem Vorteil ausgenutzt werden kann. Die Versicherung besitzt sowohl gegenüber dem Patienten, als auch gegenüber dem Arzt Erfahrungswerte über Nachfrageverhalten und Abrechnungswerte, die es ihr erlauben, deren Handlungsspielräume einzuengen. Dadurch wird eine falsche Abrechnung durch den Arzt ebenso erschwert wie ungerechtfertigte Anträge, die der Patient an seine Versicherung stellt. Darüber hinaus besitzt sie Erfahrungen bei der Schadensabwicklung. Der Patient kann seinerseits den Heilungserfolg durch sein gesundheitsrelevantes Verhalten beeinflussen. Dazu gehören bspw. sein Lebensstil, die Einhaltung von Therapievorschriften oder auch sportliche Betätigung. Weiterhin kennt er im Gegensatz zur Versicherung sein eigenes Krankheitsrisiko genauer. Aufgrund seines Verhaltens und seiner Erfahrungen aus der Vergangenheit ist es ihm eher möglich, etwaige Risiken abzuschätzen als einer Versicherung, die daher bei der Vertragsgestaltung auf Standardtarife angewiesen ist. Der Arzt schließlich verfügt über detaillierte medizinische Kenntnisse in der Diagnose und Therapie und kann diese gegenüber dem Patienten bei der Verordnung von Leistungen und gegenüber der Versicherung bei der Abrechnung ausnutzen, ohne daß diese genau kontrollieren können, inwie-

weit die Maßnahmen des Arztes eine Berechtigung besitzen. Inwiefern diese Handlungsspielräume konkret ausgenutzt werden, hängt von den Rahmenbedingungen des Gesundheitssystems, d.h. der Ausgestaltung durch den Gesetzgeber und den korporativen Strukturen im Gesundheitswesen, sowie von den konkreten Informationsbeziehungen ab.

Abbildung 4.6: Informationsbeziehungen im ambulanten Sektor



4.3.2 Anforderungen des Gesundheitswesens an Agency-Modelle

Ob und inwieweit sich der im Abschnitt 4.2 vorgestellte Ansatz weiterentwickeln läßt oder ob andere Ansätze eher zur Darstellung der interessierenden Beziehungen im Gesundheitswesen geeignet sind, kann a priori nicht abschließend beurteilt werden. Das Hauptproblem besteht darin, daß die Beziehung zwischen Arzt und Patient zwar den Grundbaustein des Gesundheitswesens bildet, aus Sicht der Principal-Agent-Theorie jedoch Schwächen besitzt (vgl. hierzu Zweifel 1994, S. 20ff. bzw. Pfaff und Zweifel 1998, S. 189f.). Dazu zählen die Vernachlässigung der beiderseitigen Informationsvorteile (Double Moral Hazard)⁵⁵, die Frage, welcher der Akteure die beschriebenen Handlungen unternimmt, sowie die Reduzierung der Betrachtung auf lediglich zwei Akteure. Hieraus resultieren sogenannte ergänzende Sachwalter-Beziehungen (vgl.

⁵⁵ Der Begriff Double Moral Hazard stammt aus der Literatur über die Gewährleistung bei Defekten eines Produktes und bezieht sich auf das Verhalten des Käufers und des Verkäufers (vgl. Cooper und Ross 1985, S. 103ff.)

Zweifel 1994, S. 20). Damit sind Beziehungen zwischen Akteuren bzw. Gruppen von Akteuren gemeint, die in den Prozeß involviert sind und das Marktversagen zwischen Arzt und Patient verhindern bzw. kompensieren sollen. Zu diesen ergänzenden Sachwaltern zählen Verbände, Privat- und Sozialversicherer, Arbeitgeber und der Staat. Ihre Aufgabenbereiche ergeben sich aus den Schwachpunkten des in Abschnitt 4.2 vorgestellten Modells und lassen sich in die beiden Grundbereiche Informationsvermittlung und Vertragsabschluß zusammenfassen (vgl. zur folgenden Darstellung Zweifel 1994 S. 20).

Die Funktion der *Informationsvermittlung* als erster Bereich folgt aus dem Mangel an Detailwissen über den medizinischen Behandlungsprozeß und seine Auswirkungen auf den Gesundheitszustand. Vor allem die Messung der Effektivität der Behandlung erscheint dabei zentral. Ob ergänzende Sachwalter in der Lage sind, diese Auskünfte ohne erhebliche Mehrkosten bei der Informationsbeschaffung bereitzustellen, erscheint zweifelhaft. In aller Regel dürften sie zwar dazu eher in der Lage sein als der Patient; da sie jedoch den Behandlungsprozeß als solchen nicht direkt begleiten und außer den Mitteilungen des Arztes über keinerlei direkte Information verfügen, bleiben ihnen nur die Erfahrungen aus vergangenen Fällen sowie eine kostspielige Überprüfung der medizinischen Behandlung. Der zweite Bereich betrifft das *Aushandeln und den Abschluß von Verträgen*, die für den Patienten selbst mit erheblichem Aufwand verbunden sind. Problematisch ist hierbei, daß der ergänzende Sachwalter ebenfalls die relevanten Informationen benötigt. Welche dies für den konkreten Fall sind, kann auch der ergänzende Sachwalter ex ante nur schwer abschätzen. Weiterhin müßte sichergestellt sein, daß er sich als perfekter Agent des Patienten verhält und keine Eigeninteressen verfolgt.⁵⁶

In der Realität treten ergänzende Sachwalter in der Arzt-Patient-Beziehung häufig auf, vor allem die Versicherer sind hier zu nennen, da sie sowohl mit dem Patienten als auch mit dem Arzt direkt in Kontakt treten. Durch ihre Anwesenheit wandelt sich die direkte Beziehung zwischen Arzt und Patient zu drei bilateralen Kontakten. Diese treten nahezu zeitgleich auf, und die daraus resultierenden Handlungen und Vereinbarungen besitzen Auswirkungen auf die Aktionen

⁵⁶ Zweifel et al. (2001) untersuchen die Wahl eines ergänzenden Sachwalters durch den Konsumenten im Rahmen der Beziehung zwischen Arzt und Patient. Bezüglich der Aufgabenbereiche der ergänzenden Sachwalter schließen sie, daß aus Sicht der Principal-Agent-Theorie zuerst die Informationsvermittlung zu nennen ist. Bezieht man in die Überlegungen jedoch die Kosten der Implementierung eines Vergütungssystems mit ein, so ergibt sich, daß dies für den Patienten selbst kaum zu erreichen ist, und daher die Verhandlungen über die Vergütung die Hauptaufgabe des ergänzenden Sachwalters darstellt.

und Anreize der dritten, nicht involvierten Partei sowie für den gesamten medizinischen Prozeß.⁵⁷

Arzt-Patient

Betrachtet man zunächst die Entscheidung über den Erstkontakt, so treten im ambulanten Sektor originär die Patienten als Nachfrager für medizinische Leistungen auf. Schließt man Vorsorgeuntersuchungen von der Betrachtung aus, ist diese Nachfrage allerdings nicht beständig und tritt zudem nur im Krankheitsfall auf (vgl. Arrow 1963, S. 948).⁵⁸ Das nachgefragte Gut Gesundheitsleistung ist sehr heterogen, d.h. es unterscheidet sich von Patient zu Patient und von Behandlung zu Behandlung (vgl. Gaynor 1994, S. 224). Auch die Präferenzen der Patienten fallen auseinander, so daß einige eine intensive, andere dagegen eine minimale Beratung und Behandlung wünschen. Die Qualität des Gutes medizinische Behandlung ist für den Patienten sowohl vor als auch nach der Behandlung nicht unbedingt meßbar (vgl. Arrow 1963, S. 949 und Richard 1996, S. 201). Dies trifft vor allem dann zu, wenn die Patienten „über den Ablauf medizinischer Prozesse nicht informiert sind und deshalb nicht in der Lage sind, selbst aus Krankheitssymptomen auf eine Therapie zu schließen“ (Richard 1993, S. 2). Entsprechend allen Dienstleistungsproduktionen gilt auch für die Produktion im Gesundheitswesen und für das damit verbundene Produkt das *uno-actu-Prinzip*, und der Umstand, daß Diagnose und Therapie i.d.R. von ein und demselben Arzt vorgenommen werden, ermöglichen dem Anbieter einen diskretionären Handlungsspielraum (vgl. Arrow 1963, S. 949).⁵⁹ Der Arzt kann demnach die Menge der medizinischen Leistungen durch Therapievorschlüsse so steuern, daß sie für ihn gewinnmaximal wird (vgl. Gaynor 1994, S. 229ff. und Richard 1993, S. 2).

Der dem Arzt zur Verfügung stehende Handlungsspielraum hängt entscheidend von der Ausgestaltung des Vergütungssystems und von dem Umstand ab, daß der Arzt zumindest temporär ein Behandlungsmonopol besitzt (vgl. Kortendieck 1993, S. 186). Weiterhin wird der Erfolg der Behandlung auch durch die Compliance des Patienten beeinflusst (vgl. Wille und Ulrich 1991, S. 27ff. und Abschnitt 6.1). Sowohl seine natürlichen Heilungskräfte als auch seine

⁵⁷ Es ist zwar theoretisch durchaus vorstellbar, daß eine direkte (trilaterale) Beziehung zwischen allen drei Akteuren besteht. In der Realität ist eine solche Konstellation allerdings nicht zu erwarten und wird daher auch nicht weiter untersucht.

⁵⁸ Die Probleme der Entscheidung, ob ein Patient einen Arzt aufsucht und die Suche nach dem geeigneten Arzt werden an dieser Stelle von der Betrachtung ausgeklammert.

⁵⁹ Das *uno-actu-Prinzip* besagt, daß bei einer Dienstleistung Produzent und Konsument bei der Erstellung eng zusammenarbeiten. Ökonomisch bedeutet dies, daß Produktion und Konsum räumlich und zeitlich zusammenfallen (vgl. Herder-Dorneich 1994, S. 638).

Unterstützung der medizinischen Behandlung sind hierbei anzuführen. In dieser wechselseitig asymmetrischen Informationsbeziehung zwischen Arzt und Patient tritt letzterer somit als Konsument auf, ist jedoch gleichzeitig auch ein Produktionsfaktor. Die Wirkung einer Behandlung ergibt sich demnach als „joint product“ (Wille und Ulrich 1991, S. 27) zwischen ärztlicher Leistung und dem Verhalten des Patienten. Es existiert also ein diskretionärer Handlungsspielraum eines Akteurs, der sowohl beim Arzt als auch beim Patienten anfallen kann (vgl. Wille und Ulrich 1991, S. 27).

Weiterhin herrscht auf dem Markt für ärztliche Leistungen eine große Unsicherheit bezüglich der Anbieter und der Nachfrager. „Patients are uncertain about their condition, the accuracy of the physician’s diagnosis, his honesty, and the amount of effort or quality he has expected on their case. [...] Physicians do not know the patient’s condition, they are uncertain about the technology of proceeding health from health care, and they do not (generally) know the patient’s reservation price“ (Gaynor, 1994, S. 225).

Patient-Versicherung

Eine weitere wechselseitig asymmetrische Informationsbeziehung besteht zwischen Patient und Versicherung. Hierbei überträgt der Patient die Finanzierung an einen (Versicherungs-) Agenten und leitet somit die finanzielle Kontrolle des Arztes an seine Versicherung weiter (vgl. Kortendieck 1993, S. 187ff.). Vor Vertragsabschluß besitzt die Versicherung lediglich unpräzise Informationen über das tatsächliche Risiko des Patienten. Zwar verfügt sie über allgemeine Erfahrungswerte, jedoch erstrecken sich diese nur auf vergangene Informationen, so daß bei langfristigen Verträgen unerwartete Ereignisse die Unabhängigkeit der Risiken beeinflussen und damit eine zentrale Voraussetzung für die Funktionsfähigkeit des Risikopooling nicht vorliegt (vgl. Kortendieck 1993, S. 188f. und Arrow 1963, S. 960). Der Patient kennt vor Vertragsabschluß nicht die Solvenz des Versicherungsunternehmens und besitzt auch keine Transparenz über die Versicherungsbedingungen.

Nach Vertragsabschluß besteht für die Versicherung das Problem, daß sie den Gesundheitszustand des Patienten und auch dessen gesundheitsrelevantes Verhalten nicht beobachten kann und somit die Kosten der medizinischen Behandlung als gegeben hinnehmen muß. Dieses „verifiability problem“ (Stiglitz 1983, S. 5) spiegelt damit die Unkenntnis über das Auftreten und die Ausprägung des Versicherungsfalles wider. Grundsätzlich lassen sich zwei Formen des Versichertenverhaltens unterscheiden (vgl. Breyer und Zweifel 1999, S. 186 und Pauly 1974, S. 48). Zum einen besitzt der Patient die Möglichkeit, die Wahrscheinlichkeit des Eintritts einer Erkrankung durch Vorbeugung und allgemei-

nen Lebenswandel zu beeinflussen (ex ante Moral Hazard). Zum anderen kann der Patient auf den Genesungsverlauf und die Wahl der Behandlungsmethode Einfluß nehmen (ex post Moral Hazard).

Für den Patienten besteht das Informationsdefizit hauptsächlich in bezug auf die Schadensabwicklung, da hier der langjährigen Erfahrung der Versicherung ein großes Gewicht zukommt (vgl. hierzu Kortendieck 1993, S. 188). Es existiert daher ein Anreizproblem für den Patienten, das durch die Anwesenheit der drei Faktoren Risiko, Versicherung und unvollständige asymmetrische Information entsteht.⁶⁰ Bei einer (Voll-)Versicherung konsumieren die Patienten die medizinische Leistung zu Grenzkosten von Null (vgl. Pauly 1968, S. 532ff.). Falls die Preiselastizität der Nachfrage nicht vollkommen unelastisch ist, führt dies zu einer höheren Nachfrage, die grundsätzlich eine höhere Prämie zur Folge haben müßte. Vergleichen läßt sich diese Situation mit der des prisoners' dilemma (vgl. Eisen 1976, S. 199). Zwar ist die prämiensteigende Auswirkung eines erhöhten Konsums der medizinischen Leistung allen Individuen bekannt, niemand wird sich jedoch einschränken, da der individuelle Nutzen des Mehrkonsums die zusätzlich anfallenden Kosten, die sich auf alle Versicherten verteilen, übersteigt und das Individuum somit nur einen Anteil trägt. Daher dominiert für den Patienten die Strategie der exzessiven Nachfrage die Strategie der Zurückhaltung des Konsums aufgrund der damit verbundenen Prämien erhöhungen.

Die Versicherung besitzt in der Beziehung zum Patienten Informationsvorteile bezüglich der Auslegung der im Versicherungsvertrag enthaltenen Bedingungen, die notwendigerweise nicht vollständig für alle zukünftigen Ereignisse spezifiziert werden können. Da in einem Versicherungsmarkt der Versicherungsabschluß und die Prämienzahlungen sowie die eventuellen Leistungen zeitlich auseinanderfallen, muß sich der Vertrag auf eine Vielzahl von zukünftigen Schäden anwenden lassen. Durch eine zu genaue Spezifizierung wird dieser späteren möglichen Risiken nicht mehr gerecht, es entstehen im Zeitablauf zunehmend Versicherungslücken. Aufgrund dieses Umstandes wird die Unsicherheit des Patienten durch die Versicherung nicht vollständig beseitigt, sondern durch neue Unsicherheit bezüglich des Verhaltens der Versicherung abgelöst.

⁶⁰ Im Falle der symmetrischen Information würde ein Versicherungsvertrag alle möglichen Aktionen mit einbeziehen und es ergäbe sich daher kein Anreizproblem. Bei risikoaversen Patienten, für die das Risiko keine Rolle spielt, existiert keine Versicherung (vgl. Stiglitz 1983, S. 6).

Arzt-Versicherung

Auch das Verhältnis von Arzt und Versicherung ist durch eine asymmetrische Informationsbeziehung gekennzeichnet. Diese betrifft sowohl die Verhandlungen zwischen den Ärzteverbänden und den Vertretern der Krankenkassen als auch die Finanzierung der Leistungen durch die Versicherung, die an die Stelle des Patienten tritt. Die Kontrolle der vom Arzt erbrachten Leistungen und der Abrechnungen kann die Versicherung nur aufgrund ihrer Erfahrungswerte vornehmen, nicht jedoch aufgrund des Gesundheitszustands des Patienten. Man kann sich zwei Extremfälle der Arzthonorierung vorstellen (vgl. Breyer und Zweifel 1999, S. 259ff.). Zum einen wird der Arzt unabhängig von dem erzielten Ergebnis und der erbrachten Leistung entlohnt. In diesem Fall erhält er für jeden behandelten Patienten eine Fallpauschale. Zum anderen besteht die Möglichkeit der Einzelleistungsvergütung. Hierbei wird das Ergebnis ebenfalls nicht in die Entlohnung mit einbezogen, die Vergütung erstreckt sich nur auf den Umfang der erbrachten medizinischen Leistung. Die Fallpauschale bewirkt nun, daß der Arzt möglichst viele Patienten behandeln möchte, diese allerdings nicht besonders intensiv. Bei der Einzelleistungsvergütung liegt ein Anreiz zur Ausweitung der Menge an medizinischen Leistungen und damit auch der Behandlungsausgaben vor, da die Vergütung an der Zahl der erbrachten Einzelleistungen ansetzt, unabhängig von ihrer Wirksamkeit.

Es zeigt sich, daß zwischen den Akteuren unterschiedliche Informationsbeziehungen bestehen, daß aber der Arzt in den Beziehungen zu Patient und Versicherung eine herausragende Stellung einnimmt. Das bedeutet, daß seine diskretionären Handlungsspielräume trotz der Einschränkungen durch die anderen Akteure dominieren. Er handelt sowohl für den Patienten als auch für die Versicherung als Agent, so daß eine Beziehung vorliegt, die sich als „double agency“ (Blomqvist 1991, S. 412, vgl. auch Abschnitt 5.3) bezeichnen läßt.

4.3.3 Interdependenzen und Anreize

Die Existenz ergänzender Sachwalter schränkt die Darstellung der Arzt-Patient-Beziehung in einem Zwei-Personen-Modell stark ein. Um die Wirkungsweise der einzelnen Beziehungen abbilden zu können, muß der bisherige Ansatz erweitert werden. Einen möglichen Ausgangspunkt stellen die drei grundlegenden Beziehungen zwischen Arzt, Patient und Versicherung dar, die das Ergebnis des medizinischen Leistungserstellungsprozesses sowie dessen Finanzierung determinieren (siehe hierzu Abbildung 4.7).

In der Beziehung zwischen Patient und Arzt entscheidet der Patient über den Erstkontakt, das heißt über die Frage, ob er einen Arzt aufsuchen soll. Hier

spielen vor allem zwei Faktoren eine entscheidende Rolle, der Gesundheitszustand und die Kosten eines Arztbesuchs. Zu letzteren zählen die Höhe der Behandlungskosten, die der Patient selbst zu tragen hat, die Zeit- und Wegekosten sowie die Opportunitätskosten, die durch andere Tätigkeiten entstehen. Je schlechter der Gesundheitszustand, desto dringlicher ist die Entscheidung über den Erstkontakt und desto eher findet ein Erstkontakt statt. Der Arzt trifft die Entscheidung über die Frequenz eines Arztbesuches, d.h. über die Häufigkeit und Intensität einer Behandlung.⁶¹ Er stellt die Diagnose und empfiehlt bzw. verordnet eine Therapie, die u.U. auch mit Hilfe des Patienten aus mehreren Vorschlägen ausgewählt wird. Bei der Festlegung der medizinischen Leistung entscheidet der Arzt nach einem Kosten-Nutzen-Kalkül, in das Entlohnungsaspekte, der entstehende Aufwand und sein Berufsethos einfließen.⁶² Die ethische Komponente unterstellt, daß der Arzt an einem hohen Gesundheitszustand des Patienten interessiert ist und keine gesundheitsschädigenden Leistungen erbringt. Einen ähnlichen Effekt besitzt ein Reputationsmechanismus, der den Arzt anhält, gute Leistungen zu erbringen, damit der Patient ihn im Krankheitsfall erneut aufsucht und nicht zu einem Kollegen wechselt.

Der Erfolg einer Behandlung hängt jedoch nicht alleine von der ausgewählten medizinischen Leistung und der Intensität der Behandlung ab. Es bestehen auch Effekte, die sowohl der Arzt als auch der Patient nicht beeinflussen können - hierunter fallen insbesondere Umweltzustände und die Selbstheilungskräfte des Patienten. Zum Erfolg einer Behandlung trägt zusätzlich auch das Verhalten des Patienten bei. Dazu gehört beispielsweise das Einhalten von Dosierungsvorschriften für Medikamente oder die Befolgung der ärztlichen Anweisungen.⁶³

Die Beziehung zwischen Arzt und Patient determiniert somit die medizinische Leistung, die Intensität der Behandlung, die Höhe der Behandlungsausgaben und letztlich auch den Gesundheitszustand des Patienten. Die Beziehung zwischen der Versicherung bzw. Krankenkasse und dem Patienten bestimmt das Versicherungsverhältnis. Hier greift der Staat durch zahlreiche Vorschriften in diese Beziehung ein. Grundsätzlich wird über die Höhe der Prämien- bzw. Beitragszahlungen und über die Auszahlung im Schadensfall, d.h. Kostenerstattung oder Sachleistung, entschieden. Hinzu kommen eventuelle Selbstbeteiligungen

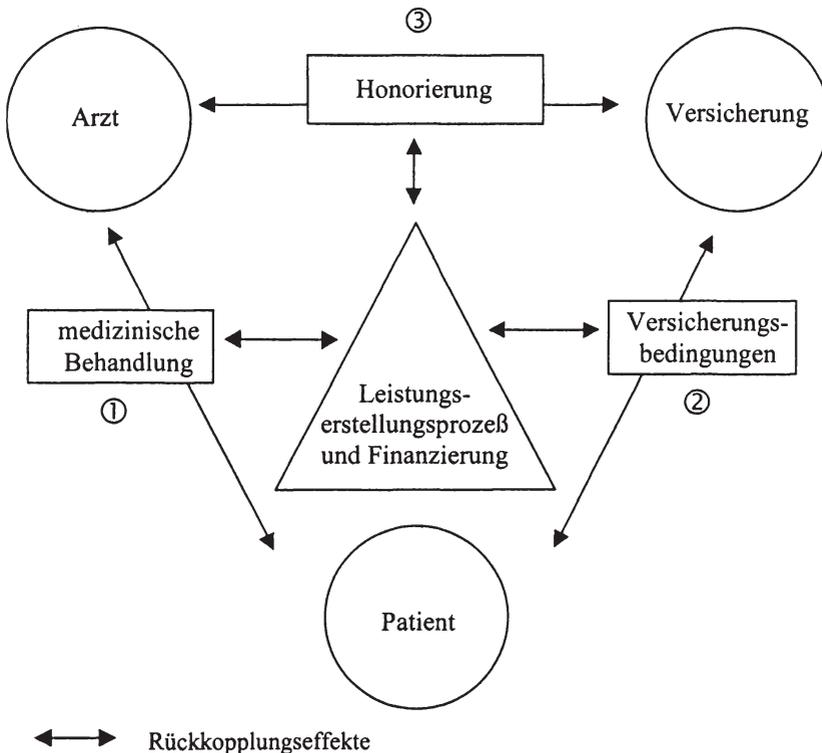
⁶¹ Zu einer Untersuchung über den Entscheidungsprozeß bei der Nachfrage nach medizinischen Leistungen siehe Pohlmeier und Ulrich (1995).

⁶² Siehe hierzu bspw. die Arbeit von Zweifel (1982) oder den Beitrag von McCullough (1988).

⁶³ Verhält sich der Patient nicht entsprechend der ärztlichen Anordnungen oder führt sein Verhalten zu einer Verzögerung des Heilungserfolges, entspricht dies dem Fall des ex post Moral Hazard (vgl. Breyer und Zweifel 1999, S. 186 und Pauly 1974, S. 48).

und Selbstbehalte. Diese Regelungen beeinflussen die Kosten für den Patienten, die entweder ständig oder nur im Krankheitsfall auf ihn zukommen. Ebenfalls werden die Entscheidung des Patienten über den Erstkontakt und die möglichen Alternativen eines Arztbesuches, die beispielsweise die Selbstmedikation betreffen, beeinflusst. Es bleibt also festzuhalten, daß die Versicherungsbedingungen die Inanspruchnahme der medizinischen Leistungen durch den Patienten und die Finanzierung bestimmen, darüber hinaus aber auch das Risikoverhalten und die Vorsorge des Patienten beeinflussen (ex ante Moral Hazard).

Abbildung 4.7: Ein dreiteiliges Modell des Gesundheitsprozesses



Ärzte und Versicherung handeln den Leistungskatalog aus und setzen staatlich festgelegte Vergütungsformen der ärztlichen Leistungen um bzw. regeln die Honorierung der erbrachten Leistungen. Der Gesetzgeber hat in dieser Beziehung eine Vielzahl von Regulierungen erlassen. So schreibt er bspw. in der ge-

setzlichen Krankenversicherung die abrechnungsfähigen Leistungen vor und damit auch den Anteil, den der Patient direkt zu übernehmen hat. Somit legt er fest, in welche Bereichen des Gesundheitswesens die Ausgaben durch Budgets gedeckelt sind. Bei den Verhandlungen muß die Versicherung aufgrund ihrer Erfahrungswerte die Aufwendungen der Ärzte und die daraus resultierenden Ausgaben abschätzen. Die Art der Vergütung beeinflusst auch das Verhalten des Arztes in bezug auf den Umfang der medizinischen Leistungen. So besteht der bereits beschriebene Anreiz, bei einer Einzelleistungsvergütung die Menge auszuweiten, während durch eine Fallpauschale die Anzahl der behandelten Patienten steigt. Bei einer Kostenbeteiligung des Arztes besitzt er dagegen kaum einen Anreiz, unnötige medizinische Leistungen zu erbringen. Der Honorierungsform kommen daher mittelbar Auswirkungen auf die Behandlung des Patienten zu. Die daraus resultierenden Behandlungsausgaben determinieren weiterhin auch die Höhe der zukünftigen Prämien und Beiträge.

Es zeigt sich also, daß zwischen den Akteuren des Gesundheitswesens und ihren vertraglichen Beziehungen Interdependenzen bestehen, die bei der formalen Ausgestaltung eines Principal-Agent-Modells mit einbezogen werden müssen. In einem erweiterten Ansatz sind daher sowohl (1) Aspekte der medizinischen Behandlung, (2) die Versicherungsbedingungen und (3) die Honorierung der erbrachten Leistung zu berücksichtigen. Werden einzelne Bereiche vernachlässigt, kann die Beziehung zwischen Arzt, Patient und Versicherung nicht mehr als Ganzes analysiert werden.

4.4 Kritische Würdigung der Agency-Theorie im Gesundheitswesen

Die Anwendung der Principal-Agent-Theorie auf die Beziehung zwischen Patient und Arzt macht deutlich, daß eine konventionelle Analyse den Strukturen des Gesundheitswesens nicht gerecht wird. Zum einen treten neben der Beziehung zwischen Patient und Arzt, die den Grundbaustein des Gesundheitswesens bildet, ergänzende Sachwalter hinzu. Diese sind in den Prozeß der Gesundheitsproduktion involviert und sollen dazu beitragen, das Marktversagen zwischen Arzt und Patient zu verhindern bzw. zu kompensieren. Sie sollen zwei grundlegende Funktionen erfüllen: Zum einen betrifft dies das Aushandeln und den Abschluß von Verträgen, zum anderen die Vermittlung von Informationen. Eine besondere Stellung unter ihnen kommt der Versicherung bzw. der Krankenkasse zu, da diese mit beiden Akteuren, Patient und Arzt, direkt in Kontakt treten. Als Beispiel hierfür ist die Vergütung ärztlicher Leistungen oder die Absicherung des finanziellen Krankheitsrisikos des Patienten gegen eine Prämienzahlung zu nennen. Darüber hinaus bestehen zwischen den Akteuren des Gesundheitswesens und ihren Handlungen Interdependenzen. Die Entscheidung des Patienten, einen Arzt aufzusuchen und damit medizinische Leistungen in Anspruch zu

nehmen, beeinflusst bspw. auch das Angebotsverhalten des Arztes, die Abrechnung mit der Versicherung oder Krankenkasse und damit auch den Versicherungsvertrag des Patienten. Es ist daher notwendig, bei einer Analyse des Prozesses der Gesundheitsproduktion diese Rückkopplungen zwischen den einzelnen Akteuren und der Abhängigkeit ihrer Handlungen von den vertraglichen Beziehung zu berücksichtigen.

5 Forschungsansätze über das Verhältnis Arzt - Patient - Versicherung

5.1 Übersicht

Für eine Weiterentwicklung des dargestellten Ansatzes der Übertragung der Principal-Agent-Theorie auf das Gesundheitswesen und insbesondere für die Integration der in Abschnitt 4.3 beschriebenen Ansatzpunkte der theoretischen Untersuchung der Beziehung spielen die vielfältigen gesundheitsökonomischen Modellansätze der Beziehung zwischen Arzt, Patient sowie einer Versicherung eine besondere Rolle. Vor diesem Hintergrund erfolgt ein umfassender Überblick über bisherige Studien, beginnend mit den Auswirkungen asymmetrischer Information zwischen den Vertragsparteien im Gesundheitswesen, Untersuchungen, die sich speziell an ein Versicherungssystem, sei es die amerikanische Health Maintenance Organization oder der britische National Health Service, anlehnen, sowie Ansätzen, die direkt den Arzt als Agenten des Patienten betrachten. Im weiteren Verlauf dieses Kapitels werden dann vier ausgewählte Arbeiten näher vorgestellt. Bei diesen handelt es sich um Ansätze, die direkte Anknüpfungspunkte zu den Problemen der Umsetzung der Agency-Theorie auf das Gesundheitswesen bieten oder in Verbindung zu dem im nächsten Kapitel präsentierten Ansatz stehen. Dargestellt werden die Modelle von Ellis und McGuire (1990), Blomqvist (1991), Ma und McGuire (1997) sowie Leonard und Zivin (2001).

In der ökonomischen Literatur existiert eine Vielzahl von Untersuchungen über die Agency-Beziehungen im Gesundheitswesen. Eine allgemeine Analyse über die Auswirkungen von Informationsasymmetrie und Unsicherheit im Gesundheitssektor liefert Rochaix (1997). Zeckhauser (1969) analysiert die Anreize und die Risikoaufteilung bei Verträgen zwischen Patienten und Versicherer. Spence und Zeckhauser (1971) untersuchen das Zusammenspiel zwischen individuellen Aktionen der Versicherten, der Natur und dem Informationsstand der Versicherung bei der Bestimmung optimaler Verträge. Das Hauptergebnis ist, daß nur in Fällen, in denen der Versicherer den Umweltzustand beobachten kann, eine herkömmliche Versicherung optimal ist. Pauly (1980) stellt den Arzt in seiner Funktion als Agent des Patienten dar und untersucht weiterhin die Rolle von Informationen und der Nachfrage des Konsumenten für das Verhalten des Arztes. Dionne und Contandriopoulos (1983) stellen in einem Übersichtsartikel ein theoretisches Grundmodell für die von Pauly behandelten Punkte auf. Insbesondere gehen sie auf Punkte wie Prestige und Berufsethik des Arztes und die Beziehung zwischen Versicherung, Krankenhaus, Konsument und Arzt ein.

Zweifel (1982) beschreibt ein ökonomisches Modell des ärztlichen Verhaltens und geht dabei insbesondere auf Aspekte der Berufsethik ein.

Meza (1983) behandelt den Zusammenhang zwischen Krankenversicherung und der Nachfrage nach medizinischen Leistungen unter Vernachlässigung der Moral Hazard-Effekte. Er kommt zu dem Schluß, daß es auch bei optimalem Versicherungsschutz zu einem Anstieg der Nachfrage nach medizinischen Leistungen kommt. Meza stellt daher das Resultat von Feldstein (1973) in Frage, der die exzessive Nachfrage nach medizinischen Leistungen dem Moral Hazard-Effekt zuordnet und schließt daher, daß die Auswirkungen einer Versicherung auf die Wohlfahrt überschätzt werden. Ellis und McGuire (1993) untersuchen Selbstbeteiligungen auf Anbieter- und Nachfragerseite und die daraus resultierenden Anreize. Clayton et al. (1993) analysieren die Arzt-Patient-Beziehung vor dem Hintergrund der Kommunikation und Interaktion zwischen beiden. Des weiteren untersucht Chiu (1997) Versicherungen und Konsumentenwohlfahrt in einem Modell, in dem sich Ärzte als perfekte Agenten verhalten und in dem diese immer im Interesse ihrer Patienten handeln. Die Einführung einer konventionellen Versicherung führt dann zu einem Wohlfahrtsverlust für die Konsumenten, wenn das Angebot an Gesundheitsleistungen hinreichend preisunelastisch ist. Eigene Handlungen des Patienten werden von Giuffrida und Gravelle (1998) mit in die Überlegungen einbezogen mit dem Ergebnis, daß finanzielle Anreize in manchen Fällen die Compliance des Patienten erhöhen können.

Ein Teil der Arbeiten befaßt sich speziell mit der Health Maintenance Organization (HMO) in den USA. So untersucht Selden (1991) Capitation-Verträge⁶⁴ und kommt dabei zu dem Ergebnis, daß die optimale Versorgung eine Vollversicherung des Patienten mit einem gemischten Vergütungssystem kombiniert. Kwon (1997) analysiert die Determinanten der Vergütungsstrukturen bei Verträgen zwischen Leistungserbringung und einer Versicherung. Die Struktur der Honorierungsverträge und ihre Verteilung im HMO-Markt hängt von zwei Faktoren ab: Zum einen von der Bedeutung der Qualität der Leistungen bei verschiedenen Vergütungssystemen für den Konsumenten und zum anderen von dem Grad, zu dem die Ärzte bereit sind, finanzielle Risiken zu akzeptieren. Im Fall von Managed Care Organisationen stellen Stone und Mantese (1999) den Einfluß von Managed Care Systemen auf verschiedene Formen der Principal-Agent-Beziehung zwischen Arzt und Patient dar. Bereits Ma und Riordan (1997) zeigen, daß durch Managed Care eine First-best-Lösung erreicht und die exzessive Nachfrage durch Moral Hazard in der Krankenversicherung vermindert wird.

⁶⁴ Unter Capitation-Verträgen versteht man Vergütungssysteme, bei denen der Arzt eine Fallpauschale für die Behandlung erhält.

Im Rahmen des britischen National Health Service (NHS) haben einige Studien die vertraglichen Beziehungen untersucht. Der Ansatz von Levaggi (1996) beschreibt die Schaffung eines ‚internen Marktes für Gesundheitsleistungen‘ durch die Trennung der Verantwortlichkeit für den Kauf der Leistungen und ihre Erbringung. Mit Hilfe eines Principal-Agent-Modells ermittelt Levaggi, daß es durch den Wettbewerb der Leistungserbringer dem Käufer möglich ist, private Informationen zu erhalten. Durch die strikte Einhaltung der Budgets ändert sich die Risikoaufteilung und der First-best-Vertrag im Vergleich zur Standardliteratur. Chalkey und Malcomson (1996) beschreiben die Eigenschaften von NHS-Verträgen und untersuchen in einer weiteren Studie (Chalkey und Malcomson 1998) NHS-Verträge, bei denen der Arzt nur nach der Anzahl der behandelten Personen bezahlt wird, wenn die Qualität der erbrachten Leistungen nicht beobachtet werden kann. Weiterhin zeigen Chalkey und Malcomson (1998a) für den Fall, daß die Nachfrage des Patienten nicht die Qualität der erbrachten Leistungen des Anbieters wiedergibt, daß für teilweise wohlwollende Anbieter eine Kostenbeteiligung optimal ist. Fraja (2000) entwickelt ein Modell, indem der NHS den Anbietern medizinischer Leistungen einen Vertrag anbietet, ohne die genauen Kosten der Anbieter zu kennen. Hauptergebnis der Analyse ist, daß der Preis pro Untersuchung mit der Effizienz des Anbieters steigen sollte.

Einen Ansatz der Arzt-Patient-Interaktion präsentiert Rochaix (1989), in dem die Informationsasymmetrie zwischen Arzt und Patient durch die Unsicherheit bezüglich des Gesundheitszustands des Patienten modelliert ist. Es wird gezeigt, daß durch die Suche der Patienten nach adäquater Behandlung die Ärzte sich als ‚bessere Agenten‘ verhalten. Clark und Olsen (1994) betrachten ein Modell, in dem der Arzt als perfekter Agent für eine Gruppe von Patienten tätig ist. Dabei werden die Gesundheitsleistungen von einer Gruppe von Kostenträgern finanziert. Es ergibt sich ein endogenes Gesundheitsbudget, das von den Entscheidungen des Arztes abhängt. Lee (1995) präsentiert ein Modell der Arzt-Patient Interaktion unter asymmetrischer Information. In diesem Rahmen wird die Erstkontakt-Entscheidung des Patienten explizit modelliert. Eine Versicherung tritt allerdings nicht explizit als Akteur auf. Jelovac (2001) untersucht Honorierungsverträge für Ärzte, wenn weder der Aufwand des Arztes bei der Beschaffung von Informationen über den Gesundheitszustand noch der aktuelle Gesundheitszustand vertraglich festgeschrieben werden können. Für wiederholte Inanspruchnahme medizinischer Leistungen durch den Patienten resultiert eine Selbstbeteiligung auf Anbieterseite als optimaler Vertrag, durch den der Arzt mit den nötigen Anreizen ausgestattet wird.

Vor dem Hintergrund der Principal-Agent-Beziehung zwischen Arzt und Patient stellt Mooney (1991) verschiedene Ansätze in der Gesundheitsökonomie vor mit dem Ergebnis, daß nicht nur Gesundheit oder Information des Patienten im Zentrum des Interesses sein sollten, sondern daß das Hauptaugenmerk auf den Nutzen des Patienten gelegt werden sollte. Mooney und Ryan (1993) üben Kritik an der Anwendung der Principal-Agent-Theorie, da ihrer Ansicht nach Aspekte wie die Interaktionen zwischen den Nutzenfunktionen oder Anreizkompatibilität nicht ausreichend berücksichtigt werden. Ryan (1994) erweitert ökonomische Aspekte der Arzt-Patient-Beziehung im Rahmen der Agency-Theorie um soziologische Blickpunkte. Vick und Scott (1995), Vick und Scott (1998) und Scott und Vick (1999) untersuchen die Agency-Beziehung zwischen Arzt und Patient mit Hilfe empirischer Auswertungen von Fragebogenstudien hinsichtlich der Ursachen der Mängel in dieser Beziehung und den Präferenzen des Patienten bezüglich dieser. Hauptergebnis der Untersuchungen ist die Notwendigkeit einer verbesserten Ausbildung der Ärzte in bezug auf ihre Kommunikation mit den Patienten.

5.2 *Optimale Vergütungs- und Versicherungssysteme*

Ellis und McGuire (1990) behandeln die Problematik der optimalen Wahl eines Versicherungs- und Vergütungssystems im Gesundheitswesen. Dort kann der Fall eintreten, daß nicht alle Zahlungssysteme zu einer Markträumung führen. Darüber hinaus muß die angebotene Menge an medizinischer Leistung des Arztes nicht der nachgefragten Menge des Patienten entsprechen (vgl. Ellis und McGuire 1990, S. 376ff.). Mit Hilfe eines Konfliktlösungsansatzes wird der Fall untersucht, in dem Arzt und Patient nicht über den Umfang der medizinischen Behandlung übereinstimmen, und es werden Bedingungen für das Versicherungs- und Vergütungssystem ermittelt; die Präferenzen hängen von den komplexen Versicherungs- und Vergütungssystemen ab. Darüber hinaus existieren keine Preise, die zu einer Markträumung führen. Es stellt sich also die Frage, wie eine Übereinkunft über den Umfang der medizinischen Leistungen erzielt wird und wie das Versicherungs- und Vergütungssystem ausgestaltet sein sollte.

Modellierung der Konflikt-Lösung

Generell bestehen mehrere Möglichkeiten, die abweichenden Auffassungen zwischen Arzt und Patient zu behandeln. Eine Annahme ist, daß bei Divergenz zwischen Angebot und Nachfrage das Prinzip der kürzeren Marktseite gilt. Dies bedeutet, falls X_D^* und X_S^* die gewünschten nachgefragten und angebotenen Mengen bezeichnen, ergibt sich die realisierte Menge als $X = \min(X_D^*, X_S^*)$. Allerdings erscheint eine solche Lösung für das Gesundheitswesen zu einfach, da die Relevanz der Arzt-Patient-Beziehung Druck auf beide Seiten ausübt, eine

Einigung zu erzielen. So wissen die Patienten um die speziellen Kenntnisse und Vorteile des Arztes und sind von daher bereit, ihm entgegen zu kommen. Für den Arzt existieren rechtliche, ethische und kommerzielle Gründe, die Zufriedenheit des Patienten zu berücksichtigen und die Menge an medizinischer Leistung an dessen Wünschen zu orientieren. Eine Variation dieses Ansatzes liegt darin, eine Gewichtung durch eine einfache Funktion wie bspw. $X = \delta X_D^* + (1-\delta) X_S^*$ zu spezifizieren, mit $0 \leq \delta \leq 1$. Eine fehlende theoretische Fundierung dieses Modells in reduzierter Form stellt jedoch einen Nachteil einer solchen Formulierung dar.

Ellis und McGuire verwenden statt dessen ein Verhandlungsmodell. Dabei bezeichnet der von ihnen definierte ‚Punkt minimaler Erwartungen‘ eines Akteurs diejenige Menge an medizinischer Leistung, die der andere Akteur präferiert (X_D^* aus Sicht des Arztes, X_S^* aus Sicht des Patienten). Mit Hilfe dieses Ansatzes kann das Ergebnis des Verhandlungsprozesses als Lösung folgender Cobb-Douglas Wohlfahrtsfunktion dargestellt werden:

$$\max_x \left[U(X) - U(X_S^*) \right]^{(1-\gamma)} \left[V(X) - V(X_D^*) \right]^\gamma, \quad (5.1)$$

wobei $U(X)$ und $V(X)$ die Nutzenfunktionen des Patienten und des Arztes darstellen und $(1-\gamma)$ und γ das jeweilige Verhandlungsgewicht, mit $0 \leq \gamma \leq 1$. Unter der Annahme von konkaven, quadratischen Nutzenfunktionen und gleicher Verhandlungsgewichte ($\gamma = 1/2$) ergibt sich die Menge an medizinischer Leistung als Durchschnitt der beiden gewünschten Mengen:

$$X = 1/2 (X_D^* + X_S^*). \quad (5.2)$$

Moral Hazard und Agency-Beziehungen

Die Anwendung des dargestellten Konzeptes zur Konfliktlösung erfolgt auf Situationen, in denen der Patient krank ist und den Arzt aufsucht. Dabei beeinflusst Moral Hazard die Lage der Nachfragekurve des Patienten, der Versicherungsumfang bestimmt den gewünschten Punkt auf der Nachfragekurve. Die Ausprägung bzw. die Art der Agency-Beziehung hat Einfluß auf die Lage der Angebotsfunktion des Arztes, und durch das Vergütungssystem wird die gewünschte Menge bestimmt.

Für den Patienten wird angenommen, daß er, obwohl versichert, risikoneutral ist. Falls der Patient gesund ist, hängt sein Nutzen nur von einem nicht gesund-

heitsbezogenen Gut N ab, im Krankheitsfall zusätzlich von den konsumierten Gesundheitsleistungen X . Letztere werden monetär ausgedrückt, und die Kosten pro Einheit sind auf Eins normiert. Sein Einkommen sei Y , die Versicherungsprämie P und c die Höhe der Selbstbeteiligung, mit $0 \leq c \leq 1$. Daraus ergibt sich, daß der Nutzen bei Gesundheit $U_H = N = Y - P$ ist. Für den Krankheitsfall folgt hieraus:

$$U_I = N + B(X) - K = Y - P - cX + B(X) - K. \quad (5.3)$$

Der Gesamtnutzen aus der medizinischen Leistung wird mit $B(X)$ bezeichnet, und K stellt eine Konstante dar, stellvertretend für die fixen Kosten der Krankheit. Um die Ergebnisse der Konflikt-Lösung anzuwenden, ist es nötig, daß die Gesundheitsertragsfunktion $B(X)$ konkav und quadratisch verläuft:

$$B(X) = aX - \frac{1}{2}bX^2, \quad (5.4)$$

mit $a > 0$ und $b > 0$. Für eine gegebene fixe Prämie wählt der Patient die Menge X , ohne die Rückwirkung auf die Höhe der Prämie zu beachten. Durch Substitution von (5.4) in (5.3) ergibt sich für eine gegebene Selbstbeteiligung die Menge, die den Nutzen des Patienten maximiert:

$$X_D^* = \begin{cases} (a - c) / b, & c < a \\ 0, & \text{sonst} \end{cases} \quad (5.5)$$

Der Arzt zieht Nutzen aus seinem Behandlungsgewinn π und aus den Gesundheitserträgen des Patienten B . Letzteres bedeutet, daß der Arzt als Agent des Patienten handelt. Beide Komponenten hängen von der erbrachten medizinischen Leistung X ab. Die Zielfunktion des Arztes stellt sich damit dar als

$$V(X) = \alpha B(X) + \Pi(X). \quad (5.6)$$

Die Grenzrate der Substitution α zwischen Gewinnen und Gesundheitserträgen ist konstant und beschreibt den Grad der Agency-Beziehung zwischen Arzt und Patient. Für $\alpha = 1$ ist der Arzt ein perfekter Agent, er gewichtet die Gesundheitserträge und seinen Gewinn gleich. Eine unvollkommene Agentenrolle wird durch $\alpha < 1$ gekennzeichnet.⁶⁵ Das Vergütungssystem und damit der Gewinn des

⁶⁵ Weiterhin wird in dem Artikel der Fall einer ‚super agency‘ betrachtet, mit $\alpha \geq 1$ (vgl. Ellis und McGuire 1990, S. 382).

Arztes besteht aus zwei Komponenten: erstens einer Pauschale R , die unabhängig von der erbrachten medizinischen Leistung bezahlt wird und zweitens einer auf den Kosten basierenden Komponente $(1-s)X$. Dabei gibt s die Kostenbeteiligung auf der Anbieterseite an. Unter der Annahme konstanter Grenzkosten von Eins ergibt sich damit als Gewinnfunktion:

$$\Pi(X) = R - sX. \tag{5.7}$$

Ein solches Vergütungssystem funktioniert in Analogie zum Versicherungssystem des Patienten. Es lassen sich dabei zwei Extremfälle identifizieren. Der erste Extremfall äußert sich in der prospektiven Vergütung, die durch $s = 1$ gekennzeichnet ist. Der zweite Fall entspricht einem an den Kosten orientierten System mit $R = 0$ und $s = 0$. Setzt man Gleichung (5.7) in (5.6) ein, so ergibt sich nach Optimierung die gewünschte Menge an medizinischer Behandlung des Arztes:

$$X_s^* = \begin{cases} (a - s/\alpha) / b, & s < \alpha a \\ 0, & \text{sonst.} \end{cases} \tag{5.8}$$

Man erkennt, daß für eine unvollkommene Agency-Beziehung und damit für niedrige Werte von α der Arzt stärker auf die Anreize des Vergütungssystems durch die Kostenbeteiligung s reagiert.

Aus den gewünschten medizinischen Leistungen für Patient (5.5) und Arzt (5.8) läßt sich nun die Verhandlungslösung analog zu Gleichung (5.2) bestimmen. Man erhält für symmetrische Verhandlungsmacht:

$$X(c, s) = a/b - (c + s/\alpha) / 2b. \tag{5.9}$$

Bei dieser Menge an medizinischer Leistung sind die sozialen Grenzkosten gleich dem sozialen Grenznutzen. Ellis und McGuire zeigen, daß die soziale Wohlfahrt identisch mit dem Nutzen des Patienten ist.⁶⁶ Dies resultiert daraus, daß die Wohlfahrt des Arztes sich lediglich aus seinem Behandlungsgewinn ergibt und der erwartete Gewinn vollständig durch die Pauschale abgeschöpft

⁶⁶ Das soziale Optimum ist nach Ellis und McGuire durch $B'(X) = 1$ gekennzeichnet, d.h. der Grenznutzen der medizinischen Leistung ist konstant. Die rechte Seite resultiert aus den konstanten Kosten je Einheit medizinischer Leistung. Aus $B'(X) = 1$ und (5.4) ergibt sich, daß $X^D = (a-1)/b$ ist und daß diese Menge optimal ist, wenn für die Zahlungssysteme $(c+s/\alpha)/2 = 1$ gilt (vgl. Ellis und McGuire 1990, S. 383f.)

werden kann, indem diese gleich den erwarteten Kosten gesetzt wird. Weiterhin gilt, daß ein soziales Optimum nie erreicht wird, wenn $\alpha > 1$ ist. Dann wird lediglich eine Second-best-Lösung erreicht mit $c = s = 1$, d.h. der Patient ist nicht versichert und es existiert ein prospektives Vergütungssystem für den Arzt. Für $\alpha < 1$ existieren multiple Gleichgewichte, die das soziale Optimum erreichen.⁶⁷

Weitere Implikationen ergeben sich aus der Lösung des Verhandlungsprozesses. Erstens kann es zu unverhältnismäßig hohen Ausgaben für medizinische Leistungen kommen, selbst dann, wenn der Patient nicht versichert ist, da ohne Kostenbeteiligung auf Anbieterseite die konsumierte Menge an medizinischer Leistung zu hoch ausfällt. Zweitens beeinflussen die Ärzte den Nutzen aus den medizinischen Leistungen, ohne die Nachfrage zu ändern. Die angebotene Menge kann daher abseits der Nachfragekurve liegen. Dies stellt eine alternative Interpretation der anbieterinduzierten Nachfrage dar. Drittens ergibt sich auch in diesem einfachen Modell ohne Risikoaversion mit zwei Instrumenten (c, s), daß die First-best-Situation nicht immer erreicht werden kann, selbst wenn als soziales Ziel der effiziente Konsum medizinischer Leistungen formuliert wird.

Risikoaversion des Patienten

In einer Erweiterung des Modells im Rahmen der Erwartungsnutzen-Theorie ist der Patient mit zwei unterschiedlichen Zuständen der Welt konfrontiert, die mit seinem Gesundheitszustand und dem Grenznutzen des Einkommens variieren. Der Patient ist mit der Wahrscheinlichkeit $1 - \delta$ gesund und mit Wahrscheinlichkeit δ krank, mit $0 < \delta < 1$. Ein gesunder Akteur hat einen Grenznutzen des Einkommens von η , der eines Kranken beträgt λ . Er ist risikoavers, wenn der Grenznutzen des Einkommens bei Gesundheit kleiner als der Grenznutzen bei Krankheit ist, d.h. ($\eta < \lambda$). In diesem Fall wird er eine aktuarisch faire Versicherung abschließen. Falls er risikofreudig ist, also $\lambda < \eta$ gilt, wird er keine Versicherung vereinbaren. Für den Erwartungsnutzen ergibt sich dann in Analogie zu Gleichung (5.3):

$$\begin{aligned} EU &= (1 - \delta) \eta (Y - P) + \delta [\lambda (Y - P - cX) + B(X) - K] \\ &= Y - P + \delta [B(X) - \lambda cX - K]. \end{aligned} \tag{5.10}$$

⁶⁷ Da das gewünschte Angebot des Arztes mit abnehmender Agency α auch sinkt, muß dann die Kostenbeteiligung s auch gesenkt werden, um die unvollkommene Agency-Beziehung zu kompensieren (vgl. Ellis und McGuire 1990, S. 384).

Die zweite Zeile ergibt sich, wenn der erwartete Grenznutzen des Einkommens gleich Eins gesetzt wird.⁶⁸ Für den Fall $\lambda > 1$ liegt dann Risikoaversion vor. Für die Versicherungsprämie gilt dann in dieser Modellvariante:

$$P = \delta(1 - c)X. \tag{5.11}$$

Die nachgefragte Menge an medizinischer Leistung ergibt sich aus der Maximierung des Nutzens bei Krankheit unter konstanter Prämie:

$$X_D^* = \begin{cases} (a - c\lambda)/b, & c < a/\lambda \\ 0, & \text{sonst.} \end{cases} \tag{5.12}$$

Man erkennt dabei den Unterschied zu Gleichung (5.5), da der Grenznutzen des Einkommens im Krankheitsfall die nachgefragte Menge beeinflusst. Konkret bedeutet dies, daß bei unvollständiger Versicherung und damit einem positiven Selbstbeteiligungssatz ($c > 0$) durch einen höheren Grenznutzen des Einkommens bei Krankheit λ weniger medizinische Leistungen nachgefragt werden.⁶⁹

Für die Verhandlungslösung ergibt sich nun in einer analogen Vorgehensweise zu Gleichung (5.9):

$$X(c, s) = a/b - (\lambda c + s/\alpha)/2b. \tag{5.13}$$

Die neue Effizienzbedingung weicht lediglich durch den Grad der Risikoaversion λ von Gleichung (5.9) ab. In diesem Fall ergibt sich für das soziale Optimum, daß die Selbstbeteiligung gleich Null ist und der Grenzertrag der medizinischen Leistung gleich Eins. Für einen risikoaversen Patienten ergibt sich als sozial optimale Menge $X_{SO}^* = (a - 1)/b$. Liegt hingegen Risikofreudigkeit vor, so wird das soziale Optimum ohne Versicherung erreicht.

Unter der Annahme gleicher Verhandlungsmacht folgt bei risikoaversen Patienten, daß das First-best-Zahlungssystem durch Vollversicherung für den Patienten ($c^* = 0$) und durch ein gemischtes Vergütungssystem für den Arzt ($R > 0$ und $s = 2\alpha \leq 1$) gekennzeichnet ist. Dies bedeutet, daß der Arzt nicht als perfekter Agent des Patienten agiert, so daß eine niedrige durch den Arzt präferierte

⁶⁸ Es gilt in diesem Fall $(1-\delta)\eta + \delta\lambda \equiv 1$ (vgl. Ellis und McGuire 1990, S. 385).

⁶⁹ Ein ansteigender Grenznutzen des Einkommens im Krankheitsfall geht mit einem Einkommensrückgang einher, woraus sich die geringere Nachfrage nach Gesundheitsleistungen ableiten läßt.

Menge an medizinischen Leistungen eine hohe durch den Patienten präferierte Menge kompensiert. Falls die First-best-Lösung nicht erreicht werden kann, ergibt sich ein rein prospektives Vergütungssystem mit $R > 0$ und $s = 1$. Ein wichtiges Ergebnis ist dabei, daß Systeme mit einer nur auf den Kosten basierenden Vergütung und teilweiser Absicherung des Patienten ($s = 0$ und $0 < c < 1$) nie zu einem optimalen Ergebnis führen.

Erweiterungen

Die in der Versicherungsliteratur wichtigen Moral Hazard-Effekte können ebenfalls innerhalb des dargestellten Modellrahmens analysiert werden. Das Hauptergebnis der Untersuchungen zu diesem Bereich ist, daß mit Zunahme des Moral Hazard-Effektes der Versicherungsumfang reduziert werden sollte. In dem vorgestellten Modell ergibt sich, daß unter gewissen Bedingungen die Anreize auf der Angebotsseite hinreichend stark sind, damit Moral Hazard irrelevant für das Versicherungs- und Vergütungssystem ist. Falls die Anreize allerdings sehr schwach sind, stellt Moral Hazard bei Second-best-Lösungen durchaus ein Problem dar. Der Moral Hazard-Effekt wird durch den Parameter a , der Konstanten aus der Nachfragekurve, repräsentiert. Je näher a dem Wert Eins ist, desto stärker ist das Moral Hazard-Problem, da die Nachfrage dann für einen gegebenen Preis elastischer ist. Ellis und McGuire untersuchen den Zusammenhang zwischen Moral Hazard a und Agency α , gegeben symmetrische Verhandlungsmacht. Die Hauptergebnisse sind, daß, wenn der Arzt als unvollkommener Agent des Patienten handelt, das Zahlungssystem unabhängig von der Ausprägung des Moral Hazard ist. Weiterhin ist ein System ohne Versicherung selbst bei schwacher Risikoaversion und Moral Hazard optimal, falls der Arzt sich als ‚super agent‘ verhält, der zuviel Leistungen verordnet. Eine Zunahme des Umfangs der Agency bei konstantem Moral Hazard führt dazu, daß das Vergütungssystem von einem gemischten System ($R > 0$, $0 < s < 1$) übergeht zu einem rein prospektiven System ($R > 0$, $s = 1$).

Die bisherigen Ergebnisse wurden unter der Annahme gleicher Verhandlungsstärke abgeleitet. In der Realität hingegen dürfte eine andere Gewichtung vorliegen. Falls der Patient souverän über die Behandlungsmenge entscheiden kann, ergibt sich, daß nur bei Risikoneutralität oder –freudigkeit eine First-best-Lösung erreicht wird, allerdings ohne den Abschluß einer Versicherung.⁷⁰ Bei Moral Hazard und Risikoaversion ist jedoch nur eine Second-best-Lösung möglich; mit steigender Risikoaversion tendiert das System zur Vollversicherung. Entscheidet hingegen der Arzt alleine über die medizinische Leistung, so hat die

⁷⁰ In diesem Fall ist die Behandlungsmenge unabhängig von der Ausgestaltung des Vergütungssystems, insbesondere der Kostenbeteiligung.

Selbstbeteiligung c keinen Einfluß auf die optimale Menge. Sie kann so gesetzt werden, daß das optimale Versicherungsniveau erreicht wird, d.h. $c = 0$ bei Risikoaversion und $c = 1$ bei Risikoneutralität. Der Vergütungsparameter s wird dann derart gesetzt, daß immer das optimale Niveau an medizinischen Leistungen gewählt wird.⁷¹

Kritische Würdigung

Das präsentierte Modell von Ellis und McGuire setzt bei unterschiedlichen Vorstellung zwischen Arzt und Patient über das Niveau der medizinischen Leistung an. Der Ansatz zur Konfliktlösung zeigt allerdings insofern Schwächen, als nur für symmetrische Verhandlungsstärke analytisch verwertbare Ergebnisse herzuleiten sind. Die Darstellung der Risikoaversion des Patienten weicht von der üblichen Vorgehensweise mit einer konkaven Nutzenfunktion ab. Die gewählte Form ergibt auch bei Vollversicherung keinen konstanten Grenznutzen des Einkommens, da dieser vom Zustand der Welt abhängt. Diese Formulierung vernachlässigt jedoch die Auswirkung der Gesundheitserträge auf das Einkommen. Außer Acht gelassen werden weiterhin Probleme asymmetrischer Information zwischen Arzt und Patient und zwischen Versicherung, Arzt und Patient.

5.3 Die Rolle des Arztes als doppelter Agent

Ausgangspunkt der Untersuchung von Blomqvist (1991) ist die Zustandsabhängigkeit der Nachfrage nach Gesundheitsleistungen. Dies bedeutet, daß der Patient je nach Ausprägung seines Gesundheitszustands eine bestimmte Menge medizinischer Leistungen nachfragt. Je schlechter der Gesundheitszustand, desto mehr Leistungen werden nachgefragt. Werden die Kosten für die medizinische Behandlung zum Großteil durch (private oder soziale) Versicherungen gedeckt, ergibt sich dann zusammen mit dem Problem der asymmetrischen Information zwischen den Anbietern von Gesundheitsleistungen und den Patienten, daß die Organisation eines Systems zur Bereitstellung von Krankenversicherung und medizinischer Behandlung zentral von der Ausgestaltung der vertraglichen Beziehungen abhängt (vgl. Blomqvist 1991, S. 413f.).

Den Ärzten als Erbringer der medizinischen Leistung kommt in einem solchen System eine zentrale Rolle zu. Einerseits besitzen die Patienten ein Interesse daran, daß der Arzt als (perfekter) Agent in ihrem Sinne arbeitet. Dies betrifft sowohl die Bereitstellung von Informationen als auch die eigentliche Behand-

⁷¹ Falls der Patient risikoneutral ist, so erreicht ein System mit $s^* = \alpha$ das soziale Optimum, für $\alpha \leq 1$. Für $\alpha \geq 1$ und risikoaverse Patienten ergibt sich $s^* = 1$ und $c^* = 0$.

lungsleistung. Andererseits ist es im Sinne der Krankenversicherer, wenn die Ärzte effizient in der Verwendung der Leistungen handeln, da dadurch die zu erwartenden Ansprüche an die Versicherungen sinken. In einem effizient organisierten System handeln daher die Ärzte als „double agents“ (Blomqvist 1991, S. 412).

Aus Sicht der Patienten sollen die Ärzte durch die Bereitstellung von Informationen und Leistungen wie in einem konventionellen Principal-Agent Problem in ihrem Interesse handeln. Falls aber ein Teil oder die gesamten anfallenden Kosten von dritter Seite (z.B. einer Versicherung) übernommen werden, verlangt die Effizienz eines Gesundheitssystems, daß die Ärzte auch die gesamten Kosten der Gesundheitsleistungen in Betracht ziehen, d.h. die Interessen der Versicherer berücksichtigen. Die Situation der double agency kann daher als Kombination aus Informationsasymmetrien und der Finanzierung der Gesundheitsleistungen von dritter Seite aufgefaßt werden.

Daher werden die Eigenschaften alternativer vertraglicher Vereinbarungen unter asymmetrischer Information sowohl auf den Märkten für Krankenversicherungen als auch auf dem Markt für ärztliche Leistungen untersucht. Dem verwendeten Modell liegt folgende Situation zugrunde: Der Arzt erhält durch die Untersuchung des Patienten private Informationen über den Gesundheitszustand, die weder dem Patienten selbst noch der Versicherung bekannt sind. Diese Situation wird für zwei unterschiedliche Versicherungssysteme betrachtet. Zum einen wird eine konventionelle Versicherung mit einer Einzelleistungsvergütung (fee-for-service) untersucht, zum anderen ist das System einer Health Maintenance Organization (HMO) Gegenstand der Analyse.⁷²

Vollständige Information

Im Fall der vollständigen Information maximiert ein repräsentativer Konsument seinen Erwartungsnutzen E über die Versicherungsprämie und die medizinische Leistung. Die Nutzenfunktion in Zustand i ist additiv-separabel im Nutzen aus dem Konsum $u(c_i)$ und dem Gesundheitszustand h_i . Letzterer hängt negativ vom jeweiligen Umweltzustand θ_i ab, der als Signal für die Schwere der Erkrankung dient, und positiv von den konsumierten medizinischen Leistungen z_i . Dabei bezeichnet ein höherer Index i eine schwerere Krankheit und π_i die Wahrscheinlichkeit eines Zustandes:⁷³

⁷² Da letztere nicht Bestandteil der vorliegenden Arbeit ist, beschränkt sich die Darstellung auf ein konventionelles Versicherungssystem.

⁷³ Bezüglich der Eigenschaften der Funktionen wird Konkavität von $u(\cdot)$ und $h(\cdot)$ unterstellt.

$$\begin{aligned}
 E &= \sum \pi_i \left(h_i + u \left(c_i \right) \right), \quad i = 0, 1, \dots, n \\
 h_0 &= \bar{h} \\
 h_i &= h \left(z_i - \theta_i \right) < \bar{h}.
 \end{aligned}
 \tag{5.14}$$

Falls der Patient gesund ist, ist sein Gesundheitszustand \bar{h} . Das Signal θ_i kann im Fall der vollständigen Information von allen Beteiligten kostenlos beobachtet werden, so daß es möglich ist, Versicherungsverträge aufzusetzen, die die Menge der medizinischen Leistung für einen bestimmten Umwelt- bzw. Krankheitszustand festlegen. Bezüglich des Versicherungsmarktes und des Marktes für Gesundheitsleistungen wird angenommen, daß beide Märkte kompetitiv sind, so daß die medizinischen Leistungen zum Preis von Eins angeboten werden und auf dem Versicherungsmarkt kein Load-Faktor erhoben wird.⁷⁴ Die resultierende Versicherungsprämie m ist dann gleich den erwarteten Ausgaben für Gesundheitsleistungen:

$$m = \sum \pi_i z_i ; \tag{5.15}$$

für den Konsum ergibt sich, daß die Menge an konsumierten Gütern c_i gleich dem verfügbaren Einkommen y abzüglich der Prämienzahlungen ist: $c_i = y - m$.

Im Gleichgewicht wählt der Konsument einen Vertrag, der die medizinische Leistung in Abhängigkeit von der Erkrankung festlegt ($z_i = z(\theta_i)$) und dem Versicherer einen Erwartungsgewinn von Null sichert. Weiterhin ist der Patient immer vollversichert, so daß sein Grenznutzen in allen Gesundheitszuständen gleich ist (vgl. Blomqvist 1991, S. 415).

Die Situation bei asymmetrischer Information

Durch die Diagnose der Krankheit erhält der Arzt kostenlose private Informationen θ über den Gesundheitszustand des Patienten. Dieser weiß zwar, ob er krank ist, kann aber nicht den genauen Wert von θ ermitteln. Letzteres gilt auch für den Versicherer, der ebenfalls über keine krankheitsbezogenen Informationen verfügt. Aus diesem Grund ist ein Versicherungsvertrag, der die medizinischen Leistungen in Abhängigkeit des Krankheitszustandes des Patienten festlegt, nicht durchsetzbar. Der Grund hierfür liegt in dem Interesse des Patienten, daß

⁷⁴ Dies bedeutet, daß bspw. keine Provision, Schadensabwicklungskosten oder sonstige Verwaltungskosten anfallen (vgl. hierzu Breyer und Zweifel 1999, S. 195).

der Arzt die Ernsthaftigkeit der Erkrankung gegenüber dem Versicherer überschätzt, so daß er möglichst umfassende medizinische Leistungen erhält. Bei Einzelleistungsvergütung führt der Wettbewerb unter den Leistungserbringern dann dazu, daß Ärzte nur im Sinne ihrer Patienten handeln und nicht als perfekter Agent für die Versicherung tätig sind. In dieser Situation besteht eine Möglichkeit der Vertragsgestaltung in einem konventionellen Versicherungsvertrag mit einer Prämie m , die es dem Patienten ermöglicht, die Menge z_i an medizinischer Leistung in jedem Zustand zu konsumieren, wobei die Versicherung dabei nur einen Anteil $1 - \sigma$ an den tatsächlich anfallenden Kosten trägt. Der Parameter σ ist folglich die Selbstbeteiligungsrate des Patienten.

Zur Vereinfachung wird aber zunächst unterstellt, daß der Arzt dem Patienten immer die Wahrheit über den Krankheitszustand θ mitteilt. Damit handeln die Ärzte in einer ihrer beiden Agency-Beziehungen als perfekte Agenten. Gegeben die Prämie und den Selbstbeteiligungssatz ergibt sich für den Konsum des Patienten $c_i = y - m - \sigma z_i$. In einem kompetitiven Versicherungsmarkt folgt für die Prämie:

$$m = (1 - \sigma) \cdot \sum \pi_i z_i (m, \sigma). \quad (5.16)$$

Diese entspricht den erwarteten Ausgaben für die Behandlung, abzüglich der Selbstbeteiligung des Patienten. Im Gleichgewicht wählt der Konsument dann einen Vertrag über m und σ , der, unter Berücksichtigung der optimalen Auswahl der medizinischen Leistung z_i seinen Erwartungsnutzen E maximiert. Weiterhin zeigt Blomqvist (vgl. 1991, S. 430f.), daß die Lösung einen positiven Selbstbeteiligungssatz mit $0 < \sigma < 1$ beinhaltet, so daß der Patient nicht vollständig versichert ist. Der Grund hierfür liegt darin, daß zum einen Verträge, die auf den Gesundheitszustand bezogen sind, nicht existieren. Zum anderen müssen bei konventioneller Versicherung die Nutzensteigerungen durch eine vollständigere Versicherung gegen den Moral Hazard-Effekt abgewogen werden.⁷⁵

Die Annahme, daß sich der Arzt als perfekter Agent des Patienten verhält, entspricht in der Regel jedoch nicht dem tatsächlichen Verhalten. In einem System der Einzelleistungsvergütung besteht für die Ärzte ein Anreiz, den Patienten nicht mit den korrekten Informationen zu versorgen und dadurch u.U. die Nachfrage nach ihren medizinischen Leistungen zu steigern. Dies rührt daher, daß sich die Ärzte lediglich in einem Wettbewerbsmarkt als Preisnehmer ver-

⁷⁵ Es bleibt anzumerken, daß die Nichtexistenz kontingenter Verträge eine Konsequenz der Arzt-Versicherungs-Beziehung ist. Falls ein Mechanismus gefunden werden kann, durch den die Ärzte die Versicherer korrekt informieren, wären auch kontingente Verträge möglich.

halten und dort keinen Anreiz besitzen, die Nachfrage zu manipulieren. Verfügen sie dagegen zu einem gewissen Grad über Monopolmacht, kann es zu einer Induzierung der Nachfrage kommen, wenn durch die falsche Darstellung der medizinischen Notwendigkeit keine Kosten, psychologisch oder physisch, entstehen. Nach Blomqvist liegt ein weiterer Grund, den Gesundheitszustand zu schlecht einzustufen und die Nachfrage zu beeinflussen, in möglichen Haftungsforderungen. Falls ein Arzt für Schäden haftbar gemacht wird, die dem Patienten deshalb entstehen, daß der Arzt ihm nicht von der Dringlichkeit einer medizinischen Maßnahme überzeugen konnte, ist die beste Strategie der Ärzte, dem Patienten immer den schlechtesten Zustand mitzuteilen.

Nehmen die Patienten weiterhin an, daß der Arzt ihnen die Wahrheit über ihren Gesundheitszustand mitteilt, werden sie immer als medizinische Leistung z_n bevorzugen, die Nachfrage beim schlechtesten Gesundheitszustand. Im Gleichgewicht wird der Vertrag (m, σ) gewählt, der den Erwartungsnutzen E maximiert. Für die Nullgewinnbedingung der Versicherung und damit für die Prämie gilt nun:

$$m = (1 - \sigma) (1 - \pi_0) z_n . \quad (5.17)$$

Die Prämie m entspricht den Ausgaben für die medizinische Leistung z_n , die mit der Erkrankungswahrscheinlichkeit und dem Anteil der Ausgaben, der von der Versicherung übernommen wird, multipliziert werden. Letzterer nimmt mit zunehmendem Selbstbeteiligungssatz des Patienten ab. In der beschriebenen Entscheidungssituation führt dies dazu, daß immer die umfangreichste medizinische Leistung z_n gewählt wird, sobald der Patient erkrankt. Es bleibt hier anzumerken, daß Blomqvist darauf verweist, daß bei konventioneller Versicherung der Patient die medizinische Leistung wählt und der Arzt dem Patienten lediglich die Informationen für diese Wahl mitteilt. In einem solchen Szenario hat der Arzt keinen direkten Einfluß auf die nachgefragte Menge, sondern kann nur durch die Weitergabe von Informationen bezüglich der Erkrankung auf die medizinische Leistung einwirken.

Ein anderer Fall liegt vor, wenn der Patient hingegen die Strategie des Arztes erkennt und seiner Mitteilung über den Krankheitszustand θ keinen Glauben schenkt. Angenommen, der Patient kennt die wahre Wahrscheinlichkeitsverteilung von θ , so steht er vor dem Problem, die folgende Erwartungsnutzenfunktion hinsichtlich der medizinischen Leistung z zu maximieren:

$$E = \pi_0 [\bar{h} + u(y - m)] + \sum \pi_i [h(z - \theta_i) + u(y - m - \sigma z)]. \quad (5.18)$$

Der erste Term in Gleichung (5.18) steht für die Wahrscheinlichkeit des Zustandes gesund, der zweite für die Wahrscheinlichkeit, krank zu sein. Aus der Maximierung folgt eine Nachfragefunktion nach medizinischen Leistungen $z = z(m, \sigma)$. Wenn die Annahmen der Versicherer über das Verhalten des Patienten richtig sind, d.h. wenn sie dessen Nachfragefunktion kennen, dann führt der Wettbewerb zu Verträgen über m, σ , die folgendes Optimierungsproblem erfüllen müssen:

$$\max_{m, \sigma} E \quad \text{u.d.N.} \quad m - (1 - \sigma) \left(1 - \pi_0\right) z = 0. \quad (5.19)$$

Wie in der Situation, in welcher der Arzt als perfekter Agent des Patienten handelt, gilt ein positiver Selbstbeteiligungssatz $0 < \sigma < 1$. Die Prämie berechnet sich auch hier als Anteil der Versicherung an den Ausgaben (z), multipliziert mit der Wahrscheinlichkeit zu erkranken.

Kritische Würdigung

Das Modell von Blomqvist stellt besonders auf die hervorgehobene Rolle des Arztes als doppelter Agent, sowohl gegenüber dem Patienten als auch gegenüber der Versicherung, ab. Die dabei unterstellten Verhaltensmuster des Arztes vernachlässigen allerdings seine Rolle gegenüber der Versicherung und die Auswirkungen auf die Vertragsgestaltung, da im Fall der asymmetrischen Information kein Wissen über den Gesundheitszustand des Patienten weitergegeben wird. Darüber hinaus ist es fraglich, ob der Patient tatsächlich die entscheidende Instanz bei der Wahl des Umfangs der medizinischen Leistung ist. Es ist eher anzunehmen, daß, falls der Patient wirklich die Leistungen festlegt, der Arzt zumindest eine Vorselektion getroffen hat. Selbst bei einer unterstellten Vorauswahl dürfte es dem Patienten an dem nötigen Wissen fehlen, die Diagnose des Arztes zu verifizieren und unterschiedliche Therapiemöglichkeiten zu beurteilen. Weiterhin wurden Aspekte der Arzthonorierung und der Entscheidung über Informationsweitergabe oder auch Therapiemöglichkeiten vernachlässigt, so daß das Verhalten des Arztes lediglich auf a priori Annahmen denn auf ökonomisch rationalem Verhalten beruht. Da die Entscheidungen des Arztes letzten Endes jedoch den Heilungserfolg und die Kosten der Leistungserstellung zumindest mit beeinflussen dürften, sollten diese in ein Modell aufgenommen werden.

5.4 *Optimale Krankenversicherungsverträge und die Entlohnung ärztlicher Leistungen*

Die ökonomische Theorie interpretiert nach Ma und McGuire (1997) die Entwicklung sozialer Institutionen, wie bspw. Gesundheitssysteme, als Antwort auf vorhandenes Marktversagen. Gesundheitsmärkte werden als „organized delivery systems“ (Ma und McGuire, 1997, S. 685) betrachtet. Im Gesundheitssektor fehlt jedoch ein wichtiger Markt, nämlich der, auf dem Versicherungsverträge auf dem Output basieren, d.h. daß die Auszahlung im Schadensfall von der Erkrankung und dem Gesundheitszustand abhängt (siehe hierzu auch Arrow 1963). Da es aber sehr kostspielig ist, den Gesundheitszustand zu verifizieren, basiert der Versicherungsschutz i.d.R. lediglich auf der gewählten Behandlungsmenge des Patienten. Nach Pauly (1968) und Zeckhauser (1970) existiert in der Krankenversicherung dabei ein trade-off zwischen Risk-Sharing und Moral Hazard. Die optimale Selbstbeteiligung des Patienten für eine Einheit der medizinischen Leistung setzt den Patienten ex ante einem Risiko aus, reduziert dabei aber zum Teil die ex post bestehenden Anreize, Gesundheitsleistungen übermäßig nachzufragen.

Nach Ma und McGuire sollte eine Theorie in der Interpretation komplexer vertraglicher Vereinbarungen über die reine Betrachtung von Risk-Sharing- und Moral Hazard-Problemen hinausgehen. Zum einen bestehen Interaktionen zwischen Versicherern, Ärzten und Patienten, die relevant sind für die Herleitung von optimalen Versicherungs- und Vergütungsverträgen. Zum anderen werden i.d.R. zwei weitere Probleme in der Vertragsgestaltung im Gesundheitswesen nicht erfaßt, da für sie keine Märkte existieren. Erstens kann die Behandlungsmenge vertraglich nicht festgelegt werden, auch nicht ex post. Die Verträge werden statt dessen auf Basis von Benachrichtigungen über die Behandlungsmenge geschlossen, da eine Verifizierung zu kostspielig wäre. Es existieren dabei zwei Anreizprobleme: Auf der Seite des Patienten, da er, für den Fall, daß er einen Teil der Kosten zu tragen hat, den Arzt bitten kann, eine zu geringe Menge zu berichten. Auf der Seite des Arztes, falls auch für ihn eine Kostenbeteiligung pro Menge medizinischer Leistung existiert.

Zweitens werden einige Bestandteile der Behandlung nie der Versicherung mitgeteilt, z.B. der Aufwand des Arztes. Die Idee hinter dieser Besonderheit ist folgende: Visiten, Krankenhaustage oder Kosten vermitteln lediglich einen Teil der Ressourcenaufwendungen für die Behandlung. Der Aufwand des Arztes ist folglich ein kostenloser Bestandteil der Behandlung und kann daher nicht vertraglich fixiert werden. Es besteht somit ein Moral Hazard-Problem auf der Seite des Arztes (Hidden Action). Die Arbeit von Ma und McGuire analysiert in die-

sem Zusammenhang die Interaktionen zwischen den beiden fehlenden Märkten für die Behandlungsmenge und die Entlohnung des ärztlichen Aufwands.

Grundmodell

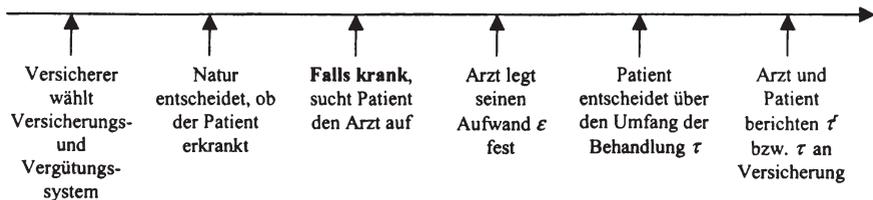
Das verwendete Modell beschreibt eine Agency-Beziehung zwischen Arzt und Patient. Es wird dabei auf den Behandlungsprozeß abgestellt, d.h., daß im Fall einer Krankheit der Schaden durch medizinische Behandlung gemildert bzw. beseitigt werden kann. Der Suchprozeß des Patienten und die Entscheidung über eine Arztkonsultation werden nicht betrachtet. Die Gesundheitsproduktion wird dabei als Prozeß mit den zwei Inputfaktoren Umfang der Behandlung τ und Aufwand des Arztes ε beschrieben. Unter dem ersten Faktor werden Gesundheitsleistungen im herkömmlichen Sinn zusammengefaßt, so z.B. die Anzahl der Arztbesuche oder die Zahl der Krankenhaustage. Deren Umfang kann gemessen und ex post verifiziert werden. Der Faktor Aufwand umfaßt andere Maßnahmen des Arztes, durch die die Intensität oder Qualität der Behandlung gesteigert wird. Der Aufwand repräsentiert daher „any costly activity that affects the patient’s valuation of services he receives, including dimensions of convenience, comfort, as well as some narrowly defined ‘clinical’ quality of care“ (Ma und McGuire 1997, S. 688). Im einfachsten Fall umfaßt dieser Faktor den Zeitaufwand pro Visite.

Bezüglich des Honorars werden folgende Annahmen getroffen: Die Ärzte werden anhand des Umfangs der Behandlung bezahlt, d.h. nicht nach ihrem Zeitaufwand. Die Entscheidung über den Aufwand steht jedoch im Zusammenhang mit dem Vergütungssystem. Die Verbesserung des Gesundheitszustands kann nicht vertraglich festgeschrieben werden, und der Umfang der Behandlung läßt sich ex ante nicht bestimmen, so daß nur der berichtete Umfang Gegenstand des Vertrages wird.

Die extensive Form des Modells besteht aus fünf Spielstufen (vgl. Abbildung 5.1). Auf der ersten Stufe wählt der Versicherer die Elemente des Versicherungs- und Vergütungssystems. Auf Stufe zwei entscheidet die Natur mit Wahrscheinlichkeit p , ob der Patient erkrankt. Falls er gesund ist, endet das Spiel, ansonsten sucht der Patient einen Arzt auf. Letzterer wählt im nächsten Schritt seinen Aufwand ε . Auf der vierten Spielstufe entscheidet der Patient über den Umfang der Behandlung τ , nachdem er den Aufwand des Arztes beobachtet hat. Auf der letzten Stufe geht es darum, welcher Umfang der Behandlung dem Versicherer mitgeteilt wird. Der Arzt und der Patient spielen ein ‚Berichtsspiel‘ (reporting subgame). Dabei schlägt der Arzt dem Patienten einen Umfang der Behandlung τ' vor, der nicht identisch mit τ sein muß. Falls der Patient

einverstanden ist, wird τ' berichtet, ansonsten τ . Dabei ist es für beide möglich, den tatsächlichen Umfang der Behandlung, der dem Versicherer mitgeteilt wird, zu bestimmen, und zwar für den Arzt dadurch, daß er den korrekten Umfang vorschlägt und für den Patienten, indem er einen abweichenden Vorschlag ablehnt. Die Idee dahinter ist, daß ein abweichender Bericht nur dann möglich ist, wenn er sowohl im Interesse des Arztes als auch im Interesse des Patienten ist.⁷⁶

Abbildung 5.1: Spielstufen des Modells von Ma und McGuire



Der Erwartungsnutzen des Patienten hängt von seiner Gesundheit, dem Nutzen aus den konsumierten Gesundheitsleistungen, dem verfügbaren Einkommen und seinem Kostenanteil an der Behandlung ab. Seine Anfangsausstattung mit Einkommen sei mit w bezeichnet. Er erkrankt mit Wahrscheinlichkeit p und erleidet einen negativen Gesundheitsschock s . Die Gesundheitsproduktion ist durch eine Funktion $F(\tau, \epsilon)$ gegeben, abhängig vom Umfang der Behandlung und dem Aufwand des Arztes.⁷⁷ Die Zuzahlung des Patienten pro Behandlung beträgt β , mit $0 \leq \beta \leq c$, wobei c die konstanten Grenzkosten der Behandlung darstellen. Weiterhin zahlt dieser dem Versicherer eine Prämie in Höhe von α . Der Patient ist risikoavers. Für seinen Erwartungsnutzen ergibt sich dementsprechend:

$$EU = p U(w - \alpha - \beta \tau' - s + F(\tau, \epsilon)) + (1 - p) U(w - \alpha). \tag{5.20}$$

Die Nutzenfunktion des Arztes ist separabel in Einkommen und Aufwand. Er ist risikoneutral bezüglich seines Einkommens. Sein Reservationsnutzen ist auf Null normiert. Die Behandlungskosten pro Einheit sind c und werden vom Arzt

⁷⁶ Damit die tatsächliche Behandlungsmenge der Versicherung mitgeteilt wird (truthful reporting), muß gelten, daß die Vergütung des Umfangs der Behandlung positiv ist, d.h., daß keine Kostenbeteiligung auf Anbieterseite vorliegt.

⁷⁷ Die Funktion F ist zunehmend in τ und ϵ , konkav und auf das Intervall $(0, s)$ beschränkt, d.h. durch die Gesundheitsproduktion kann maximal der Gesundheitsschock s kompensiert werden.

übernommen. Die Vergütung besteht aus zwei Komponenten, einer Pauschale ρ , die als prospektive Vergütung interpretiert werden kann, und einer Vergütung für den berichteten Umfang der Behandlung in Höhe von $\delta + c$ pro Einheit. Der Vergütungsparameter δ gibt dabei den Auf- oder Abschlag auf die Grenzkosten an. Die Kosten des Aufwands sind entsprechend der konvexen Funktion $G(\varepsilon)$:

$$V = p \left[\rho + (\delta + c) \tau - c \tau - G(\varepsilon) \right]. \quad (5.21)$$

Die Versicherung maximiert dann den Erwartungsnutzen des Patienten unter der Nebenbedingung eines ausgeglichenen Budgets:

$$\alpha = p \left[\rho (\delta + c - \beta) \tau \right]. \quad (5.22)$$

Wahl von Aufwand und Behandlungsmenge

Die Aktionen von Arzt und Patient werden unter der Annahme festgelegt, daß der Versicherer auf der ersten Stufe die Parameter α , β , ρ und γ bestimmt. Auf Stufe zwei ermittelt die Natur den Gesundheitszustand des Patienten, und auf Stufe drei wählt der Arzt seinen Aufwand. Auf Stufe vier entscheidet der Patient nun über den Umfang der Behandlung τ , der seinen Erwartungsnutzen maximiert. Als Bedingung erster Ordnung ergibt sich durch Ableiten von Gleichung (5.20) nach τ :

$$\beta = F_{\tau}(\tau, \varepsilon). \quad (5.23)$$

Diese Gleichung besagt, daß der Umfang der Behandlung so gewählt wird, daß die Selbstbeteiligung pro Einheit der medizinischen Behandlung gleich dem Grenznutzen aus der Behandlung ist. Die Versicherungsprämie α beeinflusst den Umfang der medizinischen Leistung nicht. Aus der Bedingung (5.23) kann die Reaktionsfunktion des Patienten auf die Aufwandsentscheidung des Arztes hergeleitet werden:

$$\frac{d\tau}{d\varepsilon} = \frac{F_{\tau\varepsilon}}{F_{\tau\tau}}. \quad (5.24)$$

Das Vorzeichen und damit die Steigung der Reaktionsfunktion hängt von der Kreuzableitung der Gesundheitsproduktionsfunktion ab. Falls der Aufwand des Arztes und der Umfang der Behandlung Komplemente sind, wählt der Patient

bei einer Erhöhung des Aufwands des Arztes einen größeren Umfang der Behandlung. Falls Substitute vorliegen, wird der Patient einen geringeren Umfang wählen.

Entscheidung des Arztes

Der Arzt antizipiert bei seiner Entscheidung über den Aufwand die Wahl des Umfangs der Behandlung durch den Patienten. Dies bedeutet, daß er die Reaktion des Patienten kennt und damit die Reaktionsfunktion in sein Entscheidungskalkül mit einbezieht. Das resultierende Stackelberg-Gleichgewicht auf der dritten Stufe $\tau(\epsilon)$ kann durch folgenden Optimierungsansatz dargestellt werden:

$$\begin{aligned} \max_{\epsilon, \tau} \quad & \rho + \delta \tau - G(\epsilon) \\ \text{u.d.N.} \quad & \beta = F_{\tau}(\tau, \epsilon). \end{aligned} \tag{5.25}$$

Die Wahl des Arztes hängt auch davon ab, ob Aufwand und Umfang der Behandlung in der Gesundheitsproduktion Substitute oder Komplemente sind. Falls der Arzt einen Aufschlag für jede Einheit medizinischer Leistung erhält ($\delta > 0$), steigt sein Nutzen nur durch einen höheren Umfang der Behandlung an, da er für den erbrachten Aufwand nicht entlohnt wird. Wenn weiterhin Substitute vorliegen, dann kann es nur zu einer Erhöhung des Behandlungsumfangs kommen, wenn der Arzt seinen Aufwand reduziert und somit der Patient einen höheren Behandlungsumfang wählt. Im Fall der Substitute muß der Arzt, damit er ein positives Aufwandsniveau wählt, einen Teil der Behandlungskosten tragen. Dagegen darf, falls ϵ und τ Komplemente sind, keine Kostenbeteiligung auf der Anbieterseite vorliegen ($\delta > 0$), damit der Arzt einen positiven Aufwand wählt.

Optimales Versicherungs- und Vergütungssystem

Auf der ersten Stufe des Modells werden die Versicherungs- bzw. Vergütungsparameter festgelegt. Dies geschieht durch Maximierung des Erwartungsnutzens des Patienten unter folgenden Nebenbedingungen: Die Versicherungsprämie ist aktuarisch fair, der Erwartungsnutzen des Arztes entspricht seinem Reservationsnutzen und die Wahl des Aufwands ϵ und des Umfangs der Behandlung τ bildet ein Gleichgewicht auf der dritten Stufe. Aufgrund des Versicherungssystems ergibt sich, daß der Arzt lediglich seinen Reservationsnutzen erhält. Weiterhin gilt, daß die Pauschale nicht Gegenstand der Optimie-

rung ist, da sie immer so gesetzt werden kann, daß sie gleich den Kosten des Aufwands abzüglich der Vergütung für die medizinische Behandlung ist.

Das Referenzszenario ergibt sich aus der Annahme, daß der Aufwand des Arztes in einem Vertrag festgeschrieben werden kann und dadurch auch die Vergütung am Aufwand orientiert wird. Der Arzt berichtet in diesem Fall den korrekten Umfang der Behandlung, da seine Wohlfahrt unabhängig von der Wahl des Patienten ist. Dieser Fall ist analog zu der in Zeckhauser (1970) analysierten Situation. Dort unternimmt der Arzt allerdings keine Handlungen, so daß der Erstattungsparameter δ nicht berücksichtigt werden muß. Dieses Referenzszenario bezeichnet die Second-best-Lösung, da es eine gelockerte Version des von Ma und McGuire behandelten Problems bei unbeobachtbarem Aufwand des Arztes für die Versicherung darstellt, was als Third-best-Gleichgewicht bezeichnet werden kann. Bei diesem ist es nicht möglich, den Umfang der Behandlung und den Aufwand des Arztes ex ante vertraglich festzulegen. Weiterhin muß der Erstattungsparameter δ mindestens gleich den negativen Grenzkosten ($-c$) sein, damit der korrekte Umfang der Behandlung mitgeteilt wird.

Generell gilt, daß die Second-best-Lösung auch im Third-best-Fall erreichbar ist, wenn die Second-best-Niveaus von Aufwand und Umfang der Behandlung auch ein Gleichgewicht im Third-best-Fall bilden. Dies bedeutet, daß der Erstattungsparameter δ so gewählt werden kann, daß der Arzt die gleichen Anreize bezüglich seines Aufwands wie im Second-best-Fall besitzt. Mit anderen Worten kann durch die geeignete Wahl von δ ein Second-best-Gleichgewicht erreicht werden. Falls die Restriktion für ‚truthful reporting‘ greift, d.h. ($\delta \geq -c$), kann ein solches Gleichgewicht jedoch nicht erreicht werden, da in diesem Fall der Erwartungsnutzen des Patienten strikt inferior zur Second-best-Lösung ist. Zusammenfassend läßt sich festhalten, daß, selbst wenn eine gewünschte Input-Kombination implementiert werden könnte, die Anforderungen an eine korrekte Mitteilung eine Umsetzung verhindern.

Ethisches Verhalten des Arztes

Eine Vielzahl von Modellen unterstellt für den Arzt eine Berufsethik oder altruistisches Verhalten. Die Modellierung hierfür ist jedoch nicht einheitlich. Ma und McGuire nehmen für die Berufsethik des Arztes an, daß der Nutzen aus der Behandlung mindestens ein bestimmtes Niveau erreichen muß. Dies wird durch eine zusätzliche Bedingung eingeführt:

$$F(\tau, \varepsilon) \geq \bar{F}. \quad (5.26)$$

Unter dieser Bedingung wählt der Arzt auf Stufe drei nun ein Aufwandsniveau, das zusammen mit der Wahl des Patienten mindestens den Nutzen \bar{F} erbringt. Dabei ergibt sich für den Fall der Substitute, daß, wenn die Iso-Nutzenlinie, die sich aus (5.26) ergibt, steiler als die Reaktionsfunktion des Patienten verläuft, der Erwartungsnutzen des Patienten mit einer Zunahme der Selbstbeteiligung ansteigt. Durch die höhere Selbstbeteiligung wird der Patient einen geringeren Umfang der Behandlung wählen. Damit der Nutzen aus der Behandlung größer als \bar{F} ist, muß der Arzt seinen Aufwand erhöhen. Dieser wird dem Arzt allerdings nicht vergütet. Dies bedeutet nun, daß in Situationen, in denen die Second-best-Lösung nicht erreichbar ist, durch eine Berücksichtigung der Berufsethik der Erwartungsnutzen des Patienten verbessert werden kann. Im Vergleich zum Third-best-Gleichgewicht mit $\delta = -c$ kann der Versicherer die Berufsethik dadurch fördern, daß er die Selbstbeteiligung anhebt.

Kritische Würdigung

Die Arbeit von Ma und McGuire untersucht Versicherungs- und Vergütungssysteme unter Berücksichtigung der Interaktionen zwischen Arzt und Patient. Weiterhin wird unterstellt, daß Abrechnungen der Versicherung nur auf dem berichteten Umfang der Behandlung basieren und daß die Restriktion, die sich aus der Bedingung für korrekte Mitteilungen ergibt, die Vertragsparameter beeinflusst. Die Interaktionen zwischen Arzt und Patient werden mittels einer Nachfrage-Reaktion modelliert. Dabei wählt der Patient den Umfang der Behandlung, nachdem er den Aufwand des Arztes beobachtet hat. Dieses Vorgehen ist kritisch hinsichtlich zweier Punkte: Erstens ist es nicht notwendiger Weise der Patient, der den Umfang der Behandlung bestimmt. Diese Entscheidung wird in der Regel auch durch den Arzt (mit) getroffen. Zweitens stellt die Kenntnis des ärztlichen Aufwands eine problematische Annahme dar. Zwar kann der Patient den Zeitaufwand beurteilen, jedoch fehlt ihm das nötige Wissen, um die Qualität und die Sorgfalt beurteilen zu können. Weiterhin abstrahiert die Modellierung der Gesundheitsproduktion als ein deterministischer Prozeß von zufälligen, nicht beeinflussbaren Faktoren wie Umweltzuständen oder Selbstheilungskräften, die neben den Inputfaktoren auch das Ergebnis des Gesundheitsprozesses beeinflussen.

5.5 Die Vergütung ärztlicher Leistungen bei unbeobachtbarem Aufwand des Arztes und des Patienten

Der Ansatzpunkt für die Untersuchung von Leonard und Zivin (2001) ist die Suche nach einem Mechanismus, der die Ärzte dazu bewegt, die notwendigen medizinischen Leistungen zu erbringen. In diesem Zusammenhang stellt sich die Frage nach der Art der Entlohnung der ärztlichen Leistungen. Leonard und Zivin

analysieren diese Fragestellung anhand eines Modells, das den Gesundheitsprozeß als „joint production“ (Leonard und Zivin 2001, S. 2) abbildet. Dies bedeutet, daß neben dem diagnostischen Aufwand des Arztes auch die Handlungen des Patienten das Ergebnis des Leistungserstellungsprozesses beeinflussen (siehe hierzu auch Abschnitt 6). In diesem Fall liegt in der Principal-Agent-Beziehung zwischen Arzt und Patient ein Double Moral Hazard-Problem vor. Für die Analyse werden zwei Arten von Honorierungssystemen betrachtet: Zum einen eine aufwandsbezogene Vergütung (effort-contingent compensation), wie sie in Form einer Einzelleistungsvergütung in zahlreichen Gesundheitssystemen vorliegt, zum anderen eine ergebnisorientierte Vergütung (outcome-contingent compensation).

Die Informationsstruktur des Modells spielt dabei wie in allen informationsökonomischen Modellen eine zentrale Rolle. Es wird unterstellt, daß ein Patient nicht die Aktionen des Arztes und umgekehrt der Arzt nicht die Handlungen des Patienten beobachten kann. Die Patienten können die diagnostische Leistung des Arztes, obwohl anwesend, nicht überprüfen, da der Arzt bei der Behandlung einen Wissensvorsprung besitzt. Auf der anderen Seite sind Ärzte bei der Ausübung des Aufwands der Patienten in aller Regel nicht anwesend, so daß sie dessen Einfluß auf die Gesundheit nicht abschätzen können. Als zusätzlicher Akteur wird ein sozialer Planer eingeführt, der den Aufwand des Arztes und das Ergebnis der Behandlung beobachten kann, nicht aber den Aufwand des Patienten.⁷⁸ In diesem Modellrahmen werden aufwandsbezogene Verträge als System angesehen, in dem eine dritte Partei, die sowohl das realisierte als auch das angemessene Niveau des medizinischen Aufwands ermitteln kann, das Honorierungssystem bestimmt. Diese Partei ist bspw. der Arbeitgeber des Arztes, eine Versicherung oder auch ein Regulierer.

Ein Problem bei einer ergebnisorientierten Vergütung ist, daß Ärzten nicht die notwendigen Anreize geben werden können, den benötigten Aufwand zu erbringen. Sie würden ungerechtfertigterweise bestraft, wenn der Patient nach einer Behandlung stirbt, ohne daß sie dafür verantwortlich sind. Hingegen vernachlässigen aufwandsbezogene Honorierungssysteme die Rolle des Patienten für die eigene Gesundheit. Falls medizinischer Aufwand und der des Patienten Komplemente sind, so hängt der Grenznutzen der ärztlichen Handlungen auch von den Aktionen des Patienten ab. Letztere beeinflussen somit den Gesundheitszustand nicht nur direkt, sondern auch indirekt über die medizinischen Leistungen.

⁷⁸ Bezüglich des sozialen Planers wird angenommen, daß dieser medizinische Kenntnisse besitzt, Budgets überschreiten kann und die Interessen des Patienten und des Arztes verfolgt (vgl. Leonard und Zivin 2001, S. 3).

Welchem der beiden Vergütungssysteme der Vorzug zu geben ist, hängt von den Charakteristika der zu behandelnden Krankheit ab. Für den Fall, daß zwischen ärztlichem und Patientenaufwand Komplementarität besteht, stellt eine ergebnisorientierte Vergütung das bessere System dar. Im anderen Fall, daß der Aufwand kaum komplementär ist, ist eine aufwandsbezogene Honorierung vorzuziehen. Der zweite Fall ist dann relevant, wenn bspw. bei einer Operation der Erfolg nicht oder nur unwesentlich vom Verhalten des Patienten abhängt. Dagegen kann im Zusammenhang z.B. mit Rückenbeschwerden, die auch vom Aufwand des Patienten beeinflußt werden, von hoher Komplementarität ausgegangen werden.

Theoretische Modellierung

Der erwartete Nettowert der Nachfrage nach medizinischer Behandlung gibt die Veränderung des Erwartungsnutzens des Patienten wieder. Er ist eine Funktion der Opportunitätskosten der gesund verbrachten Zeit ω und des erwarteten Anstiegs der Gesundheit h , abzüglich der Kosten der Behandlung C und des negativen Nutzens, der sich beim Patienten durch den Aufwand der Gesundheitsproduktion $c(p)$ einstellt, wobei p den Aufwand per se symbolisiert:⁷⁹

$$\Delta EU = \omega h - C - c(p). \quad (5.27)$$

Der erwartete Anstieg der Gesundheit h entspricht dabei dem Erwartungswert des Anstiegs des Gesundheitszustands H aufgrund der Inanspruchnahme von Gesundheitsleistungen. Bezüglich des Ausgangszustands wird angenommen, daß der Patient krank ist. Eine Verbesserung des Gesundheitszustands wird durch einen positiven Wert von H ausgedrückt. Dieser hängt positiv vom jeweiligen Aufwand des Arztes und des Patienten ab, wobei es nicht möglich ist, von der realisierten Verbesserung des Gesundheitszustands auf den Aufwand Rückschlüsse zu ziehen.⁸⁰ Der erwartete Anstieg der Gesundheit h kann mittels einer Produktionsfunktion dargestellt werden, die von den folgenden Faktoren abhängt: dem medizinischen Aufwand, dem Patientenaufwand, den medizinischen Fertigkeiten und der Fähigkeit des Patienten (Effizienz), die Gesundheitsinputs in Gesundheit zu transformieren. Ein Anstieg eines dieser Faktoren erhöht c.p.

⁷⁹ Nach Leonard und Zivin (vgl. 2001, S. 6) kann diese Spezifikation im Sinne eines Gesundheitskapitalstocks analog zu dem Modell von Grossman (1972) interpretiert werden. Dabei ist ω der Wert einer zusätzlichen gesund verbrachten Stunde, d.h. der Lohnsatz bzw. die Opportunitätskosten der Freizeit.

⁸⁰ Dies bedeutet, daß jedes Niveau von H mit jeder Kombination von Patienten- und Arztaufwand erreichbar ist, so daß auch der Patient aus der Kenntnis des eigenen Aufwands nicht auf den des Arztes schließen kann (vgl. Leonard und Zivin 2001, S. 6).

die Wahrscheinlichkeit der Genesung. Die Produktionsfunktion ist vom Cobb-Douglas-Typ und lautet:

$$h = \pi p^\alpha m^\beta. \quad (5.28)$$

Dabei stellt π einen Produktivitätsfaktor dar, der positiv von den Fähigkeiten des Arztes und des Patienten abhängt und nicht näher spezifiziert wird. Der Aufwand des Patienten ist durch p gegeben, der des Arztes durch m . Die Parameter α und β sind die Elastizitäten des Outputs, bezogen auf den jeweiligen Inputfaktor. Die Funktion besitzt abnehmende Skalenerträge in der Gesundheitsproduktion, und es wird angenommen, daß $0 < \alpha < 1$, $0 < \beta < 1$ und $0 < \alpha + \beta < 1$. Zur Vereinfachung wird das Produkt aus dem Produktivitätsfaktor π und den Opportunitätskosten der gesund verbrachten Zeit ω als A bezeichnet. Der negative Nutzen aus dem Aufwand ist eine lineare Funktion von p für den Patienten bzw. m für den Arzt. Der Koeffizient sei für den Patienten auf Eins normalisiert, für den Arzt auf D . Für den Nutzen des Patienten läßt sich daher schreiben:⁸¹

$$U = A p^\alpha m^\beta - p - C. \quad (5.29)$$

Vollständige Information

Als Referenz für die Analyse der Entscheidungen bei asymmetrischer Information dient die Nutzenmaximierung im Fall der vollständigen Information. Dabei können sowohl der Arzt als auch der Patient die Handlung des anderen beobachten, folglich entsteht kein Koordinationsproblem. Die soziale Wohlfahrt ist die Differenz aus dem Nutzen aus der Gesundheit und den Kosten des Aufwands:

$$W = A p^\alpha m^\beta - p - D m. \quad (5.30)$$

Diese Funktion wird nun nach dem Aufwand des Patienten p und dem Aufwand des Arztes m abgeleitet, die daraus resultierenden Niveaus sind mit p_{FI} und m_{FI} bezeichnet. Mit Hilfe dieser Aktionen können dann der Nutzen des Patienten U_{FI} und die soziale Wohlfahrt W_{FI} hergeleitet werden.

⁸¹ Anstelle von ΔEU in Gleichung (5.27) wird der erwartete Nettowert der Nachfrage nach medizinischer Behandlung im weiteren mit U bezeichnet.

*Gemeinsame Produktion mit unbeobachtbarem Aufwand*⁸²

Für den Fall, daß beide Akteure die Aktion des anderen nicht beobachten können, wird ein sozialer Planer eingeführt, der die Vergütungen implementiert und dessen Ziel die Wohlfahrtsmaximierung ist. Dabei wird unterstellt, daß der Planer zwar den medizinischen Aufwand des Arztes und das Ergebnis des Leistungserstellungsprozesses beobachten kann, daß er allerdings keine Informationen über den Aufwand des Patienten besitzt. Unter diesen Umständen legt der soziale Planer das Vergütungsschema für ein aufwandsbezogenes System der Entlohnung fest. Alternativ werden Arzt und Patient nach dem Ergebnis der Gesundheitsproduktion honoriert. Unter beiden Vertragsmöglichkeiten ist es ihm möglich, das soziale Optimum in der Situation vollständiger Information zu erreichen.

In einer Second-best-Welt besitzt der soziale Planer keine Informationen über die Interaktionen zwischen Arzt und Patient. Er kann diese somit nicht steuern und die Vergütungssysteme nicht dazu benutzen, die vollständigen Konsequenzen aus der Gesundheitsproduktion in Form von Nutzen und Kosten an Arzt oder Patient zu transferieren. Daher sind die Anreize für Arzt und Patient, den maximalen Aufwand zu erbringen, zu gering. Es existiert des weiteren kein Vergütungssystem, das ein anderes in allen Situationen vollständig dominiert. Die soziale Wohlfahrt wird für bestimmte Krankheiten durch ergebnisorientierte Verträge maximiert, für andere Krankheiten geschieht dies durch aufwandsbezogene Systeme.

Da der soziale Planer keine genaue Kenntnis über das Verhalten des Patienten besitzt, nimmt er bezüglich des Aufwands des Patienten an, daß dieser invariant gegenüber dem medizinischen Aufwand des Arztes ist. Es wird unterstellt, daß aufgrund der Tatsache, daß der Planer beide Akteure nicht dazu bringen kann, die vollständigen Konsequenzen aus ihren Handlungen zu tragen, die Vergütung für Arzt oder Patient nicht dem Wert des Ergebnisses der Gesundheitsproduktion entspricht.

Aufwandsbezogene vs. erfolgsorientierte Verträge

Im Fall der aufwandsbezogenen Verträge wählt der soziale Planer das Niveau des medizinischen Aufwands, ohne Rückschlüsse auf den Aufwand des Pa-

⁸² Im weiteren Verlauf wird auf die formale Darstellung der Optimalwerte bei vollständiger und asymmetrischer Information verzichtet und lediglich die ökonomische Intuition und die Zusammenhänge aufgezeigt. Für eine formale Darstellung siehe die Arbeit von Leonard und Zivin (2001).

tienten ziehen zu können. Falls beide Aktionen Komplemente darstellen, hängt das optimale medizinische Niveau auch vom Aufwand des Patienten ab, d.h., der Planer kann den medizinischen Aufwand nicht optimal bestimmen. Er unterstellt daher, daß das Aufwandsniveau des Patienten fix ist, also $p = p^?$. Der Planer maximiert somit den folgenden Ausdruck bezüglich des medizinischen Aufwands m :

$$A \bar{p}^\alpha m^\beta - \bar{p} - D m. \quad (5.31)$$

Der Patient reagiert nun seinerseits auf die Aktion des Arztes mit der Wahl des eigenen Aufwands. Die Reaktionsfunktion des Patienten ist durch die Bedingung erster Ordnung für den Nutzen des Patienten gegeben (partielle Ableitung von Gleichung (5.29) nach p). Diese ist identisch mit der Optimalitätsbedingung unter vollständiger Information. Die Optimalwerte für den Fall einer aufwandsbezogenen Vergütung sind durch p_E , m_E , U_E und W_E gegeben.

Für den Fall, daß der Aufwand des Patienten demjenigen bei vollständiger Information entspricht ($p^? = p_{FI}$), wird für den medizinischen Aufwand die Lösung bei vollständiger Information erreicht. Für $p^? < p_{FI}$ nimmt der medizinische Aufwand ab, im anderen Fall ($p^? > p_{FI}$) steigt er hingegen an, d.h. der Arzt arbeitet zu viel. Ein ähnliches Bild ergibt sich für die soziale Wohlfahrt. Falls der erwartete Aufwand des Patienten demjenigen bei vollständiger Information entspricht, so folgt $W_E = W_{FI}$. Für die anderen Fälle ist die Wohlfahrt strikt niedriger als bei vollständiger Information. Im Gegensatz dazu kann der Nutzen des Patienten größer als unter vollständiger Information sein, da der Patient den Arzt nicht für zu harte Arbeit entschädigen muß.

Bei ergebnisorientierten Verträgen erhalten Arzt und Patient eine Bezahlung als Funktion des Outputs. Dabei wird der Anteil des Patienten an der Vergütung mit s_p und der des Arztes mit s_m bezeichnet. Der Planer maximiert nun die Wohlfahrt, indem er die optimalen Werte dieser Anteile festlegt.⁸³ Der soziale Planer setzt die Anteile auf ihr maximal mögliches, anreizkompatibles Niveau, und der Patient und der Arzt maximieren ihren jeweiligen Nutzen. Für den Patienten ist dieser gegeben durch:

$$s_p A p^\alpha m^\beta - p \quad (5.32)$$

⁸³ Wenn Arzt und Patient über genügend Anreize verfügen, würden sie die vollen Anteile erhalten. Da diese Anreize jedoch nicht implementiert werden können, sind die Anteile kleiner (vgl. Leonard und Zivin 2001, S. 11). Dies bedeutet, daß nicht die vollen Konsequenzen aus den Handlungen gezogen werden.

und für den Arzt in Form von:

$$s_m A p^\alpha m^\beta - D m. \quad (5.33)$$

Das resultierende Nash-Gleichgewicht entspricht dem Punkt, an dem der gewählte Aufwand eines Akteurs den Erwartungen des anderen entspricht. Der Planer hat außer hinsichtlich der Festlegung der Anteile keine weiteren Funktionen. Die optimalen Werte bei ergebnisorientierter Entlohnung sind durch p_o , m_o , U_o und W_o gegeben. Wenn entweder der Arzt oder der Patient nicht den vollen Anteil erhält, so bedeutet dies, daß einer der beiden Akteure zu wenig Aufwand betreibt. Daher sind der Nutzen des Patienten und die soziale Wohlfahrt strikt inferior zu der Lösungen bei vollständiger Information.

Ein Vergleich der Lösungen bei aufwandsbezogenen und ergebnisorientierten Verträgen läßt sich in drei Folgerungen zusammenfassen:

Folgerung 1: In einer Second-best-Welt hängt das von den Patienten präferierte Vergütungssystem der Ärzte vom Krankheitszustand ab. Ergebnisorientierte Vergütungen sind besser als aufwandsbezogene für Krankheiten, bei denen die Grenzproduktivitäten von Arzt und Patient (α und β) hoch sind und folglich Komplemente vorliegen. Im Gegensatz dazu ist eine aufwandsbezogene Vergütung in dem Fall besser, in dem die Grenzproduktivität nur für einen Faktor hoch ist.

Diese Folgerung impliziert, daß die Wahl des ärztlichen Vergütungssystems von der Art der zu behandelnden Krankheit abhängt. Für den Fall, daß der Patient medizinische Leistungen in Anspruch nimmt und der Aufwand von Arzt und Patient für die behandelte Krankheit komplementär ist, sind die soziale Wohlfahrt und der Nutzen des Patienten bei ergebnisorientierter Vergütung höher. Auf der anderen Seite gilt für Krankheiten, in denen die Aufwendungen keine Komplemente darstellen, daß die aufwandsbezogene Vergütung dominiert.

Folgerung 2: Falls der Nutzen des Patienten bei ergebnisorientierter Vergütung gleich dem bei aufwandsbezogener Vergütung ist, ist die Wohlfahrt im Fall der ergebnisorientierten Vergütung höher.

Folgerung 3: Falls der Nutzen des Patienten bei ergebnisorientierter Vergütung größer als bei aufwandsbezogener Vergütung ist, so dominiert die Wohlfahrt bei ergebnisorientierter Vergütung immer diejenige bei aufwandsbezogenen Systemen.

Die letzten beiden Folgerungen beschreiben den Zusammenhang zwischen dem Nutzen des Patienten und der sozialen Wohlfahrt und deren Abhängigkeit vom vorliegenden Vergütungssystem. Die Wohlfahrt bei ergebnisorientierter Vergütung ist demnach mindestens so hoch wie bei aufwandsbezogener Vergütung. Für den Fall, daß der Nutzen des Patienten bei ergebnisorientierter Vergütung größer ist, ist auch die Wohlfahrt in diesem Vergütungssystem höher.

Kritische Würdigung

Das vorgestellte Modell von Leonard und Zivin beschreibt die Wahl von Vergütungssystemen, wenn die Gesundheitsproduktion nicht nur durch den ärztlichen Aufwand beeinflusst wird, sondern auch die Handlungen des Patienten nehmen Einfluß. Während die aufwandsbezogene Vergütung in der Praxis als Einzelleistungsvergütung (fee-for-service) angewendet wird, finden sich für die erfolgsorientierte Vergütung kaum Anwendungsbeispiele. Die Annahme, daß der soziale Planer den medizinischen Aufwand und das Ergebnis des Behandlungsprozesses beobachten kann, erscheint problematisch. Da die Verbesserung des Gesundheitszustands durch die Behandlung einem Patienten direkt zugeordnet werden kann, handelt es sich hierbei um individuelles Wissen des Patienten, so daß nicht eindeutig ist, inwiefern der soziale Planer über Informationen darüber verfügen sollte. Weiterhin wird die Gesundheitsproduktion als deterministischer Prozeß modelliert, wobei auch eher davon auszugehen ist, daß auch Faktoren den Prozeß beeinflussen, die weder durch den Arzt noch durch den Patienten gesteuert werden kann.⁸⁴ Die Interaktionen und damit der Zusammenhang zwischen ärztlichem Aufwand und dem des Patienten werden nur für den Fall der Komplemente analysiert, während die Fälle Unabhängigkeit und Substitute von der Analyse ausgeklammert werden. Inwiefern die abgeleiteten Ergebnisse auch für letztere Zusammenhänge zwischen den Aktionen gelten, geht aus den Ausführungen von Leonard und Zivin nicht hervor. Da nicht ersichtlich ist, inwiefern die Honorierung im Zusammenspiel mit anderen Faktoren, bspw. der Berufsethik, die Entscheidung des Arztes über den medizinischen Aufwand determiniert, ist die explizite Ausklammerung des ärztlichen Nutzens bei der aufwandsbezogenen Honorierung problematisch. Als letzter Kritikpunkt ist die Vernachlässigung einer Versicherung zu nennen, so daß die Auswirkungen der ärztlichen Vergütung auf den Patienten über die Versicherungsprämie oder eine Selbstbeteiligung unberücksichtigt bleiben.

⁸⁴ Die Annahme, keine funktionale Form für den Produktivitätsfaktor π einzuführen, bedingt an sich keine Stochastik für die Gesundheitsproduktion, solange dieser Faktor lediglich durch die Fähigkeiten des Arztes und des Patienten beeinflusst wird.

6 Double Moral Hazard zwischen Arzt und Patient

6.1 Zusammenhang zwischen ärztlicher Leistung und Compliance des Patienten

Ein zentraler Kritikpunkt an der Übertragung eines Standardmodells der Principal-Agent-Theorie stellt die Vernachlässigung der beidseitigen Informationsasymmetrie zwischen Arzt und Patient dar. Aus denen im vorangegangenen Abschnitt dargestellten Forschungsarbeiten geht zudem hervor, daß auch in bisherigen Ansätzen eine Umsetzung dieser Problematik nicht stattgefunden hat. Es erweist sich folglich als notwendig, den Zusammenhang zwischen den durch die Vertragspartner unbeobachteten Aktionen von Arzt und Patient zu verdeutlichen. Dabei werden die verschiedenen Formen des Kontaktes zwischen Arzt und Patient näher betrachtet sowie die Auswirkungen von simultanen und sequentiellen Aktionen untersucht. Die theoretische Modellierung der Situation des Double Moral Hazard erfolgt sowohl auf der Behandlungs- als auch auf der Vertragsstufe. Durch eine Trennung in beide Modellstufen wird erreicht, daß die Effekte zuerst nur im Fall einer Behandlung des Patienten durch den Arzt analysiert werden, bevor in einem nächsten Schritt die Auswirkungen der beidseitigen asymmetrischen Information für die Vertragsgestaltung untersucht und schließlich auch auf den Fall ausgeweitet werden, in dem keine Arztkonsultation vorliegt.

Wie in Abschnitt 4.3.1 dargestellt und aus den Ausführungen des vorangegangenen Kapitels ersichtlich, sind die Ausprägungen der Informationsasymmetrien im Gesundheitswesen häufig nicht nur einseitig, sondern wechselseitig, wie etwa zwischen Patient und Arzt. Der Patient kann i.d.R. weder die Wirkung der Leistungen des Arztes genau einordnen, noch besitzt der Arzt genauere Informationen über das behandlungsbegleitende Verhalten des Patienten.⁸⁵ Letzteres wird als Compliance bezeichnet, die die gesundheitsfördernden Anstrengungen des Patienten, die er zusätzlich zur medizinischen Behandlung durch den Arzt unternimmt, beschreibt (vgl. Wille und Ulrich 1991, S. 27). Besondere Bedeutung kommt in diesem Zusammenhang dem direkten Verhältnis zwischen den ärztlichen Leistungen und der Compliance des Patienten zu. Dieses Verhältnis kann eine substitutive Beziehung, aber auch komplementär sowie unabhängig voneinander sein. Ziel dieses Kapitels ist es, das Verhalten von Arzt und Patient bei wechselseitiger asymmetrischer Information zu

⁸⁵ Der Patient nimmt daher sowohl die Rolle des Konsumenten als auch die eines Produktionsfaktors ein, so daß sich das Behandlungsergebnis als „joint product“ aus medizinischer Leistung und individueller Nutzung ergibt (vgl. Wille und Ulrich 1991, S. 27).

analysieren und Auswirkungen von vertraglichen Regelungen auf die Handlungen der Akteure zu untersuchen.

Der Gesundheitszustand eines Patienten wird, wie bereits beschrieben, nicht nur durch die vom behandelnden Arzt verabreichte medizinische Leistung beeinflusst. Neben Umweltfaktoren, die als exogen angesehen werden können, trägt auch das eigene Verhalten des Patienten zur Veränderung des Gesundheitszustands bei. Dieses gesundheitsrelevante Verhalten umfaßt das Befolgen von Therapievorschriften, bestimmte Freizeitaktivitäten oder den individuellen Lebensstil des Patienten, wie Trink- und Rauchverhalten sowie Ernährungsweise.⁸⁶ Der Patient entscheidet selbständig über seine Handlungen und führt diese anschließend durch. Sie haben Auswirkungen auf seinen Nutzen durch ein gesteigertes Wohlbefinden, können gleichzeitig aber auch seiner Gesundheit abträglich sein (vgl. hierzu Cutler 1996, S. 6ff.). Darüber hinaus sind sie mit Kosten verbunden, sei es in direkter Form wie Konsumausgaben oder eher indirekter Art durch den verursachten Aufwand. Dies bedeutet, daß der Patient bei jeder seiner Handlungen Nutzen und Kosten abwägen muß, gemäß seines Informationsstands über die etwaigen Folgen.⁸⁷

Im Kontext einer Analyse der Beziehung zwischen Arzt und Patient erweist sich das angesprochene Zusammenspiel von ärztlicher (medizinischer) Leistung und der Compliance des Patienten als besonders relevant. Zwar werden sich medizinische Leistung und gesundheitsförderndes Verhalten jeweils positiv auf den Gesundheitszustand des Patienten auswirken, unklar bleibt allerdings der Zusammenhang beider Aktionen sowie deren Zusammenspiel. Auf der einen Seite werden medizinische Leistungen durch den Abschluß einer Krankenversicherung für den Patienten günstiger. Bei Vollversicherung bedeutet dies, daß der Patient die Leistungen des Arztes zu Grenzkosten von Null in Anspruch nehmen kann, so daß die Anreize, sich im Hinblick auf die eingesetzte medizinische Leistung und damit auf seinen Gesundheitszustand optimal zu verhalten, sei es bezogen auf die Prävention oder seine Compliance, nur gering sind. Betrachtet man lediglich die dem Patienten entstehenden direkten monetären Kosten, so wäre das Ergebnis eine substitutive Beziehung zwischen ärztlicher Leistung und Patientenverhalten, da die eigenen Anstrengungen mit Kosten verbunden sind und durch die kostenlosen medizinischen Leistungen ersetzt werden können. Umgekehrt geht die Erhöhung der Kosten für den Patienten durch eine Ein-

⁸⁶ Zu den Auswirkungen des persönlichen Lebensstils auf die Gesundheit siehe Marschall (1999).

⁸⁷ Generell müßte dabei die Zeitpräferenz des Patienten berücksichtigt werden, da Nutzen und Kosten der Handlungen nicht gleichzeitig auftreten. Zur Vereinfachung wird dies jedoch ausgeschlossen.

führung einer Selbstbeteiligung mit höheren Anstrengungen einher, da diese im Vergleich zur medizinischen Leistung relativ billiger werden.

Auf der anderen Seite stehen die Produktivitätsaspekte der individuellen Handlungen, die bei einer Beschränkung der Betrachtung auf entstehende Kosten vollständig vernachlässigt werden (siehe Abschnitt 4.2). Wenn sowohl ärztliche Leistung als auch die Compliance positiv auf den Gesundheitszustand wirken, dann ist für eine detaillierte Untersuchung relevant, wie diese Inputs in die Gesundheitsproduktion im Verhältnis zueinander stehen. Die Beziehung zwischen der Compliance des Patienten und der medizinischen Leistung des Arztes läßt sich durch die Begriffe strategische Komplemente, strategische Substitute bzw. strategische Unabhängigkeit kennzeichnen (vgl. Bulow, Geanakoplos und Klemperer (1985)). Im ersten Fall führt eine Erhöhung der Aktivität eines Akteurs zu einer Verringerung der Grenzproduktivität des anderen (strategische Komplemente). Im Gegensatz dazu sind im zweiten Fall Aktionen in ihrer Beziehung strategisch substitutiv, wenn eine Erhöhung der Aktivität eines Akteurs zu einer Erhöhung der Grenzproduktivität des anderen führt. Bei Unabhängigkeit hat die Erhöhung keinen Einfluß auf die Grenzproduktivität des anderen Akteurs. Der Zusammenhang zwischen beiden Begriffen wird in der späteren theoretischen Analyse ein zentrales Element sein (vgl. Abschnitt 6.5).

6.2 Kontakte zwischen Arzt und Patient

Prinzipiell kann eine Einteilung des Kontaktes nach zwei Kriterien vorgenommen werden. Zum einen stellt sich die Frage, ob dem Arztkontakt ein Suchprozeß vorausgeht, zum anderen, ob es sich um eine langfristige oder kurzfristige Beziehung handelt. Suchprozesse resultieren aus der Tatsache, daß der Patient nicht oder nur ungenügend über die Fähigkeiten der Ärzte informiert ist und zusätzlich noch keine vorherige Konsultation des Arztes vorliegt (Erstkontakt). Wiederholte Arztbesuche werden hingegen nicht von einem Suchprozeß begleitet, es sei denn, der Patient ist mit dem Erfolg der Behandlung nicht zufrieden und möchte noch einen weiteren Arzt konsultieren. Für den Fall, daß der Patient immer den gleichen Arzt aufsucht, spielen Suchprozesse keine Rolle.

Hinsichtlich der zeitlichen Dimension des Kontaktes ist festzustellen, daß ein langfristiger Kontakt an den Behandlungserfolg und darüber hinaus an ein Vertrauensverhältnis zwischen Arzt und Patient geknüpft ist. Ein Beispiel für letzteres ist die Beziehung zu einem Hausarzt, der für den Patienten i.d.R. den ersten Ansprechpartner darstellt und dessen medizinische Fähigkeiten er am ehesten einzuschätzen vermag. Ein kurzfristiger Kontakt hingegen hat seine Ursachen entweder im fehlenden Vertrauen zum behandelnden Arzt oder in der

Unzufriedenheit mit dem Ergebnis der Behandlung. In diesem Fall wird der Patient bei einer erneuten Erkrankung nicht mehr denselben Arzt aufsuchen, sondern es kommt zu einem erneuten Suchprozeß.

Die vorliegende Arbeit konzentriert sich auf kurzfristige Kontakte zwischen Arzt und Patient, so daß wiederholte Konsultationen und Lerneffekte nicht betrachtet werden. Auch Suchprozesse werden im weiteren von der Betrachtung ausgeschlossen, da sie für die untersuchte Beziehung von ärztlicher Leistung und Compliance keine Relevanz besitzen.

6.3 *Simultane versus sequentielle Entscheidungen*

Bei der Inanspruchnahme medizinischer Leistungen besitzt die zeitliche Abfolge der Handlungen von Arzt und Patient, für den Fall, daß sich letzterer zu einer Behandlung entschlossen hat, Auswirkungen auf das Ergebnis des Behandlungsprozesses. Dabei können zwei Fälle als Ansatzpunkte für eine zeitliche Betrachtung unterschieden werden: Einmal die Situation, in der beide, Arzt und Patient, ihre Aktionen unabhängig von einander treffen und in der sie über keinerlei Informationen über die Entscheidungen des anderen verfügen, und zudem die Konstellation, in der ein Akteur die Handlungen des anderen beobachten kann, bevor er die Entscheidung über die eigene Aktion trifft. Diese Unterscheidung hängt unmittelbar von der Informationsverteilung zwischen beiden Akteuren ab, ob also vollständige Information vorliegt oder ob immer noch Informationsdefizite und damit Asymmetrien bestehen.⁸⁸ Das resultierende Gleichgewicht der Handlungsstrukturen zwischen Arzt und Patient ist dann vom Bayes-Nash-Typ. Die Pay-offs der beiden Spieler nach Vollzug der Aktionen hängen dann u.a. von der Ausprägung der Handlungen beider Akteure ab.

Die erste mögliche Zeitstruktur wird durch die Situation bei unvollständiger Information abgebildet, in der beide Akteure ihre Handlungen *simultan* wählen. Zur Vereinfachung wird dabei angenommen, daß beide nur jeweils zwischen zwei Aktionsniveaus wählen können, einem hohen Aufwand für die medizinische Leistung m bzw. Compliance a , gekennzeichnet mit dem Index h , und einem niedrigen mit dem Index l . Abbildung 6.1 zeigt die Situation anhand zweier Entscheidungsbäume. Im ersten Schaubild (6.1.I) wird angenommen, daß der Arzt als erster seine Aktion wählt und anschließend der Patient. Die Simultaneität der Entscheidungen wird durch die gestrichelte Linie dargestellt. Sie deutet an, daß der Patient nicht zwischen den Entscheidungen des Arztes unterscheiden kann. Speziell folgt daraus, daß der Patient nicht weiß, an welchem der

⁸⁸ Banerjee und Beggs (1992) präsentieren eine Methode, die sequentiellen und simultanen Handlungen in einem Principal-Agent-Modell zu untersuchen.

beiden Entscheidungspunkte er sich befindet, so daß diese zu einer Informationsmenge zusammengefaßt werden (vgl. Gibbons 1992, S. 120 und Eichberger 1993, S. 12). Schaubild (6.1.II) zeigt die Situation, in der zuerst der Patient seine Aktion wählt und erst danach der Arzt. Auch der Arzt kennt die Aktion des Patienten nicht und besitzt daher keine Information über den Entscheidungspunkt. Beide Schaubilder repräsentieren somit dieselbe Spielstruktur, da keiner der Spieler die Aktion des anderen kennt. Daher entspricht dies der Situation mit simultanen und unabhängigen Entscheidungen der beiden Akteure. Es ist in diesem Fall nicht relevant, ob der Arzt oder der Patient zuerst seine Aktion wählt und somit am Beginn der skizzierten Entscheidungsbäume steht.

Die resultierenden Pay-offs bzw. Erwartungsnutzen (U für den Patienten und V für den Arzt) hängen neben den gewählten Aktionen auch von deren Einfluß auf die Gesundheit und damit von dem resultierenden Gesundheitszustand des Patienten ab.⁸⁹ Welcher der Pay-offs von den Akteuren bevorzugt wird, richtet sich nach dem Verhältnis der Compliance und der medizinischen Leistung, d.h., ob es sich um Komplemente oder um Substitute handelt (vgl. auch Ma und McGuire 1997, S. 692). Falls die Compliance und die medizinische Leistung Substitute sind, dann wäre es angebracht, bei gegebener medizinischer Leistung die Compliance zu reduzieren, um einen höheren Erwartungsnutzen zu erzielen. Auf der anderen Seite würde im Fall der Komplemente ein höheres Niveau der medizinischen Leistung eine hohe Compliance fordern, damit ein hoher Erwartungsnutzen erreicht wird. Da die Aktionen aufgrund der beidseitigen asymmetrischen Information jedoch nicht als optimale Reaktion auf die Aktion des anderen Spielers gewählt werden können, kann eine bestimmte Konstellation der Erwartungsnutzen nicht gezielt erreicht werden.

Im Gegensatz dazu steht die Situation, in der die Aktionen *sequentiell* gewählt werden und bei der keine beidseitige Informationsasymmetrie besteht. In diesem Fall ist es entscheidend, wer als erstes seine Handlungen auswählt.⁹⁰ Geht man dabei vom Ablauf der Arztkonsultation aus, so wird der Patient, nachdem die Natur über seinen Gesundheitszustand entschieden hat (krank oder gesund), seinen Arzt aufsuchen. Dieser führt Untersuchungen durch, spricht mit dem Patienten und entscheidet dann über die Therapie bzw. schlägt dem Patienten eine Therapie vor. Daran anschließend fällt der Patient seine Entscheidung über sei-

⁸⁹ Dabei gilt zu beachten, daß der Gesundheitszustand eine Zufallsvariable darstellt, deren Verteilung von den individuellen Aktionen von Arzt und Patient beeinflußt wird (vgl. Kapitel 4.2).

⁹⁰ Dabei wird der Fall wiederholter Aktionen nicht betrachtet. Es besteht jedoch bei sequentiellen Aktionen eine einseitige Informationsasymmetrie, da der beginnende Akteur keine Informationen über die Wahl des nachfolgenden Spielers besitzt.

nen behandlungsbegleitenden Aufwand (Compliance). Im weiteren Verlauf der Untersuchung wird der Arzt den ersten Zug unternehmen und der Patient folgen.

Abbildung 6.1: Simultane Entscheidungen unter unvollständiger Information

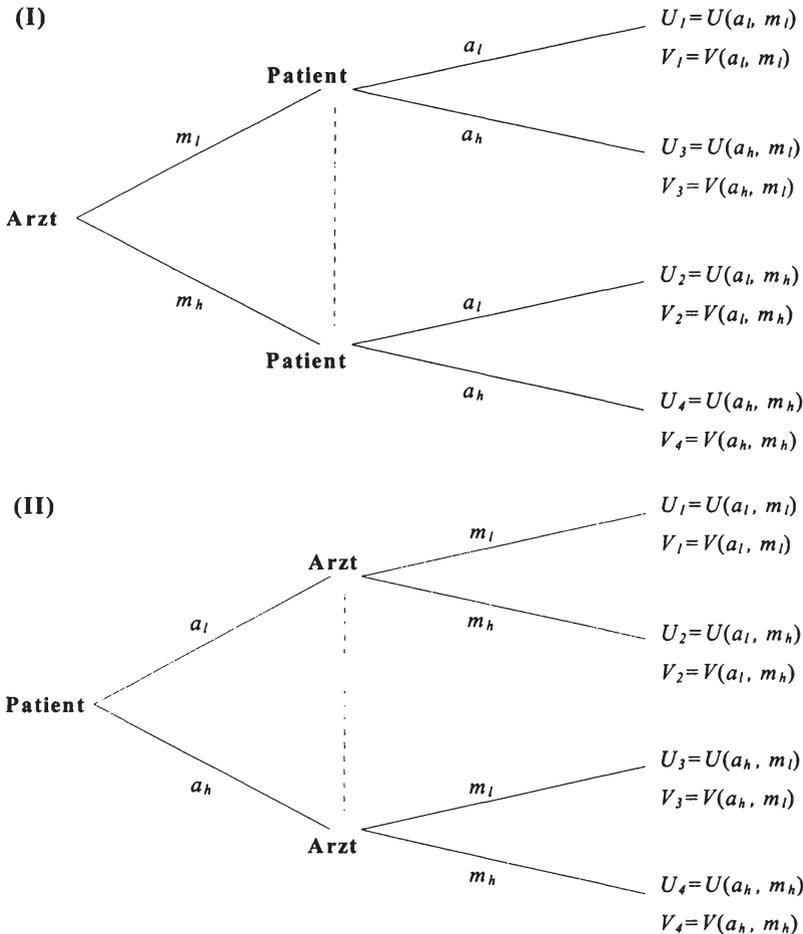
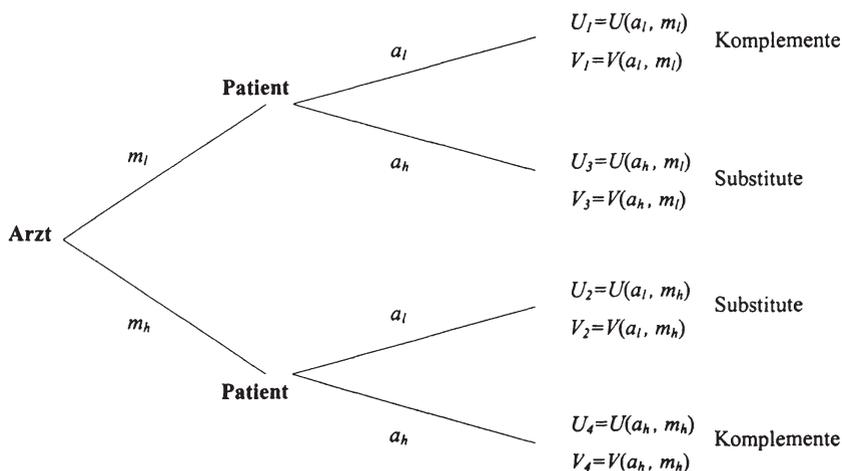


Abbildung 6.2 zeigt die Zeitstruktur bei sequentiellen Entscheidungen. Beginnend mit der Wahl der medizinischen Leistung durch den Arzt, sieht sich der Patient normalerweise dem Problem gegenüber zu beurteilen, ob der Arzt nun ein hohes oder ein niedriges Niveau der medizinischen Leistung erbracht hat. Zuerst wird zur Vereinfachung unterstellt, daß keine Unsicherheit über die ge-

wählte Aktion herrscht. Er wird daher das aus seiner Sicht optimale Niveau der Compliance wählen. Dies bedeutet, daß er, falls eine komplementäre Beziehung zwischen den Aktionen vorliegt, im Fall eines hohen Niveaus der medizinischen Leistung auch eine hohe Compliance wählen wird und vice versa. Auf der anderen Seite wählt er, falls er von einer hohen medizinischen Leistung ausgehen kann und Substitute vorliegen, ein niedriges Niveau der Compliance.

Abbildung 6.2: Sequentielle Entscheidungen ohne Unsicherheit



Im Gegensatz zum Fall simultaner Entscheidungen beeinflusst die Unterscheidung in Komplemente und Substitute nicht nur die resultierenden Pay-offs bzw. Erwartungsnutzen, sondern auch die Wahl der Aktion auf der zweiten Stufe durch den Patienten. Daher sind in diesem Fall sowohl die zeitliche Reihenfolge der Entscheidungen als auch das individuelle Wissen von großer Bedeutung für das Ergebnis. Der erste Punkt zielt auf den bereits mehrfach angesprochenen Aspekt, in welcher Reihenfolge die Aktionen gewählt werden, ab. Der zweite Aspekt bezieht sich auf die Situation, in der eine beidseitige Informationsasymmetrie zwischen Arzt und Patient vorliegt. Solange weder der Arzt noch der Patient zusätzliche Information über die Handlungen des anderen Akteurs erhalten, entscheiden sie auch bei sequentiellen Strukturen unter Unsicherheit.⁹¹ Für eine Betrachtung dieser Situation reicht es dann aus, von einer simultanen Struktur auszugehen, da hier Überlegungen, welcher Akteur die erste Handlung unternimmt, vernachlässigt werden können und sich folglich die Modellstruktur

⁹¹ Holmström (1979) diskutiert den Einfluß zusätzlicher Signale im Rahmen eines Principal-Agent-Problems.

vereinfacht. Geht man hingegen davon aus, daß der Patient die medizinischen Leistungen des Arztes beobachten und einschätzen vermag, so verringert sich bei sequentiellen Strukturen die Unsicherheit über die resultierenden Pay-offs, da die Aktionen auf der zweiten Stufe unter Berücksichtigung der Interaktionen gewählt werden können. Das Wissen um Komplemente oder Substitute ist daher im Fall der sequentiellen Handlungen bei vollständiger Information des Patienten über die Aktionen des Arztes von besonderem Wert, hingegen spielt derartiges Wissen bei simultanen Aktionen für die Wahl der Aktionen aus Sicht des Handelnden eine untergeordnete Rolle.

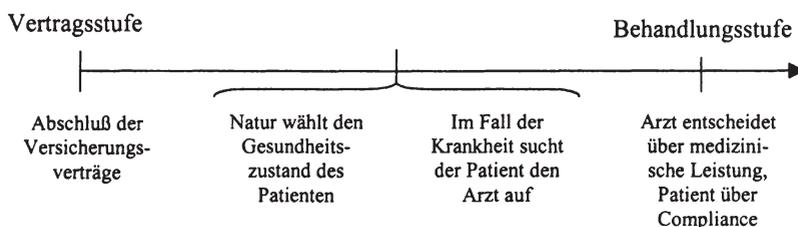
6.4 Modellstufen des Double Moral Hazard-Problems

Bei der Übertragung der Principal-Agent-Theorie wird, wie in einer Vielzahl von Arzt-Patient-Modellen, zum einen der Umstand vernachlässigt, daß neben dem Arzt als Agenten auch der Patient in der Rolle des Principals Handlungen unternimmt. Diese beeinflussen das Ergebnis des Behandlungsprozesses und sind für den Arzt als Agenten nicht beobachtbar (vgl. hierzu Abschnitt 4.3.1 und Schneider (1998)). Die Compliance des Patienten stellt dabei für den Heilungserfolg medizinischer Maßnahmen einen nicht zu vernachlässigenden Faktor dar. Zum anderen sind finanzielle Transaktionen zwischen Arzt und Patient in den meisten Gesundheitssystemen auf verschiedene Formen der Selbstbeteiligung beschränkt und erfolgen nicht direkt (vgl. Zweifel (1994) oder Schneider et al. (1995)). Dieser Umstand beruht auf der Tatsache, daß im Gesundheitswesen unter zahlreichen „ergänzenden Sachwaltern“ (Zweifel 1994, S. 20) besonders die Krankenversicherung eine bedeutende Stellung einnimmt, da diese mit beiden Parteien in vertraglichen Beziehungen steht (vgl. Kortendieck (1993), Gaynor (1994), Börsch-Supan (1998) oder Cutler und Zeckhauser (2000) sowie Abschnitt 4.3.1). Gegen eine Prämie versichert sie den Patienten gegen die im Krankheitsfall entstehenden Kosten, abzüglich einer eventuell anfallenden Selbstbeteiligung. Der Arzt erhält seine Vergütung von der Versicherung ausbezahlt. Die Versicherung nimmt daher die Rolle eines Intermediärs ein, der die finanziellen Transaktionen im Krankheitsfall regelt und dem Patienten Schutz gegen das monetäre Krankheitsrisiko gewährt. In einer theoretischen Betrachtung müßten diese beiden Faktoren in ein Modell integriert werden.

Wichtig bei der Analyse ist eine Trennung des Modells in zwei Stufen, eine Vertragsstufe und eine Behandlungsstufe (siehe Abbildung 6.3). Zunächst wird auf der Vertragsstufe die Versicherung jeweils mit dem Patienten und dem Arzt einen separaten Vertrag abschließen. Diese Verträge sind allerdings nicht unabhängig voneinander, sondern hängen von den Vertragsbestimmungen und damit auch von den Handlungen des anderen ab. Daher kann man diese Verträge als

„tripartite contracts“ (Brown und Wolfstetter 1989, S. 292) bezeichnen. Nach Abschluß der Verträge entscheidet ein unabhängiger Spieler, die Natur, ob der Patient erkrankt. Tritt dieses Ereignis ein, sucht der Patient den Arzt auf. Auf der dann folgenden Behandlungsstufe wählen Arzt und Patient ihre Handlungen, die neben dem Gesundheitszustand auch von den Vertragsparametern beeinflusst werden. Dies bedeutet, daß die Wahl der Prämie, der Selbstbeteiligung und die Form der Arzthonorierung das Ergebnis der Behandlung durchaus beeinflussen. Voraussetzung für die Behandlungsstufe ist allerdings, daß sich der Patient zum Erstkontakt entschlossen hat, d.h. den Arzt konsultiert.

Abbildung 6.3: Stufen des Modells



In der theoretischen Modellierung wird zuerst die Behandlungsstufe analysiert. Die dabei herausgearbeiteten Ergebnisse über den Zusammenhang zwischen medizinischer Leistung und der Compliance spielen in der anschließenden Untersuchung der Vertragsstufe im Fall der asymmetrischen Information eine Rolle, da die dort festgelegten Vertragsparameter die Aktionen beeinflussen.

6.5 Das theoretische Modell der Behandlung

6.5.1 Grundstruktur

Nachdem sich der Patient zu einer Behandlung entschlossen hat, entscheiden der Arzt über die medizinische Leistung und der Patient über seine Compliance simultan.⁹² Der Patient zieht Nutzen aus seinem verfügbaren Einkommen (Nettoeinkommen). Dies entspricht seinem Bruttoeinkommen y abzüglich der Versicherungsprämie σ und den Kosten einer prozentualen Selbstbeteiligung β an den Ausgaben für die medizinische Leistung m , mit $0 < \beta < 1$, so daß sich als Ausgaben des Patienten für die medizinische Behandlung βm ergibt. Dabei wird angenommen, daß für die medizinische Leistung konstante Grenzkosten von

⁹² Das Modell lehnt sich an eine Arbeit von Cooper und Ross (1985) über Garantieleistungen zwischen Käufer und Produzent auf einem Gütermarkt an.

Eins existieren. Das Bruttoeinkommen hängt positiv vom nach der Behandlung erreichten Gesundheitszustand G ab.⁹³ Dies bedeutet, daß der Patient bei guter Gesundheit in der Lage ist, ein hohes Einkommen zu erzielen; hingegen schränkt ein schlechter Gesundheitszustand seine produktiven Möglichkeiten ein, so daß sein Erwerbseinkommen niedriger ausfällt.⁹⁴ Die Prämie hat der Versicherte in jedem Fall zu entrichten, eine Selbstbeteiligung fällt nur dann an, wenn er den Arzt konsultiert und dieser ihn behandelt.

Zur weiteren Vereinfachung wird für den medizinischen Leistungserstellungsprozeß folgendes angenommen: Nach der Behandlung nimmt der Gesundheitszustand des Patienten zwei Ausprägungen an. Entweder führt die Behandlung zum Erfolg und der Patient gesundet, oder er ist nach wie vor krank. Für den ersten Fall eines guten Gesundheitszustands nimmt G den hohen Wert G_1 an, hingegen den niedrigen Wert G_0 , falls der Gesundheitszustand schlecht ist, mit $G_1 > G_0$. Der gute Gesundheitszustand wird mit Wahrscheinlichkeit $p \in (0,1)$ realisiert, ein schlechter mit der Wahrscheinlichkeit $1-p$. Die Genesung ist damit kein deterministischer Prozeß, sondern unterliegt der Stochastik. Dabei bezeichnet $p(a, m)$ die Wahrscheinlichkeit für einen guten Gesundheitszustand, gegeben die Compliance des Patienten a und die medizinische Leistung des Arztes m . Diese läßt sich auch im Sinne einer Gesundheitsproduktionsfunktion interpretieren. Bezüglich der Inputs a und m wird angenommen, daß die Erhöhung eines Inputs die Wahrscheinlichkeit für einen guten Gesundheitszustand positiv beeinflusst, wenn auch mit einer abnehmenden Rate. Beide Variablen besitzen damit eine positive, aber abnehmende Grenzproduktivität. Für die Gesundheitsproduktion folgt, daß schlechte Gesundheitszustände mit besserer medizinischer Leistung oder höherer Compliance unwahrscheinlicher werden, aber nicht ganz auszuschließen sind (vgl. Schneider (1998)).

Die Erwartungsnutzenfunktion des Patienten ist additiv-separabel in Nutzen aus dem Nettoeinkommen $U(\cdot)$ und Aufwand der Compliance $D(a)$, d.h., je mehr der Patient die Behandlung unterstützt, desto höher ist der damit verbundene

⁹³ Es wird hierbei unterstellt, daß das Bruttoeinkommen hinreichend groß und das Nettoeinkommen immer strikt positiv ist.

⁹⁴ Man kann sich dies am Beispiel des Krankengeldes verdeutlichen. Falls der Patient krankheitsbedingt nicht mehr seiner Arbeit nachgeht, dann erhält er von seiner Versicherung eine Transferleistung, die i.d.R. niedriger ausfällt als sein vertraglich vereinbartes Einkommen. Besonders einleuchten ist dies im Fall eines Selbständigen, dessen Einkommen von seiner Produktivität bestimmt wird, die wiederum direkt von seinem Gesundheitszustand abhängt. Von anderen Einkommensarten wird an dieser Stelle abstrahiert.

Aufwand.⁹⁵ Der Versicherte ist risikoavers im Nettoeinkommen, der zugehörige Teil der Nutzenfunktion ist somit streng konkav, während für den Aufwand angenommen wird, daß der daraus resultierende negative Nutzen mit Zunahme der Compliance weiter ansteigt. Die Funktion $D(\cdot)$ ist daher konvex und als Erwartungsnutzen folgt:

$$EU = p(a, m) U(y[G_1] - \sigma - \beta m) + (1 - p(a, m)) U(y[G_0] - \sigma - \beta m) - D(a). \quad (6.1)$$

Im weiteren bezeichnet U_1 den resultierenden Nutzen aus einem hohen Gesundheitszustand G_1 und U_0 den Nutzen aus einem niedrigen Gesundheitszustand G_0 :

$$U_0 = U(y[G_0] - \sigma - \beta m),$$

$$U_1 = U(y[G_1] - \sigma - \beta m).$$

Der Nutzen des Arztes ist additiv-separabel in Einkommen, Berufsethik und Aufwand, wobei der Arzt risikoneutral im Einkommen ist, da er sein Einkommensrisiko auf eine Vielzahl von Patienten verteilen kann. Für den Arzt gilt, daß er für seine erbrachte medizinische Leistung m eine Honorierung in Form einer Fallpauschale $\omega > 0$ erhält, zuzüglich einer Zahlung δ pro Einheit der medizinischen Leistung. Falls dem Arzt nur ein Teil seiner entstehenden Kosten vergütet wird liegt eine Kostenbeteiligung auf Anbieterseite vor (vgl. Ellis und McGuire (1990) und (1993)). Im anderen Fall beschreibt δ den Aufschlag auf seine Grenzkosten, die ihm vergütet werden. Der Arzt erhält somit pro Einheit medizinischer Leistung ein Honorar abzüglich seiner Kosten.⁹⁶ Weiterhin wird angenommen, daß der Arzt nach Abschluß der Behandlung der Versicherung die tatsächlich in Anspruch genommene medizinische Leistung mitteilt. Die medizinische Leistung verursacht ihm neben monetären Kosten auch Aufwand in Höhe von $C(m)$ mit den Ableitungen $C' > 0$ und $C'' > 0$. Die Berufsethik, derer sich der Arzt verpflichtet, läßt sich dadurch beschreiben, daß der erwartete Gesundheitszustand des Patienten positiv in die folgende Erwartungsnutzenfunktion eingeht:

⁹⁵ Bei einer additiv-separablen Nutzenfunktion variiert der Grad der Risikoaversion des Nutzens aus dem Einkommen U nicht mit den Anstrengungen a (vgl. Macho-Stadler und Pérez-Castrillo 1997, S. 19).

⁹⁶ Nimmt man an, daß die medizinische Leistung zu Grenzkosten von Eins erbracht werden kann, so folgt für die Honorierung $\omega + (\delta + 1)m - m = \omega + \delta m$ (siehe Gleichung (6.2)).

$$EV = [\omega + \delta m] + \varepsilon [p G_1 + [1 - p] G_0] - C[m] \geq \bar{V}. \quad (6.2)$$

Der Parameter $\varepsilon \in [0,1]$ gibt die Intensität der Berufsethik an. Das bedeutet, daß ein Arzt für $\varepsilon=0$ den Ausdruck in der eckigen Klammer, d.h. den erwarteten Gesundheitszustand des Patienten, nicht in seine Entscheidung mit einbezieht und somit keiner Berufsethik unterliegt. Ein Arzt mit $\varepsilon=1$ hingegen gewichtet den Gesundheitszustand des Patienten hoch. Der Arzt verhält sich dann als perfekter Agent des Patienten. Eine Verbesserung des Gesundheitszustands wirkt um so stärker, je höher der Wert des Parameters ε ist.

Die Versicherung bezieht vom Versicherten die Prämie σ und leistet im Fall einer Konsultation eine Zahlung an den Arzt in Höhe seines Honorars abzüglich der Selbstbeteiligung des Patienten:

$$\sigma = E(\omega + [\delta - \beta] m). \quad (6.3)$$

$E(\cdot)$ gibt dabei den Erwartungswert einer Auszahlung der Versicherung an den Arzt an, die lediglich im Falle einer Konsultation wirksam wird. Für den Versicherungsmarkt wird unterstellt, daß vollständiger Wettbewerb herrscht und daß die Versicherung zu fairen Prämien angeboten wird. Es werden hier sowohl für den Versicherten als auch für den Arzt nur lineare Verträge betrachtet.

Für das Modell werden folgende Annahmen bezüglich der Informationsverteilung getroffen: Der erreichte Gesundheitszustand ist sowohl dem Arzt als auch dem Patienten bekannt, nicht jedoch der Versicherung. Daher ist es ihr nicht möglich, die in den Verträgen vereinbarten Zahlungen auf den Gesundheitszustand zu beziehen. Die Gesundheitsproduktion ist Arzt und Patient ex ante bekannt, d.h. sie wissen um den grundsätzlichen Einfluß der Aktionen auf die Genesungswahrscheinlichkeit. Die Akteure können allerdings von ihrer eigenen Aktion keine Rückschlüsse auf das Verhalten des Mitspielers ziehen. Da es sich hierbei um ein einmaliges Spiel handelt, ist es unerheblich, ob sie ihre Aktionen ex post wechselseitig ermitteln können.

6.5.2 Kooperative Lösung

Grundsätzlich besteht zwischen dem Arzt und dem Patienten eine wechselseitige Informationsasymmetrie. Da sowohl die medizinische Leistung als auch die Anstrengungen des Patienten den Wert des Arztbesuches für die jeweils andere Partei beeinflussen, kennzeichnet dies die Situation des doppelten Moral

Hazard.⁹⁷ Um den Effekt des Double Moral Hazard auf die Compliance und die medizinische Leistung abschätzen zu können, ist es zunächst notwendig, die Behandlungssituation bei vollständiger Information zu analysieren.

In einer solchen Situation wählen beide Akteure ihre Handlungen so, daß die Summe der erwarteten Nutzen maximiert wird. Die First-best-Lösung ergibt sich dann durch folgenden Ansatz:

$$\begin{aligned} \max_{a,m} S = & p(a,m) [U_1 + \varepsilon G_1] + (1 - p(a,m)) [U_0 + \varepsilon G_0] \\ & + (\omega + \delta m) - D(a) - C(m). \end{aligned} \tag{6.4}$$

Für die Bedingungen erster Ordnung folgt dann:

$$\frac{\partial S}{\partial a} : p_a [U_1 - U_0 + \varepsilon (G_1 - G_0)] = D', \tag{6.5}$$

$$\frac{\partial S}{\partial m} : p_m [U_1 - U_0 + \varepsilon (G_1 - G_0)] + \delta = C' + \beta [p U_1' + (1 - p) U_0']. \tag{6.6}$$

Gleichung (6.5) besagt, daß der Patient seine Anstrengung so wählt, daß der gemeinsame erwartete Grenznutzen⁹⁸ auf der linken Seite gleich dem Grenzaufwand ist. Mit anderen Worten, die Summe der erwarteten Grenznutzen bei Arzt und Patient durch eine Zunahme der Compliance des Patienten muß im Optimum dem Grenzaufwand des Patienten entsprechen.

Für Gleichung (6.6) gilt, daß der gemeinsame erwartete Grenznutzen aus der Behandlung gleich den erwarteten Grenzkosten ist, bestehend aus dem Grenzaufwand der medizinischen Leistung und den Grenzkosten der Selbstbeteiligung. Da beide Gleichungen von jeweils beiden Variablen (medizinische Leistung und Compliance des Patienten) abhängen, bilden diese die Reaktionsfunktionen des Patienten und des Arztes ab. Die resultierende First-best-Lösung wird im weiteren mit (a^*, m^*) bezeichnet.

⁹⁷ Vgl. u.a. Demski und Sappington (1991), Bhattacharyya und Lafontaine (1995), Cooper und Ross (1995), Al Najjar (1997) sowie Kim und Wang (1998).

⁹⁸ Der gemeinsame erwartete Grenznutzen bezeichnet die Summe aus den erwarteten Grenznutzen des Arztes und des Patienten, aufgrund einer Erhöhung der Compliance oder der medizinischen Leistung.

6.5.3 Nichtkooperative Lösung

Im Fall des doppelten Moral Hazard können beide Akteure die Handlungen des jeweils anderen nicht direkt beobachten. Hier maximieren der Patient und der Arzt ihren Nutzen, ohne die Auswirkungen auf den jeweils anderen zu berücksichtigen. Der Patient kann nicht aus dem erreichten Gesundheitszustand Rückschlüsse auf die medizinische Leistung des Arztes schließen. Analog ist es dem Arzt nicht möglich, die Anstrengungen des Patienten zu beurteilen. Für den Patienten folgt aus dieser Situation asymmetrischer Information, daß lediglich sein eigener Erwartungsnutzen in den Maximierungsansatz eingeht:

$$\max_a p(a, m) U_1 + (1 - p(a, m)) U_0 - D(a). \quad (6.7)$$

Daraus ergibt sich als Bedingung erster Ordnung, daß

$$p_a(a, m)(U_1 - U_0) = D'(a). \quad (6.8)$$

Gleichung (6.8) zeigt, daß der Patient seine gesundheitsrelevante Anstrengung derart wählt, daß der erwartete Grenznutzen der Compliance gleich dem Grenzaufwand ist. Bezüglich der Versicherungsparameter ist es von Interesse, wie eine Erhöhung des Selbstbeteiligungssatzes auf die Anstrengungen des Patienten wirkt. Mit Hilfe des Impliziten Funktionentheorems erhält man aus Gleichung (6.8):

$$\frac{da}{d\beta} = \frac{p_a(U_1' - U_0') \left(m + \frac{\partial \sigma}{\partial \beta} \right)}{p_{aa}(U_1 - U_0) - D''} > 0. \quad (6.9)$$

Der Nenner ist hierbei mit der hinreichenden Bedingung für ein Nutzenmaximum identisch und daher unter den getroffenen Annahmen negativ. Der Zähler ist ebenfalls negativ, da die Differenz der Grenznutzen des Einkommens ($U_1' - U_0'$) aufgrund der Konkavität der Nutzenfunktion kleiner als Null ist. Hieraus folgt, daß der negative Effekt der Selbstbeteiligung auf die Prämie ($\partial \sigma / \partial \beta$) kleiner als der Effekt der Gesundheitsausgaben (m) ist.⁹⁹ Als Gesamteffekt ergibt sich, daß ein Anstieg des Selbstbeteiligungssatzes die Anstrengungen des Patienten für eine gegebene medizinische Leistung ebenfalls erhöht. Ähnlich

⁹⁹ Für die Ableitung der Versicherungsprämie nach dem Selbstbeteiligungssatz ergibt sich $|\partial \sigma / \partial \beta| = E(\beta m) = \beta E(m) < m$, da der Selbstbeteiligungssatz kleiner als Eins ist ($\beta < 1$).

können empirische Untersuchungen für die Schätzung der Preiselastizität der Nachfrage nach medizinischen Leistungen interpretiert werden. Ein in diesem Zusammenhang ermittelter Wert von -0,2 für die Elastizität führt zu einer sinkenden Nachfrage mit steigendem Preis (vgl. Cutler und Zeckhauser 2000, S. 584ff.). Als Folge dieser Zusammenhänge können in der Situation einer negativen Preiselastizität der Nachfrage medizinische Leistungen aufgrund des gesunkenen Relativpreises für eigene Anstrengungen durch eine höhere Compliance des Patienten substituiert werden.

Der Arzt maximiert simultan seinen Erwartungsnutzen über die von ihm erbrachte medizinische Leistung. Für den Optimierungsansatz ergibt sich

$$\max_m \left[\omega + \delta m \right] + \varepsilon \left(p(a, m) G_1 + (1 - p(a, m)) G_0 \right) - C(m) \quad (6.10)$$

und als Bedingung erster Ordnung für ein Nutzenmaximum:

$$\delta + p_m(a, m) \varepsilon (G_1 - G_0) = C' \quad (6.11)$$

Kennzeichnend für die aus Sicht des Arztes optimale Menge an medizinischer Leistung ist die Gleichheit von Grenzaufwand für die Bereitstellung der Leistung C' und Grenznutzen der medizinischen Leistung, bestehend aus dem monetären Grenznutzen δ und dem Grenznutzen eines besseren Gesundheitszustands.

Analog zum Fall des Patienten läßt sich auch hier untersuchen, wie die Honorierungsparameter ω und δ die medizinische Leistung beeinflussen. Aufgrund der Risikoneutralität hängt die Entscheidung des Arztes nicht von der Fallpauschale ω sondern lediglich von der Höhe der Kostenerstattung δ ab. Für eine Erhöhung der Erstattung ergibt sich:

$$\frac{dm}{d\delta} = - \frac{1}{\varepsilon p_{mm} (G_1 - G_0) - C''} > 0 \quad (6.12)$$

Der Nenner stellt die hinreichende Bedingung für ein Maximum dar und ist daher negativ, so daß sich ein positiver Gesamteffekt ergibt. Die Folge einer höheren Kostenerstattung bzw. einer geringeren Kostenbeteiligung des Arztes ist daher ein Anreiz zur Ausdehnung der medizinischen Leistung.

Für die weitere Analyse stehen die Gleichungen (6.8) und (6.11) im Mittelpunkt der Betrachtung. Sie spezifizieren die Handlungen eines Akteurs als Funktion der Aktionen des anderen Spielers. Der Patient wählt daher seinen Aufwand als Reaktion auf die medizinische Leistung des Arztes und umgekehrt. Diese Reaktionsfunktionen spezifizieren ein Nash-Gleichgewicht, die nichtkooperative Lösung (a^N, m^N).

6.5.4 Vergleich beider Lösungen

Durch den Vergleich der Lösungen bei vollständiger und bei asymmetrischer Information erhält man einen Anhaltspunkt über die Auswirkung der Informationsasymmetrie auf die Reaktionsfunktionen. Für den Patienten zeigt eine direkte Gegenüberstellung der Gleichungen (6.5) und (6.8), daß sein Aufwand im Fall der vollständigen Information genau dann höher ist als bei unvollständiger Information, wenn die Grenzkosten höher sind. Es gilt, daß $a^*(m) > a^N(m)$, wenn aufgrund der Monotonie des Aufwands folgt, daß $D'(a^*) > D'(a^N)$. Damit muß auch die linke Seite von Gleichung (6.5) größer als die von Gleichung (6.8) sein, das bedeutet, daß

$$a^*(m) > a^N(m) \Leftrightarrow p_{a^*} [U_1 - U_0 + \varepsilon (G_1 - G_0)] > p_{a^N} [U_1 - U_0]. \quad (6.13)$$

Gleichung (6.13) besagt, daß bei Kooperation genau dann ein höheres Niveau der Compliance gewählt wird, wenn der erwartete Grenznutzen der Compliance bei Kooperation größer als bei Nicht-Kooperation ist. In diesem Fall unterscheidet sich die kooperative First-best-Lösung von der nichtkooperativen Lösung. Mit anderen Worten: Realisiert der Patient, daß der Arzt von einem höheren Gesundheitszustand profitiert, so wird er seine Anstrengungen für alle Werte von m erhöhen. Für eine gegebene medizinische Leistung wird der Patient im Fall der vollkommenen Information mehr Anstrengungen betreiben als in der Situation asymmetrischer Information, es sei denn, der Arzt besitzt keine Berufsethik.¹⁰⁰ Für den Verlauf der beiden Reaktionsfunktionen des Patienten folgt, daß die Funktion bei kooperativer Lösung für alle Werte von m über derjenigen bei nichtkooperativer Lösung verläuft.

Für den Arzt ergibt der Vergleich von kooperativer und nichtkooperativer Lösung folgendes Bild: Angenommen, die medizinische Leistung bei vollkommener Information sei größer als bei asymmetrischer Information ($m^*(a) > m^N(a)$),

¹⁰⁰ Es läßt sich dann aber anzweifeln, ob der Patient einen Arzt aufsuchen würde, der keine Berufsethik besitzt. Gegen eine solche Argumentation läßt sich einwenden, daß dem Patienten nur begrenzte Informationen über die Motivation des Arztes zur Verfügung stehen.

dann folgt daraus, daß der Grenzaufwand der Behandlung auch höher sein muß, unter der Bedingung, daß die Funktion des Aufwandes monoton verläuft. In diesem Fall muß die linke Seite von Gleichung (6.6) abzüglich der Selbstbeteiligung größer sein als die linke Seite der Gleichung (6.11):

$$\begin{aligned}
 m^*(a) > m^N(a) &\Leftrightarrow \\
 p_m^* \left[U_1 - U_0 + \varepsilon (G_1 - G_0) \right] + \delta - \beta \left[p U_1' + (1-p) U_0' \right] & \\
 > \delta + p_{m^N} \varepsilon (G_1 - G_0). & \quad (6.14)
 \end{aligned}$$

Der erste Teil der Ungleichung entspricht der optimalen Menge an medizinischer Leistung bei Kooperation (m^*). Im Vergleich zur medizinischen Leistung bei Nicht-Kooperation ist die medizinische Leistung in der kooperativen Lösung dann höher, wenn der erwartete Grenznutzen der medizinischen Leistung abzüglich der erwarteten marginalen Selbstbeteiligung des Patienten größer als der erwartete Grenznutzen bei Nicht-Kooperation ist. In diesem Fall verläuft die Reaktionsfunktion des Arztes bei Kooperation für alle Werte von a oberhalb der Reaktionsfunktion bei Nicht-Kooperation. Durch die Effekte einer Selbstbeteiligung ist es aber möglich, daß die Reaktionsfunktionen sich schneiden.¹⁰¹ Weiterhin gilt, daß der erste Teil der Ungleichung (6.14) für $\varepsilon = 0$ (keine Berufsethik des Arztes) die optimale medizinische Leistung aus Sicht des Patienten spezifiziert.

6.6 Das Verhältnis zwischen Compliance des Patienten und ärztlicher Behandlungsleistung

6.6.1 Substitutions- versus Komplementaritätsbeziehung

Das Ergebnis des Behandlungsprozesses wird durch die Wahl der medizinischen Leistung und der Anstrengungen beeinflusst, da diese positiv auf die Wahrscheinlichkeit eines guten Gesundheitszustands wirken. Das Niveau der medizinischen Leistung und der Compliance wird dabei durch die Reaktionsfunktionen determiniert. Im weiteren Verlauf wird angenommen, daß für die Reaktionsfunktionen im Fall der Informationsasymmetrie ein eindeutiges und stabiles Nash-Gleichgewicht existiert. Während bislang bestimmt wurde, wie sich die

¹⁰¹ Der Fall, in dem die medizinische Leistung bei Informationsasymmetrie höher ausfällt als bei vollständiger Information, impliziert höhere Behandlungsgrenzkosten als Behandlungsgrenznutzen. Das bedeutet, daß die Selbstbeteiligung für die letzte Einheit medizinische Leistung höher als ihr Nutzen ist. Im weiteren Verlauf wird dies ausgeschlossen, da dann der Patient bei vollständiger Information den Arzt nicht aufsuchen würde.

Reaktionsfunktionen bei vollständiger und asymmetrischer Information unterscheiden, wurden die Auswirkungen der Verläufe auf das Niveau der medizinischen Leistung und der Compliance ausgeklammert. Dies betrifft vor allem den bereits diskutierten Zusammenhang zwischen beiden Variablen (vgl. Abschnitt 6.1). Im Mittelpunkt der Analyse steht aus diesem Grund die Auswirkung des Verhältnisses der beiden Aktionen zueinander, also inwiefern komplementäre oder substitutive Zusammenhänge die Ergebnisse beeinflussen.

In der Analyse wird angenommen, daß ein Akteur auch dann eine positive Aktion ($a, m > 0$) wählt, wenn der andere Akteur sich dazu entschließt, nichts zu unternehmen, da ansonsten u.U. kein Gleichgewicht mit positiven Aktionen existiert (vgl. Cooper und Ross 1985, S. 106). In Termini des Modells bedeutet dies, daß $a(m = 0) > 0$ und $m(a = 0) > 0$. Um den Verlauf der Reaktionsfunktionen zu ermitteln, werden die Bedingungen erster Ordnung für den Arzt und den Patienten bei Nicht-Kooperation entsprechend des Impliziten Funktionentheorems total differenziert und nach da/dm aufgelöst, so daß für den Patienten folgt:

$$\frac{da}{dm} = \frac{p_a (U'_1 - U'_0) - p_{am} (U_1 - U_0)}{p_{aa} (U_1 - U_0) - D''} \quad (6.15)$$

Der Nenner dieses Bruches ist die hinreichende Bedingung für ein Nutzenmaximum des Patienten und daher negativ. Das Vorzeichen des Zählers hängt von zwei Faktoren ab: Der erste Term ist nur dann ungleich Null, wenn ein positiver Selbstbeteiligungssatz vorliegt (Selbstbeteiligungskomponente). Der zweite Term ist die strategische Komponente, deren Effekt vom Vorzeichen der Kreuzableitung der Wahrscheinlichkeit für einen hohen Gesundheitszustand (p_{am}) abhängt. Diese Kreuzableitung gibt an, wie sich die Grenzproduktivität der Compliance ändert, wenn sich die medizinische Leistung des Arztes erhöht. Im weiteren Verlauf werden die Begriffe strategische Komplemente bzw. strategische Substitute verwendet (vgl. Bulow, Geanakoplos und Klemperer (1985)). Im Unterschied zu Komplementen und Substituten, die eine direkte Beziehung zwischen zwei Variablen beschreiben, beziehen sich die Begriffe strategische Komplemente und Substitute auf den Einfluß einer Variablen auf die Grenzproduktivität einer anderen Variablen (vgl. Bulow, Geanakoplos und Klemperer 1985, S. 494).¹⁰² Im vorliegenden Fall bedeutet dies, daß zusätzliche medizinische Leistung die Grenzproduktivität der gesundheitsfördernden Anstrengungen

¹⁰² Es gilt hierbei zu beachten, daß eine Erhöhung eines der beiden Faktoren nur die Wahrscheinlichkeit eines besseren Gesundheitszustands erhöht und daher im strengen Sinn nicht von Grenzproduktivität gesprochen werden kann. Im weiteren Verlauf wird zur Vereinfachung jedoch durchgängig der Begriff Grenzproduktivität verwendet.

erhöht und daß damit die Wahrscheinlichkeit einer Genesung steigt (strategische Komplemente). Der umgekehrte Fall besagt, daß durch gesundheitsfördernde Anstrengungen die Grenzproduktivität der medizinischen Leistung sinkt, da bspw. sportliche Betätigung die Wirkung einer Therapie u.U. verzögert.

Analog zur Steigung der Reaktionsfunktion des Patienten ergibt sich für den Arzt, daß die Veränderung der medizinischen Leistung aufgrund einer marginalen Erhöhung der Compliance durch folgenden Ausdruck gegeben ist:

$$\frac{d m}{d a} = - \frac{\varepsilon p_{am} (G_1 - G_0)}{\varepsilon p_{mm} (G_1 - G_0) - C''} . \quad (6.16)$$

Auch hier entspricht der Nenner der hinreichenden Bedingung für ein Nutzenmaximum des Arztes und ist daher negativ. Man erkennt, daß die Steigung in diesem Fall nur von der Existenz strategischer Komplemente bzw. Substitute abhängt. Die Steigung ist immer Null, unabhängig vom Vorzeichen der Kreuzableitung p_{am} , falls der Arzt den Gesundheitszustand des Patienten vernachlässigt ($\varepsilon=0$).

6.6.2 Resultate ohne Selbstbeteiligung

Ohne Berücksichtigung der Effekte der Einführung einer Selbstbeteiligung hängen die Steigungen der Reaktionsfunktionen lediglich davon ab, ob die Aktionen des Arztes und des Patienten die Eigenschaft strategischer Substitute oder Komplemente besitzen. Das bedeutet, daß für die Gleichung (6.15) folgt:

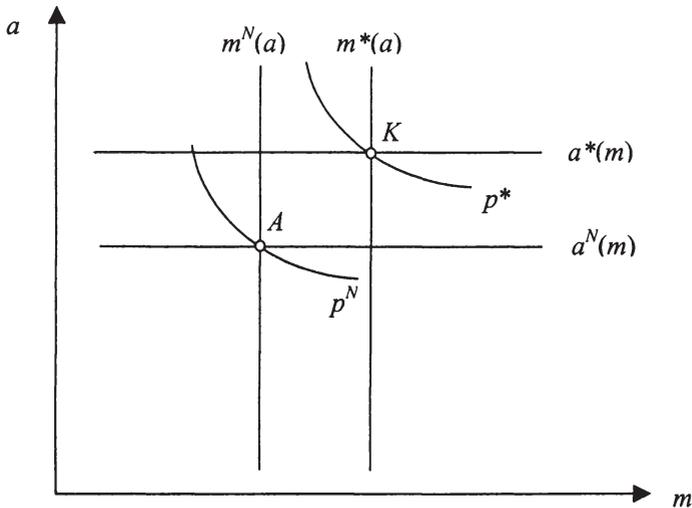
$$\frac{d a}{d m} = - \frac{p_{am} (U_1 - U_0)}{p_{aa} (U_1 - U_0) - D''} . \quad (6.17)$$

Gleichung (6.17) gibt die Steigung der Reaktionsfunktion des Patienten ohne Selbstbeteiligung wieder. Für den Arzt gilt unverändert die Reaktionsfunktion aus Gleichung (6.16), da eine Selbstbeteiligung des Patienten die Steigung seiner Reaktionsfunktion nicht beeinflusst:

$$\frac{d m}{d a} = - \frac{\varepsilon p_{am} (G_1 - G_0)}{\varepsilon p_{mm} (G_1 - G_0) - C''} . \quad (6.18)$$

Die Nenner beider Gleichungen sind negativ. Für den Fall, daß die Grenzproduktivitäten der ärztlichen Leistung und der Compliance vom jeweils anderen Faktor unabhängig sind ($p_{am} = 0$), ergibt sich in Abbildung 6.4, daß beide Reaktionsfunktionen unelastisch gegenüber den Aktionen des anderen Akteurs sind und damit für den Patienten horizontal, für den Arzt vertikal verlaufen.

Abbildung 6.4: Unabhängigkeit von Compliance und ärztlicher Leistung

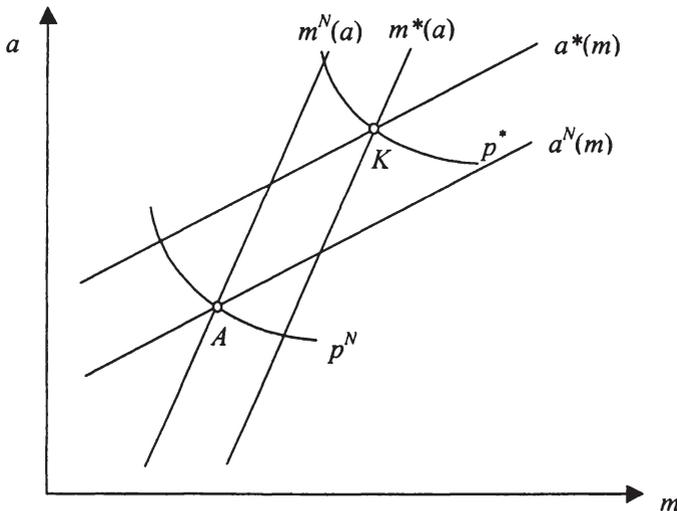


Quelle: Erstellt in Anlehnung an Cooper und Ross (1985).

Die Achsen bezeichnen die Menge an medizinischer Leistung m und die Anstrengung des Patienten a . Die Reaktionsfunktionen a^* und m^* kennzeichnen die First-best-Lösungen. Man erkennt im Vergleich zur Lösung bei Informationsasymmetrie a^N und m^N , daß für ein gegebenes Niveau von a bzw. m die gewählte Aktion jeweils höher liegt. Der Punkt K kennzeichnet die First-best-Lösung im Schnittpunkt der Reaktionsfunktionen. Im Vergleich mit der nichtkooperativen Lösung (Punkt A) fällt sowohl die medizinische Leistung als auch die Compliance höher aus. Die Kurven p^* und p^N bezeichnen die Iso-Wahrscheinlichkeitskurven für die First-best- und die nichtkooperative Lösung. Die negative Steigung erhält man durch totales Differenzieren von $p(a, m)$. Je weiter die Kurve vom Ursprung entfernt ist, desto höher ist c.p. das Niveau mindestens eines Inputs und desto höher ist die Wahrscheinlichkeit, gesund zu werden (vgl. Lanoie 1991, S. 87).

Falls die Erhöhung der medizinischen Leistung die Grenzproduktivität der Anstrengung des Patienten erhöht bzw. falls umgekehrt eine Erhöhung der Anstrengungen die Produktivität der medizinischen Leistung steigert, liegen strategische Komplemente vor ($p_{am} > 0$). In diesem Fall ist der Zähler in den Gleichungen (6.17) und (6.18) jeweils negativ, so daß sich in Verbindung mit dem negativen Wert des Nenners als Gesamteffekt eine Zunahme der Compliance des Patienten mit steigender medizinischer Leistung ergibt, d.h. ein höheres Niveau der medizinischen Leistung durch eine höhere Compliance erreicht wird. Die Steigungen beider Reaktionsfunktionen sind positiv, so daß dann $da/dm > 0$ und $dm/da > 0$. Dieser Zusammenhang ist aus Abbildung 6.5 ersichtlich.¹⁰³

Abbildung 6.5: Compliance und medizinische Leistung bei strategischen Komplementen



Quelle: Erstellt in Anlehnung an Cooper und Ross (1985).

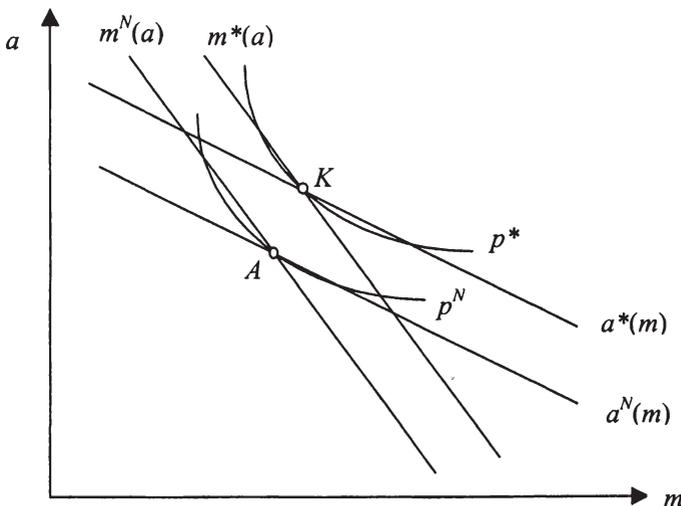
Die Geraden a^* und m^* bezeichnen wieder die Reaktionsfunktionen bei vollständiger Information. Die resultierende First-best-Lösung K zeigt im Vergleich zur nichtkooperativen Lösung A , daß durch die asymmetrische Information der Arzt eine geringere medizinische Leistung anbietet und die Compliance des

¹⁰³ Zur Vereinfachung sind im folgenden die Reaktionsfunktionen als Geraden eingezeichnet, was aber keine Rückschlüsse auf deren tatsächlichen Verlauf zuläßt. Dieser hängt von der Nutzenfunktion und der Dichte ab. Eine weitere Vereinfachung ist der parallele Verlauf der Reaktionsfunktionen bei kooperativer und nichtkooperativer Lösung.

Patienten sinkt. Die Anreizproblematik hat demnach zur Folge, daß die Akteure in Nichtkenntnis der Auswirkungen ihrer Handlungen ein im Vergleich zur vollkommenen Information zu niedriges Niveau wählen.

Strategische Substitute liegen dann vor, wenn die marginale Produktivität der Anstrengungen des Patienten mit steigender medizinischer Leistung sinkt ($p_{am} < 0$) und umgekehrt die Compliance mit zurückgehender medizinischer Leistung steigt. Letzteres ist beispielsweise dann der Fall, wenn sportliche Betätigung den Heilungsprozeß verlangsamt. Für die Steigungen der Reaktionsfunktionen folgt, daß dann jeweils der Zähler in den Gleichungen (6.17) und (6.18) positiv ist, so daß sich als Gesamteffekte $da/dm < 0$ und $dm/da < 0$ ergeben. Die Steigung der Reaktionsfunktionen ist folglich negativ. Die medizinische Leistung des Arztes und die Anstrengungen des Patienten sind daher Substitute. Im Gegensatz zu den Fällen der strategischen Unabhängigkeit und der strategischen Komplemente sind die resultierenden Niveaus der medizinischen Leistung und der Anstrengung des Patienten hier nicht eindeutig. Abbildung 6.6 und Abbildung 6.7 veranschaulichen dieses Ergebnis.

Abbildung 6.6: Gleichgerichtete Effekte bei strategischen Substituten



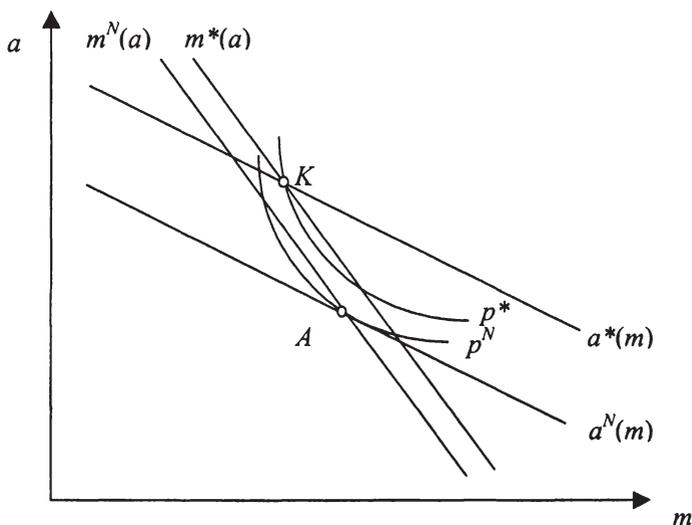
Quelle: Erstellt in Anlehnung an Cooper und Ross (1985).

Im ersten Fall (Abbildung 6.6) kommt es durch die Informationsasymmetrie zu einer Reduktion beider Niveaus. Die Reaktionsfunktionen bei der nichtkoopera-

tiven Lösung liegen unterhalb derer bei vollkommener Information. Durch die asymmetrische Information kommt es zu der bereits erwähnten Verschiebung. Diese Ergebnisse sind analog zum Fall der strategischen Komplemente.

Ein abweichendes Ergebnis ist in Abbildung 6.7 dargestellt. Hier liegen zwar die Reaktionsfunktionen bei asymmetrischer Information wiederum unter denen bei vollständiger Information, doch während es zu einer Absenkung der Compliance kommt, steigt das Niveau der medizinischen Leistung über das First-best-Niveau an (vgl. hierzu die Punkte K und A). Die Erklärung für dieses Ergebnis ist intuitiv (vgl. hierzu Cooper und Ross 1985, S. 108f. und Yavas 1995, S. 252f.). Ausgehend von der kooperativen Lösung ergibt sich als erster Effekt durch die Informationsasymmetrie ein Rückgang beider Inputs bei gegebenen Anstrengungen der anderen Partei. Dies betrifft die angesprochene Verschiebung der Reaktionsfunktionen.

Abbildung 6.7: Gegenläufige Effekte bei strategischen Substituten



Quelle: Erstellt in Anlehnung an Cooper und Ross (1985).

Als zweiter Effekt führt eine Reduktion bspw. der Compliance des Patienten dazu, daß sich die medizinischen Leistungen produktiver auf die Wahrscheinlichkeit eines höheren Gesundheitszustands auswirken. Daher wird der Arzt seine Behandlungsleistung erhöhen. Die gleiche Logik gilt auch umgekehrt für die Compliance, die bei Reduktion der medizinischen Leistung ebenfalls pro-

duktiver wirkt. Es ist nun möglich, daß der Effekt höherer Grenzproduktivitäten den ersten Effekt, der durch die Informationsasymmetrie entsteht, für entweder den Arzt oder den Patienten dominiert. Dies gilt jedoch nicht für beide Akteure. Das First-best-Niveau kann somit für den Arzt oder den Patienten, sogar überschritten werden, jedoch nicht für beide. Erbringt der Arzt als Folge mehr medizinische Leistung als bei vollständiger Information, kann dies als alternative Interpretation der anbieterinduzierten Nachfrage betrachtet werden.¹⁰⁴

6.6.3 Zur Wirkung einer Selbstbeteiligung¹⁰⁵

Die aus Sicht des Patienten optimale medizinische Leistung wird nun durch die ihm entstehenden Kosten beschränkt. Die Ableitung seines Erwartungsnutzens nach der medizinischen Leistung ergibt:

$$p_m (U_1 - U_0) = \beta [p U_1' + (1-p) U_0'] \quad (6.19)$$

Für $\beta = 0$ ist die rechte Seite gleich Null. Der Patient würde theoretisch die medizinische Leistung unendlich ausdehnen, da sie ihm keine Kosten verursacht. Die Selbstbeteiligung führt nun dazu, daß die optimale Menge aus der Sicht des Patienten begrenzt ist.

Die Einführung einer Selbstbeteiligung beeinflusst daher das resultierende Nash-Gleichgewicht. Es ergibt sich, daß die Steigung der Reaktionsfunktion für den Patienten nicht mehr in allen Fällen eindeutig bestimmt werden kann. Wie Gleichung (6.15) zeigt, hängt die Steigung mit der Ausprägung des Selbstbeteiligungsanteils und dem Grad der strategischen Komponente zusammen.

Es ist ersichtlich, daß neben einer positiven Selbstbeteiligungskomponente auch der Nutzen und der Grenznutzen des Patienten von einer Einführung der Selbstbeteiligung betroffen sind. Im weiteren Verlauf wird jedoch zur Vereinfachung der Analyse angenommen, daß das Nettoeinkommen des Patienten nach Einführung der Selbstbeteiligung unverändert bleibt.¹⁰⁶ Diese Annahme erlaubt den

¹⁰⁴ Allerdings wird in diesem Modell eine steigende Arztdichte nicht betrachtet.

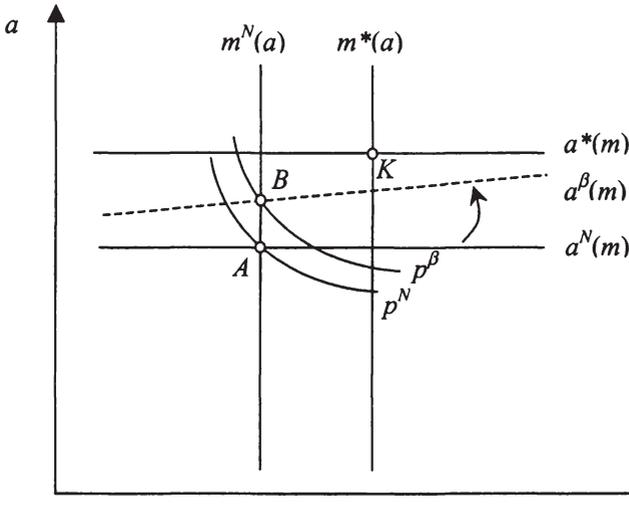
¹⁰⁵ Es wird an dieser Stelle nur der Effekt einer Selbstbeteiligung auf der Patientenseite betrachtet, da aufgrund der Risikoneutralität des Arztes der Erstattungsparameter δ lediglich linear in die Bedingung erster Ordnung des Arztes (Gleichung (6.11)) eingeht bzw. keine Auswirkung auf die Steigung seiner Reaktionsfunktion hat (Gleichung (6.16)).

¹⁰⁶ Durch die Selbstbeteiligung verringert sich die Prämie, so daß das Nettoeinkommen nicht um den vollen Betrag der Selbstbeteiligung zurückgeht. Allerdings bleibt zu beachten, daß ohne eine Einkommenswirkung die Selbstbeteiligung keine Anreize auf die Nachfrage nach medizinischen Leistungen hätte.

direkten Vergleich mit den im vorherigen Abschnitt abgeleiteten Ergebnissen. Weiterhin ergibt Gleichung (6.9), daß c.p. die Anstrengungen des Patienten bei einem höheren Selbstbeteiligungssatz steigen, so daß die neue Reaktionsfunktion für alle Werte von m oberhalb der Reaktionsfunktion ohne Selbstbeteiligung verläuft.

Im Fall der Unabhängigkeit der Anstrengungen von der medizinischen Leistung ($p_{am} = 0$) hat die Einführung einer Selbstbeteiligung nun eine positive Steigung der Reaktionsfunktion des Patienten zur Konsequenz (gestrichelte Linie $a^\beta(m)$ in Abbildung 6.8). Aus Sicht des Patienten liegt nun eine komplementäre Beziehung zwischen der Leistung des Arztes und der Compliance des Patienten vor (vgl. Abbildung 6.8).

Abbildung 6.8: Unabhängigkeit und Selbstbeteiligung

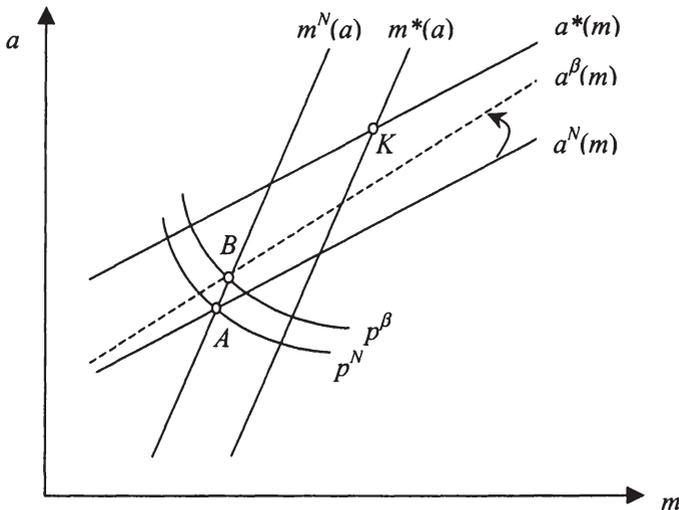


Eine Erhöhung der medizinischen Leistung führt somit auch zu einer Erhöhung der Compliance. Für jedes Niveau der ärztlichen Leistung erbringt der Patient bei Vorliegen einer Selbstbeteiligung höhere Anstrengungen, so daß sich die Wahrscheinlichkeit eines besseren Gesundheitszustands erhöht. Anhand der Iso-Wahrscheinlichkeitskurven erkennt man, daß die Wahrscheinlichkeit der Genesung durch die höhere Compliance steigt (p^β im Vergleich zu p^N). Falls der Erwartungsnutzen aus dem Einkommen für den Patienten stärker zunimmt als der Aufwand durch die höhere Compliance, erhöht sich sein gesamter Erwartungsnutzen durch die Einführung der Selbstbeteiligung. Als Ergebnis folgt,

daß durch die Selbstbeteiligung die Anstrengungen des Patienten bei gleichbleibender medizinischer Leistung steigen.

Liegen strategische Komplemente vor ($p_{am} > 0$), so resultiert aus der Einführung einer Selbstbeteiligung, daß eine Erhöhung der medizinischen Leistung eine höhere Compliance des Patienten bedingt. Die Wirkung der strategischen Komponente wird durch die Selbstbeteiligungskomponente noch verstärkt. Die Reaktionsfunktion des Patienten verläuft über derjenigen der nichtkooperativen Lösung. Folglich liegen bei gegebenem Niveau der medizinischen Leistung seine Anstrengungen über denen ohne Selbstbeteiligung (gestrichelte Linie).¹⁰⁷ In Abbildung 6.9 erkennt man den geschilderten Effekt.

Abbildung 6.9: Strategische Komplemente und Selbstbeteiligung



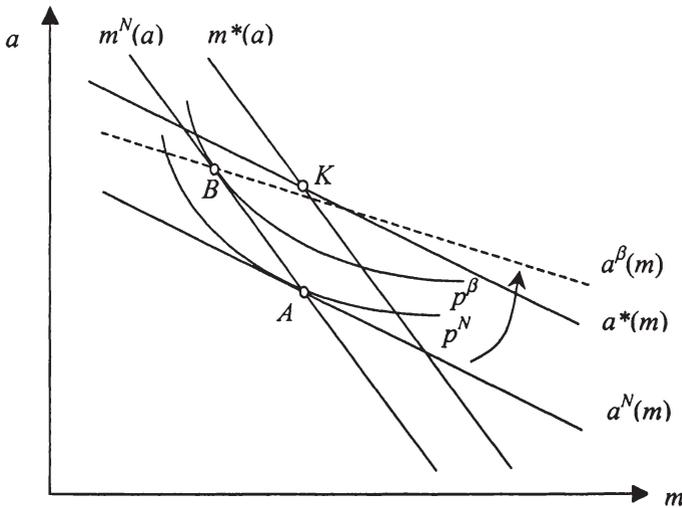
Die gestrichelte Linie gibt nun den neuen Verlauf der Reaktionsfunktion des Patienten wieder, während die Reaktionsfunktion des Arztes unverändert bleibt. Es ist ersichtlich, daß im Schnittpunkt mit der Reaktionsfunktion m^N (Punkt B) sowohl das Niveau der medizinischen Leistung als auch das der Compliance des Patienten über dem Niveau ohne Selbstbeteiligung liegt. Mit anderen Worten: In dem Fall der strategischen Komplemente führt eine Einführung der

¹⁰⁷ Die tatsächliche Steigung der Reaktionsfunktion a^β ist für die qualitativen Ergebnisse irrelevant. Die in Abbildung 6.9 und Abbildung 6.10 dargestellte Änderung der Steigung steht stellvertretend für die Auswirkungen der Selbstbeteiligung auf die strategische Komponente.

Selbstbeteiligung des Patienten zu einem höheren Niveau der Anstrengungen und der medizinischen Leistungen, wodurch wiederum die Wahrscheinlichkeit einer Genesung erhöht wird. Dies erkennt man an der Iso-Wahrscheinlichkeitskurve p^β , die ein höheres Niveau aufweist als im Nash-Gleichgewicht ohne Selbstbeteiligung (p^N). Der Erwartungsnutzen des Patienten nimmt auch in diesem Fall zu, wenn der Erwartungsnutzen des Nettoeinkommens stärker ansteigt als der Aufwand des Patienten. Welcher der beiden Inputs der Gesundheitsproduktion stärker zunimmt, hängt dabei letztlich von den Steigungen der Reaktionsfunktionen ab.

Anders verhält es sich im Fall der strategischen Substitute, bei dem eine Erhöhung der medizinischen Leistung die Grenzproduktivität der Compliance absenkt. Hierbei führt die Einführung einer Selbstbeteiligung für den Patienten u.U. zu einem flacheren Verlauf der Reaktionsfunktion (vgl. Abbildung 6.10).

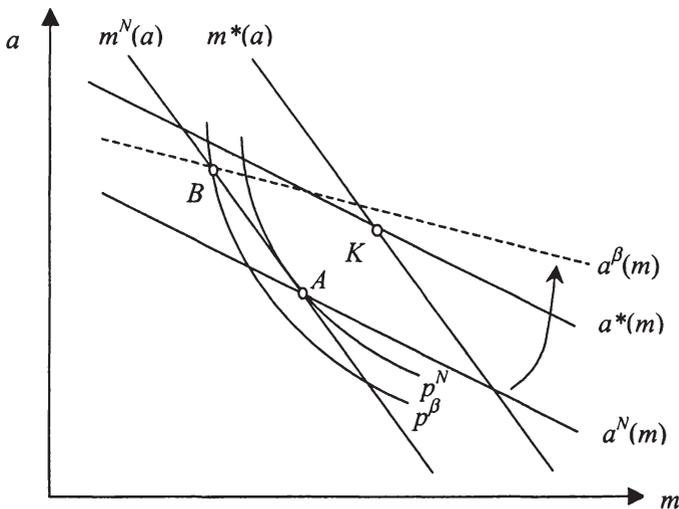
Abbildung 6.10: Strategische Substitute und erhöhte Genesungswahrscheinlichkeit



Bei gegebener medizinischer Leistung des Arztes sind dann c.p. die Anstrengungen des Patienten höher. Dabei ist es möglich, daß die Wirkung der strategischen Komponente durch die Selbstbeteiligungskomponente überlagert wird und die Reaktionsfunktion sogar einen steigenden Verlauf hat. Es ist ersichtlich, daß durch die neue Lage der Reaktionsfunktion des Patienten (Linie $a^\beta(m)$) die medizinische Leistung des Arztes zurückgeht, während hingegen die Com-

pliance zunimmt (Punkt B). Hierbei ist es möglich, daß die Anstrengungen des Patienten das First-best-Niveau übersteigen. Durch die Einführung einer Selbstbeteiligung im Fall der strategischen Substitute wird die medizinische Leistung durch die Compliance des Patienten substituiert. Als Wohlfahrtseffekt für den Patienten ergibt sich auch in diesem Fall, daß die Genesungswahrscheinlichkeit steigt. Der Erwartungsnutzen des Patienten erhöht sich, wenn der Erwartungsnutzen des Nettoeinkommens stärker ansteigt als der Aufwand des Patienten. Allerdings hängt dieses Ergebnis von der aus der Gesundheitsproduktionsfunktion resultierenden Iso-Wahrscheinlichkeitskurve ab (siehe hierzu Abbildung 6.11).

Abbildung 6.11: Strategische Substitute und verminderte Genesungswahrscheinlichkeit



Grundsätzlich hängt die Grenzrate der Substitution zwischen Compliance und medizinischer Leistung von den Grenzproduktivitäten der beiden Inputs ab und wird durch die Steigung der Isoquanten dargestellt. Bei einem steilen Verlauf der Iso-Wahrscheinlichkeitskurve kann der Fall eintreten, daß die Genesungswahrscheinlichkeit sinkt. Somit können der Anstieg der Compliance und die damit verbundene Erhöhung der Wahrscheinlichkeit den Rückgang der Genesungswahrscheinlichkeit durch die geringeren medizinischen Leistungen nicht kompensieren. Damit besteht die Möglichkeit, daß der Erwartungsnutzen des Patienten sogar zurückgehen kann.

Zusammenfassend läßt sich für alle Formen der strategischen Komponente feststellen, daß sich die Compliance des Patienten durch die Einführung einer Selbstbeteiligung erhöht. Die intuitive Begründung hierfür liegt darin, daß sich sowohl die Anstrengungen des Patienten als auch die medizinische Leistung des Arztes positiv auf die Genesungswahrscheinlichkeit auswirken. Könnte zuvor die medizinische Leistung zu Grenzkosten von Null konsumiert werden, so erhöhen sich die Behandlungskosten durch die Einführung einer Selbstbeteiligung. Die Compliance wird daher im Vergleich zur medizinischen Behandlung relativ billiger. Die Änderung des Erwartungsnutzens des Patienten hängt neben der Compliance und der medizinischen Leistung auch von der Genesungswahrscheinlichkeit ab. Genaue Aussagen sind daher nicht möglich, jedoch deutet eine zunehmende Genesungswahrscheinlichkeit auf einen höheren Erwartungsnutzen hin.

6.7 Die Vertragsstufe

Für die Vertragsgestaltung sind die Ergebnisse der nichtkooperativen Lösung auf der Behandlungsstufe relevant, da diese von der Versicherung bei Vertragsabschluß berücksichtigt werden müssen. Geschieht dies nicht, ergeben sich aus den vereinbarten Versicherungs- und Vergütungsparametern u.U. fehlergerichtete Anreize. Für den zeitlichen Ablauf unter Berücksichtigung der Vertragsgestaltung werden folgende Annahmen getroffen: Die Natur wählt den Gesundheitszustand des Patienten. Mit der Wahrscheinlichkeit π ist er krank, mit Wahrscheinlichkeit $1-\pi$ bleibt er gesund.¹⁰⁸ Dabei wird unterstellt, daß diese Wahrscheinlichkeiten allgemein bekannt sind und Versicherung, Arzt sowie Patient keinen Informationsvorsprung besitzen. Im Krankheitsfall sucht der Patient den Arzt zwecks einer Behandlung auf. Die Versicherung maximiert den Erwartungsnutzen des Patienten über die Versicherungsparameter Selbstbeteiligungssatz β , Fallpauschale des Arztes ω und Kostenbeteiligung δ .¹⁰⁹ Die Prämie errechnet sich dann aus der Nullgewinnbedingung der Versicherung. Die Maximierung erfolgt unter verschiedenen Nebenbedingungen, und zwar der Nullgewinnbedingung der Versicherung, die Teilnahmebedingung des Arztes sowie die Anreizbedingungen des Arztes und des Patienten, die sich aus der Behandlungsstufe ergeben.

¹⁰⁸ Es wird hierbei zur Vereinfachung angenommen, daß das Verhalten des Patienten nicht die Wahrscheinlichkeit zu erkranken beeinflusst, d.h., der Fall des ex ante Moral Hazard wird nicht untersucht.

¹⁰⁹ Analog zu Stewart (vgl. 1994, S. 196) maximiert der Versicherer den Erwartungsnutzen des Patienten unter Nebenbedingungen.

Die Erwartungsnutzenfunktion des Patienten besteht aus den beiden Komponenten Erwartungsnutzen aus der Behandlung, gewichtet mit der Wahrscheinlichkeit zu erkranken und dem Erwartungsnutzen, falls keine Krankheit auftritt und er nicht den Arzt aufsucht:

$$EU = \pi \left\{ p(a, m) U(y [G_1] - \sigma - \beta m) + (1 - p(a, m)) U(y [G_0] - \sigma - \beta m) - D(a) \right\} + (1 - \pi) U(y [G_1] - \sigma). \quad (6.20)$$

Dabei bezeichnet (wie bereits in Abschnitt 6.5 dargestellt) G_1 den Zustand gesund, während G_0 einen schlechten Gesundheitszustand widerspiegelt, so daß $G_1 > G_0$. Zur Vereinfachung bezeichnet im weiteren Verlauf U_0 den resultierenden Nutzen eines schlechten Gesundheitszustands nach der Behandlung, U_1 den Nutzen eines guten Gesundheitszustands nach einer Behandlung und U_2 den Nutzen ohne eine Behandlung, mit $U_0 < U_1 < U_2$:

$$\begin{aligned} U_0 &= U(y [G_0] - \sigma - \beta m), \\ U_1 &= U(y [G_1] - \sigma - \beta m), \\ U_2 &= U(y [G_1] - \sigma). \end{aligned}$$

Bei der Nullgewinnbedingung der Versicherung entspricht die erhaltene Prämie den erwarteten Behandlungsausgaben. Diese setzen sich zusammen aus dem ärztlichen Honorar (ω und δm) abzüglich der Selbstbeteiligung, die der Patient zu tragen hat βm , multipliziert mit der Wahrscheinlichkeit eines Krankheitsfalls und damit der Behandlung:

$$\sigma = \pi (\omega + [\delta - \beta] m). \quad (6.21)$$

Die Teilnahmebedingung des Arztes besagt, daß im Falle einer Behandlung der Arzt mindestens seinen Reservationsnutzen \bar{V} erhalten muß, damit er dem Vertrag zustimmt:

$$\pi \left([\omega + \delta m] + \varepsilon [p G_1 + [1 - p] G_0] - C[m] \right) \geq \bar{V}. \quad (6.22)$$

In diesem Zusammenhang steht ω für eine Fallpauschale, die der Arzt für die Behandlung erhält. Weiterhin existiert eine von den entstandenen Kosten abhängige Komponente δm , wobei der Parameter δ sowohl positiv sein kann, was einer Einzelleistungsvergütung entspricht, als auch negativ, in Form einer Kostenbeteiligung. Dem Arzt entsteht durch die medizinische Behandlung des Patienten Aufwand in Höhe von $C(m)$, wobei die Funktion konvex in der medizinischen Leistung ist. Weiterhin wird für den Arzt eine Berufsethik unterstellt, die dadurch gekennzeichnet ist, daß eine Verbesserung des erwarteten Gesundheitszustands des Patienten den Erwartungsnutzen des Arztes positiv beeinflusst.

Falls weder auf der Seite des Arztes noch auf der Seite des Patienten Moral Hazard vorliegt, besteht das Optimierungsproblem aus den Gleichungen (6.20), (6.21) und (6.22). Um die Effekte der asymmetrischen Information bei Nicht-Kooperation zu berücksichtigen, ist es daher notwendig, die Anreize der geschlossenen Verträge auf die individuellen Handlungen des Arztes und des Patienten in den Ansatz mit einzubeziehen. Im einzelnen bedeutet dies, daß für den Arzt zu berücksichtigen ist, welchen Einfluß die Wahl der Vertragsparameter auf seine optimale Entscheidung über die medizinische Leistung hat. Die Anreizbedingung des Arztes ergibt sich aus der Maximierung seines Erwartungsnutzens über die medizinische Leistung als Bedingung erster Ordnung:

$$\delta + p_m(a, m) \varepsilon(G_1 - G_0) - C' = 0. \quad (6.23)$$

Äquivalent gilt für die Anreizbedingung des Patienten, daß auch die Auswirkung des Vertragsabschlusses auf die Wahl der Compliance abzubilden ist. Die Bedingung erster Ordnung für den Patienten erhält man durch die Maximierung seines Erwartungsnutzens bezüglich seiner Compliance:

$$p_a(a, m) (U_1 - U_0) - D' = 0. \quad (6.24)$$

Aus diesen Bedingungen erster Ordnung, die gleichzeitig die Reaktionsfunktionen bezüglich der Aktion des anderen Spielers angeben, lassen sich folgende Zusammenhänge bezüglich der Vertragsparameter herleiten:

$$a = a(\beta, \delta),$$

$$m = m(\delta).$$

Man erkennt, daß das Niveau der Compliance von der Selbstbeteiligung abhängt (siehe auch Gleichung (6.9)). Weiterhin beeinflusst der Kostenerstattungssatz des

Arztes das Verhalten des Patienten, da er direkt auf das Niveau der medizinischen Leistung und indirekt auf den Nutzen durch die Selbstbeteiligung des Patienten wirkt (Gleichung (6.12)). Dagegen ergibt sich für die medizinische Leistung, daß lediglich der Kostenerstattungssatz für die Höhe eine Rolle spielt. Die Fallpauschale geht nicht ein, da der Arzt risikoneutral im Einkommen ist. Der Selbstbeteiligungssatz spielt direkt ebenfalls keine Rolle, da nicht das Einkommen des Patienten, sondern lediglich der Gesundheitszustand in die Nutzenfunktion des Arztes eingeht.¹¹⁰

6.7.1 First-best-Lösung

Die Berücksichtigung der Gleichungen (6.23) und (6.24), den Anreizbedingungen für Arzt und Patient, entfällt, da die Wahl der medizinischen Leistung und der Compliance von der Versicherung beobachtet werden kann. Im Fall der vollständigen Information besteht das Maximierungsproblem somit aus den Gleichungen (6.20) bis (6.22). Unter Berücksichtigung der Tatsache, daß sich die Versicherungsprämie des Patienten durch die Wahl der Selbstbeteiligung, der Fallpauschale und der Kostenerstattung des Arztes ergibt, läßt sich das Problem durch Einsetzen von Gleichung (6.21) in die Zielfunktion des Patienten zu der in Anhang 6.A dargestellten Lagrange-Funktion (A.1) umformulieren.¹¹¹

Die notwendigen Bedingungen ergeben sich durch partielles Ableiten nach den Entscheidungsvariablen Compliance und medizinische Leistung sowie den Vertragsparametern. Bezüglich des gesundheitsrelevanten Verhaltens des Patienten a folgt:

$$\frac{\partial L}{\partial a} : \quad \pi \left[p_a (U_1 - U_0) \right] - \pi D' + \pi \lambda p_a \varepsilon (G_1 - G_0) = 0. \quad (6.25)$$

Der erste Teil $\pi [\dots]$ gibt den erwarteten Grenznutzen einer höheren Compliance an. Der zweite Term ist der erwartete Grenzaufwand, und der dritte ist Ausdruck des erwarteten Grenznutzens des Arztes durch seine Berufsethik. Analog zu Zweifel (1994) bezeichnet der Lagrange-Parameter λ die marginale Änderung

¹¹⁰ Auf indirekte Effekte der Vertragsparameter auf die Compliance oder die medizinische Leistung über die strategischen Komponenten (vgl. Abschnitt (6.5)) wird an dieser Stelle nicht näher eingegangen.

¹¹¹ Diese Funktion läßt sich auch wie folgt interpretieren: λ sei nicht mehr der Lagrange-Parameter, sondern eine fixe Gewichtung des Erwartungsnutzens des Arztes (vgl. Rees 1985, S. 21). In diesem Fall wird nach einem Pareto-Optimum unter Beachtung der Anreizbedingungen gesucht. Eine mögliche Interpretation ist, daß der Patient und der Arzt effizient über die Verträge verhandeln.

der Zielfunktion durch die marginale Änderung der Nebenbedingung. Dieser Teilnahmeeffekt spiegelt die Opportunitätskosten des Arztes durch seinen Reservationsnutzen bzw. den Schattenpreis des Einkommens des Arztes wider. Im Vergleich zu Gleichung (6.24) erkennt man, daß bei kooperativer Lösung (First-best) die Nutzenwirkungen einer höheren Compliance auf den Arzt mit in die Überlegungen einbezogen werden. Daraus folgt, daß sich das Niveau der Compliance bei Kooperation erhöht, wenn der erwartete gemeinsame Grenznutzen größer ist als der erwartete Grenznutzen des Patienten bei Nicht-Kooperation (vgl. Schneider 1999, S. 14f.).

Bezüglich der partiellen Ableitung nach der medizinischen Leistung des Arztes folgt:

$$\frac{\partial L}{\partial m} : \quad \pi p_m (U_1 - U_0) + \pi [p U_1' (\beta \pi - \beta) + (1 - p) U_0' (\beta \pi - \beta)] + (1 - \pi) U_2' (\beta \pi) + \lambda \pi [\delta + p_m \varepsilon (G_1 - G_0) - C'] = 0. \quad (6.26)$$

Die ersten drei Ausdrücke geben die Nutzenwirkungen einer marginalen Erhöhung der medizinischen Leistung auf den Patienten an. Der erste kennzeichnet den erwarteten direkten Grenznutzen durch eine Erhöhung der medizinischen Leistung. Die beiden anderen Ausdrücke stellen den indirekten Prämien- und Selbstbeteiligungseffekt der Ausdehnung und die damit verbundenen erwarteten Grenzkosten dar, einmal für den Fall der Behandlung sowie für den Fall ohne Behandlung. Der vierte Term entspricht dem direkten erwarteten Effekt auf den Arzt, bestehend aus dem Grenzertrag einer zusätzlichen Einheit medizinischer Leistung δ , dem Grenznutzen der Berufsethik sowie dem mit der medizinischen Leistung verbundenen Grenzaufwand. Bei Nicht-Kooperation besteht die Bedingung erster Ordnung lediglich aus dem letzten Ausdruck (vgl. Gleichung (6.23)). Falls der erwartete Grenznutzen bei Kooperation abzüglich der erwarteten marginalen Selbstbeteiligung größer ist als der erwartete Grenznutzen bei Nicht-Kooperation, so ist die medizinische Leistung größer.

Für die Höhe des Selbstbeteiligungssatzes läßt sich die Bedingung erster Ordnung (A.4) umformen zu:

$$U_2' = p U_1' + (1 - p) U_0'. \quad (6.27)$$

Gleichung (6.27) besagt, daß die Differenz zwischen dem Grenznutzen ohne Behandlung und dem erwarteten Grenznutzen aus der Behandlung gleich Null ist. Dies ist nur dann der Fall, wenn der Patient durch die Behandlung vollstän-

dig gesundet ($p = 1$) und darüber hinaus keine Ausgaben für die medizinische Leistungen tätigen muß ($\beta = 0$). Letzteres steht für den Fall der vollständigen Versicherung, in dem der Patient keine Zuzahlungen zur Behandlung zu entrichten hat. Da das Verhalten des Patienten beobachtet werden kann, tritt Moral Hazard in diesem Fall nicht auf, und somit ist Vollversicherung optimal. Folglich erreicht der Patient unabhängig von seinem Gesundheitszustand immer dasselbe Nutzenniveau durch die Vollversicherung.

Aus der Bedingung (A.5) für die Fallpauschale des Arztes folgt:

$$\pi [p U'_1 + (1-p) U'_0] + (1-\pi) U'_2 = \lambda. \quad (6.28)$$

Die rechte Seite zeigt den direkten Effekt einer Erhöhung der Fallpauschale und ist hier konstant, da der Arzt risikoneutral im Einkommen und daher der Grenznutzen konstant ist. Die linke Seite von Gleichung (6.29) gibt den indirekten Effekt auf den Patienten durch die Erhöhung der Prämie an. Im Optimum wird die Fallpauschale gewählt, bei der die erwarteten Grenzkosten für den Patienten dem Grenznutzen des Einkommens des Arztes entsprechen. Man erkennt, daß der Lagrange-Parameter λ positiv ist, da eine Erhöhung des Reservationsnutzens des Arztes sich negativ auf die Zielfunktion auswirkt und damit die Nebenbedingung bindend ist.

Aus Gleichung (A.6) für die Kostenerstattung des Arztes folgt:

$$\pi [p U'_1 + (1-p) U'_0] + (1-\pi) U'_2 = \lambda. \quad (6.29)$$

Auch hier ist ersichtlich, daß die Kostenerstattung so gewählt wird, daß die erwarteten marginalen Auswirkungen einer Erhöhung auf den Patienten konstant sind. Die linke Seite spiegelt den direkten Effekt einer Erhöhung der Kostenerstattung für den Patienten durch eine Erhöhung der Prämie wieder. Die rechte Seite zeigt den Effekt für den Arzt auf. Der Grenzertrag einer Erhöhung ist dabei abhängig vom Niveau der medizinischen Leistung (vgl. Gleichung (A.6)).

6.7.2 Lösung bei asymmetrischer Information

Das Maximierungsproblem besteht in diesem Fall aus den Gleichungen (6.20), (6.21) und (6.22) unter Berücksichtigung des Einflusses der Vertragsparameter auf die Wahl der Aktionen a und m bei asymmetrischer Information (Gleichungen (6.23) und (6.24)). Da sich die Versicherungsprämie des Patienten aus der Wahl der Selbstbeteiligung und der Kostenerstattung des Arztes ergibt, läßt sich

das Problem durch Einsetzen von Gleichung (6.21) in die Zielfunktion des Patienten zu der Lagrange-Funktion (B.1) im Anhang umformulieren.

Die notwendigen Bedingungen ergeben sich durch partielles Ableiten nach der medizinischen Leistung, der Compliance und nach den Vertragsparametern.¹¹²

Für den Selbstbeteiligungssatz folgt aus Gleichung (B.2) nach geeigneter Umformung unter Berücksichtigung von Gleichung (6.24):

$$U_2' = p U_1' + (1 - p) U_0' - \frac{a_\beta}{m(1 - \pi)} \lambda \varepsilon p_a [G_1 - G_0]. \quad (6.30)$$

Im Gegensatz zu Gleichung (6.27) ist aus Gleichung (6.30) ersichtlich, daß der Grenznutzen ohne Behandlung und der erwartete Grenznutzen mit Behandlung nun nicht mehr identisch sind. Dies bedeutet, daß der Patient nicht vollständig versichert ist und daß ein positiver Selbstbeteiligungssatz vorliegt, da der Grenznutzen ohne Behandlung kleiner als bei Behandlung ist und daraus aufgrund der Konkavität von U ein höheres Nutzenniveau resultiert. Dies wird durch den dritten Ausdruck auf der rechten Seite der Gleichung (6.30) dargestellt, der die Wirkung des Double Moral Hazard wiedergibt. Er besitzt ein positives Vorzeichen und wird vom Erwartungsnutzen des Patienten aus der Behandlung subtrahiert. Der Ausdruck berücksichtigt den Grenznutzen einer Erhöhung des Selbstbeteiligungssatzes auf die Berufsethik des Arztes. Dieser ist positiv, da ein höherer Selbstbeteiligungssatz die Compliance erhöht und diese dann die Genesungswahrscheinlichkeit steigert. Der Nenner ist ebenfalls positiv. Die Wirkung der asymmetrischen Information ist um so größer, je kleiner das Niveau der medizinischen Leistung und je kleiner die Wahrscheinlichkeit ist, gesund zu bleiben bzw. je größer die Wahrscheinlichkeit eines Arztbesuchs ist. Die Erklärung hierfür ist, daß eine Selbstbeteiligung nur anfällt, wenn es zu einer Behandlung kommt. Eine höhere medizinische Leistung reduziert den negativen Effekt der asymmetrischen Information, da so das Niveau der Genesungswahrscheinlichkeit näher an dem Niveau bei vollständiger Information liegt und somit der erwartete Nutzen aus der Behandlung für den Patienten ansteigt.

Die Bedingung erster Ordnung für die Fallpauschale des Arztes (B.3) lautet nach Umformung:

¹¹² Die Ableitungen nach der medizinischen Leistung sowie nach der Compliance sind durch die Gleichungen (6.23) und (6.24) gegeben.

$$\pi [p U_1' + (1-p) U_0'] + (1-\pi) U_2' = \lambda. \quad (6.31)$$

Gleichung (6.31) ist identisch mit Gleichung (6.27), der Optimalbedingung für die Fallpauschale bei vollständiger Information, da die Fallpauschale die Wahl der Aktionen nicht beeinflusst. Die Gleichung gibt den erwarteten Grenznutzen aus einer Erhöhung der Fallpauschale wieder. Dieser ist konstant und besteht aus dem erwarteten Grenznutzen sowohl mit als auch ohne Behandlung. In beiden Fällen resultiert die Veränderung des Erwartungsnutzens durch den indirekten Effekt aus der Erhöhung der Versicherungsprämie. Der Lagrange-Parameter λ resultiert aus dem konstanten Grenzertrag einer erhöhten Fallpauschale auf das Nettoeinkommen und den Nutzen des Arztes.

Die dritte partielle Ableitung der Lagrange-Funktion ergibt die Bedingung erster Ordnung für die Kostenerstattung des Arztes (B.4), und durch Umformung unter Berücksichtigung der Gleichungen (6.23) und (6.24) ergibt sich:

$$\pi [p U_1' + (1-p) U_0'] + (1-\pi) U_2' = \frac{1}{E} [p_m m_\delta (U_1 - U_0) + \lambda m + \lambda \varepsilon p_a \alpha_\delta (G_1 - G_0) - SB], \quad (6.32)$$

mit $E := (\delta - \beta) m_\delta + m,$

$$SB := \beta m_\delta [p U_1' + (1-p) U_0'].$$

Die linke Seite ist identisch mit der linken von Gleichung (6.29). Während bei jener aber der erwartete Grenznutzen des Einkommens konstant ist, ist dies in Gleichung (6.32) nicht der Fall. Der Ausdruck auf der rechten Seite kann größer, kleiner oder gleich dem Parameter λ sein, je nachdem, welcher der nachfolgend erklärten Effekte überwiegt. Der erste Term in der Klammer entspricht dem Erwartungsgrenznutzen des Patienten. Dieser verändert sich durch den Anstieg der Wahrscheinlichkeit eines guten Gesundheitszustands aufgrund des Effektes der gestiegenen Kostenerstattung δ auf die medizinische Leistung. Der zweite Ausdruck gibt den Lagrange-Parameter der Nebenbedingung wieder, multipliziert mit der medizinischen Leistung. Der dritte Ausdruck ist der Effekt einer höheren Kostenerstattung auf den erwarteten ärztlichen Grenznutzen der Berufsethik, resultierend aus einem Anstieg der Compliance. Von den ersten drei Teilen wird der Ausdruck SB subtrahiert, der die marginalen Kosten einer gestiegenen absoluten Selbstbeteiligung des Patienten β in Nutzeinheiten widerspiegelt. Für den Ausdruck in eckigen Klammern wird angenommen, daß er immer posi-

tiv ist, daß also die positiven Nutzenwirkungen die marginalen Kosten durch die Selbstbeteiligung des Patienten übersteigen. Dieser Ausdruck wird durch den Term E geteilt, der den marginalen Effekt einer höheren Kostenerstattung auf die Prämie angibt.¹¹³ Der Gesamteffekt der asymmetrischen Information ist a priori unklar. Je nachdem, welche Einflüsse überwiegen, kann es sein, daß der Erwartungsnutzen des Patienten im Vergleich zur Situation bei vollständiger Information zu- oder abnimmt. So nimmt der erwartete Grenznutzen mit zunehmender Produktivität sowohl der medizinischen Leistung als auch der Compliance zu, hingegen mit zunehmender absoluter Selbstbeteiligung ab. Für den Fall, daß der Ausdruck SB hinreichend klein ist, so daß der Ausdruck in eckigen Klammern auf der linken Seite von Gleichung (6.32) positiv ist, liegt der erwartete Grenznutzen über dem First-best-Niveau und damit der Erwartungsnutzen darunter. Nimmt man darüber hinaus an, daß der erstattete Anteil der Kosten höher als der Anteil der Zuzahlungen des Patienten ist, so folgt hieraus, daß der Ausdruck E positiv ist und der erwartete Grenznutzen des Patienten damit zurückgeht. Dieser Effekt ist um so größer, je mehr medizinische Leistung verordnet wird, da E positiv von m abhängt. Weiterhin verringert sich für den Fall einer höheren medizinischen Leistung bei Nichtkooperation c.p. der Unterschied zwischen dem Niveau bei vollständiger und bei asymmetrischer Information, so daß die Wahrscheinlichkeit einer Genesung des Patienten ansteigt.

6.8 Kritische Würdigung des Double Moral Hazard-Problems

Das Verhältnis zwischen der Leistung des Arztes und der Compliance des Patienten beeinflußt das Gleichgewicht auf der Ebene der Behandlung. Die Inanspruchnahme medizinischer Leistung und das Niveau der Anstrengungen des Patienten hängen dabei davon ab, ob sie sich in ihrer Produktivität als unabhängig, strategische Komplemente oder strategische Substitute klassifizieren lassen. Im Gegensatz zu den anderen Fällen führt die nichtkooperative Lösung bei strategischen Substituten nicht notwendigerweise zu einer Absenkung von medizinischer Leistung und Compliance gegenüber der First-best-Lösung. Statt dessen kann der Fall eintreten, daß das Niveau eines Inputs in die Gesundheitsproduktion über dem First-best-Niveau liegt, während das Niveau des anderen niedriger ist. Die Einführung einer proportionalen Selbstbeteiligung ändert die Ergebnisse insofern, als daß sich die Reaktionsfunktion des Patienten auf die Leistung des Arztes in Richtung First-best-Lösung verschiebt. Dies bedeutet, daß der Patient in allen Fällen mehr gesundheitsfördernde Anstrengungen unternimmt. Während im Fall der Unabhängigkeit der Inputs die medizinische Lei-

¹¹³ Es wird dabei unterstellt, daß die Kostenerstattung pro Einheit medizinischer Leistung stets höher ist als der Selbstbeteiligungssatz, so daß der erwartete Grenznutzen des Einkommens für den Patienten immer positiv ist.

stung unverändert bleibt, wird sie im Fall strategischer Substitute reduziert. Bei strategischen Komplementen hingegen steigt sowohl das Niveau der Anstrengungen als auch das Niveau der medizinischen Leistungen an.

Durch das Auftreten beidseitiger asymmetrischer Information kommt es beim Vertragsabschluß zu Abweichungen von der First-best-Lösung. Der Patient ‚wählt‘ in dieser Situation eine positive Selbstbeteiligung und ist damit nicht mehr vollständig versichert. Für die Vergütungsparameter des Arztes folgt, daß die an den entstandenen Kosten orientierte Komponente so gewählt wird, daß nun der erwartete Grenznutzen des Einkommens für den Patienten und damit sein Einkommensnutzen nicht mehr konstant ist. Bei der Fallpauschale ergeben sich keine Veränderungen, da sie weder das Niveau der Compliance noch der medizinischen Leistung beeinflusst. Als schwierigster Fall erweist sich somit die Festlegung der Kostenerstattung des Arztes, da diese nicht nur Auswirkungen auf die Versicherungsprämie hat sondern auch auf die Wahl der medizinischen Leistung bzw. der Compliance und damit auch auf die Genesungswahrscheinlichkeit. Über die Entwicklung des Grenznutzens des Einkommens des Patienten liegen in diesem Fall keine gesicherten Erkenntnisse vor. Eine durch die höhere Kostenerstattung gestiegene medizinische Leistung des Arztes besitzt c.p. sowohl Effekte auf die Gesundheit durch eine höhere Genesungswahrscheinlichkeit als auch auf die Ausgaben des Patienten durch eine höhere Prämie und eine höhere Zuzahlung.

Problematisch bei einer solchen Analyse sind die Annahmen über das Wissen von Arzt und Patient. Die Ergebnisse hängen zentral davon ab, welche Information der Arzt und Patient hinsichtlich der strategischen Zusammenhänge bzw. der Auswirkungen ihres Handelns besitzen. Falls dem Patienten keine Informationen über den Effekt einer Erhöhung seiner Anstrengungen auf die Grenzproduktivität der medizinischen Leistungen zur Verfügung stehen, führt dies zu Fehlentscheidungen. Es kann daher auch als Aufgabe des Arztes angesehen werden, den Patienten mit Informationen über die medizinische Leistung und den Zusammenhang mit seinen eigenen Handlungen zu versorgen, so daß dieser die Informationen bei der Wahl seiner Compliance berücksichtigt. Der aufgeklärte Patient ist in diesem Zusammenhang nicht als Kontrollinstanz des Arztes zu sehen. Der Arzt besitzt ein berechtigtes Interesse daran, daß der Patient über die Folgen seines Handelns und den Zusammenhang mit der Inanspruchnahme medizinischer Leistungen informiert ist.

6.A Anhang zu Kapitel 6

(A) Vollständige Information

Lagrange-Funktion:

$$\begin{aligned}
 L = & \pi \left\{ p(a, m) U \left[y(G_1) - \pi [\omega + (\delta - \beta) m] - \beta m \right] \right. \\
 & + (1 - p(a, m)) U \left[y(G_0) - \pi [\omega + (\delta - \beta) m] - \beta m \right] - D(a) \left. \right\} \\
 & + (1 - \pi) U \left[y(G_1) - \pi [\omega + (\delta - \beta) m] \right] \\
 & + \lambda \left\{ \pi \left[(\omega + \delta m) + \varepsilon (p G_1 + (1 - p) G_0) - C(m) \right] - \bar{V} \right\}.
 \end{aligned} \tag{A.1}$$

Partielle Ableitung nach der Compliance a :

$$\frac{\partial L}{\partial a} : \quad \pi \left[p_a (U_1 - U_0) \right] - \pi D' + \pi \lambda p_a \varepsilon (G_1 - G_0) = 0. \tag{A.2}$$

Partielle Ableitung nach der medizinischen Leistung m :

$$\begin{aligned}
 \frac{\partial L}{\partial m} : \quad & \pi p_m (U_1 - U_0) + \pi \left[p U_1' (\beta \pi - \beta) + (1 - p) U_0' (\beta \pi - \beta) \right] \\
 & + (1 - \pi) U_2' (\beta \pi) + \lambda \pi \left[\delta + p_m \varepsilon (G_1 - G_0) - C' \right] = 0.
 \end{aligned} \tag{A.3}$$

Partielle Ableitung nach dem Selbstbeteiligungssatz β :

$$\frac{\partial L}{\partial \beta} : \quad \pi \left[m (\pi - 1) (p U_1' + (1 - p) U_0') \right] + (1 - \pi) \pi m U_2' = 0. \tag{A.4}$$

Partielle Ableitung nach der Fallpauschale ω :

$$\frac{\partial L}{\partial \omega} : \quad \pi \left[-\pi (p U_1' + (1 - p) U_0') \right] + [1 - \pi] (-\pi) U_2' + \lambda \pi = 0. \tag{A.5}$$

Partielle Ableitung nach der Kostenerstattung δ :

$$\frac{\partial L}{\partial \delta} : \quad \pi \left[-\pi m (p U_1' + (1 - p) U_0') \right] + (1 - \pi) (-\pi m) U_2' + \lambda \pi m = 0. \tag{A.6}$$

(B) Asymmetrische Information

Lagrange-Funktion:

$$\begin{aligned}
L = & \pi \left\{ p(a(\beta, \delta), m(\delta)) U \left[y(G_1) - \pi [\omega + (\delta - \beta) m(\delta)] - \beta m(\delta) \right] \right. \\
& + (1 - p(a(\beta, \delta), m(\delta))) U \left[y(G_0) - \pi [\omega + (\delta - \beta) m(\delta)] - \beta m(\delta) \right] \\
& \left. - D(a) \right\} + (1 - \pi) U \left(y(G_1) - \pi [\omega + (\delta - \beta) m(\delta)] \right) \\
& + \lambda \left\{ \pi \left[(\omega + \delta m(\delta)) + \varepsilon p(a(\beta, \delta), m(\delta)) G_1 \right. \right. \\
& \left. \left. + (1 - p(a(\beta, \delta), m(\delta))) G_0 \right] - C(m(\delta)) - \bar{V} \right\}.
\end{aligned} \tag{B.1}$$

Partielle Ableitung nach dem Selbstbeteiligungssatz β :

$$\begin{aligned}
\frac{\partial L}{\partial \beta} : & \pi \left[p_a a_\beta (U_1 - U_0) - D_a a_\beta \right] + \lambda \pi \varepsilon p_a a_\beta (G_1 - G_0) \\
& + \pi \left[m(\pi - 1) (p U_1' + (1 - p) U_0') \right] + (1 - \pi) \pi m U_2' = 0.
\end{aligned} \tag{B.2}$$

Partielle Ableitung nach der Fallpauschale ω :

$$\frac{\partial L}{\partial \omega} : \pi \left[-\pi (p U_1' + (1 - p) U_0') \right] + (1 - \pi) (-\pi) U_2' + \lambda \pi = 0. \tag{B.3}$$

Partielle Ableitung nach der Kostenerstattung δ :

$$\begin{aligned}
\frac{\partial L}{\partial \delta} : & \pi \left[p_a a_\delta (U_1 - U_0) - D_a a_\delta + p_m m_\delta (U_1 - U_0) \right] \\
& + \pi \left(-\pi (\delta - \beta) - \beta \right) m \delta (p U_1' + (1 - p) U_0') \\
& + \lambda \pi \left[(\delta - 1) m_\delta + \varepsilon p_a a_\delta (G_1 - G_0) \right] \\
& + \varepsilon p_m m_\delta (G_1 - G_0) - C_m m_\delta + (1 - \pi) \left(-\pi (\delta - \beta) m_\delta \right) U_2' \\
& + \pi \left[-\pi m (p U_1' + (1 - p) U_0') \right] \\
& + (1 - \pi) (-\pi m) U_2' + \lambda \pi m = 0.
\end{aligned} \tag{B.4}$$

7 Empirische Analyse der Arzt-Patient-Beziehung

7.1 Umsetzung theoretischer Befunde

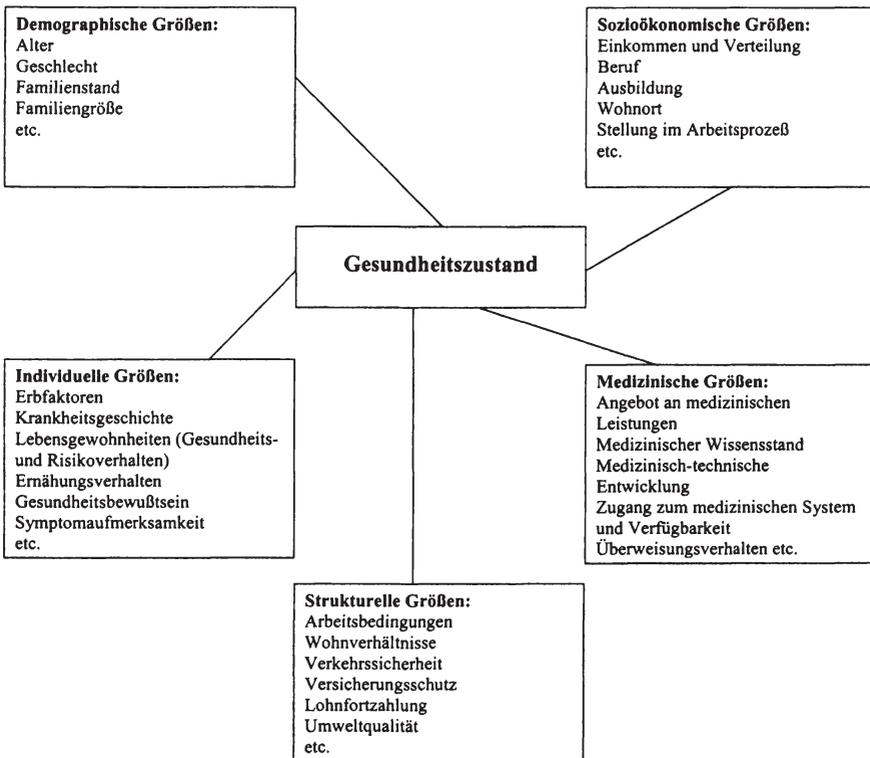
Eine empirische Analyse stellt einen wichtigen Bestandteil im Rahmen der Analyse des Verhältnisses zwischen Arzt und Patient dar. Dabei sollen die in der theoretischen Untersuchung ermittelten Ergebnisse anhand von ausgewählten Daten überprüft und gegebenenfalls kritisch hinterfragt werden. Zunächst ist es notwendig, die Möglichkeiten und Grenzen einer Umsetzung der Ergebnisse der vorherigen Kapitel genauer zu betrachten. Dabei spielen sowohl die Determinanten des Gesundheitszustands als Ansatzpunkt für folgende Analysen als auch die Datenverfügbarkeit für eine Konkretisierung der Überlegungen eine Rolle. Durch einen Überblick über bestehende empirische Studien sollen Gemeinsamkeiten und Unterschiede in der Vorgehensweise und in den Ergebnissen dargestellt werden. Grundsätzlich bieten sich für die betrachtete Thematik zwei Arten der empirischen Analyse an. Die erste Möglichkeit besteht in der Untersuchung von Querschnittsdaten, zum zweiten kann eine Panel-Analyse durchgeführt werden. Bei der ersten Variante werden die Individualdaten zu einem Zeitpunkt erfaßt, bei der zweiten spielen Effekte, die aus dem Zeitablauf resultieren, eine Rolle. Dabei erscheint es sinnvoll, zuerst auf die einzelnen Schätzmethoden einzugehen, die für die Analyse der vorhandenen Daten geeignet scheinen, bevor anschließend die Datenauswahl beschrieben und die Schätzergebnisse dargestellt werden.

Aufgrund der in den vorgestellten theoretischen Modellen (siehe Kapitel 1 und 6) herausgearbeiteten Merkmale der Arzt-Patient-Beziehung erscheint eine Darstellung der Bestimmungsfaktoren des Gesundheitszustands als Ausgangspunkt für eine empirische Untersuchung zweckmäßig (siehe Abbildung 7.1). Der Gesundheitszustand wird durch demographische, individuelle, strukturelle, sozioökonomische und medizinische Größen beeinflusst. Anhand der demographischen Größen ist ersichtlich, daß das Alter und das Geschlecht für das Individuum nicht beeinflussbare Größen darstellen, während die familienbezogenen Faktoren wie Familienstand und -größe verändert werden können.

Individuelle Größen stellen auf Faktoren ab, die bei jeder Person unterschiedlich stark ausgeprägt sind und von ihr z.T. auch beeinflusst werden können. In der Regel dürfte nicht eine einzelne Größe für den Gesundheitszustand entscheidend sein; vielmehr sind das Zusammenwirken mit anderen Faktoren und die sich ergebenden Interdependenzen von entscheidender Bedeutung. So wird die Krankheitsgeschichte sowohl durch die Lebensgewohnheiten, das Ernährungsverhalten, das Gesundheitsbewußtsein sowie durch die Erbfaktoren beeinflusst. Die

Faktoren, über deren Ausprägung das Individuum selbst entscheidet und die als gesundheitsrelevantes Verhalten beschrieben werden können, werden in dem in Abschnitt 6 präsentierten Modellansatz durch die Compliance des Patienten abgebildet. Demographische Größen sind das Alter, das Geschlecht und familienbezogene Variablen. Dabei können diese Faktoren nicht durch das Individuum beeinflusst werden bzw. beschreiben im Fall der familienbezogenen Faktoren langfristige Gegebenheiten.

Abbildung 7.1: Bestimmungsfaktoren des Gesundheitszustands



Quelle: Vgl. SVRKAiG 1994, S. 49.

Sozioökonomische Größen beschreiben die Stellung des Individuums in der Gesellschaft. Unter derartigen Größen sind das Einkommen und Vermögen, Beruf, Ausbildung, Wohnort und die Arbeit zu subsummieren. Sie können beeinflusst werden, und es bestehen starke Wechselwirkungen mit den

individuellen und strukturellen Größen. Letztere geben die Situation im Umfeld der betrachteten Person an. In der ökonomischen Theorie werden strukturelle Größen oft unter dem Schlagwort Umwelt zusammengefaßt. Bei den medizinischen Größen handelt es sich nicht nur um die Inanspruchnahme medizinischer Leistungen, sondern auch um technischen Fortschritt oder Zugangsmöglichkeiten zum System der medizinischen Versorgung, worunter bspw. auch der Krankenversicherungsschutz fällt. Über die geschilderten Zusammenhänge hinaus muß beachtet werden, daß einerseits ein Teil der aufgeführten Größen selbst durch den Gesundheitszustand beeinflusst wird. Dies betrifft die Inanspruchnahme der medizinischen Leistungen ebenso, wie sozioökonomische und individuelle Größen. Andererseits bestehen auch zwischen den einzelnen Variablengruppen starke Abhängigkeiten.

Vor diesem Hintergrund stellt sich die Frage, wie die in den Abschnitten 1 und 6 hergeleiteten Ergebnisse über die Beziehung zwischen Arzt und Patient in ein empirisch überprüfbares Modell übertragen werden können. Ein zentrales Ergebnis aller angesprochenen Ansätze ist, daß die Vertragsgestaltung in der Krankenversicherung und die Informationsbeziehungen im Gesundheitswesen die Inanspruchnahme der medizinischen Leistung durch den Patienten ebenso beeinflussen wie das Verhalten des Arztes und des Patienten und damit das Ergebnis des Leistungserstellungsprozesses. Für eine empirische Überprüfung dieser Zusammenhänge sind Individualdaten über die Vertragsgestaltung, die Inanspruchnahme medizinischer Leistungen, die Compliance des Patienten und über die Informationsverteilung zwischen den Akteuren des Gesundheitswesens nötig.¹¹⁴ Allerdings ergeben sich bei der Erhebung dieser Daten mindestens vier Probleme: Erstens sind vertragliche Regelungen in Deutschland weitgehend durch den Gesetzgeber festgelegt. Dies betrifft vor allem den Bereich der gesetzlichen Krankenversicherung und die Honorierung der ärztlichen Leistungen nach dem Einheitlichen Bemessungsmaßstab (EBM) sowie die Private Krankenversicherung und die Gebührenordnung für Ärzte (GOÄ).¹¹⁵ Im Gegensatz zur GKV sind die Verträge der PKV zwar individueller gestaltet, allerdings stehen Daten über das Versicherungsverhältnis i.d.R. nicht zur Verfügung. Letztlich kann nur die Art des Versicherungsverhältnisses unterschieden werden, d.h., ob ein Individuum gesetzlich oder privat versichert ist. Zweitens unterliegt die konkrete Inanspruchnahme medizinischer Leistungen nach Art und Umfang dem Datenschutz, so daß nur Informationen über die Häufigkeit der Inanspruch-

¹¹⁴ Eine Analyse anhand aggregierter Daten erscheint nicht sinnvoll, da die theoretischen Modelle makroökonomischer Natur sind und durch eine Aggregation zu viele relevante Informationen verloren gingen.

¹¹⁵ Siehe zur ambulanten ärztlichen Versorgung Schneider et al. 1995, S. 208ff. und Statistisches Bundesamt 1998, S. 14ff. und S. 364ff.

nahme genutzt werden können. Drittens existiert keine klare Abgrenzung des Begriffes Compliance, so daß zuerst geklärt werden muß, welche Handlungen des Patienten dazuzurechnen sind. Darüber hinaus ist unklar, ob diese Aktionen beobachtet und quantifiziert werden können. Viertens ist eine Beurteilung und Messung der Informationsverteilung im Gesundheitswesen kaum möglich.¹¹⁶

So bleibt als Ansatzpunkt für eine empirische Überprüfung der Zusammenhang zwischen medizinischer Leistung, Compliance des Patienten und seinem Gesundheitszustand. Dabei ist jedoch zu beachten, daß i.d.R. keine genauen Maße für die relevanten Faktoren zur Verfügung stehen und man somit auf eine Approximation des Verhaltens von Arzt und Patient angewiesen ist. Unter Berücksichtigung dieser Einschränkungen der Datenauswahl lassen sich aus den theoretischen Modellen folgende fünf überprüfbare Hypothesen herleiten:

Hypothese 1:

Neben der medizinischen Leistung wird der Gesundheitszustand des Patienten auch durch dessen eigene Handlungen beeinflusst. Hierzu zählt zum einen seine Entscheidung, einen Arzt aufzusuchen, zum anderen auch sein behandlungsbegleitendes Verhalten (Compliance).

Hypothese 2:

Die Compliance des Patienten und die medizinische Leistung des Arztes sind nicht unabhängig voneinander, sondern beeinflussen sich gegenseitig und damit auch den aus der Behandlung resultierenden Gesundheitszustand.

Hypothese 3:

Die Inanspruchnahme medizinischer Leistungen hängt vom Gesundheitszustand vor der Behandlung ab. In der Regel wird bei einem schlechteren Gesundheitszustand eine größere Menge an medizinischen Leistungen nachgefragt. Die Wirkung der medizinischen Leistung auf den Gesundheitszustand hängt weiterhin vom Schweregrad der Erkrankung ab.

Hypothese 4:

Die Art des Krankenversicherungsschutzes beeinflusst die Inanspruchnahme der medizinischen Leistungen durch den Patienten ebenso wie sein gesundheitsrelevantes Verhalten. Diese Faktoren determinieren wiederum den resultierenden Gesundheitszustand des Individuums.

¹¹⁶ Empirische Studien, in welchen eine asymmetrische Informationsverteilung eine zentrale Rolle spielt, finden sich bspw. bei Chiappori und Salanié (1997) und (2000), die anhand von KFZ-Versicherungsdaten vertragstheoretische Aspekte wie adverse Selektion und Moral Hazard untersuchen.

Hypothese 5:

Die medizinische Leistung ist kein homogenes Produkt, sondern setzt sich aus dem Umfang der Behandlung und dem erbrachten Aufwand des Arztes zusammen. Während der Patient über den Umfang (mit-) entscheidet, kann der Aufwand lediglich unvollständig beobachtet werden.

7.2 Überblick über empirische Studien

In den Bereichen Arzt-Patient-Beziehung und ambulanter Sektor des Gesundheitswesens sowie hinsichtlich der Ermittlung der Determinanten der Gesundheit existieren eine Vielzahl von empirischen Studien. Die Bandbreite der Untersuchungen reicht von aggregierten Studien über die Analyse der Nachfrage nach medizinischen Leistungen oder die Bestimmungsgründe des Gesundheitszustands bis hin zu Versicherungsstudien.

Auster et al. (1969) untersuchen anhand von Daten für 48 US-Bundesstaaten aus dem Jahre 1960 den Zusammenhang zwischen der Mortalität, dem Angebot an Gesundheitsleistungen bzw. den Ausgaben für Gesundheit und Umweltvariablen. Die geschätzten Koeffizienten der zweistufigen Schätzung sind jedoch sämtlich insignifikant, was die Vermutung nahelegt, daß zum einen die Mortalität nur ein schwacher Indikator für den Gesundheitszustand ist, zum anderen, daß aggregierte Schätzungen nicht in der Lage sind, die individuellen Effekte auf den Gesundheitszustand wiederzugeben. In einer Arbeit aus dem Jahr 1974 gibt Feldstein einen Überblick über ökonometrische Studien im Bereich der Gesundheitsökonomie. Dabei geht Feldstein auf den Krankenhausmarkt, den Markt für ärztliche Leistungen sowie den Krankenversicherungsbereich ein. Weiterhin beinhaltet die Arbeit einen Überblick über ökonometrische Probleme wie die Verwendung von Mikro- oder Makrodaten, Fehler in den Variablen oder auch über Probleme der Variablenauswahl bei gesundheitsökonomischen Studien.

Die Frage, welchen Einfluß nichtmonetäre Faktoren auf die Nachfrage nach Gesundheitsleistungen ausüben, ist Gegenstand der Untersuchung von Acton aus dem Jahr 1975. Mit Hilfe eines simultanen Gleichungsmodells wird ein Modell umgesetzt, in dem die Nachfrage nach frei verfügbaren und nicht frei verfügbaren medizinischen Leistungen beschrieben wird, mit dem Resultat, daß insbesondere die Entfernung zum Arzt oder Krankenhaus, gemessen durch die benötigte Zeit, die Nachfrage nach medizinischen Leistungen beeinflusst. Allerdings trifft dies nur dann zu, wenn derartige Leistungen frei verfügbar sind, d.h. keine Rationierung vorliegt. Weiterhin analysiert LeFranc (1989) anhand von bestehenden Untersuchungen die sozioökonomischen Determinanten der Gesundheit, die zusammen mit dem gesundheitsrelevanten Verhalten und der Beziehung zwischen Arzt und Patient den Gesundheitszustand beeinflussen.

Breyer und Ulrich (2000) betrachten für den Zeitraum von 1970-1995 den Einfluß des Alters und des medizinisch-technischen Fortschritts auf die Gesundheitsausgaben. Sie zeigen, daß sowohl die Altersstruktur der Bevölkerung als auch der technische Fortschritt neben dem Einkommen einen wesentlichen Beitrag zur Entwicklung der Gesundheitsausgaben in der Gesetzlichen Krankenversicherung leisten. Hart (1988) schätzt den Wohlfahrtsverlust, der sich aus den Moral Hazard- und Agency-Problemen im Bereich der Krankenversicherung ergibt. Für Australien gelangt er zu dem Ergebnis, daß für eine homogene medizinische Leistung der Wohlfahrtsverlust durch die Krankenversicherung 1,56% der Ausgaben für diese Leistung beträgt. Gaynor und Polachek (1994) untersuchen die Information auf dem Markt für medizinische Leistungen. Bei einem Vergleich zwischen dem Ausmaß der unvollständigen Information des Patienten und des Arztes ermitteln Gaynor und Polachek, daß die unvollständige Information des Patienten ca. 1,5mal so hoch ist wie auf der Seite des Arztes. Die Autoren kommen außerdem zu dem Schluß, daß mit einer Zunahme der Suchkosten des Patienten sind auch dessen Informationsdefizit ansteigt. Chiapori et al. (1998) analysieren den Zusammenhang zwischen Moral Hazard und der Nachfrage nach ärztlichen Leistungen. Das Resultat einer Untersuchung von 4578 Individuen über zwei Jahre ist, daß zwar die Praxisbesuche nicht durch die Höhe der Selbstbeteiligungsrate beeinflußt werden, daß jedoch die Hausbesuche von der Selbstbeteiligung abhängen. Sie werten dieses Ergebnis als Indiz dafür, daß neben den direkten Kosten auch nichtmonetäre Kosten der Nachfrage nach medizinischen Leistungen die Arztbesuche beeinflussen.

Holly et al. (1998) präsentieren ein ökonometrisches Modell für die Inanspruchnahme medizinischer Leistungen und der Versicherung in der Schweiz. Dabei wird ein simultanes Probit-Modell geschätzt, wobei die erste abhängige Variable die Bereitschaft des Konsumenten angibt, eine Zusatzversicherung abzuschließen, und die zweite abhängige Variable die Wahrscheinlichkeit beziffert, daß die Person bei der Inanspruchnahme medizinischer Leistungen einen Krankenhausaufenthalt aufweist. Dabei hat die Tatsache, daß ein Konsument eine Zusatzversicherung besitzt, einen positiven Effekt auf die Wahrscheinlichkeit eines Krankenhausaufenthaltes.

Die Nachfrage nach Gesundheit steht im Mittelpunkt der Untersuchung von Wagstaff (1986). Der Analyse liegt das Modell von Grossman (1972) zugrunde, daß die Nachfrage nach Gesundheitsleistungen als abgeleitete Nachfrage betrachtet. Dabei wird zwischen einem Konsummodell und einem Investitionsmodell unterschieden. Im Konsummodell werden die direkten Nutzenwirkungen der Gesundheit betrachtet. Dagegen führt im Investitionsmodell ein besserer Gesundheitszustand zu mehr gesund verbrachter Zeit, die zur Einkommenserzie-

lung verwendet werden kann. Mit diesem Einkommen ist es dem Individuum dann möglich, nutzenstiftende Güter zu erwerben. In der Arbeit von Wagstaff werden sowohl die reduzierte Form als auch die Strukturgleichung als rekursives Gleichungssystem geschätzt. Für den zweiten Fall der Strukturgleichung wird das Gesundheitskapital als latente Variable aufgefaßt. Betrachtet man lediglich die Schätzung für den Gesundheitszustand, so ergibt sich im Investitionsmodell, daß der Gesundheitszustand positiv von der Bildung, vom Lohnsatz und von der Anzahl der Jahre beim Arbeitgeber und negativ vom Alter, der Verstädterung und einigen Umweltindikatoren abhängt. Darüber hinaus ist die Gesundheit für Männer signifikant besser als für Frauen. Im Konsummodell stellen sich diese Effekte anders dar. Sowohl für die Bildungsvariablen als auch für die Familiengröße ergibt sich kein einheitlicher Effekt. Während bspw. die Anzahl der Schul- und Ausbildungsjahre einen negativen Effekt besitzt, ergibt sich für eine im Jahr 1976 erhaltene Ausbildung ein positiver Zusammenhang mit der Gesundheit. Ebenfalls positiv wird sie durch den Lohnsatz, das Vermögen und das erwartete Lebensarbeitseinkommen beeinflusst. Ein negativer Zusammenhang liegt für die Variablen Alter, Verstädterung sowie die Umweltvariablen vor.

Leu und Doppmann (1986) schätzen die Nachfrage nach Gesundheit und Gesundheitsleistungen mit Hilfe eines LISREL-Modells anhand schweizer Daten.¹¹⁷ Dabei werden neben den Gleichungen für die Nachfrage nach Gesundheit und Gesundheitsleistungen auch Gleichungen für das Erwerbs- und Transfereinkommen geschätzt. Die Gesundheit, das Erwerbs- und das Transfereinkommen stellen die latenten Variablen dar. Die Ergebnisse entsprechen weitgehend den theoretischen Erwartungen des Grossman-Modells. In einer vergleichbaren Untersuchung überprüfen Leu und Gerfin (1992) das Grossman-Modell unter Verwendung eines MIMIC-Modells.¹¹⁸ Das theoretische Modell wird allerdings nicht in allen Punkten bestätigt. Ein ähnliches Verfahren verwenden Pohlmeier und Ulrich (1992) für deutsche Daten, wobei die Ergebnisse grundsätzlich mit den erwarteten Vorzeichen des Grossman-Modells übereinstimmen.

Die Nachfrage nach medizinischen Leistungen im Lebenszyklus untersuchen Pohlmeier und Ulrich (1996) mit Hilfe einer Panelschätzung. Die Autoren betonen die Relevanz der Unterscheidung der Altersabhängigkeit der Nachfrage nach medizinischen Leistungen in drei Effekte: den physischen Alterseffekt, der

¹¹⁷ LISREL steht für linear structural relationships. Es handelt sich dabei um ein Strukturgleichungsmodell mit latenten Variablen (vgl. Leu und Doppman 1986, S. 162).

¹¹⁸ MIMIC-Modelle (Multiple Indicators Multiple Causes) stellen einen Spezialfall der LISREL-Modelle dar, wobei nur eine latente endogene Variable enthalten ist (vgl. Leu und Gerfin 1992, S. 69f.).

die Morbidität wiedergibt, den Kohorteneffekt, der den Einfluß makroökonomischer Schocks abbildet und die intertemporale Entscheidungskomponente. Für die Nachfrage nach allgemeinärztlichen Leistungen ergibt sich ein u-förmiger Zusammenhang zwischen Alter und der Nachfrage nach medizinischen Leistungen.

Der Sachverhalt, daß sich die Nachfrage nach medizinischen Leistungen in die Erstkontaktentscheidung, die durch den Patienten getroffen wird, und in die Entscheidung über die Frequenz der Behandlung, die stark arzt determiniert ist, einteilen läßt, bildet den Ansatz der Studie von Pohlmeier und Ulrich (1995). Mit Hilfe eines negativ binomial verteilten Hürdenmodells werden die zwei Stufen der Inanspruchnahme modelliert. Dabei werden die Kontakt- und Frequenzentscheidungen sowohl für Hausarzt- als auch für Facharztbesuche durch zwei unterschiedliche stochastische Prozesse bestimmt. Die Ergebnisse der Kontaktstufe entsprechen denen früherer Schätzungen ohne eine Hürdenspezifikation. Eine Schlußfolgerung hieraus ist, daß in solchen Ansätzen die zweite Stufe der Inanspruchnahme durch die erste Stufe dominiert wird.

Robst und Graham (1997) untersuchen den Einfluß der Ärzte auf die Inanspruchnahme und den Gesundheitszustand. Dabei benutzen die Autoren zwei Teilstichproben, städtische Bevölkerung und Landbevölkerung. Diese sollen den unterschiedlichen Zugang zu den medizinischen Leistungen wiedergeben. Obwohl die Variable für den Gesundheitszustand ein subjektives Maß auf einer Skala von 1 (schlecht) bis 5 (exzellent) ist, werden für die Schätzungen Ordinary-Least-Squares-Methoden angewendet. Für die Landbevölkerung folgt als Ergebnis, daß eine steigende Zahl von Ärzten einen positiven Effekt auf den Gesundheitszustand hat. Daneben profitieren alte Menschen mehr von der Erreichbarkeit der Ärzte als junge. Für die städtischen Gebiete findet sich jedoch keine klare Beziehung zwischen Gesundheit und der Anzahl der Ärzte.

Nocera und Zweifel (1998) überprüfen das Grossman-Modell mit Hilfe eines Paneldatensatzes. Die Schätzungen wurden für zwei Teilstichproben durchgeführt, da für einen Teil der Beobachtungen die Daten nur für einen bestimmten Zeitraum vorlagen.¹¹⁹ Die erste Stichprobe umfaßt Individuen, für die Informationen für die Jahre 1989 bis 1992 vorlagen. In der zweiten Stichprobe waren alle Individuen vertreten, die an beiden Befragungen teilgenommen hatten und für die die fehlende Information für die Jahre 1981 bis 1992 konstruiert werden

¹¹⁹ Die Daten entstammen aus zwei Befragungen aus den Jahren 1981 und 1993 einer Schweizerischen Krankenversicherung. Weiterhin waren Information über die Nutzung von medizinischen Leistungen in den Jahren 1981 bis 1992 und Versicherungsaufzeichnungen für die Jahre 1987 bis 1992 vorhanden.

mußte. In diesem Zusammenhang wird die Nachfrage nach Gesundheit als Random-effects-Modell (GLS) geschätzt. Alle Koeffizienten verfügen für die erste Stichprobe über das erwartete Vorzeichen, in der zweiten Stichprobe hingegen entsprechen einige Koeffizienten nicht den theoretischen Überlegungen. Die Inanspruchnahme der medizinischen Leistung wurde mit Hilfe eines Tobit-Modells für die reduzierte Form und die Strukturgleichung geschätzt. Lediglich die Koeffizienten für Alter und Gewicht sind hierbei signifikant und besitzen das erwartete Vorzeichen. Insgesamt weichen die Schätzung der reduzierten Form und die der Strukturgleichung nicht in dem Maße voneinander ab wie in der bereits angesprochenen Studie von Wagstaff (1986).

Einen anderen Ansatz zur Bestimmung der Determinanten der Gesundheit wählen Gerdtham und Johannesson (1997). In ihrer Untersuchung über den Zusammenhang zwischen allgemeiner Zufriedenheit, Gesundheit und sozioökonomischen Faktoren benutzen sie ein Ordered Probit-Modell, da sowohl das Maß für die Zufriedenheit als auch das für den Gesundheitszustand drei Kategorien aufweist. Bei der Schätzung der Gleichungen in Strukturform ergibt sich für die Variable ‚Gesundheit‘, daß Männer einen signifikant schlechteren Gesundheitszustand als Frauen aufweisen. Weiterhin nimmt die Gesundheit mit dem Alter ab und mit der Bildung zu.

Einen ähnlichen Ansatz verfolgen Gerdtham und Johannesson (1999) und Gerdtham et al. (1999) für die Schätzungen der Nachfrage nach Gesundheit. Auch in diesem Fall ist die abhängige Variable eine kategoriale Variable. Die Schätzungen erfolgen auch hier mit der Ordered Probit-Methode. Die Ergebnisse sind mit denen der Untersuchung aus dem Jahr 1997 vergleichbar.¹²⁰ Gerdtham und Johannesson (1999) überprüfen den Zusammenhang darüber hinaus noch nur für Erwerbstätige und finden einen positiven Einfluß des Lohnsatzes auf die Gesundheit. Zudem wird in der Untersuchung von Gerdtham et al. (1999) sportliche Betätigung als erklärende Variable eingeführt, die einen signifikant positiven Effekt aufweist.

Die präsentierten Studien stellen die Vielfältigkeit der Ansätze für eine empirische Untersuchung im Gesundheitswesen vor dem Hintergrund der Arzt-Patient-Beziehung dar. Für die eigene Arbeit erweisen sich die zuletzt genannten Studien von Gerdtham und Johannesson (1997) und (1999) sowie Gerdtham et al. (1999) als interessant, da in ihnen ein vergleichbarer Ansatzpunkt verfolgt wird

¹²⁰ Hierbei ist anzumerken, daß sämtliche verwendeten erklärenden Variablen mit Dummy-Variablen ausgedrückt wurden, auch wenn sie ursprünglich stetig waren. Bspw. wurde das Alter über die Zugehörigkeit zu einer bestimmten Altersklasse ausgedrückt.

(vgl. dazu die folgenden Abschnitte). Dies bezieht sich insbesondere sowohl auf die verwendete Schätzmethode als auch auf die Datenauswahl.

7.3 *Ökonometrische Untersuchung anhand von Querschnittsdaten*

7.3.1 Einführung in Modelle diskreter abhängiger Variablen

Zur Schätzung des Zusammenhangs zwischen Gesundheitszufriedenheit, Compliance des Patienten und medizinischer Leistungen wird als abhängige Variable die ‚Zufriedenheit mit der Gesundheit‘ verwendet. Sie weist elf Kategorien auf. Es bieten sich folglich als Schätzmethode Modelle für geordnete Kategorien an. Diese gehören zur Gruppe der diskreten abhängigen Variablen. In der Regressionsanalyse wird normalerweise unterstellt, daß die abhängige Variable nach eventueller Transformation durch Werte im Bereich der reellen Zahlen dargestellt werden kann (vgl. Davidson und MacKinnon 1993, S. 511). Diese Annahme führt jedoch zu Problemen, wenn die Variable einen bestimmten Wert mit einer Wahrscheinlichkeit größer Null annimmt. Ein besonderer Fall liegt vor, wenn die abhängige Variable lediglich die Werte Null oder Eins besitzt. Man spricht daher von binären abhängigen Variablen. Neben diesen Fällen lassen sich noch andere unterscheiden, in denen die Annahme einer stetigen, abhängigen Variable verletzt ist und daher speziell hierfür entwickelte Schätzverfahren benötigt werden. (vgl. hierzu Greene 2000, S. 812).¹²¹

Das lineare Wahrscheinlichkeitsmodell

Ein erster Ansatz im Umgang mit diskreten abhängigen Variablen besteht darin, daß man eine binäre abhängige Variable auf eine Reihe von erklärenden Variablen mittels Ordinary-Least-Squares-Schätzung (OLS) regressiert. Man erhält somit ein lineares Wahrscheinlichkeitsmodell (linear probability model) (siehe hierzu Long 1997, S. 35ff.). Die Strukturform dieses Modell ist durch Gleichung (7.1) gegeben:

$$y_i = x_i \beta + \varepsilon_i . \quad (7.1)$$

Hierbei bezeichnet x_i den Vektor der erklärenden Variablen für die i te Beobachtung, β den zu schätzenden Parametervektor und ε_i den Störterm, für des-

¹²¹ Es existieren eine Vielzahl von Publikationen, die sich mit diskreten, abhängigen Variablen beschäftigt, z.B. Amemiya (1981), Maddala (1983), Dhrymes (1986), Long (1997) und Gouriéroux (2000). Cramer (1991) bezieht sich speziell auf Logit-Modelle. Übersichten über diese Modellklasse finden sich darüber hinaus in Maddala (1992), Davidson und MacKinnon (1993) oder Kennedy (1998).

sen Erwartungswert $E(\varepsilon_i) = 0$ unterstellt wird. Die abhängige Variable nimmt den Wert Eins an, wenn ein bestimmtes Ereignis eintritt, bspw. wenn das befragte Individuum mit seinem Gesundheitszustand voll und ganz zufrieden ist, und den Wert Null, wenn das Ereignis nicht eintritt.

Falls y eine binäre Zufallsvariable darstellt, so ist die unbedingte Erwartung von y die Wahrscheinlichkeit Pr , daß das Ereignis eintritt:

$$E(y_i) = [1 \times Pr(y_i = 1)] + [0 \times Pr(y_i = 0)] = Pr(y_i = 1). \quad (7.2)$$

Hingegen bestimmt sich der bedingte Erwartungswert für die obige Regression (Gleichung (7.6)) als

$$E(y_i | x_i) = [1 \times Pr(y_i = 1 | x_i)] + [0 \times Pr(y_i = 0 | x_i)] = Pr(y_i = 1 | x_i). \quad (7.3)$$

Daraus folgt, daß die bedingte Erwartung von y für ein gegebenes x die Wahrscheinlichkeit für das Ereignis $y = 1$ bei gegebenem x ist. Für das lineare Wahrscheinlichkeitsmodell gilt dann:

$$Pr(y_i = 1 | x_i) = x_i \beta. \quad (7.4)$$

Durch die Verwendung einer binären abhängigen Variablen ändert sich die Interpretation der Schätzparameter in einer OLS-Schätzung im Vergleich zum klassischen linearen Regressionsmodell nicht. Ein Anstieg einer erklärenden Variablen x_k um eine Einheit erhöht die Wahrscheinlichkeit für den Eintritt eines Ereignisses c.p. um β_k . Da das verwendete Modell linear ist, erhöht sich die Wahrscheinlichkeit entsprechend der Veränderung der erklärenden Variablen.

Allerdings sind bei Verwendung des linearen Wahrscheinlichkeitsmodells bestimmte Annahmen des klassischen linearen Regressionsmodells verletzt.¹²² Erstens betrifft dies die Annahme der Homoskedastizität. Angenommen, der Erwartungswert einer binären Zufallsvariable sei μ , so folgt für die Varianz $Var(y) = \mu(1-\mu)$.¹²³ Für das lineare Wahrscheinlichkeitsmodell gilt somit:

¹²² Zu den Annahmen des klassischen linearen Regressionsmodells siehe Maddala (1992).

¹²³ Sei y eine Dummy-Variablen mit $E(y) = \mu$. Die Varianz ist definiert als $Var(y) = E[(y-\mu)^2] = E(y^2 - 2\mu y + \mu^2) = E(y^2) - 2\mu E(y) + \mu^2$. Da y entweder die Werte 1 oder 0 annimmt, folgt $E(y^2) = E(y)$. Daher gilt $Var(y) = E(y) - 2\mu E(y) + \mu^2$. Substituiert man nun $E(y) = \mu$, so erhält man $Var(y) = \mu - \mu^2 = \mu(1-\mu)$ (vgl. Long 1997, S. 265).

$$\text{Var}(y | \mathbf{x}) = \text{Pr}(y = 1 | \mathbf{x}) [1 - \text{Pr}(y = 1 | \mathbf{x})] = \mathbf{x} \beta (1 - \mathbf{x} \beta). \quad (7.5)$$

Anhand der Gleichung (7.5) erkennt man, daß die Varianz des Störterms von der Realisation der erklärenden Variablen abhängt. Somit ist das lineare Wahrscheinlichkeitsmodell heteroskedastisch und der OLS-Schätzer von β ist ineffizient. Darüber hinaus sind die Standardfehler verzerrt, so daß sich unbrauchbare Teststatistiken ergeben.

Der zweite Einwand gegen das lineare Wahrscheinlichkeitsmodell betrifft die Normalverteilung des Störterms. Geht man zur Vereinfachung von nur einer erklärenden Variable aus, so ergibt sich für das Modell:

$$y_i = \alpha + \beta x_i + \varepsilon_i. \quad (7.6)$$

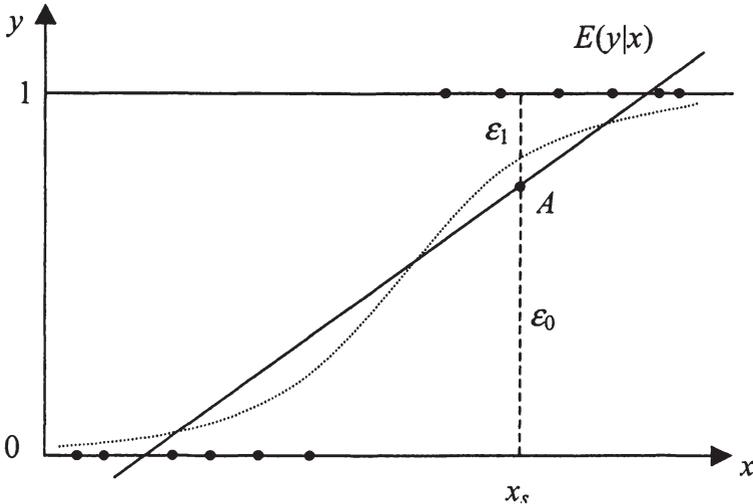
Dabei bezeichnet α die Konstante des Modells. Die bedingte Erwartung von y gegeben x ist dann $E(y | x) = \alpha + \beta x$. In Abbildung 7.2 ist sie durch die durchgezogene Linie gegeben. Auf der Abszisse ist die erklärende Variable abgetragen, auf der Ordinate die abhängige Variable. Die Beobachtungen sind durch die Punkte für $y = 0$ und $y = 1$ gegeben. Man betrachte nun einen spezifischen Wert x_s für die abhängige Variable. In Abbildung 7.2 ist die bedingte Erwartung an der Stelle x_s durch Punkt A gegeben. Der Fehler ε entspricht der Entfernung der bedingten Erwartung $E(y | x)$ vom tatsächlich beobachteten Wert. Da y nur zwei Zustände annehmen kann, ist der Fehler entweder $\varepsilon_1 = 1 - E(y | x_s)$ oder $\varepsilon_0 = 0 - E(y | x_s)$ und damit nicht normalverteilt.

Drittens kann es sein, daß man aus dem linearen Wahrscheinlichkeitsmodell vorhergesagte Werte für y erhält, die entweder negativ oder größer als Eins sind. In Abbildung 7.2 ist dies daran zu erkennen, daß die Gerade $E(y | x)$ für einige Werte von x ober- bzw. unterhalb der zugehörigen und damit zulässigen Werte von y liegt.

Der letzte Einwand betrifft die funktionale Form des Modells. Da es linear ist, resultiert ein Anstieg in x_k um eine Einheit in einer konstanten Änderung der Wahrscheinlichkeit eines Ereignisses von β_k . Das bedeutet, daß der Anstieg unabhängig vom tatsächlichen Wert von x immer derselbe ist. Realistischer ist für viele Anwendungen ein nichtlinearer Anstieg. Allgemein gilt, daß es für den Fall eines Ergebnisses in Form einer Wahrscheinlichkeit oftmals angemessener erscheint, daß die erklärenden Variablen abnehmende Effekte für vorhergesagte Werte nahe Null oder Eins haben. Für ‚binary response-Modelle‘ wird die Beziehung zwischen den erklärenden Variablen und der Wahrscheinlichkeit eines

Ereignisses als s-förmiger Verlauf modelliert. Dieser Zusammenhang ist in Abbildung 7.2 als gepunktete Kurve dargestellt.

Abbildung 7.2: Lineares Wahrscheinlichkeitsmodell für eine einzelne erklärende Variable



Quelle: Erstellt in Anlehnung an Long 1997, S. 36 und Kennedy 1998, S. 234.

Binäre Probit- und Logit-Modelle

Da sich ein lineares Wahrscheinlichkeitsmodell für die vorliegende Datenstruktur als ungeeignet erweist, benötigt man spezielle Verfahren zur Schätzung. Bevor der Fall einer kategorialen, abhängigen Variablen erläutert wird, wird als Vereinfachung wie im linearen Wahrscheinlichkeitsmodell von einer binären abhängigen Variablen ausgegangen. Zusätzlich wird unterstellt, daß diesem Fall eine latente (unbeobachtete) Variable y^* zugrunde liegt, deren Wertebereich von $-\infty$ bis $+\infty$ reicht. Für größere Werte von y^* wird $y = 1$ beobachtet, für kleinere $y = 0$ (vgl. hierzu und die folgenden Ausführungen Long 1997, S. 40ff.).¹²⁴ Für die latente Variable y^* wird unterstellt, daß eine lineare Beziehung zu den reali-

¹²⁴ Grundsätzlich ist es nicht notwendig, eine latente Variable für das Modell zu unterstellen. Long (1997) stellt ein nichtlineares Wahrscheinlichkeitsmodell dar, das nicht auf die Idee einer latenten Variablen zurückgreift.

sierten Beobachtungen der erklärenden Variablen x besteht, die sich mit folgendem Strukturmodell darstellen läßt:

$$y_i^* = x_i \beta + \varepsilon_i. \quad (7.7)$$

In diesem Zusammenhang bezeichnet x_i wieder den Vektor der erklärenden Variablen der i ten Beobachtung, β den zu schätzenden Parametervektor und ε_i den Fehlerterm. Die beobachtete Variable y_i hängt, wie in der Meßgleichung dargestellt, von der latenten Variable ab.

$$y_i = \begin{cases} 1 & \text{für } y_i^* > \tau \\ 0 & \text{für } y_i^* \leq \tau. \end{cases} \quad (7.8)$$

Der Parameter τ bezeichnet den zugehörigen Schwellenwert (engl. threshold oder cutpoint). Falls $y_i^* \leq \tau$, so folgt $y = 0$ und, wenn y_i^* den Schwellenwert übersteigt, $y = 1$.¹²⁵ Der Zusammenhang zwischen der latenten und der beobachteten Variablen kann mit Hilfe des vereinfachten Modells $y_i^* = \alpha_i + \beta x_i + \varepsilon_i$ in Abbildung 7.3 dargestellt werden. Dabei sind die latente Variable y_i^* auf der Ordinate und die erklärende Variable x auf der Abszisse abgetragen. Die bedingte Erwartung $E(y_i^* | x)$ ist durch die bei α beginnende Regressionsgerade und der Schwellenwert τ durch die gestrichelte Linie gegeben. Man erkennt drei Dichtefunktionen für y_i^* in Abhängigkeit von der Realisierung von x . Der schattierte Bereich unter der Häufigkeitsverteilung gibt den Fall an, in dem y_i^* größer als der Schwellenwert ist und für den $y = 1$ gilt. Für x_1 gilt dies in ca. 25% der Fälle, für x_2 bei ca. 90% und für x_3 weisen nahezu alle beobachteten Fälle für y den Wert Eins auf (vgl. Long 1997, S. 42).

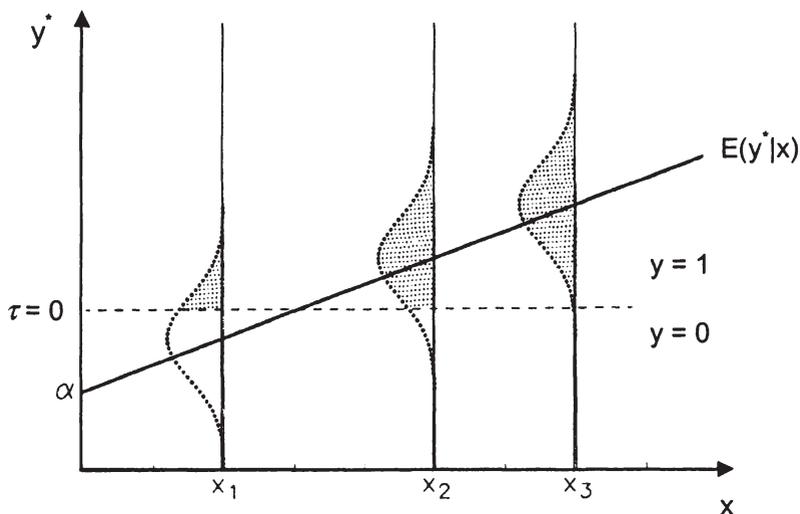
Da die latente Variable eine stetige Ausprägung besitzt, werden die Probleme des linearen Wahrscheinlichkeitsmodells vermieden. Aufgrund der fehlenden Beobachtbarkeit der Variablen, kann das Modell nicht mit OLS-Techniken geschätzt werden, sondern man verwendet einen Maximum Likelihood-Ansatz.¹²⁶ Dazu ist es jedoch notwendig, bezüglich des Fehlerterms einige Annahmen zu treffen. In der Regel betrifft dies die Unterscheidung zwischen einer normalverteilten und einer logistischen Fehlerstruktur. Im ersten Fall spricht man von einem Probit-, im zweiten von einem Logit-Modell. Wie im klassischen linearen

¹²⁵ Für die weitere Betrachtung wird als Vereinfachung $\tau = 0$ unterstellt.

¹²⁶ Eine Übersicht über Maximum Likelihood-Techniken und deren Eigenschaften findet sich in Cramer (1986).

Regressionsmodell wird auch hier unterstellt, daß die bedingte Erwartung des Fehlers gleich Null ist ($E(\varepsilon | x) = 0$).

Abbildung 7.3: Zusammenhang von latenter und beobachteter Variable



Quelle: Long 1997, S. 41.

Weiterhin kann die Varianz der Fehler nicht wie im klassischen Modell geschätzt werden. Für das Probit-Modell wird daher angenommen, daß $Var(\varepsilon | x) = 1$ und im Logit-Modell, daß $Var(\varepsilon | x) = \pi^{2/3}$. Die Normalverteilung, die im Probit-Modell für die Fehlerstruktur angenommen wird, besitzt einen Erwartungswert von Null und eine Varianz von Eins. Die Dichtefunktion ϕ ist somit

$$\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2}\right), \quad (7.9)$$

und für die Verteilung Φ folgt:

$$\Phi(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \quad (7.10)$$

Im Logit-Modell besitzen die Fehler eine logistische Verteilung mit Erwartungswert von Null und der Varianz $\pi^{2/3}$.¹²⁷ Dadurch ergeben sich recht einfache Gleichungen für die Dichtefunktion λ und die Verteilungsfunktion Λ :

$$\lambda(\varepsilon) = \frac{\exp(\varepsilon)}{[1 + \exp(\varepsilon)]^2}, \quad (7.11)$$

$$\Lambda(\varepsilon) = \frac{\exp(\varepsilon)}{1 + \exp(\varepsilon)}. \quad (7.12)$$

Die Wahrscheinlichkeit, daß die beobachtete Variable y den Wert Eins annimmt, kann in Abhängigkeit von der latenten Variablen wie folgt ausgedrückt werden:

$$Pr(y = 1 | \mathbf{x}) = Pr(y^* > 0 | \mathbf{x}). \quad (7.13)$$

Die latente Variable kann durch Gleichung (7.7) substituiert werden und man erhält:

$$Pr(y = 1 | \mathbf{x}) = Pr(\mathbf{x}\beta + \varepsilon > 0 | \mathbf{x}). \quad (7.14)$$

Subtrahiert man auf beiden Seiten $\mathbf{x}\beta$ und macht sich den Umstand zu Hilfe, daß Normal- und logistische Verteilung symmetrisch sind, so folgt, daß für $y = 1$ die Wahrscheinlichkeit für Werte größer als $-\mathbf{x}\beta$ gleich der Wahrscheinlichkeit für Werte kleiner als $\mathbf{x}\beta$ ist und damit:

$$Pr(y = 1 | \mathbf{x}) = Pr(\varepsilon \leq \mathbf{x}\beta | \mathbf{x}). \quad (7.15)$$

Die rechte Seite der Gleichung gibt die Dichtefunktion des Fehlers an der Stelle $\mathbf{x}\beta$ an. Allgemein kann dafür dann geschrieben werden:

$$Pr(y = 1 | \mathbf{x}) = F(\mathbf{x}\beta). \quad (7.16)$$

¹²⁷ Der Unterschied zwischen der Normalverteilung und der logistischen Verteilung liegt in den Enden der logistischen Funktion, die im Vergleich zur Normalverteilung stärker ausgeprägt sind (vgl. Greene 2000, S. 815).

Die zur Identifikation des Modells benötigten Annahmen sollen hier nochmals kurz zusammengefaßt werden. Erstens wird unterstellt, daß der Schwellenwert $\tau = 0$ ist. Zweitens wird für den bedingten Erwartungswert des Fehlers angenommen, daß $E(\varepsilon | x) = 0$. Drittens ist die bedingte Varianz eine Konstante, die im Probit-Modell den Wert Eins annimmt und im Logit-Modell den Wert $\pi^{2/3}$. Diese Annahmen sind in gewisser Hinsicht willkürlich, da sie nicht getestet werden können. Sie sind auf der anderen Seite aber auch notwendig, um das Modell zu identifizieren, da für eine latente Variable der Erwartungswert und die Varianz nicht bestimmt werden können.

Zur Schätzung des Modells wird ein Maximum Likelihood-Ansatz verwendet. Dabei sei p_i die Wahrscheinlichkeit, mit der ein beliebiger Wert von y für eine gegebene Beobachtung wahrgenommen wird:

$$p_i = \begin{cases} Pr(y_i = 1 | x_i) & \text{falls } y_i = 1 \text{ beobachtet} \\ 1 - Pr(y_i = 1 | x_i) & \text{falls } y_i = 0 \text{ beobachtet} . \end{cases} \quad (7.17)$$

$Pr(y_i = 1 | x_i)$ ist in Gleichung (7.16) definiert. Falls die Beobachtungen unabhängig sind, so folgt für die Likelihood-Funktion

$$L(\beta | y, X) = \prod_{i=1}^N p_i . \quad (7.18)$$

Aus Gleichung (7.18) ergibt sich zusammen mit (7.17):

$$L(\beta | y, X) = \prod_{i=1} Pr(y_i = 1 | x_i) \prod_{i=0} [1 - Pr(y_i = 1 | x_i)] , \quad (7.19)$$

wobei der Index des Produktzeichens bedeutet, daß nur die Fälle für $y = 1$ bzw. $y = 0$ miteinander multipliziert werden. Die Wahrscheinlichkeiten können nun mit Hilfe der Verteilungsfunktion analog zu Gleichung (7.16) in Abhängigkeit des zu schätzenden Parametervektors β definiert werden:

$$L(\beta | y, X) = \prod_{i=1} Pr(y_i = 1 | x_i) \prod_{i=0} [1 - Pr(y_i = 1 | x_i)] , \quad (7.20)$$

Durch Logarithmierung erhält man die Log-Likelihood-Funktion, die unter bestimmten Bedingung global konkav ist, so daß die Parameterschätzungen ein-

deutig, konsistent, asymptotisch normalverteilt und asymptotisch effizient sind:¹²⁸

$$\ln L(\beta | y, X) = \sum_{i=1} \ln F(x_i, \beta) \sum_{i=0} \ln [1 - F(x_i, \beta)]. \quad (7.21)$$

Maximiert man Gleichung (7.21) über den Parametervektor β , so erhält man die notwendigen und hinreichenden Bedingungen. Falls die Log-Likelihood-Funktion die Bedingungen globaler Konkavität erfüllt, so genügt im weiteren Verlauf die Betrachtung der notwendigen Bedingungen.¹²⁹

Zu beachten ist, daß, egal welche Verteilung man unterstellt, die Modellparameter nicht die marginalen Effekte wie im linearen klassischen Regressionsmodell wiedergeben (vgl. hierzu Greene 2000, S. 815f.). Dies läßt sich mit Hilfe des bedingten Erwartungswertes von y zeigen, der durch folgenden Ausdruck gegeben ist:

$$E(y | x) = 0 [1 - F(x\beta)] + 1 [F(x\beta)]. \quad (7.22)$$

Für die Ableitung nach dem Vektor der erklärenden Variablen erhält man

$$\frac{\partial E(y | x)}{\partial x} = \left\{ \frac{dF(x\beta)}{dx\beta} \right\} \beta = f(x\beta) \beta, \quad (7.23)$$

wobei $f(\cdot)$ die zur Verteilungsfunktion $F(\cdot)$ zugehörige Dichte darstellt.

Für das Probit-Modell, das eine Normalverteilung unterstellt, erhält man als marginalen Effekt:

$$\frac{\partial E(y | x)}{\partial x} = \phi(x\beta) \beta, \quad (7.24)$$

mit $\phi(\cdot)$ als Dichte der Standardnormalverteilung. Für eine logistische Verteilung folgt dementsprechend

¹²⁸ Globale Konkavität bedeutet, daß die zweite Ableitung der Log-Likelihood nach dem interessierenden Parameter eine negativ definite Matrix ist. Amemiya (vgl. 1985, S. 273-274) liefert einen Beweis für das Logit- und Probit Modell.

¹²⁹ Die Überprüfung der hinreichenden Bedingungen der Log-Likelihood-Funktion erfolgt i.d.R. durch das zur Schätzung verwendete Softwarepaket.

$$\frac{\partial E(y|x)}{\partial x} = \Lambda(x\beta) [1 - \Lambda(x\beta)] \beta. \quad (7.25)$$

Es ist offensichtlich, daß die marginalen Effekte mit den Werten der erklärenden Variablen variieren. Daher werden diese bei Bedarf im jeweils betrachteten Modell anhand des Mittelwerts berechnet.

7.3.2 Modelle für geordnete Kategorien

Die vorangehenden Darstellungen bezogen sich auf den einfachen Fall einer binären abhängigen Variablen. Da die in den folgenden Schätzungen verwendete abhängige Variable ‚Zufriedenheit mit der Gesundheit‘ (siehe Abschnitt 7.3.3) elf Kategorien aufweist, bieten sich als Schätzmethode Modelle für geordnete Kategorien an.¹³⁰ Bei diskreten Daten liegt dabei eine feste Ordnung bspw. von gut nach schlecht oder von niedrig nach hoch vor. Man geht davon aus, daß der abhängigen Variablen y eine latente, unbeobachtbare Variable y^* zugrunde liegt. Bezüglich der beobachteten Variablen y wird angenommen, daß sie nur unvollständige Informationen über die latente Variable liefert. Für eine Beobachtungseinheit i folgt:

$$y_i = m, \quad \text{wenn } \tau_{m-1} \leq y_i^* \leq \tau_m \quad \text{für } m = 1 \text{ bis } J. \quad (7.26)$$

Es bezeichnen τ_m die jeweiligen Schwellenwerte, durch welche die Kategorien voneinander getrennt werden. Die Randkategorien 0 und J werden durch offene Intervalle mit $\tau_0 = -\infty$ und $\tau_J = \infty$ gekennzeichnet. Für $J=2$ bezeichnet Gleichung (7.26) die Meßgleichung für binäre Modelle (siehe Gleichung (7.8)). Die Strukturgleichung des Modells ergibt sich wie folgt:

$$y_i^* = x_i \beta + \varepsilon_i. \quad (7.27)$$

x_i stellt einen Zeilenvektor dar, der in der ersten Spalte eine Konstante enthält und für x_k in der Spalte $k+1$ die i te Beobachtung. β ist der Spaltenvektor der Koeffizienten mit dem ersten Element als Konstante β_0 und ε_i der Störterm.

¹³⁰ Siehe dazu und zu den folgenden Ausführungen z. B. auch Amemyia (1981), Maddala (1983), Ronning (1991), Long (1997 und 2001) oder Tutz (2000). Jones (2000) stellt Modelle für geordnete Kategorien in einem gesundheitsökonomischen Kontext dar. Cameron und Trivedi (1986) vergleichen Zähldatenmodelle mit Modellen für geordnete Kategorien.

Ähnlich wie bei binären abhängigen Variablen kann das Modell mit Maximum Likelihood-Methoden geschätzt werden. Dazu ist es allerdings auch hier notwendig, Annahmen über die Verteilung des Störterms zu treffen. Es werden hierbei üblicherweise die Normal- und die logistische Verteilung betrachtet. Für das Ordered Probit-Modell gilt wie zuvor, daß ε normal verteilt ist mit Erwartungswert 0 und der Varianz 1. Die Wahrscheinlichkeitsdichte ϕ ist gegeben durch

$$\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2}\right) \quad (7.28)$$

und die Verteilungsfunktion Φ durch

$$\Phi(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \quad (7.29)$$

Im Ordered Logit-Modell besitzt ε eine logistische Verteilung mit dem Erwartungswert 0 und Varianz $\pi^{2/3}$. Die Dichte λ und die Verteilung Λ sind danach:

$$\lambda(\varepsilon) = \frac{\exp(\varepsilon)}{[1 + \exp(\varepsilon)]^2}, \quad (7.30)$$

$$\Lambda(\varepsilon) = \frac{\exp(\varepsilon)}{1 + \exp(\varepsilon)}. \quad (7.31)$$

Für die Wahl, welches der beiden Verteilungsmaße man für den Fehlerterm unterstellt, gibt es keine allgemeingültige Regel, da dies Wissen über den Parametervektor β voraussetzt (vgl. Greene 2000, S. 815 und Gouriéroux 2000, S. 85).¹³¹ Lediglich für den Fall, daß bspw. Mehrgleichungssysteme normal

¹³¹ Nach Madalla (1983, S. 23) müssen die geschätzten Koeffizienten eines Logit-Modells entsprechend der Varianz von $\pi^{2/3}$ mit $3^{1/2}/\pi$ multipliziert werden, damit die Resultate mit denen einer Probit-Schätzung vergleichbar sind. Amemiya (1981, S. 1487) verwendet statt dessen den empirisch ermittelten Wert $1/1,6=0,625$.

verteilte Fehlerstrukturen unterstellen, ist man auf die Normalverteilung angewiesen (vgl. Long 1997, S. 120).¹³²

Die Wahrscheinlichkeit eines beobachteten Ergebnisses $y_i = m$ für die Realisierung des Vektors x_i kann durch folgenden Ausdruck bestimmt werden:

$$Pr(y_i = m | x_i) = F(\tau_m - x_i \beta) - F(\tau_{m-1} - x_i \beta). \quad (7.32)$$

Betrachtet man die erste Kategorie der abhängigen Variablen, so ist in diesem Fall der zweite Ausdruck auf der rechten Seite gleich Null, da für die Verteilung an dieser Stelle gilt, daß $F(\tau_0 - x\beta) = F(-8 - x\beta) = 0$. Für die Wahrscheinlichkeit der letzten Kategorie ($Pr(y = J | x)$) gilt, daß der erste Ausdruck auf der rechten Seite von (7.32) dem Wert Eins entspricht, also $F(\tau_J - x\beta) = F(8 - x\beta) = 1$ ist.¹³³

Da y^* eine latente Variable darstellt, können deren Erwartungswert und die Varianz nicht geschätzt werden. Während die Varianz durch die zugrunde gelegte Verteilung bestimmt wird, kann der Erwartungswert nicht identifiziert werden. Dies führt zu bestimmten notwendigen Annahmen über die Parametrisierung des Modells, die anhand des folgenden einfachen Beispiels hergeleitet werden können.

Man betrachte das Modell $y^* = \alpha + \beta x + \varepsilon$ mit den Schwellenwerten τ_m . Dabei seien α und τ_m die wahren Parameter des datengenerierenden Prozesses. Man definiert nun eine alternative Menge von Parametern wie folgt:

$$\alpha^* = \alpha - \delta \quad \text{und} \quad \tau_m^* = \tau_m - \delta, \quad (7.33)$$

mit δ als frei wählbarer Konstante. Die Wahrscheinlichkeit, daß die abhängige Variable y den Wert m annimmt, ist identisch, unabhängig davon, ob die wahren oder alternativen Parameter unterstellt werden:

¹³² Die später präsentierten Querschnittsschätzungen legen Ordered Probit-Modell zugrunde. Diese Wahl beruht im wesentlichen darauf, daß sich für die Panel-Analyse lediglich ein Probit-Modell anbietet (siehe dazu Abschnitt 7.4.3).

¹³³ Es sei angemerkt, daß F als Platzhalter für eine der beiden Verteilungsfunktionen (Φ oder Λ) steht.

$$\begin{aligned}
Pr(y_i = m | x_i) &= F(\tau_m - \alpha - \beta x) - F(\tau_{m-1} - \alpha - \beta x) \\
&= F([\tau_m - \delta] - [\alpha - \delta] - \beta x) \\
&\quad - F([\tau_{m-1} - \delta] - [\alpha - \delta] - \beta x) \\
&= F(\tau_m^* - \alpha^* - \beta x) - F(\tau_{m-1}^* - \alpha^* - \beta x).
\end{aligned} \tag{7.34}$$

Da beide Parameter-Konstellationen den gleichen Wert für die Wahrscheinlichkeit eines beobachteten Ergebnisses für die Variable y generieren, gibt es keine Möglichkeit, zwischen den beiden Konstellationen zu wählen. Eine Veränderung der Konstanten des Strukturmodells wird dabei immer durch eine entsprechende Veränderung der Schwellenwerte kompensiert.¹³⁴ Daher werden i.d.R. bei Verwendung dieser Modelle zwei Annahmen benutzt, um das Modell vollständig identifizieren zu können:

1. Angenommen, es gilt $\tau_1^* = 0$. Dann wird in Gleichung (7.33) $\tau_1 = \delta$ gesetzt.

2. Falls $\alpha^* = 0$, so folgt in Gleichung (7.33), daß $\alpha = \delta$ ist.

Dies bedeutet, daß bei der Schätzung entweder die Konstante oder der Schwellenwert τ_1 gleich Null gesetzt werden. Durch beide Annahmen läßt sich das Modell identifizieren, da für einen der beiden Parameter eine Restriktion formuliert wird. Aufgrund dieser Annahmen ergeben sich verschiedene Parametrisierungen des Modells. Die Wahl zwischen beiden beeinflußt dabei weder die Koeffizienten β noch zugehörige Signifikanztests. Weiterhin besteht kein Effekt auf die Wahrscheinlichkeit für den Eintritt eines bestimmten Ereignisses.¹³⁵

Das vorliegende Modell wird mit Hilfe der Maximum Likelihood-Technik geschätzt. β ist der zu schätzende Parametervektor des Strukturmodells und τ der Vektor der Schwellenparameter, wobei entweder für β_0 oder für τ_1 der Wert Null angenommen wird. Aus Gleichung (7.32) folgt dann:

$$Pr(y_i = m | x_i, \beta, \tau) = F(\tau_m - x_i \beta) - F(\tau_{m-1} - x_i \beta). \tag{7.35}$$

¹³⁴ Dies bedeutet, daß das Modell nicht identifizierbar ist (vgl. Long 1997, S. 123).

¹³⁵ Wichtig sind diese unterschiedlichen Parametrisierungen vor allem deshalb, da ökonometrische Software keine einheitliche Parametrisierungen verwendet.

Die Wahrscheinlichkeit, daß die *i*te Beobachtung von *y* einen bestimmten Wert annimmt, ist dann:¹³⁶

$$p_i = \begin{cases} Pr(y_i = 1 | x_i, \beta, \tau) & \text{wenn } y = 1 \\ \vdots & \vdots \\ Pr(y_i = m | x_i, \beta, \tau) & \text{wenn } y = m \\ \vdots & \vdots \\ Pr(y_i = J | x_i, \beta, \tau) & \text{wenn } y = J \end{cases} \quad (7.36)$$

Unter der Annahme der Unabhängigkeit der Beobachtungen folgt für die Likelihood-Funktion:

$$L(\beta, \tau | y, X) = \prod_{i=1}^N p_i. \quad (7.37)$$

Zusammen mit Gleichung (7.35) ergibt sich aus (7.37):

$$\begin{aligned} L(\beta, \tau | y, X) &= \prod_{j=1}^J \prod_{y_i=j} Pr(y_i = j | x_i, \beta, \tau) \\ &= \prod_{j=1}^J \prod_{y_i=j} F(\tau_m - x_i \beta) - F(\tau_{m-1} - x_i \beta). \end{aligned} \quad (7.38)$$

Die zugehörige Log Likelihood folgt durch Logarithmierung von (7.38) als:

$$\ln L(\beta, \tau | y, X) = \sum_{j=1}^J \sum_{y_i=j} \ln \left[F(\tau_m - x_i \beta) - F(\tau_{m-1} - x_i \beta) \right]. \quad (7.39)$$

Die letzte Gleichung läßt sich dann nach den Parametern β und τ maximieren. Die resultierenden Schätzer sind konsistent, asymptotisch normalverteilt und asymptotisch effizient.

¹³⁶ Es ist hierbei zu beachten, daß es keine Rolle spielt, ob die kategoriale Variable die Wert 1 bis *J* annimmt oder wie die Zufriedenheit mit der Gesundheit im vorliegenden Datensatz 0 bis 10 bzw. 0 bis *J*-1.

Gütemaße

Ähnlich wie für die Schätzungen kann auch das herkömmliche Bestimmtheitsmaß R^2 kann für die Überprüfung der Güte einer Schätzung im Fall von qualitativen abhängigen Variablen nicht herangezogen werden. Statt dessen wurden eine Reihe sogenannter Pseudo- R^2 konstruiert, die einen Vergleich der Güte der Schätzungen erlauben.¹³⁷ In der Literatur existiert eine Vielzahl dieser Maße, so daß ein bevorzugter Gebrauch eines der Maße nicht gegeben ist. An dieser Stelle sollen zwei dieser Maße näher vorgestellt werden, die auch in der späteren Untersuchung Verwendung finden: Zum einen ist dies das Maß von McFadden, zum anderen das McKelvey/Zavoina-Maß, welches vor allem in den Sozialwissenschaften verbreitet ist. Weitere bekannte Maße sind das Maß von Aldrich und Nelson sowie das von Veall und Zimmermann. Letzteres stellt eine Modifikation des Maßes von Aldrich und Uhler dar. Darüber hinaus existieren die Maße von Maddala sowie von Cragg und Uhler (vgl. hierzu Veall und Zimmermann (1992) und (1996) und Long (1997)).

Ausgangspunkt für das Pseudo- R^2 -Maß von McFadden ist die Likelihood-Ratio-Teststatistik, die durch folgenden Ausdruck gegeben ist:

$$LRT = 2 \left(l_M - l_0 \right). \quad (7.40)$$

Dabei bezeichnet l_M den Wert der Log-Likelihood des Modells und l_0 den Wert der Log-Likelihood des Modells unter der Restriktion, daß alle Koeffizienten gleich Null sind. Das McFadden-Maß benutzt nun die beiden Log-Likelihood-Werte und entspricht:

$$R_{MF}^2 = 1 - \frac{l_M}{l_0}. \quad (7.41)$$

Dieses Maß erreicht den Wert Null, wenn der Wert von l_M nahe dem von l_0 ist, und ein Maximum von Eins für $l_M = 0$. Da dieses Maß ähnlich wie das aus der OLS-Regression bekannte R^2 mit der Anzahl der erklärenden Variablen ansteigt, wurde ein korrigiertes Maß (R_{MF}^2) entwickelt, das nur dann ansteigt, wenn l_M um mehr als Eins für jeden neuen Parameter ansteigt:

¹³⁷ In der Praxis wird jedoch häufig kein Pseudo- R^2 benutzt. Einige Autoren verweisen darauf, daß kein allgemein anerkanntes Maß vorliegt (vgl. Veall und Zimmermann, 1992, S. 333).

$$R_{MF}^2 \text{ }^{adj} = 1 - \frac{(l_M - K)}{l_0}, \quad (7.42)$$

wobei K die Gesamtzahl der geschätzten Parameter angibt. Während das Maß von McFadden der Klasse der „Significance-of-Fit“-Maße (Veall und Zimmermann 1996, S. 247) zugeordnet werden kann, gehört das Pseudo- R^2 nach McKelvey und Zavoina (vgl. 1975, S. 111-113) zur „explained variation class“ (Veall und Zimmermann 1996, S. 247). Ausgehend von der standardnormalverteilten Fehlerstruktur des Ordered Probit-Modells, ist die erwartete Summe der quadrierten Fehler gleich der Anzahl der Beobachtungen, N .¹³⁸ Die ‚Varianz‘ der Regression EV (explained sum of squares) kann berechnet werden als:

$$EV = \sum_{i=1}^N \left(\hat{Y}_i^* - \hat{\bar{Y}}_i^* \right), \quad (7.43)$$

wobei \hat{Y}_i^* die bedingte Erwartung der unbeobachteten Variablen Y_i^* ist. Für deren Mittelwert $\hat{\bar{Y}}_i^*$ gilt, daß

$$\hat{\bar{Y}}_i^* = \frac{1}{N} \sum_{i=1}^N \hat{Y}_i^*. \quad (7.44)$$

Unter Verwendung der Gleichungen (7.43) und (7.44) ergibt sich, daß das Pseudo- R^2 nach McKelvey und Zavoina (R_{MZ}^2) wie folgt:

$$R_{MZ}^2 = \frac{\sum_{i=1}^N \left(\hat{Y}_i^* - \hat{\bar{Y}}_i^* \right)^2}{\sum_{i=1}^N \left(\hat{Y}_i^* - \hat{\bar{Y}}_i^* \right)^2 + N}. \quad (7.45)$$

R_{MZ}^2 ist demnach der Quotient aus der ‚Varianz der Regression‘ und der ‚Varianz der Beobachtungen‘, d.h. der Nenner gibt die Summe aus der ‚Varianz der Regression‘ und den erwarteten quadrierten Fehlern an.

¹³⁸ Dies folgt daraus, daß die Varianz der Standardnormalverteilung gleich Eins ist. Im Fall einer Ordered-Logit-Schätzung ergibt sich die Summe der quadrierten Fehler als N multipliziert mit $\pi^{2/3}$, der Varianz der logistischen Verteilung (vgl. Long 1997, S. 105).

In einer Studie über Pseudo-R²-Maße im Fall einer Ordered Probit-Schätzung ermittelten Veall und Zimmermann (1992), daß vor allem das Maß von McKelvey und Zavoina dem R² im Fall stetiger abhängiger Variablen am nächsten kommt. Weiterhin fanden sie mit Hilfe von Simulationstechniken heraus, daß einige der Pseudo-R²-Maße das stetige R² stark unterschätzen. Speziell im Fall des Maßes von McFadden steigt die Unterschätzung mit der Anzahl der Kategorien der abhängigen Variable an.

Neben den Pseudo-R²-Maßen existieren noch sogenannte Informationsmaße, wie Akaike's Information Criterion (AIC) und das Bayesian Information Criterion (BIC), von denen auf das erste hier genauer eingegangen wird. (vgl. Cramer 1991, S. 96 und Long 1997, S. 109ff.). Für eine Verwendung dieses Maßes spricht, daß die resultierenden Erkenntnisse bezüglich der Güte der Schätzung u.U. von denen der Pseudo-R²-Maße abweichen und somit eine differenziertere Betrachtung der Schätzergebnisse ermöglichen. Akaike's Maß ist definiert als

$$AIC = \frac{-2 l_M + 2P}{N}, \quad (7.46)$$

wobei l_M die Log-Likelihood des Modells bezeichnet und P für die Anzahl der Parameter des Modells, d.h. die Koeffizienten und Schwellenwerte, steht.¹³⁹ Der Log-Likelihood-Wert l_M wird, absolut betrachtet, kleiner, je besser die Anpassungsgüte des Modells ist. Der Ausdruck $-2l_M$ hat daher einen Wertebereich von 0 bis $+\infty$. Je größer nun die Anzahl der zu schätzenden Parameter wird, desto kleiner wird dieser Ausdruck, da durch die Hinzunahme weiterer erklärender Variablen die Realisation der beobachteten Daten wahrscheinlicher wird. Der Ausdruck $2P$ wird nun als Korrektur für dieses Verhalten hinzugefügt und der gesamte Ausdruck durch die Anzahl der Beobachtungen N geteilt. Das AIC-Maß wird häufig benutzt, um Modelle über verschiedene Stichproben zu vergleichen.¹⁴⁰ Kleinere AIC-Werte geben c.p. Hinweise auf eine bessere Modelanpassung an die Daten.

¹³⁹ Amemiya (vgl. 1981, S. 1505) definiert das AIC-Maß als $-l_M + P$, d.h., er vernachlässigt die Korrektur mit dem Faktor $2/N$.

¹⁴⁰ Alternativ kann es auch verwendet werden, um nicht-genestete Modelle, die nicht mit einem Likelihood-Ratio-Test verglichen werden können, zu beurteilen (vgl. Long 1997, S. 110).

7.3.3 Daten

Zur Schätzung des Einflusses der medizinischen Leistung und der Compliance auf die Gesundheit werden Daten des Sozioökonomischen Panels (SOEP) herangezogen.¹⁴¹ Hieraus bietet sich besonders die Welle des Jahres 1986 für eine Analyse an, da neben der Nachfrage nach medizinischen Leistungen auch Variablen, die das Wohnumfeld betreffen, in dieser Welle enthalten sind. Tabelle 7.1 gibt einen Überblick über die verwendeten Variablen.

Die *abhängige Variable* ist die selbst eingeschätzte Zufriedenheit mit der Gesundheit als Proxy für die latente Variable Gesundheit. Dabei handelt es sich um eine ordinal skalierte Variable mit elf Ausprägungen, von 0 = ganz und gar nicht zufrieden bis 10 = ganz und gar zufrieden.

Die *erklärenden Variablen* lassen sich in verschiedene Gruppen einteilen. Zum einen sind dies die *prädisponierenden Variablen*. Zu dieser Gruppe gehören Variablen, die persönliche Charakteristika des Individuums aufweisen wie bspw. Alter, Geschlecht, Familienstand, Kinder und Nationalität. Bezüglich des Alters wird auch ein quadratischer Altersterm untersucht, der für einen nichtlinearen Zusammenhang zwischen dem ‚Alter‘ und der ‚Zufriedenheit mit der Gesundheit‘ steht. Es ist davon auszugehen, daß mit zunehmendem Alter die Zufriedenheit mit der eigenen Gesundheit zurückgeht. Die Dummy-Variable ‚Geschlecht‘ gibt an, ob es sich bei der befragten Person um eine Frau (1) oder einen Mann (0) handelt. Das Vorzeichen der Variable ist a priori unklar. Empirische Studien zeigen zum einen, daß bei Frauen die Zufriedenheit mit der Gesundheit geringer ist, da sie mehr medizinische Leistungen in Anspruch nehmen als Männer (vgl. Breyer und Zweifel 1999, S. 129-146 und Leu und Gerfin (1992)). Bei anderen Untersuchungen ergab sich, daß der Gesundheitszustand von Männern aufgrund der geringeren Lebenserwartung schlechter als bei Frauen ist (vgl. Gerdtham und Johannesson 1997, S. 8 und Gerdtham und Johannesson 1999, S. 1328). Die Variable ‚alleinstehend‘ nimmt für Personen, die von ihrem Partner getrennt leben (d.h. ledig, geschieden, verwitwet sind bzw. deren Ehepartner sich im Ausland befindet), den Wert Eins an. Bisherige Untersuchungen lassen vermuten, daß der Effekt auf den Gesundheitszustand negativ sein dürfte.¹⁴² Die Variable ‚Kinder‘, die angibt, ob dem Haushalt Kinder unter 16 Jahren angehören, soll den Effekt etwaiger Belastungen durch die Erziehung auf die Zufriedenheit mit der Gesundheit oder vorhandene positive

¹⁴¹ Die in dieser Arbeit verwendeten Daten des Sozioökonomischen Panels (SOEP) wurden vom Deutschen Institut für Wirtschaftsforschung (DIW) Berlin bereitgestellt.

¹⁴² So schließt Doppmann (vgl. 1985, S. 53) aus der höheren Sterblichkeitsrate Alleinstehender, daß diese einen ungesünderen Lebensstil aufweisen als Verheiratete.

Nutzenwirkungen abbilden. Für die Variable ‚Ausländer‘ kann keine Hypothese über den Effekt auf die Zufriedenheit mit der Gesundheit aufgestellt werden. Kenkel (vgl. 1995, S. 20) ermittelt für Nichtweiße einen negativen Effekt auf den Gesundheitszustand.

Tabelle 7.1: Variablenbeschreibung der Querschnittsanalyse

prädispon. Variablen	
Alter	Alter in Jahren
Alter ²	Alter in Jahren quadriert
Geschlecht	1, wenn weiblich
Kinder	1, wenn Kinder unter 16 Jahren im Haushalt
alleinstehend	1, wenn getrennt lebend, ledig, geschieden, Partner verstorben, Ehepartner im Ausland und somit keine Lebensgemeinschaft
Ausländer	1, wenn Staatsangehörigkeit nicht deutsch
sozioök. Variablen	
Einkommen	monatliches Haushaltsnettoeinkommen in 1000 DM
Schulbildung	1, wenn mindestens mittlere Reife oder Besuch einer weiterführenden Schule im Ausland
Sport	1, wenn jede Woche aktiver Sport
PKV	1, wenn Private Krankenversicherung
Gesundheitslage	
Gesundheit	Zufriedenheit mit der Gesundheit 0-10
chronisch	1, wenn leiden unter chronischen Krankheiten
keine Behinderung	1, wenn Erfüllung täglicher Aufgaben durch den Gesundheitszustand überhaupt nicht behindert
Arbeitsunfähigkeit	1, wenn 14 Tage oder länger krank geschrieben (1985)
Medizinische Inanspruchnahme	
prak. Arzt	1, wenn Besuch beim praktischen Arzt im letzten Quartal
Facharzt	1, wenn Besuch beim Facharzt im letzten Quartal
KH-Aufenthalte	1, wenn Krankenhausaufenthalt (1985)
Wohnumfeld	
Entf. Arzt	1, wenn der Hausarzt nicht zu Fuß erreichbar
Entf. Zentrum	1, wenn mehr als 25 km Entfernung zum Stadtzentrum
Luftverschmutzung	1, wenn Luftverschmutzung stark oder sehr stark
Lärmbelästigung	1, wenn Lärmbelästigung stark oder sehr stark
Arbeitsbedingungen	
Anspannung	1, wenn hohe nervliche Anspannung am Arbeitsplatz (1985)
körp. Belastung	1, wenn körperlich schwere Arbeit (1985)
Nachtschicht	1, wenn regelmäßige Nachtarbeit (1985)
Wechselschicht	1, wenn Arbeit in Wechselschicht (1985)
Umweltbelastung	1, wenn belastende Umwelteinflüsse am Arbeitsplatz (1985)

Die zweite Gruppe bilden die *sozioökonomischen Variablen*. Hierzu gehören die Schulbildung, der Krankenversicherungsschutz, das monatliche Haushalts-

nettoeinkommen und die sportliche Betätigung. Ob und inwiefern eine bessere Schulbildung sich positiv auf die Zufriedenheit mit der Gesundheit auswirkt, ist a priori unklar.¹⁴³ Man kann jedoch vermuten, daß eine bessere Schulbildung zu einem bewußteren Umgang mit der eigenen Gesundheit führt und daher im Sinne einer höheren Compliance interpretiert werden kann (vgl. hierzu auch die Untersuchung von Gerdtham und Johannesson (1999)). Die Variable ‚Schulbildung‘ nimmt dabei den Wert Eins an, wenn das Individuum mindestens einen Realschulabschluß oder den Besuch einer weiterführenden Schule im Ausland aufweisen kann. Die Variable ‚PKV‘, d.h., ob ein Individuum privat krankenversichert ist, soll überprüfen, ob durch die unterschiedliche Absicherung auch die Gesundheit beeinflußt wird.¹⁴⁴ Dies geschieht durch eine Dummy-Variable, die für PKV-Versicherte den Wert Eins annimmt. Das Haushaltsnettoeinkommen in tausend DM steht als Proxy für die Konsummöglichkeiten des Individuums.¹⁴⁵ Unklar bleibt jedoch, ob ein positiver oder negativer Einfluß des Einkommens auf die Zufriedenheit mit der Gesundheit besteht, da auch andere Schätzungen zu unterschiedlichen Ergebnissen kommen (siehe hierzu Leu und Dopmann (1986) oder auch Pohlmeier und Ulrich (1992)).

Die dritte Gruppe ist die *Gesundheitslage* des Individuums. Neben der Zufriedenheit mit dem eigenen Gesundheitszustand, die als abhängige Variable dient, sind dies erstens das Leiden unter chronischen Krankheiten, zweitens die Nichtbehinderung durch den Gesundheitszustand bei der Erfüllung täglicher Aufgaben und drittens eine mehr als 14tägige Arbeitsunfähigkeit bei Erwerbstätigen im letzten Kalenderjahr. Diese Variablengruppe dient maßgeblich als Korrekturfaktor, da ansonsten die Effekte der anderen Variablen möglicherweise überschätzt werden. Es ist davon auszugehen, daß alle Variablen bis auf die ‚Nicht-Behinderung‘ ein negatives Vorzeichen aufweisen.

¹⁴³ Einen positiven Zusammenhang zwischen der Erhöhung des Humankapitals durch die Ausbildung und der Gesundheit findet sich in Grossman (1972) und (2000). Harris und Harris (1996) argumentieren, daß Personen mit höherer Ausbildung in der Zukunft ein höheres Renteneinkommen zu erwarten haben und daß daher der Verlust an Lebenseinkommen durch einen vorzeitigen Tod hoch ist. Daneben besitzen Gebildete eine höhere interne Diskontrate, so daß sie alles zukünftige Einkommen höher gewichten als Menschen mit geringen Investitionen in das Humankapital. Daher weisen Personen mit höherer Schulbildung ein größeres Interesse an der Gesundheit auf.

¹⁴⁴ In der Regel dürfte dies über die Inanspruchnahme geschehen, jedoch ist nicht auszuschließen, daß eine bessere Absicherung per se einen positiven Effekt im Sinne eines Optionskonsums medizinischer Leistungen hat (vgl. Wille 1980, S. 143).

¹⁴⁵ Dabei wurden Beobachtungen mit einem angegebenen Haushaltsnettoeinkommen vom mehr als 20.000 DM im Monat aus dem Datensatz entfernt, um Verzerrungen durch diese Beobachtungen zu vermeiden.

Die *medizinische Inanspruchnahme* bildet die vierte Gruppe von erklärenden Variablen, zu der ein Besuch beim Haus- und Facharzt im letzten Quartal sowie die Krankenhausaufenthalte im letzten Kalenderjahr gehören. Alle drei Variablen sind binäre Variablen, die den Wert Eins annehmen, wenn im Quartal vor der Befragung bzw. für die Krankenhausaufenthalte im Jahr zuvor ein praktischer Arzt oder Facharzt aufgesucht wurde oder ein Krankenhausaufenthalt vorlag.¹⁴⁶ Für alle drei Variablen kann a priori kein eindeutiger Effekt bestimmt werden. Theoretische Überlegungen lassen den Schluß zu, daß aufgrund der erfolgreichen Inanspruchnahme medizinischer Leistungen die Zufriedenheit mit der Gesundheit hoch ist (vgl. Grossman 1972). Empirische Studien zur Gesundheitsproduktion zeigen jedoch, daß die Behandlung nicht immer zum Erfolg (Heilung) führt oder daß sie zum Zeitpunkt der Datenerhebung noch nicht abgeschlossen ist. Dann ist es möglich, daß die Zufriedenheit mit der Gesundheit trotz der Inanspruchnahme gering ist. Aufgrund der vorliegenden Daten erweist es sich als problematisch, für die medizinische Inanspruchnahme neben deren Umfang auch Indikatoren über den Behandlungsaufwand zu integrieren. Aus diesem Grund ist eine Trennung dieser Effekte nicht durchführbar.

Die fünfte Gruppe beschreibt das *Wohnumfeld*. Zum einen sind dies Lärmbelastungen und Luftverschmutzung, die auf höhere Unzufriedenheit mit der Gesundheit schließen lassen, zum anderen die Entfernung zum Hausarzt und die Entfernung zum nächsten Großstadtzentrum. Alle vier Variablen sind als Dummy-Variablen modelliert. Die erste Variable (Entfernung zum Hausarzt) mißt die Erreichbarkeit und ist damit ein Proxy für die Opportunitätskosten eines Arztbesuches. Wenn der Hausarzt nicht zu Fuß erreichbar ist, dann wird die potentielle Inanspruchnahme eingeschränkt und die Zufriedenheit mit der Gesundheit wird möglicherweise zurückgehen. Die zweite Variable, die Entfernung zum Großstadtzentrum, soll die Unterschiede zwischen Stadt- und Landbevölkerung auffangen. Zwar ist das medizinische Angebot in einer Großstadt bzw. in einem Ballungsraum besser als in ländlichen Gegenden, allerdings dürfte auch die Wohnqualität geringer sein, was sich negativ auf das Gesundheitsempfinden auswirken könnte. Die Luftverschmutzung sowie die Lärmbelastung bilden die Umweltbelastungen am Wohnort ab. Es ist davon auszugehen, daß ein negativer Einfluß auf die Gesundheit besteht.

¹⁴⁶ Alternativ zu den Krankenhausaufenthalten bzw. der Arbeitsunfähigkeit im Vorjahr sind auch Daten für 1986 aus der vierten Welle des SOEP verfügbar. Da diese allerdings das gesamte Kalenderjahr 1986 umfassen, die Befragung aber bereits im August stattfand, läge zumindest ein Teil der Aufenthalte in einem zukünftigen Zeitraum. Aus diesem Grund wurde die relativ lange zeitliche Verzögerung bevorzugt.

Die letzte Gruppe sind die *Bedingungen am Arbeitsplatz*. Dazu zählen hohe nervliche Anspannung, körperliche Belastung, Umweltbelastungen und Nacht- oder Wechselschicht. Für diese Variablen würde man erwarten, daß die Zufriedenheit mit der Gesundheit zurückgeht, je stärker diese ausgeprägt sind. Da sie aus der Welle des Jahres 1985 stammen, wurden alle Erwerbstätigen, die zwischen den beiden Erhebungen ihren Arbeitsplatz wechselten, aus der Stichprobe herausgenommen.

Die Schätzungen wurden mit drei Stichproben durchgeführt. Die erste enthält alle Befragten, umfaßt also 4880 Beobachtungen, und schließt die Arbeitsplatzvariablen und die Arbeitsunfähigkeit als erklärende Variablen aus. Die zweite Stichprobe bezieht sich auf die Nichterwerbstätigen und umfaßt 1940 Beobachtungen. Sie beinhaltet den gleichen Variablensatz wie die erste Schätzung. Die dritte Stichprobe umfaßt die Erwerbstätigen, so daß für diese Gruppe mit 2613 Beobachtungen alle Variablen in die Schätzung mit einbezogen werden konnten. Aus der Gruppe der Erwerbstätigen wurden alle diejenigen herausgerechnet, die zwischen den Befragungen von 1985 und 1986 den Arbeitsplatz gewechselt hatten. Weiterhin fehlen Arbeitslose und Befragte, die angaben, im ersten Beruf selbständig zu sein, da für diese die Arbeitsplatzcharakteristika im Datensatz entweder nicht vorliegen oder als nicht aussagekräftig zu beurteilen sind. Für diese Teilstichprobe wurde zusätzlich eine Schätzung inklusive der Arbeitsplatzcharakteristika und der Arbeitsunfähigkeit durchgeführt.

Tabelle 7.2 faßt die erwarteten Einflüsse auf den Gesundheitszustand zusammen. Dabei fällt auf, daß vor allem für die prädisponierenden Variablen keine genauen Aussagen über den erwarteten Einfluß der Variablen gemacht werden können. Mit Ausnahme des Alters gibt es keine Anhaltspunkte aus theoretischen oder empirischen Untersuchungen, aufgrund derer ein klarer Effekt abgeleitet werden könnte. Lediglich für den Einfluß der Variable ‚Geschlecht‘ auf die Zufriedenheit mit der Gesundheit finden sich in der Literatur Ergebnisse, die jedoch nicht einheitlich sind. Für die sozioökonomischen Variablen wird ein positiver Einfluß unterstellt, insbesondere vor dem Hintergrund, daß Bildung und Sport als Approximation der Compliance des Patienten aufgefaßt werden können. Für die Variablen der Gesundheitslage gilt, daß chronische Krankheiten und Arbeitsunfähigkeit einen negativen Einfluß auf die Gesundheitszufriedenheit ausüben dürften. Das Vorzeichen der medizinischen Inanspruchnahme bleibt unklar. Es finden sich sowohl Anhaltspunkte für einen positiven wie für einen negativen Effekt. Für sämtliche Variablen des Wohnumfeldes und der Arbeitsbedingungen wird für die Schätzungen ein negativer Einfluß erwartet.

Tabelle 7.2: Erwarteter Einfluß der erklärenden Variablen (Querschnittsdatensatz)

prädispon. Variablen		med. Inanspruchnahme	
Alter	-	prak. Arzt	+/-
Alter ²	?	Facharzt	+/-
Geschlecht	+/-	KH-Aufenthalte	+/-
alleinstehend	?	Wohnumfeld	
Kinder	?	Entf. Hausarzt	-
Ausländer	?	Entf. Großstadt	-
sozioök. Variablen		Luftverschmutzung	-
Einkommen	+	Lärmbelästigung	-
Schulbildung	+	Arbeitsbedingungen	
PKV	+	Anspannung	-
Sport	+	körp. Belastung	-
Gesundheitslage		Nachtschicht	-
chronisch	-	Wechselschicht	-
keine Behinderung	+	Umweltbelastung	-
Arbeitsunfähigkeit	-		

Tabelle 7.3 enthält die deskriptiven Statistiken des Querschnittsdatensatzes, aufgeschlüsselt nach den jeweiligen Stichproben. Dabei fällt zuerst das unterschiedliche Durchschnittsalter in den einzelnen Stichproben auf. Für die Gesamtstichprobe liegt es bei 42,4 Jahre, bei den Nichterwerbstätigen bei ca. 49 Jahren und bei den Erwerbstätigen bei nicht einmal 39 Jahren. Auch der Anteil der Frauen ist zwischen den Stichproben unterschiedlich. So beträgt er in der Gesamtstichprobe 48,5 %, bei den Nichterwerbstätigen jedoch knapp 69 % und bei den Erwerbstätigen nur 34,6 %.

Bei der Gruppe der sozioökonomischen Variablen fallen die Unterschiede besonders für das durchschnittliche Einkommen auf. Das höchste durchschnittliche Haushaltseinkommen besitzen die Erwerbstätigen mit 3481 DM netto im Monat, für die Nichterwerbstätigen sind es lediglich 2703 DM. Weiterhin betreibt diese Gruppe weniger aktiven Sport. Nur 18,4 % gaben an, regelmäßig sportlich aktiv zu sein, bei den Erwerbstätigen waren es mehr als 26 %. Ähnlich ist das Bild bei der Gesundheitslage. Die durchschnittliche Zufriedenheit mit der Gesundheit liegt für Nichterwerbstätige unter der für die Erwerbstätigen und der Gesamtstichprobe, weist allerdings einen höheren Anteil chronischer Gesundheitsbeschwerden auf. Bei den Erwerbstätigen gaben mehr als 70 % der Befragten an, daß keine Behinderung bei der Ausübung täglicher Arbeiten vorliegt, bei den Nichterwerbstätigen sind es nur 54,3 %. Darüber hinaus suchen Nichterwerbstätige öfter einen Arzt auf oder weisen eher einen Krankenhausaufenthalt auf.

Tabelle 7.3: Deskriptive Statistik des Querschnittsdatensatzes

Stichprobe Variable	gesamt		nicht erwerbstätig		erwerbstätig	
	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung
prädispon. Variablen						
Alter	42,447	16,238	49,018	19,211	38,707	12,001
Alter ²	2065,40	1525,16	2771,65	1910,89	1642,21	947,50
Geschlecht	0,485	0,500	0,689	0,463	0,346	0,476
alleinstehend	0,305	0,461	0,322	0,467	0,277	0,447
Kinder	0,425	0,494	0,378	0,485	0,453	0,498
Ausländer	0,243	0,429	0,202	0,401	0,272	0,445
sozioök. Variablen						
Einkommen	3,176	1,515	2,703	1,422	3,481	1,468
Schulbildung	0,314	0,464	0,270	0,444	0,336	0,472
PKV	0,091	0,287	0,079	0,270	0,098	0,298
Sport	0,234	0,423	0,184	0,387	0,263	0,440
Gesundheitslage						
Gesundheit	6,940	2,445	6,460	2,680	7,223	2,217
chronisch	0,286	0,452	0,387	0,487	0,222	0,416
keine Behinderung	0,654	0,476	0,543	0,498	0,722	0,448
Arbeitsunfähigkeit	-	-	-	-	0,178	0,383
med. Inanspruchnahme						
prak. Arzt	0,464	0,499	0,534	0,499	0,416	0,493
Facharzt	0,467	0,499	0,545	0,498	0,415	0,493
KH-Aufenthalte	0,107	0,309	0,139	0,346	0,080	0,272
Wohnumfeld						
Entf. Hausarzt	0,116	0,321	0,115	0,319	0,114	0,318
Entf. Großstadt	0,315	0,465	0,303	0,460	0,319	0,466
Luftverschmutzung	0,118	0,323	0,114	0,318	0,121	0,326
Lärmbelästigung	0,127	0,333	0,140	0,347	0,118	0,323
Arbeitsbedingungen						
Anspannung	-	-	-	-	0,238	0,426
körperliche Belastung	-	-	-	-	0,160	0,367
Nachtschicht	-	-	-	-	0,063	0,243
Wechselschicht	-	-	-	-	0,122	0,328
Umweltbelastung	-	-	-	-	0,214	0,410

7.3.4 Schätzergebnisse

Die hier präsentierten Ergebnisse beziehen sich auf die drei unterschiedlichen Stichproben ‚gesamt‘, ‚nicht erwerbstätig‘ und ‚erwerbstätig‘. Dabei handelt es sich jeweils um eine Ordered Probit-Schätzung, wie sie im Abschnitt 7.3.2 dargestellt wurde. Allen Schätzungen liegen als unabhängige Variablen die gleichen prädisponierenden und sozioökonomischen Variablen zugrunde. Darüber hinaus gehen die medizinische Inanspruchnahme, das Wohnumfeld und die Gesundheitslage mit Ausnahme der Variablen ‚Arbeitsunfähigkeit‘ in alle Schätzungen ein. Um eine Fehlspezifikation des Modells zu vermeiden, werden robuste Schätzungen mit einer heteroskedastie-konsistenten Kovarianzmatrix der Parameter durchgeführt.¹⁴⁷

Schätzung Gesamtstichprobe

Für die Gesamtstichprobe ergibt sich, daß das ‚Alter‘ einen signifikant negativen Einfluß auf den Gesundheitszustand aufweist, wenn auch mit abnehmender Rate, was aus dem positiven Koeffizienten der Variable ‚Alter²‘ hervorgeht (vgl. zu den Ergebnissen die Schätzung (1), Spalten 2 und 3 in Tabelle 7.4, S. 175). Dies bedeutet, daß mit zunehmendem Lebensalter die Zufriedenheit mit der Gesundheit bis zu einer bestimmten Grenze zurückgeht, um danach wieder anzusteigen. Es besteht somit ein u-förmiger Zusammenhang zwischen der Zufriedenheit mit der Gesundheit und dem Alter. Aus den Parameterschätzungen folgt, daß die Zufriedenheit mit der Gesundheit für die Gesamtstichprobe im Alter von 61,5 Jahren am geringsten ist, für die Nichterwerbstätigen im Alter von ca. 61,3 Jahren und für die Erwerbstätigen im Alter von ca. 53 Jahren, was u.U. die Arbeitsbelastung widerspiegelt. Dieses Ergebnis entspricht insgesamt den Erwartungen, da zunehmendes Alter auch mit Multimorbidität einhergeht.¹⁴⁸ Daß im hohen Alter der positive Effekt des quadratischen Ausdrucks überwiegt, kann seine Gründe auch in einer Gewöhnung an den Umgang mit Krankheiten, etwa in einer Anpassung des Anspruchsniveaus an die Gesundheit, haben.

¹⁴⁷ Der Vorteil von robusten Standardfehlern liegt darin begründet, daß selbst für den Fall einer falschen Verteilungsannahme die Pseudo-Maximum-Likelihood-Schätzung konsistent ist, wenn zum einen die Mittelwertfunktion richtig spezifiziert ist und zum anderen die verwendete Verteilung zur Klasse der linear exponentiellen Familie gehört (vgl. Gouriéroux, Monfort und Truong (1984)).

¹⁴⁸ Weiterhin zeigen Studien, daß mit zunehmendem Alter auch die Nachfrage nach medizinischen Leistungen zunimmt und somit die Gesundheitsausgaben ansteigen (vgl. Pohlmeier und Ulrich (1996)) bzw. daß durch den Restlebenszeiteffekt das Alter seinen Einfluß auf die Gesundheitsausgaben verliert (vgl. hierzu Zweifel, Felder und Meier (1996), Breyer (1996)). Felder et al. (2000) zeigen, daß innerhalb der letzten zwei Lebensjahre die Gesundheitsausgaben ansteigen.

Frauen sind im Vergleich zu Männern weniger mit ihrem Gesundheitszustand zufrieden, was der negative Koeffizient in Spalte 1 belegt. Allerdings erweist sich dieser Befund als nicht signifikant.¹⁴⁹ Die Variable ‚alleinstehend‘ weist ein positives Vorzeichen auf, ist aber nicht signifikant. Ausländer sind mit ihrem Gesundheitszustand zufriedener als Deutsche, was u.U. damit zusammenhängt, daß in Deutschland lebende Ausländer eine Art positive Selektion darstellen, zumindest sofern sie am Arbeitsmarkt aktiv sind.¹⁵⁰ Allerdings ist dieser Parameter nicht signifikant.

Für die sozioökonomischen Variablen ergibt sich folgendes Bild (siehe Tabelle 7.4): Das Haushaltsnettoeinkommen hat einen positiven, signifikanten Einfluß auf die Zufriedenheit mit dem Gesundheitszustand. Dafür lassen sich verschiedene Erklärungen finden. Zum einen ist das Einkommen für die Zufriedenheit mit der Gesundheit wichtig, da neben der Inanspruchnahme medizinischer Leistungen auch der Genuß gesundheitsrelevanter Güter ermöglicht wird. Der Effekt auf die Gesundheit hängt jedoch auch stark mit der subjektiven Wahrnehmung des Individuums zusammen. Weiterhin kann das monatliche Haushaltsnettoeinkommen nicht unbedingt einer Person des Haushalts zugeordnet werden, so daß u.U. kein direkter Einfluß besteht, sondern der Konsum auch über das Einkommen des Partners finanziert wird. Die Variable ‚Schulbildung‘ hat einen signifikant positiven Einfluß auf die Wahrscheinlichkeit eines besseren Gesundheitszustands. Dahinter steht, daß ein Individuum mit besserem Bildungsstand ein genaueres Wissen über gesundheitsrelevante Themen besitzt und Krankheiten möglicherweise auch früher erkennt und daher besser agiert als ein weniger gebildetes Individuum.¹⁵¹

Eine private Krankenversicherung hat keinen signifikanten Einfluß. Auch hier gilt wiederum, daß die Art der Krankenversicherung eher für die Nutzung medizinischer Leistungen relevant sein dürfte als für deren Resultat. Die *Hypothese 4* (siehe S. 142), daß die Art der Versicherung den Gesundheitszustand indirekt beeinflusst, kann somit nicht bestätigt werden. Zumindest für Deutschland ist kein signifikanter Unterschied zwischen privat oder gesetzlich Versicherten erkennbar. Die Variable ‚Sport‘, die angibt, ob wöchentlich aktiv Sport betrieben wird, ist signifikant positiv. Körperliche Ertüchtigung steigert

¹⁴⁹ Einen negativen, ebenfalls nicht signifikanten Einfluß auf den Gesundheitszustand für Frauen finden auch Nocera und Zweifel (1998).

¹⁵⁰ Die positive Selektion ergibt sich daraus, daß erwerbstätige Ausländer i.d.R. aufgrund der Arbeitsplatzchancen nach Deutschland gekommen sind und sich daher aus ihrer Sicht keine lange krankheitsbedingte Abwesenheit vom Arbeitsplatz erlauben können.

¹⁵¹ Dieses Ergebnis stimmt mit anderen Schätzungen überein (vgl. bspw. Leu und Gerfin (1992), Leu und Doppmann (1996), Nocera und Zweifel (1998), Gerdtham und Johannesson (1999) und Gerdtham et al. (1999)).

demnach die Wahrscheinlichkeit einer hohen Zufriedenheit mit dem Gesundheitszustand, unabhängig von einer eventuell höheren Verletzungs- und damit Krankheitsanfälligkeit, ein Ergebnis, das sich auch in Nocera und Zweifel (1998) wiederfindet. Insgesamt sprechen zumindest die Variablen ‚Schulbildung‘ und ‚Sport‘ für einen positiven Einfluß der Compliance des Patienten auf die Zufriedenheit mit der Gesundheit und damit für *Hypothese 1* (vgl. S. 142).

Die Variablen der Gesundheitslage weisen, wie aus Tabelle 7.4 ersichtlich, alle das erwartete Vorzeichen auf und sind signifikant auf dem 1%-Niveau. Chronische Krankheiten reduzieren die Wahrscheinlichkeit einer hohen Zufriedenheit, keine Einschränkungen bei der Erfüllung täglicher Arbeiten durch die Gesundheit erhöhen die Zufriedenheit. Ähnliches gilt für die medizinische Inanspruchnahme. Alle drei Variablen (Hausarztbesuch, Besuch beim Facharzt und Krankenhausaufenthalte) haben einen signifikant negativen Einfluß auf die Zufriedenheit mit der Gesundheit. Auf den ersten Blick mag dieses Ergebnis überraschen, da man von einem positiven Einfluß der Medizin auf den Gesundheitszustand ausgehen würde. In der Regel werden aber nur die Personen einen Arzt aufsuchen, die gesundheitliche Beschwerden aufweisen und daher nicht voll und ganz mit ihrem Gesundheitszustand zufrieden sind. Dies spricht für die Gültigkeit von *Hypothese 3* (siehe hierzu S. 142), daß die Inanspruchnahme der medizinischen Leistungen vom Gesundheitszustand vor der Behandlung abhängt. Weiterhin ist es durchaus möglich, daß trotz medizinischer Leistungen der Patient noch nicht wieder mit seiner Gesundheit voll zufrieden ist. Da die Datenerhebung nicht in medizinische Inanspruchnahmen unterscheidet, bei denen die Behandlung abgeschlossen war und solche, bei denen die Behandlung noch andauert, gehen beide Möglichkeiten in den Datensatz ein. Wünschenswert wäre es dagegen gewesen, nur abgeschlossene Behandlungen zu betrachten, um die Wirkung der medizinischen Leistung beurteilen zu können. Da auch der Behandlungsaufwand nicht als gesonderte Variable zur Verfügung steht, kann keine separate Aussage über Aufwand bzw. Qualität der Behandlung gemacht werden. *Hypothese 5* (siehe Abschnitt 7.1) erweist sich somit als nicht überprüfbar. Das negative Vorzeichen dieser Variablen ist daher eher ein Indikator dafür, daß Leute mit höherer Inanspruchnahme medizinischer Leistungen generell einen schlechteren Gesundheitszustand aufweisen als solche ohne Inanspruchnahme. Für die *Hypothese 2*, die eine gegenseitige Abhängigkeit der Compliance und der medizinischen Leistung postuliert, kann keine Aussage getroffen werden, da es hierzu notwendig wäre, mit einem System von Schätzgleichungen zu arbeiten.

Tabelle 7.4: Ergebnisse der Querschnittsschätzung für die Gesamtstichprobe und die Nichterwerbstätigen

abhängige Variable: Zufriedenheit mit der Gesundheit

erklärende Variablen	(1) gesamt		(2) nicht erwerbstätig	
	Koeffizient	P-Wert ^{a)}	Koeffizient	P-Wert ^{a)}
prädispon. Variablen				
Alter	-0,0288	0,000***	-0,0321	0,000***
Alter ²	0,0002	0,000***	0,0003	0,002***
Geschlecht	-0,0419	0,172	-0,0114	0,833
alleinstehend	0,0243	0,550	-0,0321	0,611
Kinder	-0,0330	0,340	-0,1107	0,080*
Ausländer	0,0456	0,238	-0,0437	0,520
sozioök. Variablen				
Einkommen	0,0207	0,038**	0,0326	0,072**
Schulbildung	0,0770	0,027**	0,1477	0,009***
PKV	0,0534	0,314	0,1380	0,131
Sport	0,1754	0,000***	0,2727	0,000***
Gesundheitslage				
chronisch	-0,5109	0,000***	-0,5877	0,000***
keine Behinderung	0,9091	0,000***	0,8960	0,000***
Arbeitsunfähigkeit	-	-	-	-
med. Inanspruchnahme				
prak. Arzt	-0,2377	0,000***	-0,2334	0,003***
Facharzt	-0,1848	0,000***	-0,2072	0,008***
KH-Aufenthalte	-0,2389	0,000***	-0,2560	0,000***
Wohnumfeld				
Entf. Arzt	-0,1051	0,027**	-0,0725	0,333
Entf. Zentrum	0,1428	0,000***	0,0757	0,148
Luftverschmutzung	-0,1290	0,011**	-0,1956	0,020**
Lärmbelästigung	-0,1184	0,014**	-0,1047	0,161
Log Likelihood	-9184,61		-3739,03	
Wald	2232,32	0,000	1045,90	0,000
McFadden's R ² adj.	0,114		0,125	
McKelvy/Zavoina R ²	0,404		0,458	
Akaike's Information Criterion	3,776		3,885	
N	4880		1940	

*** signifikant auf dem 1% Niveau

** signifikant auf dem 5% Niveau

* signifikant auf dem 10% Niveau

a) Der P-Wert gibt das minimale Signifikanzniveau (Fehler 1. Art) an, für das der beobachtete Wert der Teststatistik zu einer Ablehnung der Nullhypothese (Koeffizient = 0) führt (vgl. Kennedy 1998, S. 409).

Bezüglich des Wohnumfeldes kann man festhalten, daß bei einer weiten Entfernung zum Hausarzt die Wahrscheinlichkeit der Zufriedenheit mit der eigenen Gesundheit signifikant zurückgeht. Dies deutet darauf hin, daß Patienten den

Hausarzt auch im Sinne eines Optionskonsums ansehen und eine leichtere Erreichbarkeit somit ihr subjektives Empfinden des Gesundheitszustands positiv beeinflusst (vgl. Wille 1980, S. 143). Die Entfernung von einem Großstadtzentrum erhöht die Wahrscheinlichkeit für eine hohe Zufriedenheit mit der Gesundheit signifikant. Dies läßt sich darauf zurückführen, daß zwar in einer Großstadt die medizinische Infrastruktur besser als in mittleren oder Kleinstädten ist, daß aber die Lebensqualität in einem Großstadtzentrum u.U. niedriger ist. Weiterhin weisen Luftverschmutzung- und Lärmbelästigung die erwarteten negativen Einflüsse auf. Die Koeffizienten sind signifikant auf dem 5 %-Niveau. Starke Verschmutzung bzw. Lärmbelästigung wird von den untersuchten Personen als negativ im Hinblick auf ihre Einschätzung der Gesundheit empfunden. Es läßt sich vermuten, daß dies insbesondere dann der Fall ist, wenn keine Ausweichmöglichkeiten in andere Regionen bestehen. Die Qualität der Schätzung läßt sich anhand der besprochenen Gütemaße angeben, die insbesondere für den folgenden Vergleich unterschiedlicher Schätzungen relevant sind. Es ergibt sich, daß das korrigierte R^2 nach McFadden einen Wert von 0,11 annimmt. Das McKelvey-Zavoina-Maß beträgt 0,404 und das Akaike-Informationsmaß 3,776.

Schätzung Nichterwerbstätige

Bei der Schätzung für die Nichterwerbstätigen (Schätzung (2), Spalten 3 und 4 in Tabelle 7.4) treten im Vergleich zur vorangegangenen Untersuchung nur einige Unterschiede auf. Die Koeffizienten für das Alter und sowie den quadratischen Altersterm sind bei nahezu gleicher Signifikanz im wesentlichen unverändert. Bei den sozioökonomischen Variablen fällt vor allem der etwas größere Koeffizient des Haushaltsnettoeinkommens auf, der auf dem 5 %-Niveau signifikant ist. Das Haushaltsnettoeinkommen ist insbesondere für die Nichterwerbstätigen wichtig, da sie über kein eigenes Arbeitseinkommen verfügen, um den Konsum von z.B. gesundheitsrelevanten Gütern zu finanzieren.¹⁵² Der Koeffizient der Variablen ‚Schulbildung‘ ist etwas höher als bei der Gesamtstichprobe, was sich jedoch nicht durch die ökonomische Theorie erklären läßt. Der Koeffizient der Variablen ‚PKV‘ hat nun einen wesentlich höheren Wert als zuvor (0,138 im Vergleich zu 0,053), doch bleibt er auch in dieser Schätzung insignifikant. Die Ergebnisse für die Variablen der Gruppen Gesundheitslage und medizinische Inanspruchnahme entsprechen im Großen und Ganzen denen der Gesamtstichprobe (vgl. hierzu die Ergebnisse in den Spalten 1 und 2 der Tabelle 7.4)

¹⁵² Allerdings vernachlässigt eine solche Interpretation andere Einkommensarten wie bspw. Vermögenseinkommen oder Einkommen aus Vermietung und Verpachtung, die auch für Gesundheitsausgaben herangezogen werden könnten.

Dagegen sind alle Variablen für das Wohnumfeld mit Ausnahme der Luftverschmutzung insignifikant. Dies mag daran liegen, daß Menschen, die keiner Erwerbstätigkeit nachgehen, mehr Zeit zur Verfügung haben, um bspw. auch einen räumlich entfernten Arzt aufzusuchen. Diese Überlegung wird auch durch den kleineren Koeffizienten für die Entfernung zum nächsten Großstadtzentrum bestätigt. Die Opportunitätskosten sind demnach für Nichterwerbstätige geringer als für Erwerbstätige. Aus diesem Grund mag auch die Lärmbelästigung am Wohnort keinen signifikanten Einfluß auf die Zufriedenheit mit der Gesundheit ausüben. Bezüglich des signifikant negativen Koeffizienten für die Luftverschmutzung kann man festhalten, daß man in einer bestimmten Region der Verschmutzung der Luft schlechter ausweichen kann als bspw. der Belästigung durch Lärm, so daß die Luftverschmutzung signifikante Effekte auf die Zufriedenheit mit der Gesundheit besitzt.

Die Ergebnisse für das korrigierte McFadden- R^2 und McKelvey/Zavoina- R^2 liegen im Vergleich mit der Gesamtstichprobe etwas höher (0,125 bzw. 0,458), was auf eine bessere Varianzerklärung schließen läßt; gleichzeitig ist der Wert des Akaike-Informationsmaßes mit 3,885 geringfügig höher, was für eine bessere Anpassung der Gesamtstichprobe spricht (siehe Spalten 1 und 3 in Tabelle 7.4).

Schätzung Erwerbstätige

Die dritte Schätzung überprüft den Zusammenhang für die Erwerbstätigen anhand der identischen Variablenauswahl wie bei den Schätzungen für die Gesamtstichprobe und die Nichterwerbstätigen (vgl. Schätzung (3) in Tabelle 7.5, S. 180). Die Vorzeichen der signifikanten Variablen sind identisch mit denen der zuvor präsentierten Schätzungen in Tabelle 7.4. Die Koeffizienten der Altersvariable und des quadratischen Altersterms bewegen sich auf etwa dem Niveau der Gesamtstichprobe und der Nicht-Erwerbstätigen. Frauen haben eine geringere Zufriedenheit mit der Gesundheit, allerdings ist der Koeffizient auch hier nicht signifikant. Dies trifft auch für die anderen prädisponierenden Variablen zu, wobei die Variablen ‚alleinstehend‘ und ‚Ausländer‘ ein positives Vorzeichen aufweisen.

Bei der Gruppe der sozioökonomischen Variablen fällt auf, daß die Schulbildung nun keinen signifikanten Effekt mehr besitzt. Eine Erklärung liegt u.U. darin, daß für Personen, die im Erwerbsleben stehen, nicht mehr die Schulausbildung, sondern ihre berufliche Tätigkeit die Zufriedenheit mit der Gesundheit beeinflusst. Das Vorzeichen der Variablen ‚PKV‘ ist nun negativ, jedoch insignifikant. Das Einkommen ist ebenfalls nicht mehr signifikant (vgl. Tabelle 7.4). Möglicherweise liegt eine Erklärung darin, daß Erwerbstätige regelmäßig

ein Einkommen beziehen und daher nicht auf das Gesamteinkommen des Haushalts angewiesen sind. Wöchentlicher aktiver Sport erhöht wie zuvor die Zufriedenheit mit der Gesundheit signifikant. Dieses letzte Ergebnis stützt auch in dieser Schätzung die *Hypothese 1* (siehe Abschnitt 7.1), daß auch der Patient durch sein Verhalten den Gesundheitszustand beeinflussen kann.

Auch die Resultate für die Gesundheitslage entsprechen den Erwartungen. Beide, sowohl chronische Krankheiten als auch die Möglichkeit, ohne Behinderung durch den Gesundheitszustand den täglichen Arbeiten nachzugehen, besitzen das erwartete Vorzeichen und sind signifikant auf dem 1%-Niveau. Für die Variablen der medizinischen Inanspruchnahme gilt, daß durch Besuche beim praktischen Arzt, beim Facharzt und durch Krankenhausaufenthalte die Zufriedenheit mit der eigenen Gesundheit zurückgeht. Dies bestätigt die Resultate der anderen Schätzungen für die Gesamtstichprobe und die Nicht-Erwerbstätigen, sowie die Vermutung, daß die Inanspruchnahme medizinischer Leistungen mit einem schlechten Gesundheitszustand vor der Behandlung einhergeht, so daß die resultierende Zufriedenheit ebenfalls gering ist (vgl. *Hypothese 3*, Abschnitt 7.1). Bemerkenswert bei den Arztbesuchen ist, daß der Absolutwert des Schätzkoeffizienten für den praktischen Arzt geringer und für den Facharzt größer ist als bei den Schätzungen für die Gesamtstichprobe und für die Nicht-Erwerbstätigen. Bei Betrachtung der deskriptiven Statistik (siehe Tabelle 7.3) fällt auf, daß die Mittelwerte beider Variablen deutlich geringer sind als in der Stichprobe der Nicht-Erwerbstätigen, was den Schluß zuläßt, daß Erwerbstätige seltener einen Arzt aufsuchen als Nicht-Erwerbstätige.

Für die Variablen des Wohnumfeldes ergibt sich, daß der negative Effekt auf die Wahrscheinlichkeit einer hohen Zufriedenheit mit der Gesundheit bei den Erwerbstätigen höher als bei den anderen Stichproben ist, falls die Entfernung zum Hausarzt mehr als 25 Minuten zu Fuß beträgt. Eine mögliche Erklärung dafür liegt in den Opportunitätskosten, da Erwerbstätige im Regelfall lediglich außerhalb ihrer Arbeitszeit den Arzt aufsuchen können und sich dann die größere Entfernung negativ auswirkt. Dagegen erhöht eine größere Entfernung vom nächsten Großstadtzentrum die Zufriedenheit mit der Gesundheit stärker, da außerhalb des Zentrums die Lebensqualität i.d.R. höher ist. Die Luftverschmutzung am Wohnort besitzt keinen signifikanten Effekt.¹⁵³ Starke Lärmbelästigung dagegen vermindert die Zufriedenheit mit der Gesundheit signifikant. Das korrigierte R^2 nach McFadden und das Maß von McKelvey/Zavoina liegen mit Werten von 0,093 und 0,343 unter denen der Schätzungen (1) und (2). Das

¹⁵³ Dieses Ergebnis beruht u.U. darauf, daß Erwerbstätige einen großen Teil ihres Tagesablaufes nicht an ihrem Wohnort, sondern an ihrem Arbeitsplatz verbringen und die Luftverschmutzung an beiden eventuell unterschiedlich ist.

Akaike-Informationsmaß hat mit 3,723 einen niedrigeren Wert, was den Schluß nahelegt, daß das Modell für die Teilstichprobe der Erwerbstätigen eine bessere Anpassung ergibt.

Schätzung Erwerbstätige inklusive Arbeitsplatzcharakteristika und Arbeitsunfähigkeit

Für die vierte Schätzung wurden für die Erwerbstätigen arbeitsplatzbezogene Variablen sowie Arbeitsunfähigkeit in die Untersuchung mit einbezogen (Schätzung (4), siehe Spalten 3 und 4 in Tabelle 7.5). Für die Gruppe der prädisponierenden Variablen ergeben sich daraus nur geringfügige Änderungen der Parameterschätzungen und der Signifikanz. Nach wie vor sind die Altersterme signifikant, wenn auch nur noch auf dem 5 % Niveau. Alle anderen Variablen dieser Gruppe sind nach wie vor insignifikant. Ein ähnliches Bild zeigt sich bei den sozioökonomischen Variablen, bei denen lediglich wöchentlicher aktiver Sport signifikant positiv ist.

Bei der Gesundheitslage kommt nun zu den Variablen ‚chronische Krankheiten‘ und ‚keine Behinderung durch den Gesundheitszustand‘ auch noch die ‚Arbeitsunfähigkeit‘ hinzu. Die ersten beiden Variablen sind hoch signifikant und ihr Vorzeichen entspricht den Erwartungen. Falls ein Erwerbstätiger länger als 14 Tage krank geschrieben war, vermindert sich seine Zufriedenheit mit dem Gesundheitszustand signifikant. Dies läßt den Schluß zu, daß längere Arbeitsunfähigkeit den Gesundheitszustand oftmals über einen längeren Zeitraum nach der akuten Erkrankung noch negativ beeinflusst. Gleichzeitig fällt jedoch auf, daß der Koeffizient für den Krankenhausaufenthalt absolut gesehen deutlich kleiner und nicht signifikant ist.¹⁵⁴ Eine mögliche Interpretation ist, daß eine längere Arbeitsunfähigkeit oft mit einem Krankenhausaufenthalt verbunden ist bzw. daß dieser der vorübergehenden Erwerbsunfähigkeit vorausgeht. Somit geht der Effekt von der Krankenhausvariablen auf die Arbeitsunfähigkeit über.

Die Ergebnisse für das Wohnumfeld sind mit denen von Schätzung (3) vergleichbar. Auch hier besitzen die Entfernung zum Arzt und die Lärmbelästigung einen signifikant negativen, die Entfernung zum Großstadtzentrum einen signifikant positiven Einfluß. Die Luftverschmutzung besitzt auch in dieser Schätzung keinen signifikanten Effekt.

¹⁵⁴ Alternative Schätzungen mit beiden Variablen, der Arbeitsunfähigkeit und dem Krankenhausbesuch aus der vierten Welle, die sich auf das Jahr 1986 beziehen, ergaben ein ähnliches Ergebnis. Wird hingegen nur die Arbeitsunfähigkeit aus der 4. Welle entnommen, so bleibt der Krankenhausbesuch signifikant, was die Hypothese stützt, daß beide Variablen aus einer Periode hoch korreliert sind.

Tabelle 7.5: Ergebnisse der Querschnittsschätzung für Erwerbstätige

abhängige Variable: Zufriedenheit mit der Gesundheit

erklärende Variablen	(3) erwerbstätig		(4) inkl. Arbeitsbedingung	
	Koeffizient	P-Wert ^{a)}	Koeffizient	P-Wert ^{a)}
prädispon. Variablen				
Alter	-0,0326	0,008***	-0,0309	0,012**
Alter ²	0,0003	0,041**	0,0003	0,050**
Geschlecht	-0,0393	0,378	-0,0516	0,256
alleinstehend	0,0665	0,250	0,0599	0,300
Kinder	0,0070	0,846	0,0022	0,961
Ausländer	0,0481	0,340	0,0612	0,230
sozioök. Variablen				
Einkommen	0,0088	0,512	0,0105	0,432
Schulbildung	0,0307	0,518	0,0237	0,625
PKV	-0,0367	0,958	0,0158	0,822
Sport	0,1011	0,043**	0,1003	0,045**
Gesundheitslage				
chronisch	-0,4529	0,000***	-0,4183	0,000***
keine Behinderung	0,9250	0,000***	0,9031	0,000***
Arbeitsunfähigkeit	-	-	-0,3294	0,000***
med. Inanspruchnahme				
prak. Arzt	-0,1712	0,016**	-0,1428	0,044**
Facharzt	-0,2488	0,001***	-0,2498	0,000***
KH-Aufenthalte	-0,2349	0,001***	-0,0654	0,407
Wohnumfeld				
Entf. Arzt	-0,1526	0,022**	-0,1629	0,015**
Entf. Zentrum	0,1730	0,000***	0,1686	0,000***
Luftverschmutzung	-0,0964	0,161	-0,0879	0,201
Lärmbelästigung	-0,1329	0,051*	-0,1186	0,083*
Arbeitsbedingungen				
Anspannung	-	-	-0,1662	0,001***
körp. Anstrengung	-	-	0,0082	0,899
Nachtschicht	-	-	0,0018	0,986
Wechselschicht	-	-	0,1228	0,096*
Umweltbelastung	-	-	-0,0311	0,584
Log Likelihood	-4835,08		-4813,65	
Wald	979,13	0,000	1011,18	0,000
McFadden's R ²	0,093		0,095	
McKelvy/Zavoina R ²	0,343		0,355	
Akaike's Information Criterion	3,723		3,711	
N	2613		2613	

*** signifikant auf dem 1% Niveau

** signifikant auf dem 5% Niveau

* signifikant auf dem 10% Niveau

a) Der P-Wert gibt das minimale Signifikanzniveau (Fehler 1. Art) an, für das der beobachtete Wert der Teststatistik zu einer Ablehnung der Nullhypothese (Koeffizient = 0) führt (vgl. Kennedy 1998, S. 409).

In der Variablengruppe Arbeitsbedingungen besitzt die Variable ‚Anspannung‘ einen signifikant negativen Einfluß. Beruflicher Streß verringert daher die Zufriedenheit mit der eigenen Gesundheit. Weder körperliche Anstrengung noch Nachtschicht oder Umweltbelastung haben einen signifikanten Effekt. Auffällig ist der signifikant positive Koeffizient der Wechselschicht, der sich wohl nur mit einer positiven Selektion erklären läßt. Gesunde Erwerbstätige sind eher in der Lage und bereit, anstrengende Schichtarbeit auf sich zu nehmen. Insgesamt aber gilt, daß die Variablengruppe ‚Arbeitsbedingungen‘ nur bedingt einen Einfluß auf die Zufriedenheit mit der Gesundheit ausübt, sondern daß sie lediglich Ausdruck der Zufriedenheit mit dem Arbeitsplatz ist (vgl. Pohlmeier und Ulrich 1995, S. 356ff.). Das McFadden-Maß für das R^2 liegt mit 0,095 etwas über dem Wert von Schätzung (3). Das gleiche gilt für das McKelvey/Zavoina- R^2 mit einem Wert von 0,355. Das Informationsmaß nach Akaike liegt bei 3,711 und spricht daher für die beste Modellanpassung in dieser Stichprobe.

Die Analyse von Querschnittsdaten zeigt, daß neben der medizinischen Leistung auch das gesundheitsrelevante Verhalten des Patienten einen Einfluß auf die Zufriedenheit mit der Gesundheit ausübt (*Hypothese 1*). Der negative Einfluß der medizinischen Leistungen spricht für die Gültigkeit der dritten Hypothese. Hingegen besteht kein Effekt des Versicherungsstatus, so daß *Hypothese 4* nicht bestätigt wird. *Hypothese 5*, die eine Aufspaltung der Behandlung in Aufwand und Qualität voraussetzt, ist nicht überprüfbar, da der Datensatz lediglich die Anzahl der Arztbesuche als Indikator anbietet. Für die gegenseitige Abhängigkeit von medizinischer Leistung und Compliance des Patienten gilt, daß für *Hypothese 2* ein System von Schätzgleichungen notwendig wäre.

7.4 Ökonometrische Untersuchung von Paneldaten

7.4.1 Einführung in die Panel-Analyse

Die Analyse von Querschnittsdaten gibt Aufschluß darüber, welche individuellen und medizinischen Faktoren die Zufriedenheit mit der Gesundheit beeinflussen. Allerdings werden bei einer solchen Betrachtung eine Reihe von Aspekten ausgeklammert, die aus ökonomischer Sicht interessant erscheinen. Zum einen berücksichtigt eine Querschnittsanalyse nicht die Heterogenität der Individuen. Zum anderen betrifft dies vor allem die Vernachlässigung der Entwicklung im Zeitablauf. Für eine umfassendere Analyse ist es jedoch notwendig, diese Effekte näher zu betrachten. Dies ist mit Hilfe einer Untersuchung von Daten aus einem Panel möglich. Paneldaten liegen vor, wenn für eine bestimmte Untersuchungseinheit in mehreren aufeinanderfolgenden Zeitpunkten Beobachtungen für dasselbe Merkmal zur Verfügung stehen (vgl. hierzu Ronning 1991, S. 189

oder Hsiao 1986, S. 1f.).¹⁵⁵ Der Vorteil im Vergleich zu Querschnittsdaten liegt in der Beobachtung einer dynamischen Entwicklung. Außerdem können mit Paneldaten Fragestellungen untersucht werden, die mit Querschnitts- oder Zeitreihenbeobachtungen nicht analysiert werden können (vgl. dazu Baltagi 1995, S. 3ff.).

Solche Fragestellungen sind erstens die Kontrolle unbeobachteter Heterogenität. Oftmals existieren neben den erhobenen Variablen andere individual- oder zeitspezifische Einflußgrößen, die jedoch nicht erfaßt werden können (vgl. zu den folgenden Ausführungen Ronning 1996, S. 93 und Baltagi 1995, S. 3ff.). Durch Vernachlässigung dieser Effekte kann es zu Verzerrungen der Schätzung kommen. Mit Hilfe von Paneldaten ist es möglich, auf diese individual- bzw. zeitspezifischen Effekte einzugehen. Zweitens erhält man durch die Verwendung eines Panels mehr informative Daten, mehr Variabilität, weniger Kollinearität zwischen den Variablen, mehr Freiheitsgrade und damit eine höhere Effizienz der Schätzungen. Drittens können dynamische Anpassungsprozesse besser studiert werden. Veränderungen in den Daten werden durch Querschnittsdaten nicht dargestellt. Dies betrifft bspw. Arbeitsplatzwechsel, Arbeitslosigkeit oder auch Umzüge sowie Migration. Viertens ist man durch Paneldatensätze in der Lage, Effekte, die in reinen Querschnitten oder Zeitreihen nicht dargestellt werden können, zu identifizieren und zu messen. Ein Beispiel dafür ist die Erwerbsbeteiligung von Frauen. Betrachtet man lediglich einen Querschnitt und beträgt die Erwerbsbeteiligung im Durchschnitt 50 %, so kann es dafür zwei Erklärungen geben: Zum einen kann es sein, daß alle Frauen eine 50 %-Chance auf einen Arbeitsplatz besitzen, zum anderen, daß 50 % der Frauen die gesamte Zeit arbeiten, 50 % jedoch gar nicht. Im ersten Fall bedeutet dies eine hohe Arbeitsplatzfluktuation, im zweiten Fall liegt keine Fluktuation vor. Bei einem Panel ist es jedoch möglich, die Effekte der Fluktuation im Zeitablauf zu untersuchen. Fünftens erlauben es Paneldaten, kompliziertere Verhaltenszusammenhänge zu konstruieren und zu testen, und sechstens unterliegen Panel dadurch, daß sie auf der Mikroebene erhoben werden, nicht dem Problem, daß sie durch Aggregation über Firmen oder Individuen verzerrt sind.

Allerdings gibt es auch Faktoren, die die Aussagekraft von Panel-Analysen einschränken (vgl. Ronning 1996, S. 93). Die Repräsentativität eines Panels nimmt im Zeitablauf aufgrund der Panelsterblichkeit ab. Darunter versteht man das Herausfallen aus der Erhebung, entweder durch Nichtbeantwortung, durch Fortzug aus dem Erhebungsgebiet oder durch Tod. Aus diesem Grund ist das Panel nur in der ersten Welle repräsentativ. Es stellt sich daher die Frage, ob es nicht

¹⁵⁵ Eine Übersicht über die Verwendung von Paneldaten findet sich bspw. in Chamberlain (1984), Hsiao (1986), Maddala (1987), Baltagi (1995) oder Greene (2000).

sinnvoller ist, mit einer Folge von Querschnitten zu arbeiten, wobei Gruppen von Individuen (Kohorten) mit identischen Charakteristika identifiziert und die arithmetischen Mittel der Kohorten als Beobachtungswerte des Pseudo-Panels betrachtet werden. Ein derartiges Vorgehen ist repräsentativ, da für jeden Zeitpunkt eine neue Stichprobe gezogen wird. Allerdings ergeben sich dabei Meßfehler, so daß die OLS-Schätzer inkonsistent sind. Weitere Probleme, die sich im Zusammenhang mit Panel-Analysen ergeben, sind Meßfehler aufgrund einer unklaren Fragestellung, falschen Angaben der Befragten oder auch durch Beeinflussung durch den Interviewer (vgl. Baltagi 1995, S. 6). Darüber hinaus umfassen Panels häufig nur einen relativ kurzen Zeitraum. Die asymptotischen Eigenschaften hängen daher nur von der Anzahl der Individuen ab. Auf der anderen Seite ist eine Ausweitung der Zeitspanne mit erheblichen Kosten und mit Panelsterblichkeit verbunden (vgl. Baltagi 1995, S. 7).

Fixed-effects- vs. Random-effects-Spezifikation

Grundsätzlich kann für die Panel-Analyse von folgender Schätzgleichung ausgegangen werden (vgl. Greene 2000, S. 560):

$$y_{it} = \alpha_i + x_{it} \beta + \varepsilon_{it}. \quad (7.47)$$

Dabei steht der Index $i = 1, \dots, N$ für die Anzahl der beobachteten Individuen und der Index $t = 1, \dots, T$ für die untersuchten Perioden. Der Vektor x_{it} enthält K Regressoren, jedoch keine Konstante und ε_{it} bezeichnet den Störterm der Schätzgleichung. Der Koeffizient α_i ist ein über die Zeit konstanter Individualeffekt. Nimmt man an, daß dieser Individualeffekt für alle Beobachtungseinheiten gleich ist, dann ergeben sich durch eine OLS-Schätzung konsistente und effiziente Ergebnisse (vgl. zu diesen Ausführungen Greene 2000, S. 560). Dieses Grundmodell kann durch zwei Ansätze abgewandelt werden. Im ersten Fall wird angenommen, daß α_i ein gruppen- bzw. individualspezifischer konstanter Term im Regressionsmodell ist. Man spricht dann von einer Fixed-effects-Schätzung. Im Gegensatz dazu wird bei einem Random-effects-Ansatz davon ausgegangen, daß α_i ein gruppen- bzw. individualspezifischer Fehlerterm ist, der mit dem Fehler ε_{it} vergleichbar ist, nur daß er für jede betrachtete Periode identisch eingeht.

Die Fixed-effects-Schätzung nimmt an, daß Differenzen zwischen den Beobachtungseinheiten durch Unterschiede in der Konstanten ausgedrückt werden können. In Gleichung (7.47) stellen die α_i daher unbekannte Parameter dar, die im Modell geschätzt werden müssen. Dies ist mit klassischen Techniken möglich, allerdings steigt die Zahl der Parameter auf $N + K$ an. Diese Methode stellt einen guten Ansatz dar, wenn sichergestellt ist, daß die Unterschiede zwischen

den Untersuchungseinheiten als Shiftparameter der Regressionsgleichung interpretiert werden können. Dabei wird unterstellt, daß sich dieses Modell nur auf die in den Datensatz eingehenden Individuen bezieht und nicht auf die zugehörige Grundgesamtheit (vgl. hierzu Maddala 1987a, S. 309 oder Hsiao 1995, S. 388f.). Für die andere Spezifikation (Random-effects-Schätzung) wird angenommen, daß die individualspezifischen Terme zufällig über die Beobachtungseinheiten verteilt sind. Insbesondere für den Fall, daß die Stichprobe aus einer großen Population stammt, scheint eine solche Betrachtung angemessen.¹⁵⁶

Auch für die im folgenden vorgestellten Panel-Modelle gilt, daß die abhängige Variable der Schätzungen elf Kategorien aufweist, so daß sich auch hier Modelle mit geordneten Kategorien anbieten. Zuerst wird jedoch analog zur Vorgehensweise für die Querschnittsuntersuchung der einfachere Fall von binären abhängigen Variablen betrachtet, bevor auf ordinale abhängige Variablen eingegangen wird.

7.4.2 Panel-Modelle für binäre Variablen

Die in Abschnitt 7.3.1 dargestellten Modelle für diskrete abhängige Variablen für Querschnittsanalysen unterstellen, daß die Koeffizienten der Regression für alle Individuen über alle Perioden identisch sind (vgl. zur folgenden Darstellung Hamerle und Ronning 1995, S. 411ff.). Dabei ist die Annahme der Homogenität der Individuen fragwürdig, da sie individuelle Unterschiede ignoriert. Ein Ansatzpunkt für die Untersuchung von Panel-Modellen für diskrete abhängige Variablen stellt das Modell einer latenten stetigen Variable in Gleichung (7.48) dar:

$$y_{it}^* = \mathbf{x}_{it} \boldsymbol{\beta} + \varepsilon_{it} \quad \text{mit } i = 1, \dots, N \quad \text{und } t = 1, \dots, T. \quad (7.48)$$

Dabei bezeichnet \mathbf{x}_{it} den Vektor der erklärenden Variablen der i ten Beobachtung zum Zeitpunkt t und $\boldsymbol{\beta}$ den zugehörigen Parametervektor. Da y^* unobservierbar ist, nimmt man nur eine binäre Variable y_{it} wahr, die folgendermaßen definiert ist:

$$y_{it} = \begin{cases} 1 & \text{für } y_{it}^* > 0 \\ 0 & \text{sonst.} \end{cases} \quad (7.49)$$

¹⁵⁶ Da die verwendeten Daten des Sozioökonomischen Panels eine repräsentative Längsschnittstudie privater Haushalte in der Bundesrepublik Deutschland darstellen, erscheint die Verwendung von Random-effects-Modellen gegenüber Fixed-Effects-Ansätzen angemessener.

Im einfachsten Fall besteht für die Beobachtungseinheiten für den Fehlerterm u_{it} serielle Unabhängigkeit, d.h. keine Autokorrelation. Weicht man hingegen von dieser Annahme ab, so kann man sowohl individual- als auch periodenspezifische Effekte unterstellen. Für den Fehlerterm ergibt sich dann:

$$u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}. \quad (7.50)$$

In der obigen Gleichung stellt λ_t einen zeitspezifischen Parameter dar. Falls α_i als Parameter zusammen mit den Koeffizienten des Modells geschätzt wird, liegt der bereits angesprochene Fixed-effects-Ansatz vor. Wird hingegen α_i als Zufallsvariable aufgefaßt, so spricht man von einem Random-effects-Modell. Welcher der beiden Methoden der Vorzug gegeben wird, hängt von dem Ziel der Untersuchung ab (vgl. den vorherigen Abschnitt). Falls eine große Anzahl von Beobachtungseinheiten im Querschnitt vorliegt, so müssen im Fixed-effects-Modell zusätzlich N Parameter geschätzt werden, während im Random-effects-Modell nur der Mittelwert und die Varianz des zusätzlichen Fehlerterms ermittelt werden müssen. Weiterhin gilt für diesen Fall nach Maddala (vgl. 1987a, S. 309), daß die α_i dann individualspezifische Effekte messen, die ansonsten vernachlässigt würden. Gleiches gilt auch für den ‚normalen‘ Störterm ε_{it} , der die Effekte für die it e Beobachtung in der Periode t mißt. Folglich sollte, da ε_{it} als Zufallsvariable aufgefaßt wird, dies auch für α_i gelten und daher der Random-effects-Spezifikation der Vorzug gegeben werden. Der Vorteil des Fixed-effects-Modells liegt hingegen darin begründet, daß keine zusätzlichen Annahmen über die Korrelation der individualspezifischen Effekte mit den erklärenden Variablen nötig sind.

Ein weiteres Problem bei der Implementierung eines Fixed-effects-Modells besteht darin, daß es nicht möglich ist, die Heterogenität der α_i durch Differenzbildung (engl. differencing out) zu eliminieren (vgl. Hamerle und Ronning 1995, S. 413ff., Maddala 1987a, S. 315ff. und Baltagi 1995, S. 178ff.). Während für stetige Variablen bei Abwesenheit von Autokorrelation das Fixed-effects-Modell konsistente Ergebnisse für die Parameter β liefert, ist dies für binäre abhängige Variablen nicht der Fall.¹⁵⁷ Statt dessen wird ein Conditional Likelihood-Ansatz verwendet. Die grundlegende Idee dahinter ist, eine bedingte Likelihood-

¹⁵⁷ Dieses Problem wird auch als ‚incidental parameter problem‘ bezeichnet (vgl. Baltagi 1995, S. 179). Für $N \rightarrow \infty$ und eine fixe Periodenanzahl T steigt die Anzahl der Parameter α_i mit N an. Als Folge davon ergibt sich, daß die α_i nicht mehr konsistent geschätzt werden können. Weiterhin sind für diskrete abhängige Variablen die Maximum Likelihood-Schätzer für α_i und β nicht unabhängig. (vgl. Hsiao 1986, S. 159). Hingegen ist für $T \rightarrow \infty$ der Maximum Likelihood-Schätzer für α_i und β konsistent.

Funktion zu betrachten, die sich auf die hinreichende Statistik für die zusätzlichen Parameter α_i bezieht.¹⁵⁸ Dieser Ansatz kann nur für das Logit-Modell angewendet werden. In einem Probit-Modell kommt es durch den Conditional Likelihood-Ansatz nicht zu einer Vereinfachung der Berechnungen, da sich die fixen Effekte nicht eliminieren lassen, so daß alle N zusätzlichen Parameter geschätzt werden müssen. Im Fall eines Random-effects-Ansatzes erweist sich das Probit-Modell als einfacher zu handhaben.¹⁵⁹ Dabei muß für den Fall, daß der Fehlerterm α_i mit den erklärenden Variablen korreliert ist, eine Verteilung von α für gegebenes x spezifiziert werden. Im Probit-Modell wird eine multivariate Normalverteilung unterstellt.¹⁶⁰ Das Random-effects-Probit-Modell besitzt zwei wichtige Implikationen. Erstens sind die Ergebnisse im Unterschied zu einem Fixed-effects-Probit-Modell konsistent. Zweitens gilt für den Fall, daß man die Korrelationen zwischen den Fehlern vernachlässigt, daß eine einfache Probit-Schätzung mit gepoolten Daten konsistente, wenn auch ineffiziente Schätzungen liefert, die als Startwert für iterative Verfahren zur Berechnung der Maximum Likelihood-Schätzer dienen können.

Maddala (1987a) beschreibt drei Modellansätze für eine Random-effects-probit-Spezifikation. Das Heckman und Willis-Modell liefert eine effiziente Schätzung unter der Annahme, daß die Korrelation zwischen nachfolgenden Störtermen für das gleiche Individuum konstant ist ('equicorrelation'). Falls diese Annahme gelockert wird, schlagen Avery, Hansen und Hotz einen Methods of Moments-Schätzer vor. Den allgemeinsten Ansatz liefert Chamberlain, da dieser zusätzlich Korrelation zwischen dem Random-effects-Parameter und den gegenwärtigen, zukünftigen und vergangenen Werten der erklärenden Variablen erlaubt (vgl. Maddala 1987a, S. 324).

7.4.3 Schätzverfahren für Panel-Modelle mit ordinalen abhängigen Variablen

Im Gegensatz zu Modellen mit binären abhängigen Variablen bestehen nur wenige Anwendungen für Panel-Modelle mit geordneten Variablen (vgl. zu einer Übersicht Hamerle und Ronning 1995, S. 431ff.). Das und van Soest (1999)

¹⁵⁸ Dies bedeutet, eine minimal hinreichende Statistik für α_i zu finden, die nicht auf β beruht. Da die Ergebnisse einer Maximum Likelihood-Schätzung allgemeine Funktionen dieser hinreichenden Statistik sind, können diese durch Ableiten der Log-Likelihood nach α_i ermittelt werden (vgl. Baltagi 1995, S. 179f.)

¹⁵⁹ Der Vorteil der Random-effects-Spezifikation ergibt sich auch aus den verwendeten Daten (siehe Fußnote 156)

¹⁶⁰ Im Logit-Modell ist dies die multivariate logistische Funktion. Diese weist jedoch den Nachteil auf, daß die Korrelationen auf 0,5 festgesetzt sind, so daß die logistische Funktion folglich nicht die Flexibilität der Normalverteilung besitzt (vgl. Maddala 1987, S. 280 und Hamerle und Ronning 1995, S. 417f.).

schätzen das zukünftige erwartete Einkommenswachstum in Haushalten anhand eines unbalanced Panels mit Hilfe sowohl einer Random-effects- als auch einer Fixed-effects-Spezifikation. Für letztere wird, wie im Fall binärer abhängiger Variablen, eine bedingte Likelihood-Funktion unterstellt und für den Fehlerterm eine logistische Verteilung angenommen. Damit der Ansatz auf Kategorien übertragen werden kann, wird die abhängige Variable für jede Kategorie zu einer binären umgeformt. Somit erhält man $J - 1$ Schätzungen für die interessierenden Parameter, wobei J die Anzahl der Kategorien bezeichnet. Diese Resultate können dann mit Hilfe einer Minimum Distance-Schätzung zu einer einzigen Schätzung kombiniert werden.¹⁶¹ Für die Random-effects-Schätzung nehmen Das und van Soest an, daß die individualspezifischen Effekte normalverteilt mit dem Erwartungswert 0 und der Varianz σ sind. Weiterhin unterstellen sie, daß die erklärenden Variablen und die Fehlerterme unabhängig voneinander sind. Die Likelihood-Funktion stellt dann normalerweise ein t -faches Integral dar, daß sich aber unter gewissen Annahmen auf ein einfaches reduzieren läßt (vgl. Das und van Soest 1999, S. 413f.).

Die hier verwendete Methode einer Random-effects-Ordered-Probit-Schätzung ist der beschriebenen Vorgehensweise sehr ähnlich.¹⁶² Der Zusammenhang zwischen einer latenten, endogenen Variablen und den erklärenden Variablen kann anhand von Gleichung (7.51) beschrieben werden:

$$y_{it}^* = \beta' x_{it} + \varepsilon_{it} \quad \text{mit } i = 1, \dots, N \quad \text{und } t = 1, \dots, T. \quad (7.51)$$

Dabei bezeichnet der Index i die jeweilige Beobachtungseinheit und der Index t die betrachtete Periode. Für die Fehlerstruktur wird unterstellt, daß der Störterm sich in einen individualspezifischen Fehler u_i und einen klassischen Fehler v_{it} aufteilen läßt:

$$\varepsilon_{it} = v_{it} + u_i, \quad (7.52)$$

$$\text{Var}(\varepsilon_{it}) = \sigma_v^2 + \sigma_u^2 = 1 + \sigma_u^2,$$

$$\text{Corr}(\varepsilon_{it}, \varepsilon_{is}) = \rho = \frac{\text{Cov}(\varepsilon_{it}, \varepsilon_{is})}{\sigma_{\varepsilon_{it}} \cdot \sigma_{\varepsilon_{is}}} = \frac{\sigma_u^2}{1 + \sigma_u^2}.$$

¹⁶¹ Die Minimum Distance-Schätzung gehört zur Klasse der GMM-Schätzer und ist ein Weg, einen effizienten Schätzer unter nichtlinearen Restriktionen sowie eine Teststatistik für diese Restriktionen zu erhalten (vgl. Ogaki 1993, S. 463).

¹⁶² Siehe hierzu Frechette (2001), der ein Random-effects-Ordered-Probit-Modell für die Software Stata programmierte.

Die Varianz von v_{it} ist gleich Eins, die des Individualterms gleich σ_u^2 und ρ bezeichnet den Korrelationskoeffizienten zwischen zwei beliebigen Zeitpunkten t und s .¹⁶³ Dies bedeutet, daß der Fehler ε_{it} zwischen zwei beliebigen Zeitpunkten mit einem konstanten Koeffizienten ρ korreliert ist.

Anstatt der latenten Variable y^* beobachtet man die Variable y_{it} . Dabei handelt es sich um eine geordnete ordinale Variable, deren J Kategorien durch die Schwellenwerte τ_j abgegrenzt werden (siehe Gleichung (7.53)):

$$y_{it} = \begin{cases} 0 & \text{falls } y_{it}^* \leq \tau_0 \\ 1 & \text{falls } \tau_0 < y_{it}^* \leq \tau_1 \\ 2 & \text{falls } \tau_1 < y_{it}^* \leq \tau_2 \\ \vdots & \\ J & \text{falls } \tau_{J-1} < y_{it}^* \end{cases} \quad (7.53)$$

Definiert man nun $a_{it} = \tau_{j-1} - \beta'x_{it}$ und $b_{it} = \tau_j - \beta'x_{it}$, so folgt für die Likelihood-Funktion, falls $y_{it} = j$ und für die Randparameter $\tau_{-1} = -\infty$ und $\tau_J = \infty$ angenommen wird:

$$L = \sum_{i=1}^N \ln P(y_{i1}, y_{i2}, \dots, y_{iT}), \quad (7.54)$$

wobei die Wahrscheinlichkeit $P(\cdot)$ durch ein T -faches Integral gekennzeichnet ist (siehe Gleichung (7.55)). Frechette zeigt, daß dieses Integral vereinfacht werden kann, so daß folgt (vgl. Frechette 2001, S. 24):

$$\begin{aligned} P(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{a_{i1}}^{b_{i1}} \dots \int_{a_{iT}}^{b_{iT}} f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) d\varepsilon_{iT} \dots d\varepsilon_{i1} \quad (7.55) \\ &= \int_{a_{i1}}^{b_{i1}} \dots \int_{a_{iT}}^{b_{iT}} \int_{-\infty}^{\infty} \prod_{i=1}^T f(v_{it} | u_i) f(u_i) du_i d v_{iT} \dots d v_{i1} \\ &= \int_{-\infty}^{\infty} \prod_{i=1}^T F(b_{it} | u_i) - F(a_{it} | u_i) du_i. \end{aligned}$$

¹⁶³ Dieser ergibt sich daraus, daß die Kovarianz des Fehlers zwischen zwei beliebigen Zeitpunkten der Varianz des Random-effects-Parameters entspricht ($\text{Cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_u^2$). Weiterhin ist die Varianz für alle Zeitpunkte gleich, so daß im Nenner die Gesamtvarianz steht (vgl. Ronning 1991, S. 193ff.).

Dabei bezeichnet $f(\cdot)$ die Dichtefunktion und $F(\cdot)$ die Verteilungsfunktion der Normalverteilung. Das Modell wird mit Maximum Likelihood-Methoden geschätzt. Neben den geschätzten Koeffizienten für die erklärenden Variablen β werden dabei noch die geschätzten Werte für Schwellenparameter der abhängigen Variable τ sowie den Korrelationskoeffizienten ρ ermittelt.

7.4.4 Daten

Die in der Panel-Analyse verwendeten Daten stammen wie die in der Untersuchung des Querschnitts (siehe Abschnitt 7.3.3) aus dem Sozioökonomischen Panel des Deutschen Instituts für Wirtschaft. Der Datensatz basiert auf vier ausgewählten Wellen der Jahre 1984 bis 1987. Im wesentlichen stimmen die verwendeten Variablen mit denen der Querschnittsanalyse überein, es ergeben sich jedoch einige Abweichung für Fälle, in denen Variablen nicht in allen betrachteten Wellen zur Verfügung standen.

Die Daten lassen sich wiederum in fünf Gruppen unterteilen (siehe Tabelle 7.6, S. 192): die prädisponierenden und sozioökonomischen Variablen, die Gesundheitslage, die Inanspruchnahme medizinischer Leistungen und das Wohnumfeld (vgl. die Ausführungen auf S. 165ff.). In der Gruppe der *prädisponierenden Variablen* sind das Alter und das quadrierte Alter zu nennen. Sie sollen auf eine Altersabhängigkeit der Zufriedenheit mit der Gesundheit überprüfen und dabei eventuelle Nichtlinearität berücksichtigen. Die Variable ‚Geschlecht‘ nimmt den Wert Eins an, falls es sich bei der Beobachtungsperson um eine Frau handelt. Der Einfluß der Kindererziehung wird durch die Dummy-Variablen ‚Kinder‘ erfaßt. Falls im Haushalt Kinder im Alter unter sechzehn Jahren leben, nimmt diese Variable den Wert Eins an. Der Unterschied zwischen Alleinlebenden und Menschen, die mit einem Partner zusammenleben, wird durch die Variable ‚alleinstehend‘ dargestellt. Dabei ist zu beachten, daß für das Jahr 1985 keine Daten über eine außereheliche Lebensgemeinschaft zur Verfügung stehen. Diese Angaben werden jedoch benötigt, um die Daten über den Familienstand zu korrigieren. Daher wurde für alle Befragten, die sowohl im Jahr zuvor als auch im Jahr nach der Befragung mit einem Partner in einer Lebensgemeinschaft zusammen wohnten, angenommen, daß dies auch für das Jahr 1985 zutrifft. Somit nimmt für diese Gruppe der Befragten die Variable ‚alleinstehend‘ den Wert Null an. Für Befragte mit deutscher Staatsbürgerschaft ist die Variable ‚Ausländer‘ gleich Null, für alle anderen gleich Eins.

Auch für die *sozioökonomischen Variablen* ergeben sich im Vergleich zur Querschnittsanalyse nur geringfügige Änderungen. Das monatliche Haushaltsnettoeinkommen in 1.000 DM soll die Konsummöglichkeiten eines Individuums

abdecken.¹⁶⁴ Der Zusammenhang mit der abhängigen Variablen wird zunächst nicht spezifiziert, da anzunehmen ist, daß die Variable ‚Einkommen‘ auch Effekte auf sich vereint, die keine Relevanz für den Gesundheitszustand besitzen. Die Schulbildung nimmt den Wert Eins für den Abschluß mittlere Reife oder höher bzw. einen vergleichbaren Schulabschluß im Ausland an. Dabei kann sie als ein Proxy für die Compliance des Patienten angesehen werden, da eine höhere Bildung u.U. zu einem bewußteren Umgang mit der eigenen Gesundheit beiträgt. Unterschiede in der Krankenversicherung werden durch die Dummy-Variablen ‚PKV‘ dargestellt, die Aufschluß darüber gibt, ob eine private Krankenversicherung. Durch diese Variable soll überprüft werden, ob unterschiedliche Versicherungssysteme zu Unterschieden in der Gesundheitszufriedenheit führen. Eine weitere Variable, durch die die Compliance abgebildet wird, ist die Variable ‚Sport‘, die den Wert Eins annimmt, wenn regelmäßig aktiver Sport betrieben wird. Diese Variable ist jedoch nur für die ersten drei Wellen verfügbar. Daher mußte die Variable für das Jahr 1987 generiert werden. Dazu wurden die Angaben aus dem Jahr 1988 als Näherungswert herangezogen. Falls nun für das Jahr 1987 eine Arbeitsunfähigkeit von länger als sechs Wochen vorlag oder mehr als zwanzig Nächte im Krankenhaus verbracht wurden, so wurde unterstellt, daß in diesem Jahr kein wöchentlicher, aktiver Sport getrieben wurde.

Zu den Variablen der *Gesundheitslage* gehört die Zufriedenheit mit der eigenen Gesundheit als abhängige Variable. Falls die Befragungsperson mit ihrem Gesundheitszustand ganz und gar nicht zufrieden ist, so erhält sie den Wert Null, für vollständige Zufriedenheit den Wert Zehn. Die Variable besitzt somit elf Kategorien. Weiterhin gehören dieser Gruppe das Leiden unter chronischen Krankheiten und die Variable ‚keine Behinderung‘ an. Letztere gibt an, ob die Erfüllung täglicher Aufgaben durch den Gesundheitszustand beeinträchtigt wird oder nicht. Falls ein Erwerbstätiger im letzten Kalenderjahr 14 Tage oder länger krank geschrieben war, erhält die Variable ‚Arbeitsunfähigkeit‘ den Wert Eins. Für die erste Welle (1984) wurde diese Variable jedoch nicht erhoben. Daher mußte auch hier mit einer Hilfskonstruktion gearbeitet werden. Dazu wurden zuerst die Angaben für das Jahr 1985 herangezogen. Anschließend wurde für alle Befragten, die 1984 mehr als 14 Nächte im Krankenhaus verbrachten und erwerbstätig waren, angenommen, daß eine längere Arbeitsunfähigkeit vorlag.

¹⁶⁴ Dabei wurden auch hier Beobachtungen mit einem angegebenen Haushaltsnettoeinkommen von mehr als 20.000 DM im Monat aus der Stichprobe ausgeschlossen, um Verzerrungen zu vermeiden.

Falls eine Befragungsperson hingegen arbeitslos war, wurde unterstellt, daß keine Arbeitsunfähigkeit vorlag.¹⁶⁵

Da Variablen für das *Wohnumfeld* nicht für alle Wellen verfügbar sind, mußte auch hier mit einer Proxy-Variable gearbeitet werden. Es wurde daher eine Dummy-Variable kreiert, die den Wert Eins annimmt, falls das Individuum in einer Großstadt, d.h. in einer Stadt mit mehr als 100.000 Einwohnern, wohnt. Weiterhin standen auch keine Variablen für die Arbeitsplatzbedingungen zur Verfügung, so daß die Belastungen am Arbeitsplatz lediglich durch die Arbeitsunfähigkeit erfaßt werden.¹⁶⁶

Tabelle 7.6 enthält die verwendeten Variablen, Tabelle 7.7 die deskriptive Statistik des Paneldatensatzes für die einzelnen Stichproben. Für die Gesamtstichprobe stehen 3700 Beobachtungen über alle vier Wellen zur Verfügung, bei den Nichterwerbstätigen sind es 1417 und bei den Erwerbstätigen 2069.¹⁶⁷

Auffällig bei der Betrachtung der deskriptiven Statistik sind die Unterschiede zwischen den betrachteten Stichproben. Das durchschnittliche Alter in der Gesamtstichprobe beträgt 43 Jahre, für die Nichterwerbstätigen liegt es bei 49 Jahren und für die Erwerbstätigen bei 39 Jahren. Weiterhin zeigen sich Abweichungen im Anteil der Frauen in der Stichprobe (Variable ‚Geschlecht‘). Knapp 65 % beträgt der Anteil der Frauen bei den Nichterwerbstätigen, bei den Erwerbstätigen sind es weniger als 30 %. Bei den sozioökonomischen Variablen sind größere Unterschiede im Haushaltsnettoeinkommen festzustellen, das zwischen 2.591 DM für die Nichterwerbstätigen und 3.202 DM für die Erwerbstätigen liegt. Bei den Erwerbstätigen gibt auch ein höherer Anteil der Befragten an, regelmäßig Sport zu betreiben.

¹⁶⁵ Sowohl bezüglich der Variable ‚Sport‘ als auch bezüglich der Variable ‚Arbeitsunfähigkeit‘ ergaben sich keine wesentlichen Änderungen, wenn lediglich die beiden mittleren Wellen für die Schätzung herangezogen und somit generierte Variablen nicht berücksichtigt wurden.

¹⁶⁶ Die Zahl der Krankheitstage hängt jedoch auch von Arbeitsplatzmerkmalen, der beruflichen Stellung, der Arbeitsmotivation und der allgemeinen Arbeitsmarktlage ab. Die Arbeitsunfähigkeit mißt daher nicht nur die Veränderungen des Gesundheitszustandes per se, sondern auch die Variation exogener Variablen, durch die zwar der Indikator, jedoch nicht der Gesundheitszustand beeinflusst wird (vgl. Leu et al. 1986a, S. 161).

¹⁶⁷ Die Stichprobe der Erwerbstätigen umfaßt alle Beschäftigten, mit Ausnahme der Arbeitslosen und der Selbständigen.

Tabelle 7.6: Variablenbeschreibung des Paneldatensatzes

prädispon. Variablen	
Alter	Alter in Jahren
Alter ²	Alter in Jahren quadriert
Geschlecht	1, wenn weiblich
Kinder	1, wenn Kinder unter 16 Jahren im Haushalt
alleinstehend	1, wenn getrennt lebend, ledig, geschieden, Partner verstorben, Ehepartner im Ausland und somit keine Lebensgemeinschaft
Ausländer	1, wenn Staatsangehörigkeit nicht deutsch
sozioök. Variablen	
Einkommen	monatliches Haushaltsnettoeinkommen in 1000 DM
Schulbildung	1, wenn mindestens mittlere Reife oder Besuch einer weiterführenden Schule im Ausland
Sport	1, wenn jede Woche aktiver Sport
PKV	1, wenn Private Krankenversicherung
Gesundheitslage	
Gesundheit	Zufriedenheit mit der Gesundheit 0-10
chronisch	1, wenn leiden unter chronischen Krankheiten
keine Behinderung	1, wenn Erfüllung täglicher Aufgaben durch den Gesundheitszustand überhaupt nicht behindert
Arbeitsunfähigkeit	1, wenn 14 Tage oder länger krank geschrieben
Medizinische Inanspruchnahme	
prak. Arzt	1, wenn Besuch beim praktischen Arzt im letzten Quartal
Facharzt	1, wenn Besuch beim Facharzt im letzten Quartal
KH-Aufenthalte	1, wenn Krankenhausaufenthalt im letzten Jahr
Wohnumfeld	
Großstadt	1, wenn in einer Großstadt lebend

Betrachtet man die Ausprägung der Zufriedenheit mit der Gesundheit für die Stichproben, so erkennt man, daß im Vergleich mit der Gesamtstichprobe und den Erwerbstätigen die Nichterwerbstätigen eine durchschnittlich geringere Zufriedenheit aufweisen. Darüber hinaus ist besitzt diese Stichprobe den größten Anteil von Personen mit chronischen Krankheiten. Ähnlich läßt sich die Variable ‚keine Behinderung‘ interpretieren. Die Gruppe der Nichterwerbstätigen weist für alle Stichproben den niedrigsten Wert für diese Variable auf. Dies bedeutet, daß der Anteil der Befragten, die angaben, bei ihren täglichen Arbeiten durch ihren Gesundheitszustand behindert zu werden, in der Gruppe der Nichterwerbstätigen am höchsten ist (43,6 % im Vergleich zu 33,8 % bei der Gesamtstichprobe und 27,2 % bei den Erwerbstätigen). Weiterhin ist der Anteil derjenigen, die einen Arzt aufsuchen oder angaben, im letzten Jahr im Krankenhaus gewesen zu sein, auch höher. Für die restlichen Variablen sind die Unter-

schiede in den Mittelwerten nur gering, so daß keine gravierenden Abweichungen festzustellen sind.

Tabelle 7.7: Deskriptive Statistik des Paneldatensatzes

Stichprobe Variable	gesamt		nicht erwerbstätig		erwerbstätig	
	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung
prädispon. Variablen						
Alter	42,905	15,745	48,905	18,882	38,749	11,918
Alter ²	2088,72	1472,81	2747,96	1850,95	1643,42	950,62
Geschlecht	0,426	0,495	0,649	0,477	0,289	0,453
alleinstehend	0,294	0,456	0,317	0,465	0,287	0,452
Kinder	0,427	0,495	0,362	0,481	0,467	0,499
Ausländer	0,223	0,416	0,175	0,380	0,263	0,440
sozioök. Variablen						
Einkommen	2,995	1,453	2,591	1,440	3,202	1,300
Schulbildung	0,276	0,447	0,214	0,411	0,310	0,463
PKV	0,088	0,283	0,074	0,262	0,080	0,271
Sport	0,193	0,394	0,164	0,371	0,218	0,413
Gesundheitslage						
Gesundheit	7,064	2,472	6,612	2,728	7,344	2,245
chronisch	0,289	0,454	0,382	0,486	0,234	0,423
keine Behinderung	0,662	0,473	0,564	0,496	0,728	0,445
Arbeitsunfähigkeit	-	-	-	-	0,167	0,373
med. Inanspruchnahme						
prak. Arzt	0,421	0,494	0,510	0,500	0,374	0,484
Facharzt	0,422	0,494	0,511	0,500	0,375	0,484
KH-Aufenthalte	0,087	0,282	0,109	0,312	0,072	0,259
Wohnumfeld						
Großstadt	0,314	0,464	0,320	0,467	0,318	0,466

7.4.5 Schätzergebnisse

Bei den hier präsentierten Schätzungen handelt es sich um eine Random-effects-Ordered-Probit-Schätzung, wie sie in Abschnitt 7.4.3 dargestellt wurde. Analog zur Schätzung der Querschnittsdaten werden für die Stichproben ‚gesamt‘, ‚nicht erwerbstätig‘ und ‚erwerbstätig‘ dieselben erklärenden Variablen benutzt. Im einzelnen sind dies prädisponierende und sozioökonomische Variablen, weiterhin Variablen der Gesundheitslage, der medizinischen Inanspruchnahme sowie die Variable ‚Großstadt‘ für das Wohnumfeld. Für die Jahre von 1985 bis 1987 wurden Dummy-Variablen eingeführt, die auf periodenspezifische Effekte

überprüfen sollen. Zusätzlich wird für die Erwerbstätigen in einer weiteren Schätzung die Arbeitsunfähigkeit als erklärende Variable mit einbezogen. Die abhängige Variable ist in allen vier Schätzungen die selbsteingeschätzte Zufriedenheit mit der Gesundheit. Sie weist 11 Kategorien auf, von der Einschätzung des Gesundheitszustands als ganz und gar nicht zufrieden bis zur Beurteilung als ganz und gar zufriedenstellend.

Schätzung Gesamtstichprobe

Für die Gesamtstichprobe läßt sich an den periodenspezifischen Effekten erkennen, daß die Zufriedenheit mit der Gesundheit im Zeitablauf zurückgeht, unabhängig von den betrachteten Einflüssen der erklärenden Variablen (siehe hierzu Tabelle 7.8, Schätzung (1)). Für das Alter folgt, ähnlich wie in der Querschnittsschätzung, ein signifikant u-förmiger Zusammenhang mit der Gesundheitszufriedenheit. Mit zunehmendem Alter verringert sich die Zufriedenheit mit der Gesundheit zunächst, um ab einem bestimmten Alter wieder anzusteigen. Dieser Wendepunkt liegt für die Gesamtstichprobe bei 61,6 Jahren, für die Nichterwerbstätigen bei 58,7 Jahren und bei den Erwerbstätigen bei 57 Jahren. Des weiteren weisen Frauen eine signifikant geringere Zufriedenheit mit der Gesundheit auf. Alle anderen prädisponierenden Variablen sind nicht signifikant. In der Gruppe der sozioökonomischen Variablen besitzt die regelmäßige sportliche Aktivität einen positiven, signifikanten Koeffizienten. Dies spricht für die Gültigkeit von *Hypothese 1*, daß der Gesundheitszustand auch durch eigene Handlungen des Patienten beeinflusst wird. Alle anderen Variablen dieser Gruppe sind nicht signifikant, so daß dem Krankenversicherungsschutz keine Auswirkung auf die Zufriedenheit mit der Gesundheit zukommt. *Hypothese 4* kann daher nicht bestätigt werden.

Die Variablen der Gesundheitslage sind beide hochsignifikant und weisen das erwartete Vorzeichen auf. Ähnliches gilt für die medizinische Inanspruchnahme. Alle drei Koeffizienten besitzen ein negatives Vorzeichen. Der Koeffizient für einen Facharztbesuch ist jedoch nur auf dem 10 %-Niveau signifikant. Wie bereits in der Querschnittsschätzung liegt die Vermutung nahe, daß insbesondere Personen mit schlechtem Gesundheitszustand ärztliche Leistungen in Anspruch nehmen (*Hypothese 3*). Es fällt hierbei auf, daß die Größe der Koeffizienten für die medizinische Inanspruchnahme von derjenigen der Querschnittsschätzung deutlich abweicht. Das Wohnumfeld, die Tatsache also, ob eine Person in einer Großstadt wohnt, hat keinen signifikanten Einfluß auf die Zufriedenheit mit der Gesundheit. Möglicherweise vermischen sich hier Effekte, die in der Querschnittsschätzung durch verschiedene Variablen abgedeckt waren, bspw. der positive Effekt der medizinischen Infrastruktur mit negativen Effekten der

Umweltbelastung.¹⁶⁸ Für den Korrelationskoeffizienten ρ ergibt sich ein Wert von 0,3624, d.h., die Korrelation zwischen den Zeitpunkten ist nicht allzu hoch.

Für die Gütemaße der Schätzung (siehe Abschnitt 7.3.2), die dazu dienen, unterschiedliche Schätzungen miteinander zu vergleichen, läßt sich festhalten, daß das Pseudo- R^2 nach McFadden einen Wert von 0,058, das McKelvey/Zavoina-Maß ist 0,447 und das Akaike-Informationsmaß ist 3,632 besitzt. Im Vergleich zur Querschnittsschätzung ist das McFadden-Maß niedriger, wohingegen das McKelvey/Zavoina-Maß höher ausfällt. Ob nun die Querschnitts- oder die Panelschätzung die besseren Ergebnisse liefert, läßt sich durch diese beiden Maße nicht abschließend beurteilen. Das Akaike-Maß liegt jedoch unter dem Wert der Querschnittsschätzung, was für eine bessere Modellanpassung der Panel-Analyse spricht.

Schätzung Nichterwerbstätige

Im Vergleich zur ersten Schätzung weichen die Ergebnisse für die Nichterwerbstätigen leicht ab (siehe Tabelle 7.8, S. 196, Schätzung (2)). Als erstes fällt auf, daß die periodenspezifischen Effekte nicht signifikant sind. Für die betrachtete Gruppe bedeutet dies, daß keine Einflüsse im Zeitablauf ausgemacht werden können, die die Zufriedenheit mit der Gesundheit beeinflussen. Das Alter besitzt dagegen einen signifikant u-förmigen Einfluß, d.h., daß die Wahrscheinlichkeit einer hohen Zufriedenheit mit der Gesundheit mit zunehmendem Alter erst zurückgeht, jedoch dann wieder zunimmt. Im Gegensatz zur Gesamtstichprobe ergibt sich kein signifikanter Unterschied in der Gesundheitszufriedenheit zwischen Männern und Frauen. Auch die weiteren prädisponierenden Variablen sind nicht signifikant. Ein ähnliches Bild läßt sich auch für die sozioökonomischen Variablen feststellen, bei denen aktiver Sport als einziger Faktor einen signifikanten Einfluß aufweist. Dieser positive Koeffizient bestätigt auch hier die *Hypothese 1*. Hingegen besitzen weder das monatliche Haushaltseinkommen noch die Schulbildung einen signifikanten Effekt. Wie in der ersten Schätzung und in Analogie zu den Ergebnissen der Querschnittsanalyse hat das Versicherungssystem keine Auswirkungen auf die Zufriedenheit mit der Gesundheit.

¹⁶⁸ Es stellt sich dabei allerdings die Frage, ob die Variable ‚Großstadt‘ tatsächlich in der Lage ist, das Wohnumfeld abzubilden oder ob eine Verwendung dieser Variablen nicht zu wenig zwischen Wohnqualität und Konsummöglichkeiten unterscheidet.

Tabelle 7.8: Schätzergebnisse für die Gesamtstichprobe und die Nicht-erwerbstätigen (Panel-Analyse)

abhängige Variable: Zufriedenheit mit der Gesundheit

erklärende Variablen	(1) gesamt		(2) nicht erwerbstätig	
	Koeffizient	P-Wert ^{a)}	Koeffizient	P-Wert ^{a)}
periodenspez. Effekte				
Jahr 1985	-0,1260	0,011**	0,0513	0,536
Jahr 1986	-0,1758	0,000***	-0,0565	0,500
Jahr 1987	-0,1642	0,001***	-0,0134	0,876
prädispon. Variablen				
Alter	-0,0616	0,000***	-0,0822	0,000***
Alter ²	0,0005	0,000***	0,0007	0,000***
Geschlecht	-0,1389	0,028**	-0,1325	0,226
alleinstehend	-0,0438	0,578	-0,0979	0,447
Kinder	-0,0858	0,171	-0,0479	0,678
Ausländer	-0,0082	0,916	-0,1900	0,173
sozioök. Variablen				
Einkommen	0,0259	0,174	0,0201	0,529
Schulbildung	0,0970	0,178	0,1516	0,222
PKV	0,0329	0,730	-0,0122	0,943
Sport	0,1074	0,082*	0,1997	0,078*
Gesundheitslage				
chronisch	-0,5103	0,000***	-0,6028	0,000***
keine Behinderung	0,8563	0,000***	0,8476	0,000***
Arbeitsunfähigkeit	-	-	-	-
med. Inanspruchnahme				
prak. Arzt	-0,4328	0,000***	-0,4549	0,000***
Facharzt	-0,1281	0,070*	-0,1564	0,169
KH-Aufenthalte	-0,3442	0,000***	-0,4096	0,000***
Wohnumfeld				
Großstadt	-0,0621	0,343	0,0331	0,759
ρ	0,3624	0,000***	0,3935	0,000***
Log Likelihood	-6690,00		-2624,89	
LR	881,15	0,000	415,39	0,000
McFadden's R ² adj.	0,058		0,063	
McKelvy/Zavoina R ²	0,447		0,490	
Akaike's Information Criterion	3,632		3,747	
N	3700		1417	

*** signifikant auf dem 1% Niveau

** signifikant auf dem 5% Niveau

* signifikant auf dem 10% Niveau

^{a)} Der P-Wert gibt das minimale Signifikanzniveau (Fehler 1. Art) an, für das der beobachtete Wert der Teststatistik zu einer Ablehnung der Nullhypothese (Koeffizient = 0) führt (vgl. Kennedy 1998, S. 409).

Für die Gesundheitslage ergibt sich, daß chronische Krankheiten einen signifikant negativen Einfluß auf die Wahrscheinlichkeit einer hohen Zufriedenheit mit der Gesundheit besitzen. Wird weiterhin die Erfüllung täglicher Aufgaben nicht durch den Gesundheitszustand beeinträchtigt, so ist die Zufriedenheit mit dem Gesundheitszustand höher. Für die medizinische Inanspruchnahme ergeben sich allerdings Abweichungen zur ersten Schätzung und zur Querschnittsanalyse. Zwar sind die Koeffizienten für den Besuch beim praktischen Arzt ebenso wie für einen Krankenhausaufenthalt signifikant negativ, für einen Facharztbesuch ergibt sich jedoch kein signifikanter Zusammenhang. Eine mögliche Erklärung liegt darin, daß über den betrachteten Zeitraum kein stabiles Überweisungsverhalten zwischen praktischem Arzt und Facharzt vorliegt.¹⁶⁹ Wie in der Schätzung für die Gesamtstichprobe besitzt auch die Variable ‚Großstadt‘ keinen signifikanten Einfluß auf die Zufriedenheit mit der Gesundheit.

Der Korrelationskoeffizient beträgt 0,3935 und die Pseudo-R²-Maße sind im Vergleich mit der vorherigen Schätzung etwas höher. So liegt das Maß von McFadden bei 0,063 und das von McKelvey und Zavoina bei 0,490, was auf eine bessere Anpassung für die Schätzung der Nichterwerbstätigen schließen läßt. Für eine bessere Modellanpassung der Gesamtstichprobe spricht jedoch der mit 3,747 etwas höhere Wert des Akaike Informationsmaßes für die Nichterwerbstätigen.

Schätzung Erwerbstätige

Betrachtet man in Tabelle 7.9 (siehe S. 199, Schätzung (3)) die Ergebnisse für die Erwerbstätigen, so stellt man zunächst die signifikanten periodenspezifischen Effekte fest. Im Vergleich zum Basisjahr 1984 geht die Zufriedenheit mit der Gesundheit im Zeitablauf zurück, unabhängig von anderen Faktoren, die mit Hilfe erklärender Variablen untersucht werden. In der Gruppe der prädisponierenden Variablen sind das Alter und der quadratische Altersterm signifikant, so daß sich der bereits beschriebene u-förmige Zusammenhang zwischen der Gesundheit und dem Alter ergibt. Alle anderen Variablen sind nicht signifikant. Dies trifft auch für die Variable ‚Geschlecht‘ zu, deren Vorzeichen eine geringere Zufriedenheit mit der Gesundheit bei Frauen vermuten läßt, deren Signifikanz jedoch knapp über 10 % liegt. Bei den sozioökonomischen Variablen fällt auf, daß nun das Einkommen einen signifikanten, positiven Einfluß auf die Gesundheit ausübt. Weiterhin kann für den aktiven Sport kein statistisch gesicherter Zusammenhang festgestellt werden.

¹⁶⁹ Es bleibt anzumerken, daß im Gegensatz zur heutigen Situation im Gesundheitswesen ein direktes Aufsuchen des Facharztes nicht möglich war, sondern es einer Überweisung durch den Hausarzt bedurfte.

Eine mögliche Begründung hierfür liegt in den Opportunitätskosten der Zeit. Aus der deskriptiven Statistik (siehe S. 193) geht hervor, daß das durchschnittliche Einkommen der Erwerbstätigen im Vergleich zu den anderen Stichproben höher ist. Darüber hinaus wird aktiver Sport über die Wellen hinweg eher unregelmäßig betrieben.¹⁷⁰ Für die Variablen der Gesundheitslage ergeben sich mit den bisherigen Schätzungen vergleichbare Ergebnisse. Beide Koeffizienten sind signifikant und besitzen das erwartete Vorzeichen. Für die Variable ‚chronisch‘ weist der Koeffizient absolut betrachtet einen niedrigeren Wert im Vergleich zu den anderen Schätzungen auf, was auch im Fall der Querschnittsuntersuchung zutrifft. Der Hauptgrund dafür dürfte im geringeren Anteil derjenigen Personen liegen, dieangaben, unter chronischen Gesundheitsbeschwerden zu leiden. Die Variablen der medizinischen Inanspruchnahme sind bis auf die Facharztbesuche signifikant negativ, was für die *Hypothese 3* spricht. Weiterhin ist ersichtlich, daß trotz der Inanspruchnahme die Wahrscheinlichkeit einer hohen Zufriedenheit mit der Gesundheit geringer ist als ohne.¹⁷¹

Auch in dieser Schätzung besitzt die Variable ‚Großstadt‘ keinen signifikanten Einfluß, allerdings fällt der (absolut gesehen) größere Wert des Koeffizienten ins Auge. Die Korrelation des Fehlerterms über die Zeit liegt mit 0,359 im normalen Bereich. Für das McFadden R^2 -Maß beträgt der Wert 0,053, für das McKelvey/Zavoina-Maß 0,436. Das Akaike-Maß weist im Vergleich mit den anderen Stichproben mit 3,571 einen niedrigeren Wert auf, was auf eine bessere Modellanpassung für die Stichprobe der Erwerbstätigen schließen läßt.

Schätzung Erwerbstätige inklusive Arbeitsunfähigkeit

Für die letzte Schätzung wurde zu den bereits überprüften Variablen noch die längere Arbeitsunfähigkeit als Erklärung der Zufriedenheit mit der Gesundheit hinzugenommen (siehe Tabelle 7.9 auf S. 199, Schätzung (4)). Die Arbeitsunfähigkeit dient dabei als Proxy für die Arbeitsbelastungen. Die Ergebnisse für die bereits in den anderen Schätzungen enthaltenen Variablen ergeben nur wenige Änderungen. Für die periodenspezifischen Effekte und die Altersterme ergibt sich ein mit Schätzung (3) vergleichbares Bild. Frauen weisen eine signifikant geringere Zufriedenheit mit der Gesundheit als Männer auf, ein Ergebnis, das auch in der Gesamtstichprobe zu finden war.

¹⁷⁰ Dies ist daraus ersichtlich, daß die Korrelation zwischen den einzelnen Wellen für die Variable ‚Sport‘ bei den Erwerbstätigen im Schnitt bei 0,5 lag, für die Nichterwerbstätigen und die Gesamtstichprobe jedoch deutlich höher bei 0,7.

¹⁷¹ Es bleibt anzumerken, daß analog zur Querschnittsanalyse nicht in abgeschlossene und noch laufende Behandlungen unterschieden werden kann. Der Erfolg einer Behandlung kann daher mit den vorliegenden Daten nicht untersucht werden.

Tabelle 7.9: Schätzergebnisse Erwerbstätige (Panel-Analyse)

abhängige Variable: Zufriedenheit mit der Gesundheit

erklärende Variablen	(3) erwerbstätig		(4) inkl. Arbeitsunfähigkeit	
	Koeffizient	P-Wert ^{a)}	Koeffizient	P-Wert ^{a)}
periodenspez. Effekte				
Jahr 1985	-0,2449	0,000***	-0,2456	0,000***
Jahr 1986	-0,2703	0,000***	-0,2663	0,000***
Jahr 1987	-0,2549	0,000***	-0,2520	0,000***
prädispon. Variablen				
Alter	-0,0568	0,009***	-0,0575	0,008**
Alter ²	0,0005	0,058*	0,0005	0,054*
Geschlecht	-0,1376	0,117	-0,1440	0,100*
alleinstehend	-0,0531	0,612	-0,0589	0,567
Kinder	-0,0642	0,425	-0,0628	0,434
Ausländer	0,0323	0,729	0,0422	0,651
sozioök. Variablen				
Einkommen	0,0697	0,011**	0,0697	0,011**
Schulbildung	0,0652	0,475	0,0597	0,513
PKV	-0,1078	0,418	-0,1185	0,374
Sport	0,0576	0,453	0,0433	0,574
Gesundheitslage				
chronisch	-0,4265	0,000***	-0,4168	0,000***
keine Behinderung	0,9153	0,000***	0,8966	0,000***
Arbeitsunfähigkeit	-	-	-0,2174	0,013**
med. Inanspruchnahme				
prak. Arzt	-0,4205	0,000***	-0,4127	0,000***
Facharzt	-0,1266	0,188	-0,1143	0,235
KH-Aufenthalte	-0,2193	0,036**	-0,1013	0,327
Wohnumfeld				
Großstadt	-0,1105	0,185	-0,1013	0,225
ρ	0,359	0,000***	0,358	0,000***
Log Likelihood	-3664,30		-3661,21	
LR	473,83	0,000	480,02	0,000
McFadden's R ²	0,053		0,054	
McKelvy/Zavoina R ²	0,436		0,432	
Akaike's Information Criterion	3,571		3,569	
N	2069		2069	

*** signifikant auf dem 1% Niveau

** signifikant auf dem 5% Niveau

* signifikant auf dem 10% Niveau

^{a)} Der P-Wert gibt das minimale Signifikanzniveau (Fehler 1. Art) an, für das der beobachtete Wert der Teststatistik zu einer Ablehnung der Nullhypothese (Koeffizient = 0) führt (vgl. Kennedy 1998, S. 409).

Andere prädisponierende Variablen besitzen keinen statistisch abgesicherten Einfluß, ebenso wie die sozioökonomischen Variablen. Eine Ausnahme bildet

das Einkommen, das wie in Schätzung (3) signifikant ist. Aufgrund der insignifikanten Variablen ‚Schulbildung‘ und ‚Sport‘ kann *Hypothese 1* nicht bestätigt werden, bestenfalls kann der positive Effekt des Einkommens in Hinsicht eines gesundheitsfördernden Konsums interpretiert werden. Alle Variablen der Gesundheitslage sind signifikant und besitzen das erwartete Vorzeichen. Insbesondere folgt aus dem Umstand, daß eine Person vierzehn Tage oder länger arbeitsunfähig war, daß die Zufriedenheit mit der Gesundheit niedriger ist. Gleichzeitig ist der Koeffizient für einen Krankenhausaufenthalt nicht mehr signifikant. Der Zusammenhang zwischen Krankenhausaufenthalten und Arbeitsunfähigkeit, wie er auch für die Querschnittsschätzung formuliert wurde (siehe S. 179), trifft somit auch in diesem Fall zu. Besuche beim praktischen Arzt haben einen signifikant positiven Einfluß auf die Zufriedenheit mit der Gesundheit, während für Facharztbesuche kein signifikanter Effekt festgestellt werden konnte. Damit folgt, daß zum Teil *Hypothese 3* bestätigt werden kann. Auch die Ergebnisse für die Variable ‚Großstadt‘ sowie für den Korrelationskoeffizienten decken sich mit denen der vorhergehenden Schätzung.

Bezüglich der Gütemaße ist festzuhalten, daß die Werte für die R^2 -Maße nach McFadden und McKelvey/Zavoina bei 0,054 bzw. 0,432 und damit geringfügig unter denen von Schätzung (3) liegen. Das Akaike-Maß ist jedoch mit 3,569 für alle Panel-Schätzungen am niedrigsten. Dies läßt den Schluß zu, daß sich mit den in Schätzung (4) verwendeten Variablen die beste Modellanpassung ergibt.

7.5 *Kritische Würdigung*

Die empirische Umsetzung der in den vorausgegangenen Kapiteln hergeleiteten theoretischen Ergebnisse erweist sich als problematisch, da eine direkte Übertragung in ein empirisches Modell eine Vielzahl von Schwierigkeiten mit sich bringt. Dies betrifft vor allem die für eine Schätzung zur Verfügung stehenden Variablen. Zur Darstellung des Verhältnisses zwischen Arzt und Patient sowie deren Interaktionen vor dem Hintergrund der vertraglichen Situationen benötigt man neben Daten über die medizinische Inanspruchnahme und das gesundheitsrelevante Verhalten auch Informationen über die Versicherungs- und Honorierungsvereinbarungen. In der Regel stehen diese jedoch nicht im gewünschten Umfang oder sogar überhaupt nicht zur Verfügung, so daß eine Übertragung der Ergebnisse in ein empirisches Modell kaum möglich ist. Somit muß als Folge des Datenproblems der Gegenstand der Analyse auf das Verhältnis zwischen medizinischer Leistung, Compliance des Patienten und dem Gesundheitszustand beschränkt werden, das allerdings mit im Zentrum der theoretischen Analyse steht.

Aus dem verfügbaren Datenmaterial des Sozioökonomischen Panels der Jahre 1984 bis 1987 ergibt sich, daß der Focus in der empirischen Analyse der Arzt-Patient-Beziehung auf der Erklärung der Zufriedenheit des Patienten mit seiner Gesundheit liegt. Als Schätzmethode werden dabei sowohl für die Querschnitts- als auch für die Panel-Analyse Modelle für geordnete Kategorien verwendet. Dazu ist es zunächst notwendig, auf die Grundlagen der Analyse binärer abhängiger Variablen einzugehen, bevor Modelle für geordnete Kategorien erläutert werden. Die Schätzungen werden für drei Teilstichproben des Datensatzes (gesamt, nichterwerbstätig und erwerbstätig) durchgeführt. Die Ergebnisse der Querschnittsschätzungen signalisieren, daß das Verhalten des Patienten in nicht unwesentlichem Maße zur Gesundheitszufriedenheit beiträgt. Dies betrifft insbesondere regelmäßigen aktiven Sport oder auch die Bildung. Einschränkungen des täglichen Lebens durch den Gesundheitszustand verringern die Wahrscheinlichkeit einer hohen Zufriedenheit mit der Gesundheit. Weiterhin ergibt sich, daß die Inanspruchnahme medizinischer Leistungen mit einer geringen Zufriedenheit mit dem eigenen Gesundheitszustand einhergeht. Die Opportunitätskosten eines Arztbesuchs, gemessen durch die Entfernung zum Hausarzt, verringern die Zufriedenheit mit der Gesundheit. Darüber hinaus nimmt die Zufriedenheit mit der Entfernung des Wohnsitzes von einem Großstadtzentrum zu und mit der Luftverschmutzung und der Lärmbelästigung ab. Die Ergebnisse unterscheiden sich allerdings leicht für die einzelnen Stichproben. Für die Erwerbstätigen ergibt sich zusätzlich, daß sich eine längere Arbeitsunfähigkeit negativ auf die Gesundheitszufriedenheit auswirkt und daß weiterhin Arbeitscharakteristika eher die Belastung am Arbeitsplatz widerspiegeln als die Zufriedenheit mit der Gesundheit.

Für die Panelschätzung standen nicht alle in der Querschnittsuntersuchung verwendeten Variablen zur Verfügung, so daß zum Teil mit anderen Größen gearbeitet werden mußte, zum Teil fehlende Daten für einzelne Wellen generiert wurden. Die Ergebnisse der Querschnittsschätzungen wurden im großen und ganzen durch die Panel-Untersuchung bestätigt. Die sportliche Betätigung und die Schulbildung weichen jedoch in einigen Schätzungen davon ab. Weiterhin erweist sich der Besuch bei Fachärzten in drei der vier Schätzungen als nicht signifikant, was mit dem Überweisungsverhalten in den achtziger Jahren zusammenhängen dürfte. Insgesamt gesehen unterstreichen die Panelschätzungen jedoch die in den Querschnittsschätzungen ermittelten Resultate und bestätigen zumindest teilweise den Zusammenhang zwischen der Compliance des Patienten, der Inanspruchnahme medizinischer Leistungen und der Zufriedenheit mit der Gesundheit.

8 Zusammenfassung und Ausblick

Die Beziehung zwischen Patient und behandelndem Arzt stellt die Kernbeziehung innerhalb des deutschen Gesundheitswesens dar. Auf den Märkten für medizinische Leistungen existieren jedoch Informationsasymmetrien in einer Vielzahl, wie sie in kaum einem anderen Bereich einer Ökonomie anzutreffen sind. Diese Phänomene, die im Gesundheitswesen auch durch die Delegation von Aufgaben auftreten, werden in der ökonomischen Theorie in Principal-Agent-Modellen behandelt. Prinzipiell bietet sich daher die Übertragung vorhandener Ansätze auf die Arzt-Patient-Beziehung an. Letztendlich scheidet dies jedoch an den komplizierten Rahmenbedingungen des Gesundheitswesens. So sind neben Arzt und Patient noch zahlreiche ergänzende Sachwalter in das Gesundheitswesen involviert. Dazu zählen der Gesetzgeber, die Krankenversicherer, die Arbeitgeber und die Verbände des Gesundheitswesens. Von diesen sind insbesondere die Krankenversicherer für eine weitere Analyse interessant, da sie sowohl mit dem Arzt als auch mit dem Patienten in direktem Kontakt stehen. Dabei kommen den Versicherern als Aufgaben die Absicherung des monetären Krankheitsrisiko des Patienten gegen eine vereinbarte Prämie und die Vergütung der ärztlichen Leistungen zu. Aufgrund der Anwesenheit einer Versicherung als ergänzender Sachwalter bestehen zwischen den Aktionen der Beteiligten des Gesundheitswesens Interdependenzen, die im Rahmen einer theoretischen Modellierung zu berücksichtigen sind.

Betrachtet man zunächst vorhandene Modelle der Beziehung zwischen Arzt und Patient, die eine Versicherung explizit in die Struktur des Modells aufnehmen, so stellt man fest, daß diese bspw. eigene Handlungen des Patienten oder das Entscheidungskalkül des Arztes vernachlässigen. Weiterhin werden in diesen Ansätzen die Interdependenzen zwischen den Handlungen von Arzt und Patient nur zum Teil berücksichtigt. Darüber hinaus muß das Double Moral Hazard-Problem, das zwischen Arzt und Patient besteht, mit in die Überlegungen einbezogen werden. Darunter versteht man, daß weder der Patient die Aktionen des Arztes und noch der Arzt die gesundheitsrelevanten Anstrengungen des Patienten (Compliance) beobachten oder steuern kann. In einer solchen Konstellation ergibt sich, daß das Niveau der Handlungen der Akteure sich nicht nur gegenseitig beeinflusst, sondern darüber hinaus auch noch von den strategischen Interaktionen zwischen den Handlungen abhängt. Daraus folgt, daß es für das Ergebnis des Leistungserstellungsprozesses wichtig ist, ob es sich bei der medizinischen Leistung des Arztes und bei der Compliance des Patienten um strategische Komplemente oder strategische Substitute handelt oder ob beide Inputfaktoren der Gesundheitsproduktion voneinander unabhängig sind. Daneben beeinflusst auch die Wahl der Versicherungs- und Honorierungsparameter das re-

sultierende Gleichgewicht. Die Einführung einer Selbstbeteiligung des Patienten erhöht im Modell das gesundheitsrelevante Verhalten des Patienten, da es nun im Vergleich zur medizinischen Leistung relativ billiger ist. Die aus der Einführung der Selbstbeteiligung resultierenden Veränderungen der medizinischen Leistung des Arztes hängen dann von den strategischen Interaktionen zwischen den Handlungen ab. Falls die medizinische Leistung und die Compliance unabhängig voneinander sind, ändert sich das Niveau der ärztlichen Aktionen nicht. Falls strategische Komplemente vorliegen, erhöht sich neben dem gesundheitsrelevanten Verhalten auch die medizinische Leistung. Deren Niveau geht jedoch bei strategischen Substituten durch die Einführung einer Selbstbeteiligung zurück.

Bezieht man diese Ergebnisse in die Vertragsgestaltung mit ein, so ist aufgrund der beidseitigen asymmetrischen Information keine vollständige Versicherung des Patienten mehr möglich. Es existiert folglich für ihn eine positive Selbstbeteiligung. Die Honorierungsparameter des Arztes werden bei vollständiger Information so gewählt, daß der direkte Effekt einer Erhöhung der Fallpauschale oder der Kostenerstattung auf den Erwartungsnutzen des Patienten konstant ist. Im Fall des Double Moral Hazard trifft dies nur für die Fallpauschale zu, da sie die Aktionen des Arztes und des Patienten nicht beeinflusst. Der direkte Effekt einer Erhöhung der Kostenerstattung auf den Erwartungsnutzen des Patienten bleibt bei asymmetrischer Information unklar und hängt u.a. von den marginalen Kosten einer gestiegenen Selbstbeteiligung des Patienten oder dem marginalen Effekt der Kostenerstattung auf die Versicherungsprämie ab.

Eine direkte Übertragung der theoretischen Ansätze in ein empirisches Modell erweist sich vor allem aus Gründen der Datenverfügbarkeit als nicht praktikabel. Aus den theoretischen Befunden lassen sich für die empirische Überprüfung der Beziehung zwischen Arzt und Patient Hypothesen ableiten. Diese werden sowohl mittels einer Querschnittsanalyse als auch in einer Panel-Untersuchung überprüft. Als Daten stehen vier Wellen des Sozioökonomischen Panels des Deutschen Instituts für Wirtschaftsforschung zur Verfügung. Die selbsteingeschätzte Zufriedenheit mit der Gesundheit ist dabei die endogene Variable, die durch prädisponierende und sozioökonomische Variablen sowie Variablen der Gesundheitslage, der medizinischen Inanspruchnahme, des Wohnumfeldes und der Arbeitsbelastung erklärt werden soll. Als Schätzmethode wird eine Ordered-Probit-Schätzung für die Querschnittsdaten und ein Random-effects-Ordered-Probit-Modell für die Panel-Analyse verwendet. Hauptergebnisse dieser Untersuchungen sind, daß neben dem Alter auch das Verhalten des Patienten die Zufriedenheit mit der Gesundheit beeinflusst. Weiterhin ergibt sich, daß zwischen der Gesundheitszufriedenheit und der medizinischen Inanspruchnahme ein ne-

gativer Zusammenhang besteht. Dies erklärt sich dadurch, daß diese Leistungen nur in Anspruch genommen werden, wenn der Gesundheitszustand schlecht ist. Da weiterhin die Daten nicht in abgeschlossene und andauernde Behandlungen unterscheiden, muß der Patient zum Zeitpunkt der Behandlung noch nicht vollständig genesen sein, so daß die Zufriedenheit mit der Gesundheit auch nach einer Behandlung gering ist. Für das Wohnumfeld, für das die Daten nur in der Querschnittsuntersuchung zur Verfügung standen, ergibt sich, daß eine größere Entfernung zum Hausarzt die Zufriedenheit mit der Gesundheit negativ, eine größere Entfernung zum Großstadtzentrum diese positiv beeinflusst. Luftverschmutzung und Lärmbelästigung verringern die Wahrscheinlichkeit einer hohen Zufriedenheit mit der Gesundheit. Bezieht man Variablen der Arbeitsbelastung mit in die Analyse ein, so kann allerdings nur für die nervliche Anspannung am Arbeitsplatz ein signifikanter Zusammenhang mit der Gesundheit festgestellt werden.

Für weitere Untersuchungen des Verhältnisses zwischen Arzt und Patient vor dem Hintergrund der bestehenden Informationsasymmetrien bietet es sich an, den hier vorgestellten Ansatz des Kontaktes zwischen Arzt und Patient auf wiederholte Konsultationen zu erweitern, so daß auch die Effekte einer Selbstbeteiligung auf die Inanspruchnahme der medizinischen Leistung modelliert werden. Dazu bedarf es jedoch der konkreten Formulierung der Kontaktentscheidung des Patienten bzw. eines vorausgehenden Suchprozesses. Ein weiterer Ansatzpunkt ist die Ausweitung des gesundheitsrelevanten Verhaltens auch auf die Situation im Vorfeld einer Arztkonsultation, so daß aus Sicht des Arztes nicht nur ein *ex post*, sondern auch ein *ex ante* Moral Hazard des Patienten vorliegt. Dies ist insbesondere im Rahmen der Gestaltung des Versicherungsvertrages von Bedeutung. Für zukünftige Arbeiten im Bereich der empirischen Umsetzung liegt ein wichtiger Aspekt in den Möglichkeiten, die simultane Entscheidung von Arzt und Patient über ihre Aktionen zu modellieren, so daß die theoretischen Aussagen über die Interaktionen auch überprüft werden können.

Die Analyse der Beziehung zwischen Arzt und Patient unter Berücksichtigung der bestehenden Informationsasymmetrien untermauert, daß im Gesundheitswesen erhebliche Anreiz- und Steuerungsprobleme vorliegen. Auch die Anwesenheit ergänzender Sachwalter kann diese nur zu einem Teil beseitigen. Wichtig in den zukünftigen Reformbemühungen des Gesetzgebers dürfte es sein, die Eigenverantwortung des Patienten zu stärken und den Arzt mit den nötigen Anreizen auszustatten, den Patienten in diesem Unterfangen zu unterstützen. Dazu ist es aber notwendig, die gegenseitigen Interdependenzen der Handlungen des Arztes und des Patienten zu erkennen und bei der Ausgestaltung der Versicherungs- und Vergütungssysteme zu berücksichtigen.

Literaturverzeichnis

- Acton, J. P.** (1975), Nonmonetary Factors in the Demand for Medical Services: Some Empirical Evidence, in: *Journal of Political Economy*, 83, 595-614.
- Akerlof, G. A.** (1970), The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, in: *Quarterly Journal of Economics*, 84, 488-500.
- Al-Najjar, N. I.** (1997), Incentive Contracts in Two-Sided Moral Hazards with Multiple Agents, in: *Journal of Economic Theory*, 74, 174-195.
- Amemiya, T.** (1981), Qualitative Response Models: A Survey, in: *Journal of Economic Literature*, 19, 1483-1536.
- Amemiya, T.** (1985), *Advanced Econometrics*, Basil Blackwell, Oxford.
- Arnott, R. J. und J. E. Stiglitz** (1988), The Basic Analytics of Moral Hazard, in: *Scandinavian Journal of Economics*, 90, 383-413.
- Arrow, K. J.** (1963), Uncertainty and the Welfare Economics of Medical Care, in: *American Economic Review*, 53, 941-973.
- Arrow, K. J.** (1970), *Essays in the Theory of Risk Bearing*, North Holland, Amsterdam et al.
- Arrow, K. J.** (1985), The Economics of Agency, in: Pratt, J. W. und R. J. Zeckhauser (Hrsg.), *Principals and Agents: The Structure of Business*, Harvard Business School Research Colloquium Series, Harvard Business School Press, Boston, 37-51.
- Auster, R., Leveson, I. und D. Sarachek** (1969), The Production of Health, an Exploratory Study, in: *Journal of Human Resources*, 4, 411-436.
- Baltagi, B. H.** (1995), *Econometric Analysis of Panel Data*, John Wiley & Sons, Chichester et al.
- Bamberg, G. und K. Spremann** (Hrsg.) (1989), *Agency Theory, Information, and Incentives, Corrected Reprint*, New York et al.
- Banerjee, A. und A. Beggs** (1992), *Simultaneous Versus Sequential Move Structures in Principal-Agent Models*, Oxford Applied Economics Discussion Paper Series: 148, Oxford.

- Barbier, H. D.** (2000), Der Patient muß Kunde sein. Zur Misere des Gesundheitswesens, in: *Frankfurter Allgemeine Zeitung*, 6.11.2000, Nr. 258, 17.
- Bernheim, B. D. und M. D. Whinston** (1986), Common Agency, in: *Econometrica*, 54, 923-942.
- Bhattacharyya, S. und F. Lafontaine** (1995), Double-Sided Moral Hazard and the Nature of Share Contracts, in: *Rand Journal of Economics*, 26, 761-781.
- Blomqvist, A.** (1991), The Doctor as Double Agent: Information Asymmetry, Health Insurance, and Medical Care, in: *Journal of Health Economics*, 10, 411-432.
- Börsch-Supan, A.** (1998), Anreizprobleme in der Renten- und Krankenversicherung, in: Ackermann, R., Cassel, S. und E. Denner (Hrsg.), *Offen für Reformen? Institutionelle Voraussetzungen für gesellschaftlichen Wandel im modernen Wohlfahrtsstaat*, 2. Freiburger Wirtschaftssymposium, Nomos Verlagsgesellschaft, Baden-Baden, 271-290.
- Breyer, F.** (1984), Anbieterinduzierte Nachfrage nach ärztlichen Leistungen und die Zieleinkommens-Hypothese, in: *Jahrbücher für Nationalökonomie und Statistik*, 199, 415-432.
- Breyer, F.** (1984a), *Die Nachfrage nach medizinischen Leistungen: eine empirische Analyse von Daten aus der Gesetzlichen Krankenversicherung*, Springer, Berlin et al.
- Breyer, F.** (1996), Korreferat: Demographische Alterung und Gesundheitskosten: Eine Fehlinterpretation, in: Oberender, P. (Hrsg.), *Alter und Gesundheit*, Gesundheitsökonomische Beiträge Bd. 26, Nomos Verlagsgesellschaft, Baden-Baden, 47-48.
- Breyer, F. und P. Zweifel** (1999), *Gesundheitsökonomie*, 3. Auflage, Springer, Berlin et al.
- Breyer, F. und V. Ulrich** (2000), Gesundheitsausgaben, Alter und medizinischer Fortschritt: Eine Regressionsanalyse, in: *Jahrbücher für Nationalökonomie und Statistik*, 220, 1-17.
- Brown, M. und E. Wolfstetter** (1989), Tripartite Income-Employment Contracts and Coalition Incentive Compatibility, in: *Rand Journal of Economics*, 20, 291-307.

- Bruce, N. und K. Y. Wong** (1996), Moral Hazard, Monitoring Costs, and Optimal Government, in: *Journal of Risk and Uncertainty*, 12, 77-90.
- Bulow, J. I., Geanakoplos, J. D. und P. D. Klemperer** (1985), Multimarket Oligopoly: Strategic Substitutes and Complements, in: *Journal of Political Economy*, 93, 488-511.
- Bundesministerium für Gesundheit** (2001), *Die gesetzliche Krankenversicherung. Solidarität stärken*, Bonn.
- Bundesministerium für Gesundheit** (2001a), *Rede von Bundesgesundheitsministerin Ulla Schmidt beim IKK Forum "Das Gesundheitswesen am Scheideweg - Entsolidarisierung versus Kostenexplosion?" am 15.10.01 in Berlin*, online im Internet: <http://www.bmggesundheit.de/presse/reden01/151001.htm> [Stand: 18.10.2001].
- Chalkley, M. und J. M. Malcomson** (1996), Contracts for the National Health Service, in: *Economic Journal*, 106, 1691-1701.
- Chalkley, M. und J. M. Malcomson** (1998), Contracting for Health Services with Unmonitored Quality, in: *Economic Journal*, 108, 1093-1110.
- Chalkley, M. und J. M. Malcomson** (1998a), Contracting for Health Services When Patient Demand Does Not Reflect Quality, in: *Journal of Health Economics*, 17, 1-19.
- Chamberlain, G.** (1984), Panel Data, in: Griliches, Z. und M. Intriligator (Hrsg.), *Handbook of Econometrics. Volume II*, Handbooks in Economics 2, Elsevier, Amsterdam et al, 1248-1318.
- Chiang, A. C.** (1992), *Elements of Dynamic Optimization*, McGraw-Hill, New York et al.
- Chiappori, P. A. und B. Salanié** (1997), Empirical Contract Theory: The Case of Insurance Data, in: *European Economic Review*, 41, 943-950.
- Chiappori, P. A., Durand, F. und P. Y. Geoffard** (1998), Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment, in: *European Economic Review*, 42, 499-511.
- Chiappori, P. A. und B. Salanié** (2000), Testing for Asymmetric Information in Insurance Markets, in: *Journal of Political Economy*, 108, 56-78.

- Chiu, W. H.** (1997), Health Insurance and the Welfare of Health Care Consumers, in: *Journal of Public Economics*, 64, 125-133.
- Clark, D. und J. A. Olsen** (1994), Agency in Health Care with an Endogenous Budget Constraint, in: *Journal of Health Economics*, 13, 231-251.
- Clayton, E. W., Hickson, G. B., Githens, P. B. und F. A. Sloan** (1993), Doctor-Patient Relationships, in: Sloan, F. A. (Hrsg.), *Suing for Medical Malpractice.*, University of Chicago Press, Chicago et al., 50-71.
- Cooper, R. und T. W. Ross** (1985), Product Warranties and Double Moral Hazard, in: *Rand Journal of Economics*, 16, 103-113.
- Cramer, J. S.** (1986), *Econometric Applications of Maximum Likelihood Methods*, Cambridge University Press, Cambridge et al.
- Cramer, J. S.** (1991), *The Logit Model: An Introduction for Economists*, Hodder and Stoughton, London et al.
- Cutler, D. M.** (1996), *Public Policy for Health Care*, Working Paper 5591, National Bureau of Economic Research, Cambridge, MA.
- Cutler, D. M. und R. J. Zeckhauser** (2000), The Anatomy of Health Insurance, in: Culyer, A. J. und J. P. Newhouse (Hrsg.), *Handbook of Health Economics Vol 1A*, Handbooks in Economics 17, Elsevier, Amsterdam et al., 563-643.
- Das, M. und A. van Soest** (1999), A Panel Data Model for Subjective Information on Household Income Growth, in: *Journal of Economic Behavior and Organization*, 40, 409-426.
- Davidson, R. und J. G. MacKinnon** (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford et al.
- Demski, J. S. und D. Sappington** (1984), Optimal Incentive Contracts with Multiple Agents, in: *Journal of Economic Theory*, 33, 152-171.
- Demski, J. S. und D. E. M. Sappington** (1991), Resolving Double Moral Hazard Problems with Buyout Agreements, in: *Rand Journal of Economics*, 22, 232-240.
- Deutscher Bundestag** (2001), *Gutachten 2000/2001 des Sachverständigenrates für die Konzertierte Aktion im Gesundheitswesen. Bedarfsgerechtigkeit und*

- Wirtschaftlichkeit. Band I: Zielbildung, Prävention, Nutzerorientierung und Partizipation*, Drucksache 14/5660, Bonn.
- Dewatripont, M., Jewitt, I. und J. Tirole** (2000), Multitask Agency Problems: Focus and Task Clustering, in: *European Economic Review*, 44, 869-877.
- Dionne, G. und A. P. Contandriopoulos** (1985), Doctors and Their Workshops: A Review Article, in: *Journal of Health Economics*, 4, 21-33.
- Doppmann, R. J.** (1985), *Determinanten der Nachfrage nach Gesundheit und der Inanspruchnahme medizinischer Leistungen. Eine ökonomische und ökonometrische Analyse für die Schweiz*, BSBS, Basel.
- Douglas, E. J.** (1989), The Simple Analytics of the Principal-Agent Incentive Contract, in: *Journal of Economic Education*, 20, 39-51.
- Dranove, D.** (1988), Demand Inducement and the Physician/Patient Relationship, in: *Economic Inquiry*, 26, 281-298.
- Dryhmes, P. J.** (1986), Limited Dependent Variables, in: Griliches, Z. und M. D. Intriligator (Hrsg.), *Handbook of Econometrics Vol. III*, Handbooks in Economics 2, Elsevier, Amsterdam et al., 1567-1631.
- Dutta, P. K. und R. Radner** (1994), Moral Hazard, in: Aumann, R. J. (Hrsg.), *Handbook of Game Theory with Economic Applications. Volume 2*, Handbooks in Economics 11, Elsevier, North-Holland, Amsterdam et al., 869-903.
- Eichberger, J.** (1993), *Game Theory for Economists*, Academic Press, San Diego et al.
- Eisen, R.** (1976), Unsicherheit und Information - Unkontrollierbares Verhalten und das Problem des moralischen Risikos, in: *Jahrbücher für Nationalökonomie und Statistik*, 191, 193-211.
- Eisenhardt, K. M.** (1989), Agency Theory: An Assessment and Review, in: *Academy of Management Review*, 14, 57-74.
- Ellis, R. P. und T. G. McGuire** (1990), Optimal Payment Systems for Health Services, in: *Journal of Health Economics*, 9, 375-396.
- Ellis, R. P. und T. G. McGuire** (1993), Supply-Side and Demand-Side Cost Sharing in Health Care, in: *Journal of Economic Perspectives*, 7, 135-151.

- Enthoven, A. C.** (1981), The Behavior of Health Care Agents: Provider Behavior, in: van der Gaag, J. und M. Perlman (Hrsg.), *Health, Economics, and Health Economics*, Noth Holland, Amsterdam et al., 173-188.
- Evans, R. G.** (1974), Supplier-Induced Demand: Some Empirical Evidence and Implications, in: Perlman, M. (Hrsg.), *The Economics of Health and Medical Care*, Stockton Press, New York, 162-173.
- Fama, E. F.** (1980), Agency Problems and the Theory of the Firm, in: *Journal of Political Economy*, 88, 288-307.
- Felder, S., Meier, M. und H. Schmitt** (2000), Health Care Expenditure in the Last Months of Life, in: *Journal of Health Economics*, 19, 679-695.
- Feldstein, M. S.** (1973), The Welfare Loss of Excess Health Insurance, in: *Journal of Political Economy*, 81, 251-280.
- Feldstein, M. S.** (1974), Econometric Studies of Health Economics, in: Intriligator, M. D. und D. A. Kendrick (Hrsg.), *Frontiers of Quantitative Economics, Vol. II*, North Holland, Amsterdam et al., 377-447.
- Feldstein, P. J.** (1993), *Health Care Economics, 4th ed.*, Delmar, Albany et al.
- Fischer, M.** (1995), Agency-Theorie, in: *WiSt*, 6, 320-322.
- Folland, S., Goodman, A. C. und M. Stano** (2001), *The Economics of Health and Health Care*, Prentice Hall, New Jersey.
- Fraja, G. d.** (2000), Contracts for Health Care and Asymmetric Information, in: *Journal of Health Economics*, 19, 663-677.
- Frechette, G. R.** (2001), Random-Effects Ordered Probit, in: *Stata Technical Bulletin*, 61, 247-27.
- Fudenberg, D., Hölmstrom, B. und P. Milgrom** (1990), Short-term Contracts and Long-term Agency Relationships, in: *Journal of Economic Theory*, 51, 1-31.
- Gaynor, M.** (1994), Issues in the Industrial Organization of the Market for Physician Services, in: *Journal of Economics and Management Strategy*, 3, 211-255.
- Gaynor, M. und S. W. Polachek** (1994), Measuring Information in the Market:

- An Application to Physician Services, in: *Southern Economic Journal*, 60, 815-31.
- Gerdtham, U. G. und M. Johannesson** (1997), *The Relationship between Happiness, Health and Socio-Economic Factors: Results based on Swedish Micro Data*, Working Paper Series in Economics and Finance No. 207, Department of Economics, Stockholm School of Economics.
- Gerdtham, U. und M. Johannesson** (1999), New Estimates of the Demand for Health: Results Based on a Categorical Health Measure and Swedish Micro Data, in: *Social Science and Medicine*, 49, 1325-1332.
- Gerdtham, U. G., Johannesson, M., Lundberg, L. und D. Isacson** (1999), The Demand for Health: Results from New Measures of Health Capital, in: *European Journal of Political Economy*, 15, 501-521.
- Gibbons, R.** (1992), *Game Theory for Applied Economists*, Princeton University Press, Princeton, New Jersey.
- Giuffrida, A. und H. Gravelle** (1998), Paying Patients to Comply: An Economic Analysis, in: *Health Economics*, 7, 569-579.
- Gjesdal, F.** (1982), Information and Incentives: The Agency Information Problem, in: *Review of Economic Studies*, 49, 373-390.
- Gouriéroux, C., Monfort, A. und A. Trongon** (1984), Pseudo Maximum Likelihood Methods, in: *Econometrica*, 52, 681-700.
- Gouriéroux, C.** (2000), *Econometrics of Qualitative Dependent Variables*, Cambridge University Press, Cambridge et al.
- Greene, W. H.** (2000), *Econometric Analysis 4th ed.*, Prentice Hall, New Jersey et al..
- Grossman, M.** (1972), On the Concept of Health Capital and the Demand for Health, in: *Journal of Political Economy*, 80, 223-255.
- Grossman, M.** (2000), The Human Capital Model, in: Culyer, A. J. und J. P. Newhouse (Hrsg.), *Handbook of Health Economics, Vol. 1A*, Handbooks in Economics 17, Elsevier, Amsterdam et al., 347-408.
- Grossman, S. J. und O. D. Hart** (1983), An Analysis of the Principal-Agent Problem, in: *Econometrica*, 51, 7-45.

- Hackmann, J.** (1990), Freiheit und Sicherheit in Systemen der sozialen Sicherung. Zu einer Theorie sozialer Sicherungssysteme, in: Albers, W. (Hrsg.), *Finanzierungsprobleme der sozialen Sicherung Bd. I*, Duncker und Humblot, Berlin, 41-137.
- Hamerle, A. und G. Ronning** (1995), Panel Analysis for Qualitative Variables, in: Arminger, G., Clogg, C. C. und M. E. Sobel (Hrsg.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Plenum Press, New York, 401-451.
- Harris, W. T. und L. Harris** (1996), The Decision to Quit Smoking: Theory and Evidence, in: *Journal of Socio-Economics*, 25, 601-618.
- Harris, M. und A. Raviv** (1978), Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement, in: *American Economic Review*, 68, 20-30.
- Hart, O. und B. Holmström** (1987), The Theory of Contracts, in: Bewley, T. F. (Hrsg.), *Advances in Economic Theory: Fifth World Congress*, Econometric Society Monographs Series No. 12, Cambridge University Press, Cambridge et al., 71-155.
- Hart, R. F. G.** (1988), Aspects of Health Insurance: Moral Hazard, Agency Problems, and a Welfare Loss Estimate for Australia, in: Smith, C. (Hrsg.), *Economics and Health: 1988*, Proceedings of the Tenth Australian Conference of Health Economists, Monash University, Clayton, Australia, 253-271.
- Henderson, J. M. und R. E. Quandt** (1983), *Mikroökonomische Theorie. Eine mathematische Darstellung, 5. überarbeitete Auflage*, Verlag Franz Vahlen, München.
- Herder-Dorneich, P.** (1994), *Ökonomische Theorie des Gesundheitswesens. Problemgeschichte, Problembereiche, Theoretische Grundlagen*, Nomos Verlagsgesellschaft, Baden-Baden.
- Hoffmeyer, U. K.** (1994), The Health Care System in Germany, in: Hoffmeyer, U. K. und T. R. McCarthy (Hrsg.), *Financing health care, Volume 1*, Kluwer Academic Publishers, Dordrecht et al., 419-512.
- Holly, A., Gardiol, L., Domenighetti, G. und B. Bisig** (1998), An Econometric Model of Health Care Utilization and Health Insurance in Switzerland, in:

- European Economic Review*, 42, 513-522.
- Holmström, B.** (1979), Moral Hazard and Observability, in: *Bell Journal of Economics*, 10, 74-91.
- Holmström, B.** (1982), Moral Hazard in Teams, in: *Bell Journal of Economics*, 13, 324-340.
- Holmström, B. und P. Milgrom** (1987), Aggregation and Linearity in the Provision of Intertemporal Incentives, in: *Econometrica*, 55, 303-328.
- Holmström, B. und P. Milgrom** (1991), Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, in: *Journal of Law, Economics and Organization*, 7, 24-52.
- Hsiao, C.** (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge et al.
- Hsiao, C.** (1995), Panel Analysis for Metric Data, in: Arminger, G., Clogg, C. C. und M. E. Sobel (Hrsg.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Plenum Press, New York, 361-400.
- Jelovac, I.** (2001), Physicians' Payment Contracts, Treatment Decisions and Diagnosis Accuracy, in: *Health Economics*, 10, 9-25.
- Jewitt, I.** (1988), Justifying the First-Order Approach to Principal-Agent Problems, in: *Econometrica*, 56, 1177-1190.
- Jones, A. M.** (2000), Health Econometrics, in: Culyer, A. J. und J. P. Newhouse (Hrsg.), *Handbook of Health Economics, Volume 1A*, Handbooks in Economics 17, Elsevier, Amsterdam et al., 265-344.
- Kenkel, D. S.** (1995), Should You Eat Breakfast? Estimates from Health Productions Functions, in: *Health Economics*, 4, 15-29.
- Kennedy, P. E.** (1998), *A Guide to Econometrics, 4th ed.*, Blackwell Publishers, Oxford et al.
- Kiener, S.** (1990), *Die Principal-Agent-Theorie aus informationsökonomischer Sicht*, Physica-Verlag, Heidelberg.
- Kim, S. K. und S. Wang** (1998), Linear Contracts and the Double Moral-Hazard, in: *Journal of Economic Theory*, 82, 342-378.

- Kortendieck, G.** (1993), *Gesundheitsökonomie und Wirtschaftspolitik: Neoklassische versus österreichische Markttheorie dargestellt am Beispiel des Gesundheits- und Krankenversicherungswesens*, Rudolf Haufe Verlag, Freiburg i. Br.
- Kotowitz, Y.** (1987), Moral Hazard, in: Eatwell, J., Milgare, M. und P. Newman (Hrsg.), *The New Palgrave: A Dictionary of Economics, Vol. 3*, Macmillan Press, London et al., 549-551.
- Kreps, D. M.** (1990), *A Course in Microeconomic Theory*, Harvester Wheatsheaf, New York et al.
- Lanoie, P.** (1991), Occupational Safety and Health: A Problem of Double or Single Moral Hazard, in: *Journal of Risk and Insurance*, 58, 80-100.
- Lee, C.** (1995), Optimal Medical Treatment under Asymmetric Information, in: *Journal of Health Economics*, 14, 419-441.
- LeFranc, E.** (1989), Socio-economic Determinants of Health Status, in: *Social and Economic Studies*, 38, 291-305.
- Leonard, K. L. und J. Zivin** (2001), *How to Compensate Physicians When Both Patient and Physician Effort are Unobservable*, Department of Economics, Columbia University, New York.
- Leu, R. E. und R. J. Doppmann** (1986), Die Nachfrage nach Gesundheit und Gesundheitsleistungen, in: Gäfgen, G. (Hrsg.), *Ökonomie des Gesundheitswesens*, Schriften des Vereins für Socialpolitik, N.F. Bd. 159, Duncker und Humblodt, Berlin, 161-175.
- Leu, R. E., Doppmann, R. J., Keller, T. und R. Deutschmann** (1986a), Die quantitative Erfassung von Gesundheitszustand und Lebensqualität, illustriert am Beispiel von Psoriasis-kranken, in: Horisberger, B. und W. v. Eimeren (Hrsg.), *Die Kosten-Nutzen-Analyse. Methodik und Anwendung am Beispiel von Medikamenten*, Springer Verlag, Heidelberg et al., 153-248.
- Leu, R. E. und M. Gerfin** (1992), Die Nachfrage nach Gesundheit - Ein empirischer Test des Grossmann-Modells, in: Oberender, P. (Hrsg.), *Steuerungsprobleme im Gesundheitswesen*, Gesundheitsökonomische Beiträge Bd. 14, Nomos Verlagsgesellschaft, Baden-Baden, 61-79.
- Levaggi, R.** (1996), NHS Contracts: An Agency Approach, in: *Health Economics*, 5, 341-352.

- Levinthal, D.** (1988), A Survey of Agency Models of Organizations, in: *Journal of Economic Behavior and Organization*, 9, 153-185.
- Long, J. S.** (1997), *Regression Models for Categorical and Limited Dependent Variables*, Sage, Thousand Oaks et al.
- Long, J. S. und S. Cheng** (2001), Regression Models for Categorical Outcomes, in: Hardy, M. und A. Bryman (Hrsg.), *Handbook of Data Analysis*, Sage, Thousand Oaks et al., forthcoming.
- Ma, C. A. und M. H. Riordan** (1997), *Health Insurance, Moral Hazard, and Managed Care*, Department of Economics, Boston University.
- Ma, C. A. und T. G. McGuire** (1997), Optimal Health Insurance and Provider Payment, in: *American Economic Review*, 87, 685-704.
- MacDonald, G. M.** (1984), *New Directions in the Economic Theory of Agency*.
- Macho-Stadler, I. und J. D. Pérez-Castrillo** (1993), Moral Hazard with Several Agents: The Gains from Cooperation, in: *International Journal of Industrial Organization*, 11, 73-100.
- Macho-Stadler, I. und J. D. Pérez-Castrillo** (1997), *An Introduction to the Economics of Information: Incentives and Contracts*, Oxford University Press, Oxford et al.
- Maddala, G. S.** (1983), *Limited Dependet and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge et al.
- Maddala, G. S.** (1987), Recent Developments in the Econometrics of Panel Data Analysis, in: *Transportation Research Part A-Policy and Practice*, 21A, 303-326.
- Maddala, G. S.** (1987a), Limited Dependent Variable Models Using Panel Data, in: *Journal of Human Resources*, 22, 307-338.
- Maddala, G. S.** (1992), *Introduction to Econometrics, Second Ed.*, Prentice Hall, Englewood Cliffs, New Jersey.
- Marschall, P.** (1999), *Lebensstilwandel in Ostdeutschland - Ansatzpunkte für gesundheitsökonomische Analysen*, Wirtschaftswissenschaftliche Diskussionspapiere Nr. 3/99, Rechts- und Staatswissenschaftliche Fakultät, Ernst-Moritz-Arndt-Universität Greifswald.

- Marshall, J. M.** (1976), Moral Hazard, in: *American Economic Review*, 66, 880-890.
- McCullough, L. B.** (1988), An Ethical Model for Improving the Patient-Physician Relationship, in: *Inquiry*, 25, 454-465.
- McGuire, A., Henderson, J. und G. Mooney** (1988), *The Economics of Health Care. An Introductory Text*, Routledge, London et al.
- McGuire, T. G.** (2000), Physician Agency, in: Culyer, A. J. und J. P. Newhouse (Hrsg.), *Handbook of Health Economics, Vol. 1A*, Handbooks in Economics 17, Elsevier, Amsterdam et al., 461-536.
- McKelvey, R. D. und W. Zavoina** (1975), A Statistical Model for the Analysis of Ordinal Level Dependent Variables, in: *Journal of Mathematical Sociology*, 4, 103-120.
- Meza, D. de** (1983), Health Insurance and the Demand for Medical Care, in: *Journal of Health Economics*, 2, 47-54.
- Mirrlees, J. A.** (1976), The Optimal Structure of Incentives and Authority within an Organization, in: *Bell Journal of Economics*, 7, 105-131.
- Mirrlees, J. A.** (1999), The Theory of Moral Hazard and Unobservable Behaviour: Part I, in: *Review of Economic Studies*, 66, 3-21.
- Mooney, G.** (1991), *Agency in Health Care: What's the Maximand?*, Discussion Paper 10/91, Health Economics Research Unit, University of Aberdeen.
- Mooney, G. und M. Ryan** (1993), Agency in Health Care: Getting beyond First Principles, in: *Journal of Health Economics*, 12, 125-135.
- Müller, H. M.** (1997), *The Theory of Moral Hazard*, Difo-Druck, Bamberg.
- Nell, M.** (1993), *Versicherungsinduzierte Verhaltensänderungen von Versicherungsnehmern: eine Analyse der Substitutions-, Moral Hazard- und Markteffekte unter besonderer Berücksichtigung der Krankenversicherung*, VVW, Karlsruhe.
- Neudeck, W.** (1991), Fee-for-Service and Quantity Rationing in the Physician Service Market, in: López-Casasnovas, G. (Hrsg.), *Incentives in Health Systems*, Springer, Berlin et al., 99-108.

- Nocera, S. und P. Zweifel** (1998), The Demand for Health: An Empirical Test of the Grossman Model Using Panel Data, in: Zweifel, P. (Hrsg.), *Health, the Medical Profession, and Regulation*, Kluwer Academic Publishers, Dordrecht et al., 35-49.
- o.V.** (1999), Union wendet sich gegen Primärarztmodelle, in: *Frankfurter Allgemeine Zeitung*, 10.2.1999, Nr. 34, 19.
- o.V.** (2000), Patienten werden immer unzufriedener, in: *Frankfurter Allgemeine Zeitung*, 22.12.2000, Nr. 298, 4.
- o.V.** (2001), Deutsche geben immer mehr für Gesundheit aus, in: *Frankfurter Allgemeine Zeitung*, 15.2.2001, Nr. 39, 19.
- o.V.** (2001a), Der Hausarzt soll zum Lotsen werden, in: *Die Zeit*, 10.3.2001, Nr. 10, 20.
- Okagi, A.** (1993), Generalized Methods of Moments: Econometric Applications, in: Maddala, G. S., Rao, C. R. und H. D. Vinod (Hrsg.), *Econometrics, Handbook of Statistics 11*, North Holland, Amsterdam et al., 455-488.
- Pauly, M. V.** (1968), The Economics of Moral Hazard: Comment, in: *American Economic Review*, 58, 531-537.
- Pauly, M. V.** (1970), The Welfare Economics of Community Rating, in: *Journal of Risk and Insurance*, 37, 407-418.
- Pauly, M. V.** (1974), Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection, in: *Quarterly Journal of Economics*, 88, 44-62.
- Pauly, M. V.** (1980), *Doctors and Their Workshops: Economic Models of Physician Behavior*, University of Chicago Press, Chicago et al.
- Pauly, M. V.** (1986), Taxation, Health Insurance, and Market Failure in the Medical Economy, in: *Journal of Economic Literature*, 24, 629-675.
- Pfaff, D. und P. Zweifel** (1998), Die Principal-Agent Theorie - ein fruchtbarer Beitrag der Wirtschaftstheorie zur Praxis, in: *WiSt*, 4, 184-190.
- Pohlmeier, W. und V. Ulrich** (1992), Determinanten des Gesundheitszustandes - Ein empirischer Ansatz zur Outputmessung im Gesundheitswesen bei partieller Information, in: *Zeitschrift für Wirtschafts und Sozialwissen-*

schaften, 112, 219-238.

- Pohlmeier, W. und V. Ulrich** (1995), An Econometric Model for the Two-Part Decision Process in the Demand for Health, in: *Journal of Human Resources*, 30, 339-361.
- Pohlmeier, W. und V. Ulrich** (1996), Die Nachfrage nach medizinischen Leistungen im Lebenszyklus, in: Oberender, P. (Hrsg.), *Alter und Gesundheit*, Gesundheitsökonomische Beiträge Bd. 26, Nomos Verlagsgesellschaft, Baden-Baden, 49-81.
- Pratt, J. W. und R. J. Zeckhauser** (1985), Principals and Agents: An Overview, in: Pratt, J. W. und R. J. Zeckhauser (Hrsg.), *Principals and Agents: The Structure of Business*, Harvard Business School Research Colloquium Series, Boston, 1-35.
- Rees, R.** (1985), The Theory of Principal and Agent: Part 1, in: *Bulletin of Economic Research*, 37, 3-26.
- Rees, R.** (1985a), The Theory of Principal and Agent: Part 2, in: *Bulletin of Economic Research*, 37, 75-95.
- Richard, S.** (1993), *Informationsprobleme auf dem Markt für ärztliche Leistungen*, Diskussionsbeiträge aus dem Institut für Volkswirtschaftslehre (520), Universität Hohenheim.
- Richard, S.** (1996), Information Problems in the Market for Medical Services, in: Picot, A. und E. Schlicht (Hrsg.), *Firms, Markets, and Contracts: Contributions to Neoinstitutional Economics*, Physica, Heidelberg, 200-219.
- Ried, W.** (1992), *Information und Nutzen der medizinischen Diagnostik*, Peter Lang, Frankfurt am Main et al.
- Ried, W. und V. Ulrich** (1996), Die gesundheitsökonomische Herausforderung: Wirtschaftlichkeitsanalysen im Gesundheitswesen - The State of the Art, in: Theurl, E. und J. Dézsy (Hrsg.), *Herausforderungen für die Gesundheitspolitik*, Außeninstitut der Universität Innsbruck, Innsbruck, 57-95.
- Ried, W. und E. Wille** (1997), Methodische Aspekte der Bewertung von Gesundheitsleistungen, in: *Betriebswirtschaftliche Forschung und Praxis*, 123-142.
- Robst, J. und G. G. Graham** (1997), Access to Health Care and Current Health

- Status: Do Physicians Matter?, in: *Applied Economics Letters*, 4, 45-48.
- Rochaix, L.** (1989), Information Asymmetry and Search in the Market for Physicians' Services, in: *Journal of Health Economics*, 8, 53-84.
- Rochaix, L.** (1997), Asymétries d'Information et Incertitude en Santé: Les Apports de la Théorie des Contrats, in: *Économie et Prévision*, 129-130, 11-24.
- Rogerson, W. P.** (1985), The First-Order Approach to Principal-Agent Problems, in: *Econometrica*, 53, 1357-1367.
- Ronning, G.** (1991), *Mikroökonomie*, Springer, Berlin et al.
- Ronning, G.** (1996), Ökonometrie, in: von Hagen, J., Börsch-Supan, A. und P. J. J. Welfens (Hrsg.), *Springers Handbuch der Volkswirtschaftslehre - 1 Grundlagen*, Springer, Berlin et al., 77-133.
- Ross, S. A.** (1973), The Economic Theory of Agency: The Principal's Problem, in: *American Economic Review*, 63, 134-139.
- Rothschild, M. und J. E. Stiglitz** (1976), Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, in: *Quarterly Journal of Economics*, 90, 629-649.
- Ryan, M.** (1994), Agency in Health Care: Lessons for Economists from Sociologists, in: *American Journal of Economics and Sociology*, 53, 207-217.
- Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen** (1994), *Gesundheitsversorgung und Krankenversicherung 2000. Eigenverantwortung, Subsidiarität und Solidarität bei sich ändernden Rahmenbedingungen. Sachstandsbericht 1994*, Nomos Verlagsgesellschaft, Baden-Baden.
- Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen** (1998), *Gesundheitswesen in Deutschland: Kostenfaktor und Zukunftsbranche. Bd. II: Fortschritt und Wachstumsmärkte, Finanzierung und Vergütung: Sondergutachten 1997*, Nomos Verlagsgesellschaft, Baden-Baden.
- Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung** (2000), *Chancen auf einen höheren Wachstumspfad. Jahresgutachten 2000/2001*, Metzler-Poeschel, Stuttgart.

- Salanié, B.** (1997), *The Economics of Contracts: A Primer*, MIT Press, Cambridge et al.
- Sappington, D. E. M.** (1983), Limited Liability Contracts between Principal and Agent, in: *Journal of Economic Theory*, 29, 1-21.
- Sappington, D. E. M.** (1991), Incentives in Principal-Agent Relationships, in: *Journal of Economic Perspectives*, 5, 45-66.
- Schneider, M., Biene-Dietrich, P., Gabanyi, M., Hofmann, U., Huber, M., Köse, A. und J. H. Sommer** (1995), *Gesundheitssysteme im internationalen Vergleich, Ausgabe 1994*, BASYS, Augsburg.
- Schneider, U.** (1998), *Der Arzt als Agent des Patienten - Zur Übertragbarkeit der Principal-Agent-Theorie auf die Arzt-Patient-Beziehung*, Wirtschaftswissenschaftliche Diskussionspapiere 2/98, Rechts- und Staatswissenschaftliche Fakultät, Ernst-Moritz-Arndt-Universität Greifswald.
- Schneider, U.** (1999), *Ärztliche Leistung und Compliance des Patienten - der Fall des Double Moral Hazard*, Wirtschaftswissenschaftliche Diskussionspapiere 7/99, Rechts- und Staatswissenschaftliche Fakultät, Ernst-Moritz-Arndt-Universität Greifswald.
- Schulenburg, J. G. v. d.** (1987), *Selbstbeteiligung: Theoretische und empirische Konzepte für die Analyse ihrer Allokations- und Verteilungswirkungen*, Mohr Siebeck, Tübingen.
- Schweizer, U.** (1999), *Vertragstheorie*, Mohr Siebeck, Tübingen.
- Scott, A. und S. Vick** (1999), Patients, Doctors and Contracts: An Application of Principal-Agent Theory to the Doctor-Patient Relationship, in: *Scottish Journal of Political Economy*, 46, 111-134.
- Selden, T. M.** (1990), A Model of Capitation, in: *Journal of Health Economics*, 9, 397-409.
- Shavell, S.** (1979), Risk Sharing and Incentives in the Principal and Agent Relationship, in: *Bell Journal of Economics*, 10, 55-73.
- Shmanske, S.** (1996), Information Asymmetries in Health Services: The Market Can Cope, in: *Independent Review*, 1, 191-200.
- Sobel, J.** (1993), Information Control in the Principal-Agent Problem, in: *Inter-*

- national Economic Review*, 34, 259-269.
- Spence, M. und R. Zeckhauser** (1971), Insurance, Information, and Individual Action, in: *American Economic Review*, 61, 380-387.
- Statistisches Bundesamt** (Hrsg.) (1998), *Gesundheitsbericht für Deutschland: Gesundheitsberichterstattung des Bundes*, Stuttgart.
- Statistisches Bundesamt** (1999), *Fachserie 12, Gesundheitswesen, Reihe S.2, Ausgaben für Gesundheit, 1970-1996*, Metzler-Poeschel, Stuttgart.
- Statistisches Bundesamt** (2001), *Neue Gesundheitsausgabenrechnung*, Wiesbaden.
- Stewart, J.** (1994), The Welfare Implications of Moral Hazard and Adverse Selection in Competitive Insurance Markets, in: *Economic Inquiry*, 32, 193-208.
- Stiglitz, J. E.** (1983), Risk, Incentives and Insurance: The Pure Theory of Moral Hazard, in: *Geneva Papers on Risk and Insurance*, 8, 4-33.
- Stiglitz, J. E.** (2000), The Contributions of the Economics of Information to Twentieth Century Economics, in: *Quarterly Journal of Economics*, 115, 1441-1478.
- Stone, T. T. und A. Mantese** (1999), Conflicting Values and the Patient-Provider Relationship in Managed Care, in: *Journal of Health Care Financing*, 26, 48-62.
- Strassl, W.** (1988), *Externe Effekte auf Versicherungsmärkten: eine allokatons-theoretische Begründung staatlicher Regulierung*, Mohr, Tübingen.
- Sydow, U.** (1999), *Die vertragsärztliche Abrechnung einschließlich der Honorarverteilung*, KBV-Verlag, Köln.
- Takayama, A.** (1985), *Mathematical Economics, 2nd ed.*, Cambridge University Press, Cambridge et al.
- Theilen, B.** (1996), *Optimal Contract and Organizational Design under Moral Hazard and Adverse Selection*, Verlag Josef Eul, Bergisch Gladbach et al.
- Tutz, G.** (2000), *Die Analyse kategorialer Daten: anwendungsorientierte Einführung in Logit-Modellierung und kategoriale Regression*, Oldenbourg,

München et al.

- Ulrich, V. und E. Wille** (1997), Deutschland, in: Stepan, A. (Hrsg.), *Finanzierungssysteme im Gesundheitswesen: ein internationaler Vergleich*, Manz, Wien, 33-57.
- Varian, H. R.** (1996), *Intermediate Microeconomics - a Modern Approach, 4th ed.*, Norton, New York et al.
- Veall, M. R. und K. F. Zimmermann** (1992), Pseudo-R²'s in the Ordinal Probit Model, in: *Journal of Mathematical Sociology*, 16, 333-342.
- Veall, M. R. und K. F. Zimmermann** (1996), Pseudo-R² Measures for Some Common Limited Dependent Variable Models, in: *Journal of Economic Surveys*, 10, 241-259.
- Vick, S. und A. Scott** (1995), *What makes a Perfect Agent? A Pilot Study of Patients Preferences in the Doctor-Patient Relationship*, Discussion Paper 05/95, Health Economics Research Unit, University of Aberdeen.
- Vick, S. und A. Scott** (1998), Agency in Health Care: Examining Patients' Preferences for Attributes of the Doctor-Patient Relationship, in: *Journal of Health Economics*, 17, 587-605.
- Wagstaff, A.** (1986), The Demand for Health: Some New Empirical Evidence, in: *Journal of Health Economics*, 5, 195-233.
- Wille, E.** (1980), Soziale Indikatoren als Ansatzpunkte wirtschaftspolitischer Zielbildung und Kontrolle - Verbesserung oder Verschleierung politischer Entscheidungsgrundlagen, in: *ORDO, Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft*, 31, 127-151.
- Wille, E. und V. Ulrich** (1991), Bestimmungsfaktoren der Ausgabenentwicklung in der gesetzlichen Krankenversicherung (GKV), in: Hansmeyer, K. (Hrsg.), *Finanzierungsprobleme der sozialen Sicherung II*, Schriften des Vereins für Socialpolitik, N. F., Bd. 194 II, Duncker und Humblot, Berlin, 9-115.
- Winter, R. A.** (1992), Moral Hazard and Insurance Contracts, in: Dionne, G. (Hrsg.), *Contributions to Insurance Economics*, Kluwer Academic Publishers, Boston et al, 61-96.
- Yavaş, A.** (1995), Seller-Broker Relationship as a Double Moral Hazard

- Problem, in: *Journal of Housing Economics*, 4, 244-263.
- Zeckhauser, R.** (1970), Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives, in: *Journal of Economic Theory*, 2, 10-26.
- Zifonun, N. und M. Cordes** (2001), Die neue Gesundheitsausgabenrechnung der Gesundheitsberichterstattung, in: *Wirtschaft und Statistik*, 195-202.
- Zweifel, P.** (1982), *Ein ökonomisches Modell des Arztverhaltens*, Springer, Berlin et al.
- Zweifel, P.** (1994), Eine Charakterisierung von Gesundheitssystemen: Welche sind von Vorteil bei welchen Herausforderungen?, in: Oberender, P. (Hrsg.), *Probleme der Transformation im Gesundheitswesen*, Gesundheitsökonomische Beiträge Bd. 20, Nomos Verlagsgesellschaft, Baden-Baden, 9-43.
- Zweifel, P., Felder, S. und M. Meier** (1996), Demographische Alterung und Gesundheitskosten: Eine Fehlinterpretation, in: Oberender, P. (Hrsg.), *Alter und Gesundheit*, Gesundheitsökonomische Beiträge Bd. 26, Nomos Verlagsgesellschaft, Baden-Baden, 29-46.
- Zweifel, P., Lehmann, H. und L. Steinmann** (2001), *Patching up the Physician-Patient Relationship: Insurers vs. Governments as Complementary Agents*, Paper presented at the Annual Meeting of the German Economic Association (Verein für Socialpolitik), 25.-28. September, Magdeburg.

STAATLICHE ALLOKATIONSPOLITIK IM MARKTWIRTSCHAFTLICHEN SYSTEM

- Band 1 Horst Siebert (Hrsg.): Umweltallokation im Raum. 1982.
- Band 2 Horst Siebert (Hrsg.): Global Environmental Resources. The Ozone Problem. 1982.
- Band 3 Hans-Joachim Schulz: Steuerwirkungen in einem dynamischen Unternehmensmodell. Ein Beitrag zur Dynamisierung der Steuerüberwälzungsanalyse. 1981.
- Band 4 Eberhard Wille (Hrsg.): Beiträge zur gesamtwirtschaftlichen Allokation. Allokationsprobleme im intermediären Bereich zwischen öffentlichem und privatem Wirtschaftssektor. 1983.
- Band 5 Heinz König (Hrsg.): Ausbildung und Arbeitsmarkt. 1983.
- Band 6 Horst Siebert (Hrsg.): Reaktionen auf Energiepreissteigerungen. 1982.
- Band 7 Eberhard Wille (Hrsg.): Konzeptionelle Probleme öffentlicher Planung. 1983.
- Band 8 Ingeborg Kiesewetter-Wrana: Exporterönsinstabilität. Kritische Analyse eines entwicklungspolitischen Problems. 1982.
- Band 9 Ferdinand Dudenhöfer: Mehrheitswahl-Entscheidungen über Umweltnutzungen. Eine Untersuchung von Gleichgewichtszuständen in einem mikroökonomischen Markt- und Abstimmungsmodell. 1983.
- Band 10 Horst Siebert (Hrsg.): Intertemporale Allokation. 1984.
- Band 11 Helmut Meder: Die intertemporale Allokation erschöpfbarer Naturressourcen bei fehlenden Zukunftsmärkten und institutionalisierten Marktsubstituten. 1984.
- Band 12 Ulrich Ring: Öffentliche Planungsziele und staatliche Budgets. Zur Erfüllung öffentlicher Aufgaben durch nicht-staatliche Entscheidungseinheiten. 1985.
- Band 13 Ehrentraud Graw: Informationseffizienz von Terminkontraktmärkten für Währungen. Eine empirische Untersuchung. 1984.
- Band 14 Rüdiger Pethig (Ed.): Public Goods and Public Allocation Policy. 1985.
- Band 15 Eberhard Wille (Hrsg.): Öffentliche Planung auf Landesebene. Eine Analyse von Planungskonzepten in Deutschland, Österreich und der Schweiz. 1986.
- Band 16 Helga Gebauer: Regionale Umweltnutzungen in der Zeit. Eine intertemporale Zwei-Regionen-Analyse. 1985.
- Band 17 Christine Pfitzer: Integrierte Entwicklungsplanung als Allokationsinstrument auf Landesebene. Eine Analyse der öffentlichen Planung der Länder Hessen, Bayern und Niedersachsen. 1985.
- Band 18 Heinz König (Hrsg.): Kontrolltheoretische Ansätze in makroökonomischen Modellen. 1985.
- Band 19 Theo Kempf: Theorie und Empirie betrieblicher Ausbildungsplatzangebote. 1985.
- Band 20 Eberhard Wille (Hrsg.): Konkrete Probleme öffentlicher Planung. Grundlegende Aspekte der Zielbildung, Effizienz und Kontrolle. 1986.
- Band 21 Eberhard Wille (Hrsg.): Informations- und Planungsprobleme in öffentlichen Aufgabenbereichen. Aspekte der Zielbildung und Outputmessung unter besonderer Berücksichtigung des Gesundheitswesens. 1986.
- Band 22 Bernd Gutting: Der Einfluß der Besteuerung auf die Entwicklung der Wohnungs- und Baulandmärkte. Eine intertemporale Analyse der bundesdeutschen Steuergesetze. 1986.
- Band 23 Heiner Kuhl: Umweltressourcen als Gegenstand internationaler Verhandlungen. Eine theoretische Transaktionskostenanalyse. 1987.

- Band 24 Hubert Hornbach: Besteuerung, Inflation und Kapitalallokation. Intersektorale und internationale Aspekte. 1987.
- Band 25 Peter Müller: Intertemporale Wirkungen der Staatsverschuldung. 1987.
- Band 26 Stefan Kronenberger: Die Investitionen im Rahmen der Staatsausgaben. 1988.
- Band 27 Armin-Detlef Rieß: Optimale Auslandsverschuldung bei potentiellen Schuldendienstproblemen. 1988.
- Band 28 Volker Ulrich: Preis- und Mengeneffekte im Gesundheitswesen. Eine Ausgabenanalyse von GKV-Behandlungsarten. 1988.
- Band 29 Hans-Michael Geiger: Informational Efficiency in Speculative Markets. A Theoretical Investigation. Edited by Ehrentraud Graw. 1989.
- Band 30 Karl Sputek: Zielgerichtete Ressourcenallokation. Ein Modellentwurf zur Effektivitätsanalyse praktischer Budgetplanung am Beispiel von Berlin (West). 1989.

ALLOKATION IM MARKTWIRTSCHAFTLICHEN SYSTEM

- Band 31 Wolfgang Krader: Neuere Entwicklungen linearer latenter Kovarianzstrukturmodelle mit quantitativen und qualitativen Indikatorvariablen. Theorie und Anwendung auf ein mikroempirisches Modell des Preis-, Produktions- und Lageranpassungsverhaltens von deutschen und französischen Unternehmen des verarbeitenden Gewerbes. 1991.
- Band 32 Manfred Erbsland: Die öffentlichen Personalausgaben. Eine empirische Analyse für die Bundesrepublik Deutschland. 1991.
- Band 33 Walter Ried: Information und Nutzen der medizinischen Diagnostik. 1992.
- Band 34 Anselm U. Römer: Was ist den Bürgern die Verminderung eines Risikos wert? Eine Anwendung des kontingenten Bewertungsansatzes auf das Giftmüllrisiko. 1993.
- Band 35 Eberhard Wille, Angelika Mehnert, Jan Philipp Rohweder: Zum gesellschaftlichen Nutzen pharmazeutischer Innovationen. 1994.
- Band 36 Peter Schmidt: Die Wahl des Rentenalters. Theoretische und empirische Analyse des Rentenzugangsverhaltens in West- und Ostdeutschland. 1995.
- Band 37 Michael Ohmer: Die Grundlagen der Einkommensteuer. Gerechtigkeit und Effizienz. 1997.
- Band 38 Evamaria Wagner: Risikomanagement rohstoffexportierender Entwicklungsländer. 1997.
- Band 39 Matthias Meier: Das Sparverhalten der privaten Haushalte und der demographische Wandel: Makroökonomische Auswirkungen. Eine Simulation verschiedener Reformen der Rentenversicherung. 1997.
- Band 40 Manfred Albring / Eberhard Wille (Hrsg.): Innovationen in der Arzneimitteltherapie. Definition, medizinische Umsetzung und Finanzierung. Bad Orber Gespräche über kontroverse Themen im Gesundheitswesen 25.-27.10.1996. 1997.
- Band 41 Eberhard Wille / Manfred Albring (Hrsg.): Reformoptionen im Gesundheitswesen. Bad Orber Gespräche über kontroverse Themen im Gesundheitswesen 7.-8.11.1997. 1998.
- Band 42 Manfred Albring / Eberhard Wille (Hrsg.): Szenarien im Gesundheitswesen. Bad Orber Gespräche über kontroverse Themen im Gesundheitswesen 5.-7.11.1998. 1999.
- Band 43 Eberhard Wille / Manfred Albring (Hrsg.): Rationalisierungsreserven im deutschen Gesundheitswesen. 2000.
- Band 44 Manfred Albring / Eberhard Wille (Hrsg.): Qualitätsorientierte Vergütungssysteme in der ambulanten und stationären Behandlung. 2001.

- Band 45** Martin Pfaff / Dietmar Wassener / Astrid Sterzel / Thomas Neldner: Analyse potentieller Auswirkungen einer Ausweitung des Pharmaversandes in Deutschland. 2002.
- Band 46** Eberhard Wille / Manfred Albring (Hrsg.): Konfliktfeld Arzneimittelversorgung. 2002.
- Band 47** Udo Schneider: Theorie und Empirie der Arzt-Patient-Beziehung. Zur Anwendung der Principal-Agent-Theorie auf die Gesundheitsnachfrage. 2002.

Thomas Klose

Der Wert besserer Gesundheit

Frankfurt/M., Berlin, Bern, Bruxelles, New York, Oxford, Wien, 2002.

XVI. 412 S., 16 Abb., 20 Tab.

Europäische Hochschulschriften: Reihe 5, Volks- und Betriebswirtschaft
Bd. 2845

ISBN 3-631-38874-8 · br. € 60.30*

Weltweit stoßen Gesundheitssysteme an die Grenzen ihrer Finanzierbarkeit, begründet insbesondere durch medizinisch-technologischen Fortschritt. Ökonomische Evaluationen bieten Ansätze zum rationalen Umgang mit Gesundheitstechnologien. Die Kostennutzenanalyse ist die theoretisch beste Evaluationsmethode. Ihre praktische Bedeutung ist wegen noch bestehender methodischer Probleme aber gering. Diese zu lösen, ist das primäre Ziel dieses Buches. Der erste Teil umfasst theoretische Analysen und systematische Aufarbeitungen empirischer Methoden. Diese sind Grundlage einer empirischen Studie, in der im zweiten Teil Hypothesen aus den vorhergehenden Analysen erfolgreich getestet werden. Durch zahlreiche methodische Neuerungen und innovative Ergebnisse leistet dieses Buch einen Beitrag zur Etablierung der Kostennutzenanalyse in der Evaluation von Gesundheitstechnologien.

Aus dem Inhalt: Eigenschaften der Präferenzen über Gesundheit und Geld und ihre Bedeutung für den Wert besserer Gesundheit · Anspruch und Realität der Bewertung von Gesundheitsleistungen: Brückenschläge · Qualitäts-gleiche Lebensjahre und Zahlungsbereitschaft: Unterschiede · Systematische Analyse empirischer Methoden der Bewertung von Gesundheitsleistungen · Empirische Studie zur Zahlungsbereitschaft von Diabetikern



Frankfurt/M · Berlin · Bern · Bruxelles · New York · Oxford · Wien

Auslieferung: Verlag Peter Lang AG

Moosstr. 1, CH-2542 Pieterlen

Telefax 00 41 (0) 32 / 376 17 27

*inklusive der in Deutschland gültigen Mehrwertsteuer

Preisänderungen vorbehalten

Homepage <http://www.peterlang.de>

Udo Schneider - 978-3-631-75606-5

Downloaded from PubFactory at 01/11/2019 03:10:02AM

via free access