

**MANAGED CARE IN DER SCHWEIZ:  
EINE EMPIRISCHE ANALYSE DES VERHALTENS VON  
ALLGEMEINMEDIZINERN**

Andreas Werblow

Inauguraldissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft  
durch die Fakultät für Wirtschaftswissenschaft  
der Otto-von-Guericke-Universität Magdeburg

Otto-von-Guericke-Universität Magdeburg

Fakultät für Wirtschaftswissenschaft

**Thema:       Managed Care in der Schweiz: Eine empirische Analyse des Verhaltens  
              von Allgemeinmedizinern**

**Dissertation  
zur Erlangung des akademischen Grades  
Dr. rer. pol.**

genehmigt durch  
die Fakultät für Wirtschaftswissenschaft  
der Otto-von-Guericke-Universität Magdeburg

von:         Dipl.-Vw. Andreas Werblow

geb. am:    15.06.1965

in:         Dresden

Gutachter:  Prof. Dr. Stefan Felder  
              Prof. Dr. Dr. Bodo Vogt  
              Prof. Dr. Bernhard Schipp

Eingereicht: 10. Oktober 2003

Verteidigt:  23. März 2004

# INHALTSVERZEICHNIS

## I Der Arzt im Gesundheitswesen

<b>PROLOG FRAGMENTE EINER GESCHICHTE DER ARZTVERGÜTUNG .....</b>	<b>2</b>
<b>1 <u>  </u> EINLEITUNG: DER ARZT ALS LOTSE IM GESUNDHEITSSYSTEM.....</b>	<b>5</b>
<b>2 <u>  </u> DIE VERGÜTUNG DES ARZTES .....</b>	<b>7</b>
<b>3 <u>  </u> DER ARZT UND MANAGED CARE .....</b>	<b>11</b>
3.1 ORGANISATIONSTYPEN .....	13
3.2 GESCHICHTE UND ENTWICKLUNG VON MANAGED-CARE-ORGANISATIONEN .....	16
3.3 MANAGED CARE UND QUALITÄT .....	18
<b>4 <u>  </u> DER (HAUS-) ARZT IN DER SCHWEIZ .....</b>	<b>21</b>
4.1 DER (HAUS-) ARZT IM KONVENTIONELLEN SEKTOR.....	22
4.2 DER HAUSARZT IN HAM UND HMOs .....	25
4.3 AUSBLICK: TENDENZEN BEI DER ARZTVERGÜTUNG IM SCHWEIZER GESUNDHEITSSYSTEM .....	26

## II Theorie der optimalen Arztvergütung

<b>5 <u>  </u> <u>DIENER ZWEIER HERREN:</u> <u>  DER ARZT ALS AGENT VON KRANKENKASSEN UND PATIENTEN.....</u></b>	<b>30</b>
5.1 ENTSCHEIDUNG UNTER UNSICHERHEIT.....	32
5.2 DIE PRINZIPAL-AGENT-THEORIE (GRUNDLAGEN) .....	36
5.2.1 GRUNDMODELL BEI EINSEITIG VERBORGENER HANDLUNG.....	37
5.2.2 GRUNDMODELL BEI ZWEISEITIG VERBORGENER HANDLUNG.....	40
5.3 DER KONFLIKT ZWISCHEN ARZT UND PATIENT.....	42
5.4 DER KONFLIKT ZWISCHEN KRANKENKASSE UND ARZT .....	46
5.5 DER ARZT ALS DOPPELAGENT .....	52
5.6 DER ARZT ALS EXPERTE.....	60
5.7 ZUSAMMENFASSUNG UND AUSBLICK .....	64
<b>6 <u>  </u> DIE ARZTPRAXIS UND WETTBEWERB .....</b>	<b>68</b>
6.1 PRODUKTION IN EINER ARZTPRAXIS .....	68
6.1.1 VOLLSTÄNDIGE KONKURRENZ UND MONOPOL.....	69
6.1.2 INPUT-MONOPOL BEI VERTIKALER INTEGRATION – DIE ARBEITSZEIT DES ARZTES ...	72
6.2 MONOPOLISTISCHE KONKURRENZ .....	73
6.3 ANGEBOTSSEITIGE KOSTENBETEILIGUNG, QUALITÄT UND WETTBEWERB UM PATIENTEN .....	80
6.4 AUSWIRKUNGEN UNTERSCHIEDLICHER PATIENTEN EINER PRAXIS AUF ANSTRENGUNGEN DES ARZTES .....	87
6.4.1 AUFWAND UND MANAGED-CARE-ANTEIL IN DER PRAXIS.....	89

6.4.2	ANSTRENGUNGEN UND MANAGED-CARE-KONZENTRATION AUF MARKTEBENE.....	92
6.5	<b>ZUSAMMENFASSUNG .....</b>	<b>94</b>

### III Empirische Untersuchung

<b>7</b>	<b>UNTERSUCHUNGSGEBIET UND DATEN .....</b>	<b>97</b>
7.1	<b>VERGÜTUNGSVERTRÄGE IN ZÜRICH UND GENF .....</b>	<b>98</b>
7.1.1	TARIFVERTRAG ZÜRICH.....	98
7.1.2	TARIFVERTRAG GENF .....	99
7.1.3	VERGLEICH DER TARIFE IN ZÜRICH UND GENF .....	100
7.2	<b>DIE UNTERSUCHTEN HAM UND HMOs IN DER SCHWEIZ.....</b>	<b>102</b>
7.2.1	HAM WINTIMED .....	102
7.2.2	HAM WINTERTHUR ANDELFINGEN.....	104
7.2.3	HAM GENF.....	105
7.2.4	HMO ZÜRICH.....	106
7.3	<b>ARZTSELEKTION .....</b>	<b>107</b>
7.4	<b>DATENBESCHREIBUNG, VARIABLENKONSTRUKTION UND ERSTE DESKRIPTIVE STATISTIKEN .....</b>	<b>110</b>
7.4.1	ARZT – VARIABLEN.....	110
7.4.2	PATIENT – VARIABLEN .....	115
7.4.3	MARKTSTRUKTUR – VARIABLEN .....	117
7.4.4	VERSICHERTENSTRUKTUR IN GENF UND ZÜRICH IM VERGLEICH MIT DEN HAUSARZTKOLLEKTIVEN.....	118
7.5	<b>REPRÄSENTATIVITÄT DER VERWENDETEN DATEN.....</b>	<b>120</b>
<b>8</b>	<b>METHODOLOGISCHE PROBLEME DER EMPIRISCHEN ANALYSE.....</b>	<b>121</b>
8.1	<b>PRINZIPIELLES ZUM ÜBERGANG VON DEN THEORETISCHEN MODELLEN ZU DEN EMPIRISCHEN SCHÄTZUNGEN.....</b>	<b>122</b>
8.2	<b>ÖKONOMETRISCHE PROBLEME UND LÖSUNGEN.....</b>	<b>124</b>
8.2.1	MODELLIERUNG MIT HILFE LATENTER VARIABLEN .....	124
8.2.2	ENDOGENITÄT I .....	126
8.2.3	ENDOGENITÄT II .....	128
8.2.4	HIERARCHISCHE DATENSTRUKTUR.....	131
<b>9</b>	<b>SCHÄTZUNG UND ERGEBNISSE .....</b>	<b>133</b>
9.1	<b>MARKTSTRUKTUR UND VERHALTEN DES ARZTES AUF MARKTEBENE.....</b>	<b>134</b>
9.1.1	ÖKONOMETRISCHES MODELL.....	137
9.1.2	DATEN .....	138
9.1.3	SCHÄTZUNG .....	140
9.1.4	DISKUSSION UND KRITIK .....	143
9.2	<b>EFFIZIENZ VON ALLGEMEINARZTPRAXEN .....</b>	<b>144</b>
9.2.1	EFFIZIENZMESSUNG UND STOCHASTIC FRONTIER ANALYSIS (SFA) .....	146
9.2.2	DATEN .....	152
9.2.3	SCHÄTZUNG UND DISKUSSION .....	155
9.2.4	ZUSAMMENFASSUNG UND KRITIK .....	159
9.3	<b>NACHFRAGESCHÄTZUNG.....</b>	<b>161</b>
9.3.1	MODELL .....	162
9.3.2	DATEN .....	167
9.3.3	SCHÄTZUNG .....	168
9.3.4	ZUSAMMENFASSUNG UND AUSBLICK .....	176
9.4	<b>GESAMTMODELL – ANGEBOT.....</b>	<b>178</b>
9.4.1	HERLEITUNG DES MODELLS (ALLGEMEIN) .....	178

9.4.2	HERLEITUNG DES MODELLS (IM BESONDEREN) .....	180
9.4.3	DATEN .....	183
9.4.4	ÜBERWEISUNGEN UND WEITERE VERANLASSTE LEISTUNGEN .....	190
9.4.5	KONSULTATIONEN.....	196
9.4.6	GESAMTLEISTUNGEN .....	205
9.4.7	DIREKTE ARZTLEISTUNGEN .....	212
9.4.8	DISKUSSION DER RESULTATE.....	219

**10 ZUSAMMENFASSUNG UND AUSBLICK..... 222**

**LITERATURVERZEICHNIS ..... i**

**ANHANG I..... x**

**SOZIALE KRANKENVERSICHERUNG IN DER SCHWEIZ ..... x**

**ANHANG II.....xiv**

**ANHANG III .....xv**

## **TABELLENVERZEICHNIS**

Tabelle 1	Grundleistungen und Preise in Zürich und Genf (1997-1999) in SFr.....	101
Tabelle 2	Anzahl Ärzte nach Kanton und Partnerart (volles Sample) .....	110
Tabelle 3	Codierung von Geschlecht und Qualifikationsgrad der Ärzte.....	111
Tabelle 4	Arzt- und Praxiseigenschaften .....	113
Tabelle 5	Leistungen der Ärzte <sup>a</sup> .....	114
Tabelle 6	Merkmale der Versicherten.....	116
Tabelle 7	Marktstrukturvariablen .....	118
Tabelle 8	Versicherte OKP / HAM und HMO 1997 – 1999.....	119
Tabelle 9	Vergleich Gesamtleistung – Leistung CSS (1999) <sup>a</sup> .....	121
Tabelle 10	Anzahl Ärzte und Gemeindegröße (1999) .....	139
Tabelle 11	Deskriptive Statistik der erklärenden Variablen für die Schätzung der Anzahl Ärzte (1999) für $Y = 0 - 12$ .....	140
Tabelle 12	Geordnetes Probitmodell für Anzahl Ärzte ( $Y = 0 - 12$ ).....	142
Tabelle 13	Effizienz-Schätzung: Deskriptive Statistiken .....	154
Tabelle 14	Effizienz-Schätzung: Resultate.....	157
Tabelle 15	3-stufiges Nachfragemodell .....	166
Tabelle 16	Nachfrageschätzung: Deskriptive Statistik .....	168
Tabelle 17	Nachfrageschätzung: Vertragswahl der Versicherten 1999 .....	170
Tabelle 18	Nachfrageschätzung: Arztbesuche.....	172
Tabelle 19	Nachfrageschätzung: Leistungsanspruchnahme .....	175
Tabelle 20	Gesamtmodell: Deskriptive Statistik erklärender Variablen .....	187
Tabelle 21	Gesamtmodell: Deskriptive Statistik der Leistungen .....	189
Tabelle 22	Gesamtmodell: Wahrscheinlichkeit einer Überweisung .....	193
Tabelle 23	Gesamtmodell: Leistungen pro Patient und Arzt – ohne und mit mindestens 1 Konsultation .....	197
Tabelle 24	Gesamtmodell: Wahrscheinlichkeit mindestens einer Konsultation .....	199
Tabelle 25	Gesamtmodell: Regression pos. Konsultationen (random effects).....	203
Tabelle 26	Gesamtmodell: Gesamtleistungen Arzt (random-effects).....	209
Tabelle 27	Gesamtmodell: Direkte Leistungen des Arztes (random effects).....	216
Tabelle 28	Managed-Care-Effekte im Gesamtmodell.....	220

## **ABBILDUNGSVERZEICHNIS**

Abbildung 1	Beziehungen zwischen Arzt, Krankenkasse und Patient .....	31
Abbildung 2	Gesundheitsfunktion mit 3 Gesundheitszuständen.....	54
Abbildung 3	Anteil der Allgemeinmediziner an allen Ärzten im Verhältnis zur Arztdichte (Zürich und Genf 1997) .....	63
Abbildung 4	McGuire-Modell: Monopolistische Konkurrenz.....	77
Abbildung 5	McGuire-Modell: Anzahl Patienten und Veränderung der Anzahl bei unterschiedlichen Elastizitäten in Abhängigkeit der Leistungsmenge pro Patient .....	83
Abbildung 6	McGuire-Modell: Mengelösungen.....	85
Abbildung 7	Angebot und Nachfrage bei unterschiedlichen Marktgrößen.....	136
Abbildung 8	Geschätzte Bevölkerung pro Arzt (1999).....	143
Abbildung 9	Leistungseffekt in Abhängigkeit der Regelmäßigkeit unter Berücksichtigung von Konsultationen und Leistungsnummern .....	218
Abbildung 10	Kantonale und interkantonale Unterschiede in den Erwachsenenprämien 1999 .....	xii

## ABKÜRZUNGSVERZEICHNIS UND BEGRIFFSKLÄRUNGEN

AIN	Angebotsinduzierte Nachfrage
BfS	Bundesamt für Statistik (Schweiz)
BSV	Bundesamt für Sozialversicherung (Schweiz)
Capitation (Cap)	Vergütungsform, bei der die Leistungserbringer nach der Anzahl der eingeschriebenen Versicherten in einer bestimmten Periode vergütet werden.
CSS	Christlich Soziale der Schweiz Krankenversicherung in der Schweiz
DEA	Data Envelopment Analysis
FFS	Fee for Services: Einzelleistungsvergütung
FMH	Federatio Medicarum Helveticarum Standesvertretung der Schweizer Ärzte
HAM	Hausarztmodell
HMO	Health Maintenance Organization
IPA	Independent Praxis Association
KBV	Kassenärztliche Bundesvereinigung (Deutschland)
KSK	Konkordat Schweizer Krankenversicherungen (Schweiz) Seit 2001: santé suisse
MCO	Managed Care Organization
Medicaid	Staatliche Krankenversicherung in den USA für Personen mit geringem Einkommen
Medicare	Staatliche Krankenversicherung in den USA für Personen über 65 Jahren
OLS	Ordinary Least Squares (Methode der kleinsten Quadrate)
POS	Point of Service Association
PPO	Preferred Provider Organization
SFA	Stochastic Frontier Analysis
u.d.N.	unter der Nebenbedingung

### **Hinweise zu sprachlichen Regelungen und zur Notation von Zahlen**

In allen Formulierungen, in denen es nicht explizit um geschlechtsspezifische Differenzen geht, wird aus Vereinfachungsgründen die männliche Form verwendet (z.B. Arzt statt Ärztin).

Bei der Darstellung von Zahlenwerten wird die amerikanische Schreibweise verwendet; d.h. die Dezimalstellen werden mit einem Punkt getrennt und Zifferngruppen durch ein Komma (1,522.12 SFr.).

## DANKSAGUNGEN

Ich bedanke mich für zahlreiche Hinweise und Ratschläge sowie für eingehende Diskussionen bei meinem Mentor, Prof. Dr. Stefan Felder.

Mein Dank gilt auch meinen Kolleginnen Frau Dipl. Vw. Nicola Scholz und Frau Dipl. Vw. Anja Olbrich, die mir wichtige Diskussionspartnerinnen waren.

Der Krankenversicherung CSS in Luzern danke ich für die Bereitstellung der Daten. Insbesondere bedanke ich mich für Unterstützung und Hilfe bei Dr. Konstantin Beck und Bernhard Keller.

Ohne die Unterstützung meiner Familie wäre die Fertigstellung dieser Arbeit nicht möglich gewesen. Daher gilt mein Dank Ilka und Hannah, die mich durch ihre Kraft und Liebe unterstützten.



*ICH,  
DER HERR,  
DEIN ARZT.*

Ex: 15: 26, 26 -  
Inscription über dem Portal  
des Diakonissenkrankenhauses in Dresden

## **Prolog**

### **Fragmente einer Geschichte der Arztvergütung**

Die Geschichte der Arztvergütung ist so alt wie der Arztberuf selbst. Denn Fragen der „richtigen“ Vergütung von Ärzten wurde schon immer große gesellschaftliche Bedeutung beigemessen. Versteht man unter Arztvergütung nicht nur die rein geldliche Vergütung, sondern sieht man ebenso Anerkennung und die erworbene gesellschaftliche Stellung als eine Art der Arztentlohnung an, so lassen sich von der Antike bis zur Neuzeit zahllose Beispiele herausgreifen, die diese Problematik unter künstlerischen bis hin zu philosophischen Aspekten thematisieren. Bekannte Beispiele finden sich von der Bibel über Platons „Die Republik“ bis hin zu Shaws „Doctor's Dilemma“.

Die bekannteste Darstellung der besonderen Stellung des Arztes in der Gesellschaft und der Kritik des ärztlichen Verhaltens in der deutschen Literatur ist sicherlich der Faust von Goethe. So sagt bspw. Wagner vor dem Tore zu Faust:

„Welch ein Gefühl musst du, o großer Mann,  
Bei der Verehrung dieser Menge haben!  
O glücklich, wer von seinen Gaben  
Solch einen Vorteil ziehen kann!“

Woraufhin Faust über die Hörigkeit und die Unkenntnis der Menschen klagt, die keine Ahnung über die wahren Schuldigen ihrer misslichen Lage haben. Er, Faust, kommt daher zu dem Schluss, dass die Ärzte es gar nicht verdient hätten, eine derartige Anerkennung zu erhalten.

Das Problem lässt sich als eine starke *Informationsasymmetrie* zwischen Patienten und Ärzten auffassen, die es den Ärzten erlaubt, ein viel höheres Ansehen zu genießen, als es vielleicht ihren Leistungen entsprach. Durch die Ausnutzung der Unwissenheit der Patienten konnte es den Ärzten daher gelingen, die Informationsasymmetrien in bares Geld zu verwandeln. Der Zusammenhang zwischen einem guten Ansehen in der Gesellschaft und der Vergütung wird noch deutlicher, wenn man bedenkt, dass eine gute Reputation des Arztes ihm nicht nur ein hohes Ansehen in der Gesellschaft verschafft, sondern gleichzeitig seinen „Kundenkreis“ vergrößert.

Einen interessanten Einblick in diese Problematik gewährt der Blick auf die Stellung des Hausarztes im 18. Jahrhundert. Iris Ritzmann (1999) untersucht in einem Beitrag zur Arzt-Patient-Beziehung im 18. Jahrhundert den Verhaltenskodex des Arztes anhand verschiedener zeitgenössischer Quellen. Ihre Auswahl beschränkt sich auf Anleitungen für junge Mediziner im 18. Jahrhundert, denen damals großes Interesse entgegengebracht wurde. Da das ausgehende 18. Jahrhundert oft als die Epoche der medizinischen Aufklärung erkannt wird, scheint gerade diese Zeit für eine derartige Untersuchung prädestiniert. Die Patienten, die im Laufe des 18. Jahrhunderts auch immer häufiger aus dem Bürgertum stammten, wurden gegenüber den einzelnen Behandlungsmethoden kritischer. Gleichzeitig hing das Einkommen des Arztes (insbesondere des Hausarztes) im entscheidenden Maße von der Zahlungsbereitschaft dieser Patienten ab. Die medizinische Geschichtsschreibung nennt deshalb dieses System auch gern ein „Patronagesystem“: Der gelehrte Arzt bemühte sich um die Gunst des Patienten, den er zunächst von seinen Heilkünsten überzeugen musste. Das war sicherlich nicht leicht, da den meisten Patienten auch noch andere „Berater“ zur Verfügung standen, mit denen der Arzt konkurrieren musste. In einem solchen System war es für den Arzt besonders wichtig, ein *gutes* Erscheinungsbild abzugeben. Nicht nur das Äußere sondern ebenso auch die Umgangsformen des Arztes waren davon betroffen. Daher ist es auch nicht verwunderlich, dass es in jener Zeit zahlreiche Schriften gab, die dem Arzt helfen sollten, das Vertrauen der Patienten zu gewinnen.

Für das richtige Verhalten des Arztes prägte sich in den 80er und 90er Jahren des 18. Jahrhunderts der Begriff „Ärztliche Politik“. Doch dieser Begriff geriet bald in Verruf, weil er stark mit der ökonomischen Absicherung des Arztes verbunden wurde. Man suchte also einen neuen Namen und fand in der damaligen Modesprache einen anmutigeren Begriff: „Savoir-Faire“. Indes ging es immer um das Gleiche, „das Zutrauen des Kranken zu dieser oder jener Heilmethode zu unterstützen, zu befördern“, wie es ein zeitgenössischer Autor ausdrückte.<sup>1</sup> „Kleinere und größere Kunstgriffe sind zuweilen nöthig“ und im Idealfall bleiben Ärzte „ungestört in dem Vertrauen des Kranken, wenn es auch immer schlimmer und bis zum Tode geht“<sup>2</sup>

Ritzmann kommt in ihrem Aufsatz zu dem Schluss, dass „der Verhaltenskodex der Ärzte im 18. Jahrhundert vorrangig darauf abzielte, aus ökonomischen und therapeutischen Gründen die

---

<sup>1</sup> Das Zitat stammt von dem Medizinprofessor Samuel Gottlieb Vogel (1750-1837) und wurde nach Ritzmann (1999, S. 202) wiedergegeben.

<sup>2</sup> Vgl. auch zu diesen Zitaten Ritzmann (1999, S. 202).

eigene Hilflosigkeit durch ein geschicktes, Vertrauen und Respekt einflößendes Verhalten zu überdecken.

Obwohl sich das Arzt-Patienten-Verhältnis seit jener Zeit gravierend verändert hat, weisen die Beispiele aus der Vergangenheit auf Probleme hin, die heute noch von Bedeutung sind. So ist auch in modernen Informationsgesellschaften das Verhältnis zwischen Arzt und Patient durch eine ungleiche Informationsverteilung gekennzeichnet. Denn nur der Arzt kann in aller Regel die richtige Diagnose stellen und schließlich die angemessene Behandlung einleiten. Daher wird der Arzt oft auch als Sachwalter der Patienten bezeichnet. Ferner erzielt der Arzt aus seiner Tätigkeit ein Einkommen. Je nach Ausgestaltung der Vergütung und anderer Anreizmechanismen kann es zwischen legitimen Einkommensinteressen und Sachwalterfunktion zu Abwägungsproblemen kommen, die sich in Über- oder Unterversorgung der Patienten ausdrücken können. In diesem Sinne ist die Beschäftigung mit Informationsasymmetrien in der Arzt-Patienten-Beziehung und den Einkommensinteressen der Ärzte für ein besseres Verständnis des Arztverhaltens notwendig.

## I DER ARZT IM GESUNDHEITSSYSTEM

### 1 Einleitung:

#### Der Arzt als Lotse im Gesundheitssystem

Die ambulant tätigen Ärzte nehmen bei der Erstellung und Verteilung von Gesundheitsleistungen eine Schlüsselposition ein. Die Allgemeinmediziner unter ihnen sind in der Regel die Ersten, die bei gesundheitlichen Problemen von den Patienten konsultiert werden und die gegebenenfalls die Patienten zu einer Weiterbehandlung an Spezialärzte oder ins Krankenhaus überweisen. Der Hausarzt übt somit nicht nur direkt (durch Behandlung oder Verschreibung von Medikamenten), sondern auch indirekt (durch Überweisung) einen entscheidenden Einfluss auf den Gesamtumfang an medizinischen Leistungen aus, die bei der Behandlung des Patienten zu erbringen sind. In diesem Zusammenhang wird häufig auch vom „Türhüter“, neuerdings auch vom „Lotsen“ des Gesundheitswesens gesprochen.

Aus diesem Grund gilt gerade dem Verhalten des Hausarztes das besondere Interesse in der Fachliteratur. Insbesondere geht es um die Frage, wie das Verhalten des Hausarztes zu steuern ist, damit er einerseits die Interessen seiner Patienten wahrnimmt und andererseits die auch im Gesundheitswesen knappen Ressourcen wirtschaftlich einsetzt. Ein wichtiges Steuerungsinstrument kann ein Vergütungssystem sein, welches das Verhalten der Ärzte in der beschriebenen Richtung beeinflusst. Nicht zuletzt das Aufkommen neuer Versorgungs- und Vertragsformen (*Managed Care*) hat die Diskussion über die Steuerung des Arztverhaltens neu intensiviert. In der Schweiz wurde *Managed Care* mit dem neuen Krankenversicherungsgesetz (KVG) 1996 eingeführt.<sup>3</sup> Der Begriff *Managed Care* bedeutet etwa soviel wie „gesteuerte Versorgung“. Man versteht darunter sowohl die Anwendung von Managementprinzipien auf die medizinische Versorgung als auch die Integration von Versicherung und Versorgung (vgl. Kühn 1997).

Der Fokus der vorliegenden Arbeit liegt auf dem Einfluss von *Managed Care* auf das Verhalten von Allgemeinärzten in der Schweiz. Allerdings kann eine derartige Betrachtung nicht losgelöst von den anderen Problemfeldern erfolgen, denn *Managed Care* greift zum einen direkt in die

---

<sup>3</sup> *Managed Care* in der Schweiz gab es bereits vor 1996. Seit 1990 konnten verschiedene Formen in Pilotprojekten erprobt werden. Mit dem neuen Krankenversicherungsgesetz 1996 bekamen sie dann eine allgemeine gesetzliche Grundlage.

Arzt-Patienten-Beziehung ein und verwendet zum anderen bestimmte Vergütungsformen, die das Verhalten der Ärzte steuern sollen.

Die Arbeit ist in drei große Teile gegliedert. Der erste Teil beschäftigt sich mit dem Arzt als Lotsen im Gesundheitswesen. Neben der Vorstellung der grundlegenden Vergütungssysteme (Kapitel 2) beinhaltet er einen Überblick zu Geschichte und Entwicklung von Managed Care (Kapitel 3) sowie wichtige institutionelle Rahmenbedingungen, denen sich der Schweizer Arzt gegenüber sieht (Kapitel 4).

Der zweite Teil untersucht unter Verwendung unterschiedlicher theoretischer Modellansätze das Verhalten von Ärzten. Es werden zum einen die Beziehungen des Arztes zu den Patienten, der Krankenversicherung sowie zu anderen Ärzten mit Hilfe der Prinzipal-Agent-Theorie untersucht und zum anderen wird mittels des herkömmlichen mikroökonomischen Instrumentariums die Arztpraxis als Firma betrachtet. Obwohl in der neueren Literatur zum Thema Arztverhalten oftmals die Entwicklung eines umfassenden Modells des Arztverhaltens als zentrale zukünftige Aufgabe angesehen wird, beschränkt sich der theoretische Teil weitestgehend auf die Darstellung und Verknüpfung existierender Modelle von Teilaspekten des ärztlichen Verhaltens. Vor allem die Komplexität begründet dieses Vorgehen, aber auch der Umstand, dass die Beleuchtung des Untersuchungsgegenstands von unterschiedlichen Blickwinkeln aus zu interessanten Erkenntnissen führen kann, haben das gewählte Vorgehen als ratsam erscheinen lassen. So weisen z.B. die unterschiedlichen Betrachtungsweisen im zweiten Teil (im Kapitel 5 explizite Modellierung von Informationsasymmetrien und im Kapitel 6 die Annahme vollständig informierter Parteien) auf ähnliche Eigenschaften einer optimalen Vergütung von Ärzten hin.

Der dritte Teil überprüft unter Zuhilfenahme empirischer Modelle die aus der Theorie abgeleiteten Ergebnisse. Dabei müssen die realen Bedingungen der Schweiz genauso berücksichtigt werden wie Beschränkungen aufgrund der Datenbasis. Deshalb kommt die Empirie auch nicht ohne eine weitere theoretische Basis aus. Konkret werden aus der Theorie empirische Modelle abgeleitet, die anschließend mit ökonometrischen Methoden getestet werden.

Da auf der empirischen Überprüfung der bekannten Hypothesen aus der Theorie der Schwerpunkt der Arbeit liegt, dient der zweite Teil – die Vorstellung verschiedener ökonomischer Modelle des Arztverhaltens – zur Herleitung dieser Hypothesen. Mit anderen Worten ist es nicht das Ziel dieser Arbeit, ein neues ökonomisches Modell des Arztverhaltens zu entwickeln. Vielmehr geht es um die empirische Überprüfung verschiedener ökonomischer Theorien

bezüglich des ärztlichen Verhaltens von Allgemeinmediziner bzw. „Gatekeepern“ in der Schweiz.

## 2 Die Vergütung des Arztes

Einen entscheidenden Einfluss auf das Verhalten des Arztes hat das Vergütungssystem, mittels dem er für seine Leistung honoriert wird. Der Einfluss der Vergütung auf das Verhalten der Ärzte beschränkt sich allerdings nicht nur auf den direkten Einfluss der Vergütung auf das Einkommen der Ärzte. Wie im weiteren Verlauf der Arbeit noch zu sehen sein wird, muss die Vergütung auch berücksichtigen, dass die Patienteninteressen in das Entscheidungskalkül des Arztes mit eingehen sollten. M.a.W. geht es bei der Diskussion der Arztvergütung auch um die Frage, welchen Einfluss die ärztliche Vergütung auf die *Qualität* besitzt.

Die Vergütung des Arztes kann sehr unterschiedlich erfolgen. Alle Formen der Vergütung (Honorarformen) werden u.a. von Schulenburg (1981) ausführlich behandelt. Im Folgenden wird ein Überblick über die wichtigsten Vergütungsformen gegeben, die für die Vergütung von Allgemeinärzten Verwendung finden (vgl. ebenfalls Breyer et al. 2003).

Die am häufigsten angewandte Vergütungsform stellt die **Einzelleistungsvergütung** (engl.: Fee For Service – FFS) dar. Bei der Einzelleistungsvergütung wird der Arzt nach Art und Umfang der erbrachten (einzelnen) Leistung honoriert, wobei jede Leistungsart mit einem absoluten oder relativen Preis bewertet wird.

Die der Einzelleistungsvergütung entgegengesetzte Vergütungsform ist die **Kopfpauschale** (engl.: Capitation). Diese Honorierungsform entspricht der Vergütung des Arztes nach der Anzahl der während eines bestimmten Zeitraums beim Arzt eingeschriebenen Versicherten bzw. Patienten.

Ein **Gehalt** ist schließlich die Vergütung des Arztes in Form eines festen Geldbetrages, den er in einer bestimmten Periode erhält.

Die Einzelleistungsvergütung stellt die konventionelle Vergütungsform dar, die sowohl in der Schweiz als auch in Deutschland bei der Vergütung der meisten Allgemeinmediziner angewandt wird. Ihr liegt der Gedanke zugrunde, dass der Arzt entsprechend seiner Leistung vergütet werden sollte. Bei gegebenem Preis pro Einzelleistung ergibt sich das Honorar des Arztes aus der Summe der mit dem jeweiligen Preis multiplizierten abgerechneten Menge. Rechnet der Arzt

mehr Leistungen ab, steigt dementsprechend seine Vergütung. Geht man davon aus, dass die Qualität der Versorgung eng an die Leistungsmenge gekoppelt ist, kann unter der Einzelleistungsvergütung mit einer hohen Behandlungsqualität gerechnet werden. Da die Vergütung sich auf die Leistungserstellung in der Vergangenheit bezieht (Honorierung der tatsächlich abgerechneten Leistungen), bezeichnet man sie als eine retrospektive Vergütung.

Kopfpauschalen werden häufig in Managed-Care-Organisationen verwendet. Der Arzt erhält bei dieser Vergütungsform am Anfang einer Periode für die Versorgung bei ihm eingeschriebener Versicherter für die folgende Periode einen bestimmten Geldbetrag, weswegen man hier auch von einer prospektiven Vergütung spricht (Honorierung der erwarteten Kosten). Im Unterschied zur Einzelleistungsvergütung kann der Arzt bei Kopfpauschalen durch die Ausweitung der Leistungsmenge seine Vergütung nicht steigern. Vielmehr lohnt es sich für den Arzt, möglichst wenige Leistungen anzubieten. Andererseits stellt jeder zusätzliche Patient eine potenzielle Ertragssteigerung dar. Deshalb wird der rational handelnde Arzt versuchen, Patienten möglichst zu binden, indem er eine kontinuierliche Patientenbetreuung anbietet.

Auch beim Gehalt besitzt der Arzt einen Anreiz, möglichst wenige Leistungen anzubieten, da er einen festgelegten Geldbetrag für eine Periode unabhängig von der Leistungsmenge erhält. Ein zusätzliches Problem dieser Vergütung stellt die Abkoppelung der Entlohnung von der Patientenzahl dar. Denn dadurch hat der Arzt einen Anreiz, möglichst wenige Patienten zu behandeln, was sich in langen Warteschlangen manifestieren kann (Qualitätsproblem). Großbritannien lässt sich als Beispiel anführen. Dort sind die Ärzte Angestellte des nationalen Gesundheitsdienstes. Neben den Warteschlangen sind hier noch weitere Qualitätseinschränkungen festzustellen.<sup>4</sup> Das Gehalt ist auch eine verbreitete Vergütungsform in Managed-Care-Organisationen, wo es aber in der Regel unter Einbeziehung einer strengen Qualitätskontrolle eingesetzt wird.

In der Literatur sind die Einzelleistungsvergütung und die Kopfpauschale Hauptgegenstand des Interesses. Übereinstimmend kommt man dabei zu der Ansicht, dass ein gemischtes Vergütungssystem, welches Elemente beider Vergütungsformen aufweist, die beste Anreizstruktur für den Arzt besitzt, um eine möglichst optimale Versorgung der Patienten sicherzustellen. Die Ableitung eines derartigen optimalen Vergütungssystems aus unterschiedlichen Blickwinkeln steht im Zentrum des zweiten Teils dieser Arbeit. Speziell zeigen die Kapitel 5 und 6, dass die optimale Vergütungsstruktur auf unterschiedlichen Wegen hergeleitet werden kann.

---

<sup>4</sup> Siehe zum Gesundheitsdienst in Großbritannien bspw. Whynes et al. (1999).



Eine wichtige Implikation eines gemischten Vergütungssystems sei hier schon herausgegriffen: Durch den Verzicht auf eine ausschließliche Honorierung über Kopfpauschalen wird der Gefahr der Risikoselektion durch die Ärzteschaft vorgebeugt.<sup>5</sup> Bei Kopfpauschalen besitzen die Leistungserbringer einen Anreiz, kostenintensive Fälle an andere Einrichtungen weiterzugeben, um so das eigene Budget zu entlasten. M.a.W. versuchen sie, das ihnen vollständig überantwortete Risiko der Behandlungskosten dadurch zu minimieren, dass sie die Behandlung besonders schwerer Fälle zu vermeiden suchen. Mit dem Ausmaß der Übernahme des Kostenrisikos durch die Kassen bzw. der Kassenärztlichen Vereinigungen (in Deutschland) sinkt die Wahrscheinlichkeit dieses, aus Sicht der Gemeinschaft unerwünschten Verhaltens.

Eine vollkommene Übernahme des Risikos durch die Kassenärztlichen Vereinigungen, wie sie die aktuelle Honorierung im ambulanten Bereich in Deutschland (Gesamtvergütung) vorsieht, ist andererseits auch nicht sinnvoll, weil dadurch die Anreize für die Ärzte zur Kostenbegrenzung nicht oder zu wenig ausgebildet sind (vgl. Newhouse 1996). Das in Deutschland übliche Punktesystem, nachdem jede Leistung zunächst mit einem Punktwert bewertet wird und erst nach Abrechnung sämtlicher Leistungen im Abrechnungszeitraum einen Euro-Betrag zugewiesen bekommt, erhöht noch den Anreiz für den einzelnen Arzt, möglichst viele Leistungen abzurechnen. Da das Budget fest ist, geschieht die Ausweitung der Leistungen durch den einzelnen Arzt zu Lasten anderer Ärzte, die weniger Leistungen abrechnen.

Immer häufiger kommt aber auch der Ruf nach neuen Vergütungsformen auf, die sich stärker auf das Ergebnis der Behandlung stützen sollten (vgl. z.B. Krauth et al. 1997). Dabei zeigt sich, dass ein Grundbaustein einer ergebnisorientierten bzw. erfolgsorientierten Vergütung der Mix aus Einzelleistungsvergütung und Kopfpauschale ist (Krauth et al. 1997). Allerdings weist die Analyse im Kapitel 5 darauf hin, dass eine ergebnisorientierte Vergütung nicht leicht zu implementieren ist, da sie eine exakte Messung des Gesundheitszustandes vor und nach einer Behandlung erfordert sowie die Bewertung einer Verbesserung des Gesundheitszustandes für die Patienten.

Wesentliche Elemente einer optimalen Vergütungsstruktur versuchen Managed-Care-Modelle umzusetzen. Dort bestimmen gemischte Formen der Kostenbeteiligung und der Vergütung mittels Kopfpauschalen die Anreizstruktur. Eine theoretische Begründung eines derartigen Vergütungssystems geben die Modelle, die im zweiten Teil dieser Arbeit vorgestellt werden.

---

<sup>5</sup> Vgl. zu den verschiedenen Formen der Risikoselektion bspw. Ellis (1998) oder auch Breyer et al. (2003).

Ein Blick auf andere Länder zeigt, dass die Verbreitung der unterschiedlichen Vergütungssysteme eng mit der Ausbreitung und der Entwicklung von Managed-Care-Elementen verbunden ist. So sind in den USA, wo es schon seit Jahrzehnten Managed-Care-Organisationen gibt, Systeme mit Kopfpauschalen wesentlich öfter anzutreffen als in anderen Ländern. In der Schweiz werden seit 1996 immer häufiger Modelle mit Kopfpauschalen angewendet; die vorherrschende Vergütungsform ist aber nach wie vor die Einzelleistungsvergütung. In Deutschland werden schließlich seit einigen Jahren in Pilotprojekten verschiedener Krankenkassen erste Versuche mit neuen Vergütungssystemen unternommen.<sup>6</sup> Auch hier bleibt aber das alte System der Einzelleistungsvergütung das vorherrschende.

Des Weiteren sei darauf hingewiesen, dass die Vergütungssysteme ebenfalls in Stufen eingesetzt werden können, sodass es zu einer Vermischung unterschiedlicher Vergütungssysteme kommen kann. Man denke bspw. an eine Gruppenpraxis, die mittels Kopfpauschalen vergütet wird, die ihrerseits aber die beteiligten Ärzte in der Gruppenpraxis entsprechend der Einzelleistungsvergütung honoriert oder ihnen ein festes Gehalt gewährt. Gerade die so genannten „Three-Tiered“ HMOs sind für derartige Vergütungsformen gute Beispiele.<sup>7</sup> In einem solchen System übernimmt zwar das Arztkollektiv der Gruppenpraxis oder HMO das volle Kostenrisiko, das Risiko für den einzelnen Arzt wird aber minimiert.<sup>8</sup> So ist die seit der Einführung des neuen KVGs in der Schweiz zu beobachtende Tendenz, dass immer mehr Ärzte sich in Gruppenpraxen zusammenschließen, auch als eine Reaktion auf die durch das neue Gesetz eingetretene Unsicherheit zu werten.<sup>9</sup> Die in Deutschland angewandte Budgetierung der Ärzte kann in gewisser Weise ebenfalls als eine derartige Vergütungsform verstanden werden. Denn die Kassenärztlichen Vereinigungen erhalten als Vertreter der Ärzte für einen bestimmten Zeitraum ein festes Budget (vollständige Risikoübernahme). Das Budget wird dann von den Kassenärztlichen Vereinigungen auf die einzelnen Ärzte aufgeteilt. Die Aufteilung erfolgt gemäß der Einzelleistungsvergütung, allerdings – wie oben beschrieben – über Punkte, und führt zu einer (begrenzten) Risikominimierung für den einzelnen Arzt.<sup>10</sup>

---

<sup>6</sup> So z.B. das Praxisnetz in Berlin oder die medizinische Qualitätsgemeinschaft Witten-Herdecke. Vgl. dazu Rübiger et al. (2002).

<sup>7</sup> Vgl. hierzu bspw. Hillman et al. (1992) sowie die Ausführungen zu HMOs in Kapitel 3.

<sup>8</sup> Die Anreizwirkungen von unterschiedlichen Vergütungssystemen in Abhängigkeit der Gruppengröße der Gemeinschaftspraxen werden bspw. von Gaynor und Gertler (1995) untersucht.

<sup>9</sup> Vgl. Mojon-Azzi (2001a, 2001b).

<sup>10</sup> Wobei die Risikoübernahme durch den einzelnen Arzt in den letzten Jahren durch die Einführung von Praxisbudgets für bestimmte Leistungsgruppen zugenommen hat.

Eine andere Vergütung, die nicht eindeutig den klassischen Formen zugeordnet werden kann, stellt die so genannte **ressourcenbasierte Vergütung** dar. Während die Einzelleistungsvergütung die einzelne Leistung direkt vergütet (bspw. eine Konsultation) und die Kopfpauschale nur indirekt einen Bezug zur erbrachten Leistung aufweist, bezieht sich die ressourcenbasierte Vergütung auf die Art und den Umfang der eingesetzten Ressourcen. Grundidee einer solchen Vergütung ist die Beobachtung, dass der Arzt unterschiedliche Ressourcen bei der Versorgung der Patienten verwendet. Neben einem großen Fixkostenanteil (Miete, Geräte, Apparaturen u.a.) ist eine der Hauptressourcen einer Arztpraxis die Arbeitszeit des Arztes. Greifen wir das Beispiel der Konsultation heraus und fragen uns, welche Ressourcen für eine Konsultation notwendig sind. Für eine einfache Konsultation lässt sich die Frage schnell beantworten: Außer der Arbeitszeit des Arztes werden noch Praxisräume benötigt. Allerdings ist trotzdem nicht jede Konsultation gleich der anderen. Vielmehr unterscheiden sich die Konsultationen trotz gleicher Behandlungszeit hinsichtlich des Umfangs und der Intensität der Behandlung. Beide richten sich nach der Schwere der Erkrankung des Patienten. Daher können unterschiedliche Konsultationen auch ein unterschiedliches Ausmaß an ärztlichem Aufwand sowie auch vollkommen unterschiedliche Arbeitsschritte beinhalten.

### 3 Der Arzt und Managed Care

Ein Ziel von Managed Care ist es, die hohen Kosten im Gesundheitswesen zu senken. Kostensenkungen sollen dabei nicht nur durch eine Steuerung des Verhaltens der Versicherten, sondern vor allem auch durch die gezielte Steuerung des Arztverhaltens erreicht werden.

Die Aufgaben der Managed-Care-Organisation (MCO) gehen damit über die klassische Krankenversicherung hinaus, indem sie das Angebot und die Nachfrage nach medizinischen Leistungen direkt oder indirekt beeinflusst. Während auf der Angebotsseite die Leistungserbringer durch verschiedene Integrationsformen und Anreizsysteme „gesteuert“ werden, geht es auf der Nachfrageseite darum, für die Patienten eine hohe Qualität der Versorgung bei angemessenen Versicherungsprämien sicherzustellen.

Charakteristisch in solchen Systemen sind Beteiligungs- und Anreizstrukturen für den Leistungserbringer aber auch für den Versicherungsnehmer durch vertragliche Bindungen. Dabei kommt es zu einer teilweisen Verlagerung des Versicherungsrisikos auf die Leistungserbringer durch Gewinnbeteiligung oder Kopfpauschalen. Die Versicherten in einem solchen System müssen

dagegen weitestgehend auf eine freie Arztwahl verzichten und zuerst ihren Hausarzt aufsuchen. Um den Versicherten Anreize zu setzen, sich trotzdem an derartigen Modellen zu beteiligen, sind die Prämien in MCOs günstiger als im konventionellen System (für eine Übersicht vgl. Luft 1995).

Es wird immer wieder betont, dass es aufgrund der Anreizstrukturen vor allem die Gesunden und Gutverdienenden seien, die sich in MCOs einschreiben würden. Selbst wenn davon ausgegangen werden kann, dass die Selektion der Versicherten tatsächlich zu der beschriebenen Umschichtung der Risiken führt, ist damit noch nicht die Frage beantwortet, ob Managed-Care-Organisationen überhaupt nichts zur Kosteneinsparung beitragen können.<sup>11</sup> Denn wie sowohl theoretische als auch empirische Arbeiten zeigen,<sup>12</sup> besitzen MCOs mehr Möglichkeiten als herkömmliche Versicherungen, die Patienten optimal zu versorgen. Hierbei geht es um das Problem des *Moral Hazards*, welches durch die Versicherungsdeckung hervorgerufen wird. Moral Hazard bedeutet in diesem Zusammenhang, dass die Versicherten mit Versicherungsdeckung mehr Gesundheitsleistungen nachfragen als ohne Versicherungsdeckung. Während in herkömmlichen Versicherungen für die Begrenzung des Moral-Hazard-Effektes in der Regel nur nachfrageseitige Kostenbeteiligungen zur Verfügung stehen, können MCOs auch andere Instrumente einsetzen. So führt die schon erwähnte angebotsseitige Kostenbeteiligung zu einer Kostenkontrolle der Leistungserbringung. Neben der Auswahl der eigentlichen Gatekeeper-Ärzte erfolgt auch eine gezielte Auswahl anderer Leistungserbringer außerhalb der MCO (Krankenhäuser, Spezialisten). Schließlich helfen neue Managementkonzepte wie *Guidelines* und *Utilization Reviews* – eigentlich Konzepte der Qualitätssicherung –, die Kosten der Leistungserbringung zu kontrollieren. Hinzu kommt, dass durch ein gezieltes *Case Management* versucht wird, gerade bei komplizierten und kostenintensiven Einzelfällen den Behandlungsprozess zu optimieren. Damit können nicht nur Kosteneinsparungen erzielt werden, sondern auch Qualitätsverbesserungen der Versorgung.

Aufgrund der angebotsseitigen Kostenbeteiligung scheint „Übersorgung“ in MCOs kein Problem darzustellen. Vielmehr können die Anreizsysteme dazu führen, dass die Leistungserbringer weniger als die optimale Leistungsmenge oder die Leistungsmenge zu einer geringeren Qualität anbieten, um Kosten einzusparen. Ein solches Verhalten kann einerseits durch die

---

<sup>11</sup> Einen Nachweis für Risikoselektion durch HMOs zeigen bspw. Polsky und Nicholson (2000). Vgl. zur Selektionsproblematik auch Abschnitt 3.2.

<sup>12</sup> Vgl. zu den theoretischen Ausführungen die Modelle in den Kapiteln 5 und 6 und für empirische Beispiele die Darlegungen weiter unten in diesem Abschnitt.

starken Bindungen des Arztes an den ärztlichen Kodex begrenzt werden und andererseits auch wiederum durch Maßnahmen der Qualitätssicherung. Außerdem haben die Leistungserbringer in einer MCO ein natürliches Interesse daran, dass möglichst viele Versicherte in der MCO eingeschrieben sind. Daher würde eine permanente Qualitätssenkung oder die Praxis einer permanenten Unterversorgung zur Unzufriedenheit der Versicherten und schließlich zu deren Weggang führen. In diesem Zusammenhang ist auch die folgende Beobachtung von Interesse: Aufgrund der teilweisen Risikoübernahme durch die Leistungserbringer innerhalb der MCOs werden hier mehr präventive Leistungen als im konventionellen Sektor angeboten (vgl. Glied 2000).<sup>13</sup>

Im Idealfall führt die beschriebene Steuerung der medizinischen Versorgung dazu, dass weniger Patienten in Spitäler eingewiesen, weniger Medikamente verschrieben, weniger Patienten an Spezialisten überwiesen werden und gleichzeitig eine qualitativ hochstehende Gesundheitsversorgung gewährleistet bleibt oder die Qualität der Versorgung sogar verbessert wird.

### 3.1 Organisationstypen

Dieser Abschnitt stellt kompakt die Hauptformen der Managed-Care-Organisationen (MCOs) vor. Umfassende Überblicke zum Thema bieten bspw. Amelung und Schumacher (2000) und Glied (2000). Darüber hinaus gibt es zahlreiche Darstellungen in Lehrbüchern der Gesundheitsökonomie (z.B. Breyer et al. 2003, Folland 1997 und Phelps 1997). Für die folgende Übersicht vgl. auch die Darstellung von Mühlenkamp (2001).

Üblicherweise unterscheidet man folgende drei Organisationstypen:

- a) Health Maintenance Organizations (HMOs)
- b) Preferred Provider Organizations (PPOs) und
- c) Point of Service Organizations (POSs).

Diese Einteilung stellt nur eine von mehreren möglichen dar. In letzter Zeit haben sich weitere Spielarten herausgebildet, die hier nicht weiter behandelt werden sollen. Die für die Schweiz typischen Modelle sind die von Typ *a* und von Typ *b*.

---

<sup>13</sup> Zu Qualitätsaspekten vgl. auch Abschnitt 3.3.

### a) Health Maintenance Organizations

Der Begriff Health Maintenance Organization (HMO) lässt sich sinngemäß als *Organisation zur Gesunderhaltung* ins Deutsche übersetzen. Damit wird der wesentliche Aspekt der Gesundheitsversorgung in derartigen Organisationen betont, der nicht vorrangig auf die kurative Medizin abzielt, sondern vielmehr auf die Vermeidung von Krankheiten. Gleichzeitig lässt sich mit dem Begriff auch schon der Aspekt der Kosteneinsparung assoziieren, denn durch die Erhaltung der Gesundheit können auch Kosten eingespart werden. Sicherlich hoffte man mit dieser Begriffsfindung mehr Akzeptanz derartiger Organisationen in der Öffentlichkeit zu finden.<sup>14</sup>

Für die beteiligten Ärzte bedeutet eine stärkere Betonung der Gesunderhaltung, dass sie sich einem Anreizsystem gegenübersehen, welches die Gesunderhaltung der Versicherten „belohnt“. Diese Anreize werden durch entsprechende Vergütungsregelungen und Vertragsformen realisiert. Angewandte Vergütungsregelungen sind Kopfpauschalen und das Gehalt mit Gewinnbeteiligung. Die Vertragsformen sind durch die Integration von Versicherung und Leistungserstellung gekennzeichnet.<sup>15</sup> Damit vereinigt eine HMO die Aufgaben der Krankenversicherung mit den Aufgaben der Leistungserbringung. Die Versicherten, die sich für eine HMO entscheiden, verpflichten sich, immer einen Arzt der HMO aufzusuchen. Dieser kann die Patienten dann – wenn nötig – zu einem HMO-eigenen Spezialisten oder zu einem Spezialisten außerhalb der Organisation überweisen. Im Gegenzug kommen die Versicherten in den Genuss von deutlich günstigeren Prämien als im traditionellen System. Außerdem müssen die Versicherten in der Regel keine Kostenbeteiligung in Form von absoluten Selbsthalten leisten. Proportionale Selbstbeteiligung werden dagegen von einigen HMOs erhoben.

Man unterscheidet innerhalb der HMOs zwei Hauptformen: das Staff Model und das Group Model. Das Staff Model zeichnet sich dadurch aus, dass die HMO als Träger der Versicherung ambulante Leistungen selbst erbringt und die Ärzte in der Regel Angestellte der HMO sind. Die Ärzte bekommen ein festes Gehalt, welches durch erfolgsabhängige Bonuszahlungen ergänzt wird. Beim Group Model bietet die HMO nicht selbst die ambulanten Leistungen an, sondern schließt exklusive Behandlungsverträge mit einer Gruppenpraxis ab. Die Ärztegemeinschaft der Gruppenpraxis wird hauptsächlich mittels Kopfpauschalen vergütet. Damit wird ein Teil des

---

<sup>14</sup> Vgl. Luft (1999).

<sup>15</sup> Diese Charakterisierung lässt sich allgemein auf MCOs anwenden. Sie wird hier angebracht, weil HMOs die ersten MCOs waren und die anderen Organisationsformen sich mehr oder weniger aus dieser entwickelt haben. Vgl. aber auch den nächsten Punkt zur Geschichte von Managed Care.

Versicherungsrisikos auf die Leistungserbringer übertragen. Die Group-Model-HMO ist somit eher eine Versicherung; eine Versicherung allerdings, die in der Lage ist, den Leistungserstellungsprozess zu überwachen und zu steuern.

Für die stationäre Behandlung der Versicherten schließen die HMOs Behandlungsverträge mit einzelnen Krankenhäusern ab. Manch große HMO betreibt sogar eigene Krankenhäuser.

Weitere Formen von HMOs ergeben sich aus der unterschiedlichen Art der Integration der Ärzte. Bei den Netzwerk-Modellen (Network Models) schließen Krankenversicherer sowohl Verträge mit ausgewählten Ärztegruppen als auch mit Ärzten in eigener Praxis ab. Die HMOs koordinieren, organisieren und kontrollieren die Arbeit der Gruppenpraxen und der einzelnen Ärzte. Im Gegensatz zum Group/Staff Model sind die Ärzte aber nicht exklusiv für eine HMO tätig, sondern können mehreren Ärztenetzen angehören und auch Versicherte mit traditioneller Versicherung versorgen. Netzwerkmodelle haben gegenüber den klassischen HMOs (Group oder Staff) entscheidende Vorteile. So können die Ärzte relativ schnell und ohne hohe Kosten einem derartigen Modell beitreten, da sie ihren bisherigen Patientenstamm nicht verlieren. Arztpraxen, die bisher nur eine geringe Auslastung hatten, können dagegen ihre Kapazitätsauslastung verbessern.

Eine weitere Unterform stellt die so genannte IPA dar. Die Individual Practice Associations sind Ärztezusammenschlüsse aus Allgemein- und Fachmedizineren zur ambulanten Leistungserstellung, die einen Vertrag mit einer HMO abgeschlossen haben. Die IPAs unterscheiden sich von den Netzwerken hauptsächlich durch die ausschließliche Beteiligung von Einzelpraxen.

### b) Preferred Provider Organizations

PPOs sind die zweite Hauptausprägungsform von MCOs. Sie schließen mit ausgewählten Leistungserbringern Verträge ab. Die Leistungserbringer, die in der Regel über Einzelleistungsentgelte vergütet werden, sichern den PPOs zu, die Leistungen zu günstigeren Konditionen anzubieten als bei klassischen Versicherungen. Im Gegenzug erhalten die Ärzte einen gesicherten Patientenstamm. Die Versicherten müssen sich in der Regel an den Kosten der medizinischen Versorgung beteiligen. Damit die Versicherten einen Anreiz bekommen, an solchen Systemen teilzunehmen, liegen die Kostenbeteiligungen meist deutlich unter denjenigen im traditionellen System. Die Versicherten dürfen im Unterschied zu den HMOs auch Leistungs

erbringer außerhalb des Systems aufsuchen, wenn sie bereit sind, dafür einen größeren Teil der Kosten zu übernehmen.

Verschiedentlich wird argumentiert (so z.B. Amelung und Schumacher 2000), dass die PPOs die direkte Antwort der Leistungsanbieter auf den wachsenden Druck der HMOs seien. Denn indem sich Ärzte oder Arztpraxen mit dem Ziel zusammenschließen, den Versicherungsgesellschaften ein konkurrenzfähiges Angebot zu unterbreiten, schmälern sie die Attraktivität von HMOs, können sich aber weiterhin die Vorzüge der Einzelleistungsvergütung sichern.

### c) Point of Service Organization

Die Point of Service Organizations (POS) schließlich kombinieren verschiedene Merkmale von HMOs mit denjenigen einer klassischen Versicherung. Konkret bieten sie ihren Versicherten den Leistungsumfang einer normalen HMO an, gestatten es den Patienten darüber hinaus aber auch, Leistungen fremder Leistungserbringer zu beziehen. Man könnte sagen, es handele sich hier um ein HMO-Angebot mit freier Arztwahl. Tatsächlich bieten viele HMOs auch Verträge in Richtung POS an. Die Prämien bei derartigen Verträgen liegen allerdings über den herkömmlichen HMO-Prämien, wobei aber in der Regel auch hier von den Versicherten keine oder nur geringe Kostenbeteiligungen verlangt werden.

## 3.2 Geschichte und Entwicklung von Managed-Care-Organisationen

Obwohl der Begriff Managed Care erst in den frühen 1970er Jahren aufkam, gab es auch schon vorher HMO-ähnliche Organisationen.<sup>16</sup> So wurde eine der bekanntesten HMOs, die Kaiser-Permanente, bereits 1942 in Kalifornien gegründet.<sup>17</sup> Frühe Formen von Managed Care lassen sich sogar bis in das 18. Jahrhundert zurückverfolgen. Die als *contract medicine* bekannten Formen wurden bspw. von Arbeitgebern oder aber auch von Konsumgenossenschaften (*consumer cooperatives*) gegründet und vergüteten die beteiligten Ärzte mittels Kopfpauschalen.<sup>18</sup> Anfang des 20. Jahrhunderts gab es zunächst viele Barrieren für die Entwicklung von Managed Care. Ärzte, die sich an *contract medicine* beteiligten, wurden mancherorts sogar aus den Ärztevereinigungen ausgeschlossen (Glied 2000).

---

<sup>16</sup> Der Begriff geht auf Lewis Butler und Paul Ellwood zurück, die ihn Anfang der 1970er Jahre geprägt haben (vgl. Luft 1999). Vgl. weiterführend vor allem Folland (1997) oder Frech III et al. (2000).

<sup>17</sup> Vgl. Folland (1997), S. 267.

<sup>18</sup> Bis 1888 waren z.B. in New Orleans 88% der Bevölkerung auf diese Weise versichert. Auch konnten heutige HMOs auf frühere Verträge aufbauen. Vgl. Frech III et al. (2000) und Frech III (1996).



In der zweiten Hälfte des 20. Jh. stieg die Anzahl von Managed-Care-Organisationen (MCOs) erst ab Mitte der 1970er Jahre, nachdem auch vom Staat diese Formen der Gesundheitsversorgung gefördert wurden (so z.B. erstmals durch den *Federal HMO Act* von 1973). Während die Anzahl der HMOs in den 1980er Jahren einen starken Zuwachs erfuhr und Mitte der 1990er Jahre etwas zurückging, wuchs die Anzahl von PPOs gerade in dieser Zeit sehr stark an.<sup>19</sup> Die Zahl der Versicherten in HMOs nahm in den letzten zehn Jahren nur leicht zu, während es von Beginn der ersten HMOs bis weit in die 80iger Jahre des 20. Jahrhunderts hinein zu einem sehr kräftigen Anstieg kam. Gleichzeitig erfolgte eine starke Umschichtung zwischen den einzelnen Modellarten. Am Anfang der Managed-Care-Entwicklung stand eindeutig das Group- bzw. Staff-Model im Vordergrund. In den letzten Jahren dagegen nahm deren Anteil vor allem zu Gunsten der IPAs und weiterer Mischformen ab.

Gründe für diese Entwicklung sind sicherlich vielfältig. Ein Hauptgrund scheint aber in dem auf Versicherungsmärkten bekannten Phänomen der *Adversen Selektion* zu liegen.<sup>20</sup> Demnach führte die Einführung von Managed Care zu einer Selbstselektion der Versicherten: Die guten Risiken wanderten von den traditionellen Versicherungen zu den MCOs, während die weniger guten Risiken in ihren ursprünglichen Verträgen verblieben. Dies führte zu einem zu dem beobachteten starken Anstieg von MCOs und zum anderen auch zu einem starken Anstieg der Prämien im traditionellen Versicherungsbereich (vor allem relativ zu den Prämien von Managed Care). Die steigenden Prämien im konventionellen Bereich führten dann zur Immigration von immer mehr Versicherten in Managed Care. In der Folge veränderte sich so auch das Erscheinungsbild der MCOs. Aufgrund dieser Entwicklung ist es denn auch nicht verwunderlich, dass im weiteren Verlauf gerade diejenigen MC-Modelle den größten Mitgliederzuwachs verzeichneten, die den Bedürfnissen der Neuankömmlinge am besten entsprachen (möglichst umfangreiche Deckung, geringe Begrenzung der freien Arztwahl u.a.). Auf der anderen Seite übernahm nun der konventionelle Bereich viele der Management-Prinzipien aus dem Managed-Care-Bereich. Im Endeffekt führt nach Pauly und Nicholson (1999) dieser Prozess zum Ausgangspunkt zurück, wo alle Versicherten die gleiche Versicherungsdeckung und mehr oder weniger auch gleiche Versicherungsverträge besitzen (die einen in MCOs, die sich der traditionellen Versicherung angenähert haben, die anderen in traditionellen Versicherungen, die nun ihrerseits ähnlich wie die MCOs

---

<sup>19</sup> Vgl. Folland (1997, S. 275).

<sup>20</sup> Vgl. hierzu und im Folgenden Pauly und Nicholson (1999).

arbeiten). Ist dieser Punkt erreicht, dürfte der Prozess von neuem beginnen, d.h. wieder eine starke Selektion der Risiken eintreten.

Eine mögliche Lösung, um den Prozess zu stoppen, wird in risikoadjustierten Prämien gesehen. In einem solchen System werden bestimmte Risikounterschiede der Versicherten zwischen den unterschiedlichen Versicherungsunternehmen ausgeglichen, sodass eine Selektion der Versicherten hinsichtlich dieser Risiken für die einzelne Versicherung nicht mehr möglich ist. Allerdings wurde bisher noch kein „goldener Weg“ gefunden, der Risikoselektion seitens der Versicherer unmöglich machte. Vielmehr existieren zwei grundlegend unterschiedliche Positionen zu der Thematik, wobei aber alle von der Notwendigkeit einer Risikoadjustierung ausgehen. Die eine „Schule“ setzt auf eine möglichst genaue Erfassung der Risiken – vor allem durch die Berücksichtigung von Morbiditätsfaktoren. Die andere betont dagegen die Gefahren einer zu starken Ausweitung der Risikoadjustierung.<sup>21</sup> Ein interessanter Vorschlag wurde von Glazer und McGuire (2000) unterbreitet. Sie zeigen, dass eine Über- bzw. Unterbezahlung bekannter Faktoren der Risikoadjustierung (wie Alter oder Geschlecht) genau so gut funktionieren kann, wie ausgeklügelte Risikoausgleichssysteme.

### 3.3 Managed Care und Qualität

Die Betrachtung von Qualitätsaspekten in Managed-Care-Organisationen steht im Mittelpunkt der öffentlichen Aufmerksamkeit. Wie bereits dargelegt wurde, besteht das Hauptziel von MCOs in der Kosteneinsparung durch eine stärkere Betonung der Gesunderhaltung der Versicherten. Damit ist Managed Care ein starker Motor für eine präventiv ausgerichtete Medizin. Es bestehen aber auch Anreize, schlechte Risiken fern zu halten oder sie durch Leistungsbegrenzungen zu vertreiben. Daher wird vermutet, dass es gerade die chronisch Kranken sind, die Qualitätseinbußen hinnehmen müssen.

Allerdings ist ein Widerspruch zwischen Forschungsergebnissen zu Managed Care und der öffentlichen Meinung nicht zu übersehen. Denn viele Untersuchungen stützen Ergebnisse, nach denen im MC-Bereich Kosten gespart werden und eine gleich hohe Qualität wie im traditionellen Bereich geboten wird.<sup>22</sup> So können z.B. Cutler et al. (2000) am Beispiel von Herzkrankheiten

---

<sup>21</sup> Vgl. hierzu vor allem auch die Diskussion um den deutschen Risikostrukturausgleich (RSA), der diese unterschiedlichen Positionen deutlich machte (vgl. u.a. IGES et al. (2001) sowie Felder (2001) und die darin angegebene Literatur).

<sup>22</sup> Mittlerweile existiert – gerade für die USA – eine Fülle an Untersuchungen, die die Leistungserbringung und die Qualität von MCOs im Vergleich zum konventionellen Sektor zum Gegenstand haben. Einen Überblick über

zeigen, dass die Ausgaben für HMO-Versicherte zwischen 30-40% unter den Ausgaben herkömmlich Versicherter liegen. Trotzdem nimmt die Konsumentenzufriedenheit mit Managed-Care-Organisationen immer mehr ab.<sup>23</sup> Für diese Diskrepanz kann es unterschiedliche Erklärungen geben. Eine mögliche folgt aus der im vorangegangenen Abschnitt dargestellten Entwicklung des Managed-Care-Marktes. Vor dem dargestellten Hintergrund des sich veränderten Versicherungsmarktes lässt sich ein Teil der starken Unzufriedenheit der Versicherten durch die Selektion der Versicherten selbst erklären: Die „schlechten“ Risiken sind unzufrieden, weil die MCO für sie die beste von zwei schlechten Alternativen darstellt (umfangreiche Deckung und hohe Prämie vs. geringere Deckung und niedrigere Prämie) und die guten Risiken sind unzufrieden, weil ihre Prämien wegen der veränderten Versichertenstruktur in den MCOs steigen (vgl. wieder Pauly und Nicholson 1999).

Weitere Erklärungen können darin gesehen werden, dass manche Untersuchungen die Heterogenität des Managed-Care-Marktes nicht oder nur unzureichend berücksichtigen. Hinzu kommen die Schwierigkeiten, echte Kosteneinsparungen von Selektionsgewinnen zu unterscheiden.<sup>24</sup> Außerdem wird die öffentliche Meinung in starkem Maße von Medizinern, der Pharmaindustrie und den interessierten Verbänden beeinflusst. Alle genannten Akteure haben ein natürliches Interesse an einer Begrenzung bzw. Verdrängung von Managed Care, da sie es sind, die durch die Kosteneinsparungspolitik der MCOs Einkommenseinbußen zu erleiden haben (vgl. Robinson, R. 2000). Gerade der letzte Punkt macht deutlich, dass hinter der Rhetorik gegen Managed Care oftmals handfeste Interessen bestimmter Gruppen stecken.

Viele Managed-Care-Kritiker übersehen zudem den oben beschriebenen dynamischen Charakter von Versicherungsmärkten. Schon allein aus diesem Grund scheint die Vorhersage vom baldigen Ende von Managed Care in den USA nicht in Erfüllung zu gehen, wie sie von J. Robinson in einem Artikel mit dem bezeichnenden Titel *The End of Managed Care* in der amerikanischen Ärztezeitung postuliert wurde (Robinson, J. 2001). Vielmehr werden sich die Versicherungsformen immer wieder verändern, so wie sie das in der Vergangenheit auch getan haben. Managed-Care-Prinzipien werden dabei auf unterschiedliche Art auch weiterhin eine wichtige Rolle spielen, um die Kosten zu begrenzen.

---

unterschiedliche Studien gibt bspw. Hillman (1995), der diese Studien systematisch einordnet. Eine Literaturswertung für die Jahre 1997 – 2001 bieten Miller und Luft (2002). Vgl. auch Glied (2000).

<sup>23</sup> Vgl. bspw. Miller und Luft (2002).

<sup>24</sup> Vgl. auch hierzu bspw. Hillman (1995), Hellinger (1996) oder Dudley und Luft (2001). Das Selektionsproblem wird ebenfalls bei der empirischen Untersuchung eine wichtige Rolle spielen.

Für die Mediziner stellt Managed Care eine enorme Herausforderung dar. Denn die neuen Versorgungsformen verlangen auch eine radikale Neuausrichtung des ärztlichen Selbstverständnisses. Während das konventionelle Selbstverständnis des Arztes vor allem in der Betonung der Heilung von Krankheiten des einzelnen Patienten gesehen werden kann, wird in einem Managed-Care-Umfeld vom Arzt wesentlich mehr erwartet: Neben die kurativen Bemühungen sollen hier auch verstärkt präventive Maßnahmen treten, um die Heilung und Gesunderhaltung des einzelnen Patienten zu befördern. Darüber hinaus stärkt Managed Care auch die Verantwortlichkeit des Arztes für ein gesamtes Kollektiv von Versicherten. Die neue Rolle, die der Arzt in MCOs einnehmen soll, hat damit auch weitreichende Konsequenzen auf den Umfang seiner Arbeit. Sie ist viel mehr durch nichtärztliche Tätigkeiten geprägt als im konventionellen Bereich. Gerade die Ausweitung des Aufgabenfeldes eines Arztes wird aber von vielen Ärzten abgelehnt. Aus Arztsicht wird daher ein Punkt im Kampf gegen MC immer wieder betont: Aufgrund der erhöhten administrativen Anforderungen hätten die Ärzte in MCOs immer weniger Zeit für die Patienten, sodass damit auch die Qualität der Behandlung zurückgehen würde. Entgegen dieser Vorhersagen ist aber der Anteil der Arbeitszeit, die der Arzt direkt mit dem Patienten verbringt, gleich geblieben. Die durchschnittliche Zeit pro Besuch hat sogar zugenommen (vgl. hierzu Luft 1999 und Mechanic et al. 2001). Interessant ist ferner, dass die immer stärkere Durchsetzung von Managed Care ebenfalls zu einer kostengünstigeren Behandlung konventionell versicherter Patienten führt.<sup>25</sup> Managed-Care-Instrumente strahlen daher offensichtlich auch auf den konventionellen Bereich aus.

Von institutioneller Seite werden verschiedene Steuerungselemente eingesetzt, um eine qualitätsgerechte Versorgung sicherzustellen. Die wichtigsten Instrumente sind Guidelines, Qualitätszirkel, Disease-Management-Programme und das Case Management. Während die ersten beiden Instrumente der allgemeinen Qualitätssicherung dienen, bieten die beiden letztgenannten die Möglichkeit, gerade für chronisch Kranke die Qualität der Behandlung zu steigern. In Disease-Management-Programmen werden die Behandlungsabläufe für bestimmte chronische Erkrankungen optimiert. Das Case Management kümmert sich dagegen um die Koordination komplizierter und kostenintensiver Einzelfälle.<sup>26</sup>

---

<sup>25</sup> Vgl. Baker (1999).

<sup>26</sup> Eine ausführliche Darstellung der Instrumente findet sich in Robinson und Steiner (1998) und Amelung und Schumacher (2000).

Für die Nachfrager der Leistungen einer MCO ist es wichtig, die Qualität einer MCO im Verhältnis zu anderen MCOs bewerten zu können. Daher ist eine hohe Datentransparenz notwendig, um Vergleichbarkeit zu ermöglichen. Einer der wichtigsten und größten Nachfrager sind in den USA die Arbeitgeber, da sehr viele Versicherte über ihren Arbeitgeber krankenversichert sind. Gemeinsam haben die Arbeitgeber ein Daten- und Informationssystem aufgebaut, mit dessen Hilfe die Versorgungsqualität von MCOs eingeschätzt werden kann (vgl. Amelung und Schumacher 2000). Dieses Informationssystem (HEDIS)<sup>27</sup> beinhaltet 7 Kriterienkategorien mit Untergruppen, anhand derer die Performance der HMOs eingeschätzt werden kann. So werden bspw. die Effektivität der Versorgung, die Kosten der Versorgung, der Zugang zu Versorgungsleistungen, die Zufriedenheit mit den Leistungserbringern und die generelle Inanspruchnahme von Leistungen gemessen. Das nationale Komitee zur Qualitätssicherung (NCQA) entwickelte das System weiter und verwendet es zur Einschätzung der Qualität von Krankenversicherern und MCOs. Da das System öffentlich zugänglich ist, kann es von allen Interessierten genutzt werden. Dies trifft somit auch auf die Patienten wie auf jede einzelne Arztpraxis zu, die ihre Kosten und ihre Praxisqualität im Verhältnis zu anderen Arztpraxen in der Region überprüfen kann (Sue 2000). Ein ähnliches System existiert auch für die staatlichen Versicherungen Medicare und Medicaid.<sup>28</sup> Grundvoraussetzung für derartige Vergleichsmöglichkeiten ist – wie erwähnt – eine hohe Datentransparenz. MCOs haben gerade auch auf diesem Gebiet eine Vorreiterrolle eingenommen, die aus der Notwendigkeit der eigenen Legitimation geboren wurde. Mittlerweile hat sich die Veröffentlichung wichtiger Performancekennwerte in den USA stark verbreitet, wie man z.B. an den zahlreichen Rankings von Krankenhäusern erkennen kann.<sup>29</sup>

#### 4 Der (Haus-) Arzt in der Schweiz

Der ambulant tätige Arzt in der Schweiz rechnet seine Leistungen mit privaten Krankenkassen ab, die für die Versicherten ein definiertes Grundpaket zu risikoneutralen Prämien anbieten, wobei die Grundversicherung für jeden Schweizer obligatorisch ist. In der Regel wird der Arzt direkt vom Patienten bezahlt, der dann die Rechnung bei seiner Krankenkasse zur Liquidation

---

<sup>27</sup> Vgl. HEDIS (2000).

<sup>28</sup> Medicare ist die staatliche Versicherung für die über 65-Jährigen und Medicaid die Versicherung für Personen mit geringem Einkommen. Zu dem System für die Bewertung von Leistungen dieser staatlichen Versicherungen vgl. Dudley und Luft (2001).

<sup>29</sup> Vgl. hierfür stellvertretend das Ranking der NCQA (HEDIS 2000).

einreicht. Im Unterschied zu Deutschland existiert in der Schweizer sozialen Krankenversicherung ein Kostenbeteiligungssystem auf der Nachfrageseite, welches auch indirekt auf die Anbieter einwirkt. So gibt es einen absoluten Selbstbehalt, bis zu deren Höhe die Patienten sämtliche Leistungen aus der eigenen Tasche zahlen. Diesen Umstand muss der Arzt bei seinen Therapie- und Behandlungsvorschlägen berücksichtigen. Die konkreten Selbstbeteiligungsregelungen wie auch weitere allgemeine Regelungen zum Schweizer Versicherungssystem befinden sich im Anhang. An dieser Stelle wird konkret auf relevante Bestimmungen eingegangen, die den Arzt im konventionellen und im Managed-Care-Sektor direkt betreffen.

### 4.1 Der (Haus-) Arzt im konventionellen Sektor

Der Begriff konventioneller Sektor wird zur Abgrenzung vom Managed-Care-Bereich verwendet. Er beinhaltet keine wertende Aussage. Vielmehr beschreibt er das bestehende System der durch die Krankenkasse vergüteten Arztleistung sowie die Art der Versicherung, wie sie bis vor einigen Jahren ausschließlich in der Schweiz existierte. Nach wie vor ist dieses System das bestimmende in der Schweiz wie auch in Deutschland. Darüber hinaus wird im Folgenden auch auf allgemeine Regelungen bei der Gesundheitsversorgung in der Schweiz eingegangen, die für alle Ärzte gelten, so etwa auf das System der Arztaus- und Weiterbildung und auf das Standesrecht der Ärzte.

In der Regel wird der Arzt im konventionellen Sektor nach seiner erbrachten Leistung vergütet. Häufig wird das System deshalb auch Einzelleistungsvergütung genannt (vgl. Kapitel 0). Die Leistungsvergütung erfolgt in der Schweiz nach bestimmten Taxpunkten und Taxpunktswerten. D.h., dass sich der effektive Preis für eine vergütungsrelevante Leistung aus dem Produkt der entsprechenden Punktzahl für die Leistung und dem dazugehörigen Taxpunktwert ergibt. Im Unterschied zu Deutschland existiert zur Zeit aber keine Deckelung der Ausgaben (Budget). Die Vergütungstabellen mit den einzelnen Taxpunktswerten waren im Untersuchungszeitraum in jedem Kanton unterschiedlich ausgestaltet. Das bedeutet, dass ein und dieselbe Arztleistung in einem Kanton wesentlich höher bewertet wurde als in einem anderen. So wurde bspw. eine normale Konsultation in Zürich mit 23 Punkten bewertet und eine Konsultation in Genf mit 48 Punkten. Darüber hinaus ist auch der Punktwert kantonsabhängig. Während der Punktwert für medizinische Leistungen bspw. im Kanton Zürich 0.8 SFr. beträgt, lag er im Kanton Genf bei

0.5 SFr.<sup>30</sup> Zudem wurde in dem angesprochenen Vergütungsmodell die Spezialmedizin höher als die so genannte sprechende Medizin bewertet, was unter anderem auch zu deutlichen Unterschieden der Arzteinkommen in den verschiedenen Arztgruppen geführt hat. So betrug der Durchschnittsumsatz eines Allgemeinmediziners im Jahr 1999 530,000 SFr., während er für einen Urologen bei 850,000 SFr. lag.<sup>31</sup>

Neben der direkten Leistungserbringung spielt die Selbstdispensation als eine weitere Einkunftsart von Ärzten eine wichtige Rolle. Durch den Direktverkauf von Medikamenten erzielten die Ärzte im Jahr 1997 ein durchschnittliches Jahreseinkommen von 50,000 Franken.<sup>32</sup> Da die gesetzlichen Regelungen bezüglich der Selbstdispensation sich sehr stark von Kanton zu Kanton unterscheiden, variiert das Einkommen aus dem Medikamentenverkauf ebenfalls von Kanton zu Kanton. So ist beispielsweise in den Kantonen Basel, Genf, Neuenburg, Tessin und Waadt die Selbstdispensation nicht gestattet. Logische Konsequenz ist eine unterschiedliche Verteilung der Apotheken: Offenbar gibt es eine stark negative Korrelation zwischen selbstdispensierenden Ärzten und der Apothekendichte.<sup>33</sup>

Ein charakteristisches Merkmal für den konventionellen Sektor ist die freie Arztwahl der Patienten. Sie können sowohl ihren Hausarzt als auch Spezialärzte frei wählen, solange die Leistungserbringer zugelassene Kassenärzte innerhalb des Wohnkantons sind. Die freie Arztwahl wird von vielen Versicherten als ein sehr wichtiges Gut des Gesundheitssystems aufgefasst. Allerdings wird vermutet, dass diese Freiheit erhebliche Zusatzkosten für das Gesamtsystem erzeugt. Denn aufgrund der unterschiedlichen Informationsverteilung zwischen den verschiedenen Parteien im Gesundheitssystem (Versicherte, Ärzte und Krankenkassen) kann es zu einer ineffizienten Leistungserstellung kommen, welche sich z.B. oft in Doppeluntersuchungen manifestiert.

Eng mit der freien Arztwahl der Versicherten ist die Verpflichtung für die Krankenkassen verbunden, Leistungen aller als Kassenärzte anerkannten Ärzte zu übernehmen; wobei es sich hierbei um die in einem Leistungskatalog definierten Leistungen handelt. Die Krankenversiche

---

<sup>30</sup> Dieser Taxpunktwert gilt u.a. für Allgemeinmediziner, Ärzte der allgemeinen inneren Medizin und für Kinderärzte. Den kleinsten Wert weist Genf mit 0.5 SFr auf; den höchsten Basel Stadt mit 5.60 SFr (Stand: 2001, vgl. Tabelle der Taxpunktwerte, KSK, 2001, Angaben im WWW).

<sup>31</sup> Vgl. Hänggeli et al. (2001).

<sup>32</sup> Vgl. Pharma Information (1999).

<sup>33</sup> Vgl. Pharma Information (1999).

rer haben kein Recht, Ärzte von der Leistungserbringung auszuschließen (Kontrahierungszwang).

Der Berufszulassung kommt damit eine entscheidende Bedeutung bei der Gestaltung des Angebots medizinischer Leistungen zu. Zulassung wie auch die Berufsaufsicht liegen im Zuständigkeitsbereich der Kantone. Erste Voraussetzung für eine Zulassung ist das eidgenössische Arztdiplom. Das Diplom erhält man nach einem sechsjährigen Medizinstudium, welches an fünf Fakultäten in der Schweiz absolviert werden kann. Die zweite Voraussetzung für die Aufnahme einer Praxistätigkeit ist ein Weiterbildungstitel, der im Freizügigkeitsgesetz aufgeführt sein muss.<sup>34</sup> Die Durchführung und Organisation von Weiterbildungsmaßnahmen liegen in den Händen der Standesorganisation der Schweizer Ärzte (*FMH*). Sie ist somit auch für die Ausbildung zum Facharzt zuständig. Nicht nur wegen des starken Anstiegs der erteilten Arztdiplome – zwischen 1950 und 1970 waren es jährlich ca. 250 Diplome, seit 1970 sind es zwischen 700 und 800 im Jahr<sup>35</sup> –, sondern auch angesichts einer Liberalisierung bei der Arztzulassung im Zuge der bilateralen Verträge mit den Ländern der Europäischen Union wird ein starker Anstieg der niedergelassenen Ärzteschaft befürchtet. Denn die bilateralen Verträge werden es in Zukunft ausländischen Ärzten einfacher machen, eine Praxis zu eröffnen. Allein von einem Großteil derjenigen ausländischen Ärzte, die gegenwärtig in Schweizer Spitälern arbeiten, wird erwartet, dass sie sich selbständig machen wollen.

So besteht aufgrund der beschriebenen Marktsituation ein enormer Druck auf die Ärzte, da durch die auf den Markt drängenden Ärzte die Marktanteile des Einzelnen kleiner werden. Außerdem belohnt die praktizierte Vergütungsform diejenigen Ärzte, die am meisten abrechnen. Auch aufgrund dieses Drucks kann es zu medizinisch nicht notwendigen Doppeluntersuchungen oder anderen Maßnahmen der Mengenausweitung kommen. Wegen der zu erwartenden „Ärztenschwemme“ wurden daher auf Bundesebene verschiedene Maßnahmen in Erwägung gezogen; so dachte man über die Einführung eines Zulassungsstopps oder die Aufhebung des Kontrahierungszwangs nach.<sup>36</sup>

---

<sup>34</sup> Vgl. Hänggeli et al. (2001).

<sup>35</sup> Vgl. Hänggeli et al. (2001).

<sup>36</sup> Letztendlich wurde ein Zulassungsstopp verfügt. Vgl. hierzu auch die Ausführungen in Abschnitt 4.3.



#### 4.2 Der Hausarzt in HAM und HMOs

Managed Care zeichnet sich im Verhältnis zum konventionellen Sektor vor allem durch die unterschiedlichen Verantwortlichkeiten des Arztes aus. Während der Arzt im konventionellen Sektor nur für die direkte Behandlung zuständig ist, kontrolliert der Arzt im Managed-Care-Bereich die gesamte Behandlungskette des Patienten.

Die am häufigsten vorkommenden MCOs in der Schweiz sind die versicherungseigenen HMOs und die so genannten Hausarztmodelle (kurz: HAM).<sup>37</sup> Nach der oben vorgestellten amerikanischen Einteilung lassen sich die beiden Formen als Staff-HMO sowie als Mischform aus HMO und PPO auffassen. Viele HMOs und HAM werden von mehreren Versicherern betrieben. Neben den versicherungseigenen HMOs existieren mittlerweile auch ärzteigene HMOs. Bei letzteren handelt es sich um Ärztezusammenschlüsse, die von den beteiligten Ärzten selbst geleitet werden.

Die Ärzte in den HMOs behandeln in der Regel ausschließlich HMO-Patienten. Damit hat die Beteiligung an einer HMO für den Arzt weitreichende Konsequenzen, denn er muss seine Praxis und damit seinen Patientenstamm aufgeben. Dafür erwartet ihn in der HMO ein relativ gesichertes Auskommen (Gehalt, Urlaubsanspruch u.a.).

Bei den HAM handelt es sich aus Arztsicht um Arztzusammenschlüsse von Grundversorgern, denen Patienten beteiligter Versicherungen fest zugeordnet sind. Die Ärzte arbeiten in ihren eigenen Praxen und behandeln auch konventionell versicherte Patienten. Für den Versicherten ist die Einschreibung in einem derartigen HAM mit der Aufgabe der freien Arztwahl verbunden, da er sich verpflichtet, zuerst seinen Arzt innerhalb des Hausarztnetzes aufzusuchen. Damit gewinnt die Funktion des Hausarztes als Lotse sehr stark an Bedeutung. Denn aufgrund dieser Regelung besitzen die Hausärzte die Kontrolle über den gesamten Behandlungsprozess. Allerdings müssen auch die Anreize durch das Vergütungssystem in die entsprechende Richtung weisen. Anfänglich wurden die Ärzte in den meisten Hausarztnetzen weiter mit der Einzelleistungsvergütung honoriert. Erst nach und nach entwickelten sich Kopfpauschalen-Modelle oder andere Modelle der angebotsseitigen Kostenbeteiligung.

Ein Vorteil einer Hausarztnetzteiligung des Arztes kann in der Sicherung eines bestimmten Patientenstamms gesehen werden. Wobei der Arzt nicht auf die Behandlung seines alten

---

<sup>37</sup> Vgl. im Folgenden vor allem Buchs (2001).

Patientenstamms verzichten muss. Er kann sogar – je nach Standort (Einzugsgebiet eines HAMs) – durch den Beitritt zu einem Hausarztmodell seinen Patientenstamm erweitern. Vorteile können sich die Ärzte mit Managed-Care-Beteiligung auch durch die Aufhebung des Kontrahierungszwanges erhoffen. Denn es ist sicherlich nicht sehr realistisch anzunehmen, dass die Versicherer gerade Leistungserbringer mit MC-Beteiligung ausschließen werden.<sup>38</sup>

Nicht zuletzt aufgrund der Attraktivität der HAM für die Ärzte ist das Hausarztmodell gegenwärtig das bedeutendere MC-Modell im Vergleich zu den HMOs in der Schweiz.<sup>39</sup> Die Zahl der Hausarztmodelle stieg in den letzten Jahren beständig an. Die Zahl der eingeschriebenen Patienten wuchs im Jahr 1999 um 8%. Demgegenüber fiel die Steigerung der Versichertenzahlen in den HMOs schwächer aus.<sup>40</sup> Eine genauere Charakterisierung des Managed-Care-Marktes in Zürich und Genf – den Kantonen, die für die empirische Untersuchung relevant sind – bietet Abschnitt 7, wo auch die berücksichtigten Managed-Care-Organisationen ausführlich vorgestellt werden.

Insgesamt ist ein deutlicher Abschwung der Euphorie auszumachen, die noch die Einführung der ersten Managed-Care-Modelle begleitet hatte. Allerdings scheint eine endgültige Bewertung der Wirkung von Managed Care in der Schweiz verfrüht. Denn wie die Erfahrungen aus den USA zeigen, geht die Entwicklung auf dem Krankenversicherungsmarkt immer weiter (vgl. vorhergehende Abschnitte). Während MCOs ihre starken Restriktionen abbauen, übernehmen traditionelle Versicherungen immer mehr Prinzipien aus dem Managed-Care-Bereich. Ähnliche Entwicklungen sind auch in der Schweiz zu erwarten, wie ein Blick auf die zukünftigen Reformvorhaben zeigt.<sup>41</sup>

### 4.3 Ausblick: Tendenzen bei der Arztvergütung im Schweizer Gesundheitssystem

Das wichtigste Reformvorhaben, welches sich gegenwärtig – nach langer Diskussion und zahlreichen Widerständen – in der Einführung befindet, ist ein neues Vergütungssystem freipraktizierender Ärzte. Die Entwicklung eines derartigen neuen Systems ist eine direkte Folge

---

<sup>38</sup> Allerdings gibt es auch Arztnetze, die ausschließlich gegründet werden, um eine stärkere Verhandlungsmacht gegenüber den Kassen zu besitzen und jegliche Risikoübernahme ablehnen. Vgl. Beck (2003), der bezüglich der Motivation von Ärzten in MC-Organisationen der Schweiz die Begriffe *offensive* und *defensive* Motivation einführt.

<sup>39</sup> Auch für die Versicherten bieten die HAM gegenüber den HMOs Vorteile. So können sie aus einer größeren Zahl an Grundversorgern auswählen. Im günstigsten Fall brauchen sie nur die Versicherungsform zu wechseln, aber nicht den Arzt.

<sup>40</sup> Vgl. Bundesamt für Sozialversicherung (2000b).

<sup>41</sup> Vgl. für einen weiteren Überblick auch Huber et al. (2002).

aus der Einführung des neuen KVGs im Jahre 1996. Dort wurde gefordert, dass die Vergütung der Ärzte auf eine einheitliche Grundlage gestellt werden müsse; d.h., dass die kantonale Tarifstruktur aufzugeben sei und somit in der ganzen Schweiz eine einheitliche Tarifstruktur zu gelten habe, um so Transparenz und Vergleichbarkeit zu gewährleisten. Dieses neue Vergütungsmodell TARMED soll neben der einheitlichen Bewertung der ärztlichen Leistung auch zu einer anderen Gewichtung ärztlicher Tätigkeit führen. Insbesondere wird in Zukunft die so genannte sprechende Medizin gegenüber der teuren Apparatedizin aufgewertet werden.

TARMED ist ein ressourcenorientiertes Vergütungssystem<sup>42</sup> und basiert auf der Resource-Based Relative Value Scale (RBVV) von Medicare, die von Wissenschaftlern um Hsiao an der Harvard Universität entwickelt wurde. Ziel war die Entwicklung eines ressourcen-basierten relativen Wertemaßstabs durch die Untersuchung der Inputkosten und die Bereitstellung von Methoden ihrer Messung.<sup>43</sup> Die Harvard-Studie fand heraus, dass die ärztliche Arbeit eine Funktion von vier unterschiedlichen Dimensionen sei, welche sich mit Arbeitszeit, geistigem Aufwand, technischen Fähigkeiten sowie einer Stresskomponente beschreiben lassen. Letzterer Bestandteil umfasst dabei mögliche Unsicherheiten in der Diagnose oder Behandlungsrisiken. Innerhalb der vier Dimensionen können nicht nur die unterschiedlichsten Behandlungen und Prozeduren einer Arztspezialisierung dargestellt werden, sondern auch Prozeduren vollkommen unterschiedlicher Spezialisierungsrichtungen. Man schätzte hierfür spezialisierungsspezifische Arbeitselastizitäten bezüglich jeder der vier Dimensionen. Diese Vorgehensweise erlaubt es, den Arbeitsaufwand einer kurzen, aber sehr intensiven Prozedur mit einer längeren, für den Arzt aber weniger anstrengenden Prozedur vergleichbar zu machen. Im Ergebnis sollten unterschiedliche Verrichtungen so vergütet werden, dass der Preis für den einzelnen Aufwand die unterschiedliche Arbeitsintensität widerspiegelt. Damit soll das Verhalten der Ärzte in den Preisen antizipiert werden. Konkret sollten bspw. die Preise für Verrichtungen, bei denen der Arzt eine starke Mengenausweitung betreiben kann, möglichst nahe den Grenzkosten liegen. Demgegenüber könnten Preise für Arbeiten, bei denen diese Gefahr nicht besteht, wesentlich über dem Grenzkostenpreis liegen.<sup>44</sup> Eine Diskussion der RBVVS findet sich in Frech III (Hrsg.) (1991).

Ressourcenorientiert bei TARMED bedeutet, dass die ärztliche Einzelleistung hauptsächlich als Ergebnis der Beanspruchung zweier Inputs verstanden wird: der ärztlichen Arbeitszeit und

---

<sup>42</sup> Vgl. die Ausführungen zur ressourcenbasierten Vergütung in Kapitel 2.

<sup>43</sup> Vgl. Hadley (1991).

<sup>44</sup> Vgl. Wedig (1993).

von medizinisch technischen Mitteln. Diese neue Tarifstruktur soll einerseits dem Arzt die Kosten für die technischen Mittel erstatten und andererseits die ärztliche Leistung angemessen honorieren. Um für die erste Komponente entsprechende Faktoren (Produktivitätsfaktoren) bilden zu können, wurde seit Jahren die sogenannte rollende Praxiskostenstudie durchgeführt, die über die durchschnittlichen Kosten einer Praxis Auskunft gibt. Die ärztliche Leistung selbst wird direkt über die *Zeit* und die sogenannte *Dignität* bewertet. Bei der Dignität unterscheidet man zwischen der quantitativen Dignität und der qualitativen Dignität. Letztere bezieht sich auf den erworbenen Facharztstitel und die Weiterqualifikationen. Die quantitative Dignität misst dagegen die Berufserfahrung des Arztes, ausgedrückt in zusätzlichen Weiterbildungsjahren. Alles in allem wird durch die neue Struktur ein massiver Einkommenstransfer zwischen den Fachgesellschaften ausgelöst.<sup>45</sup>

Allerdings zeigt ein Blick auf ein anderes Reformvorhaben, dass es zu Widersprüchen kommen muss. Denn die angestrebte Reform der Vertragsgestaltung zwischen Krankenversicherungen und Ärzten (Abschaffung des Kontrahierungszwangs) setzt gerade auf das freie Gestalten von Verträgen. Ein einheitlicher Tarif scheint in diesem Sinne kontraproduktiv zu sein. Im Verhältnis zur gegenwärtigen deutschen Situation wie auch im Verhältnis zum alten Taxpunktsystem der Schweiz ist TARMED aber trotzdem ein Fortschritt, weil er differenzierter die unterschiedlichen Tätigkeiten des Arztes honoriert und somit auch in der Lage ist, unterschiedliche Anreize für bestimmte Aufgaben zu setzen. Daher ist es zunächst auch zu begrüßen, dass sich die deutsche Ärzteschaft ebenfalls mit EBM 2000 Plus Gedanken über eine verbesserte Tarifstruktur im Sinne des Schweizer Vorbilds macht.

Die Diskussion über die Abschaffung bzw. Einschränkung des Kontrahierungszwangs befindet sich gegenwärtig in einer entscheidenden Phase. Verschärft wurde die Diskussion in der letzten Zeit noch durch die befürchtete Ärzteschwemme aufgrund der bilateralen Verträge. Zur Lösung dieses Problems wurde nun kurzfristig entschieden, einen allgemeinen Zulassungsstopp zu verhängen.<sup>46</sup> Inwieweit diese Maßnahme die Diskussion um den Kontrahierungszwang in Zukunft beeinflussen wird, ist gegenwärtig noch nicht klar. Der Umstand, dass es eine derartige Diskussion überhaupt gibt, zeigt aber, dass Managed Care auch in der traditionellen Versiche

---

<sup>45</sup> Vgl. zur Kritik an TARMED vor allem Preisüberwacher (1999).

<sup>46</sup> Vgl. Eidgenössisches Departement des Inneren (2002).

rungslandschaft Wirkungen hervorruft, wie wir sie aufgrund der beschriebenen Entwicklung in den USA auch erwartet hätten.

## II THEORIE DER OPTIMALEN ARZTVERGÜTUNG

### 5 Diener zweier Herren:

#### Der Arzt als Agent von Krankenkassen und Patienten

Ein grundlegendes Problem auf dem Markt für Arztdienstleistungen besteht in den Informationsasymmetrien zwischen Patient und Arzt auf der einen und zwischen Krankenkasse und Arzt auf der anderen Seite. Hinzu kommt das Informationsproblem zwischen Versichertem (Patient) und Krankenkasse, welches die Probleme in den erstgenannten Beziehungen beeinflussen kann.

Zunächst sollte der Arzt aus Sicht der Patienten ein perfekter Sachwalter sein. D.h. der Patient erwartet von seinem Arzt, dass er ihn in bestmöglicher Art behandelt. Genauso sollte auch die Versicherung den Interessen ihrer Versicherten perfekt folgen. Sie gewährt dem Versicherten – gegen Zahlung einer Prämie – die Finanzierung der Behandlungskosten im Falle einer Erkrankung. Der Arzt wiederum tritt als Agent für die Versicherung im Krankheitsfall des Patienten auf, indem er die bestmögliche und kostengünstigste Behandlung wählt. Da jede Partei einen unterschiedlichen Informationsstand als die jeweils andere Partei besitzt, treten in allen Fällen Anreizprobleme für den Auftraggeber auf. Prinzipiell kann dabei zwischen Anreizproblemen auf der Angebotsseite und Anreizproblemen auf der Nachfrageseite unterschieden werden. Während nachfrageseitige Anreizprobleme vor allem zwischen Versicherten und Versicherung bestehen, liegen sie auf der Angebotsseite sowohl zwischen Arzt und Patient als auch zwischen Arzt und Versicherung. Darüber hinaus ist zu beachten, dass die Unterscheidung in Sachwalter (Agent) und Auftraggeber (Prinzipal) nicht immer eindeutig ist. So lässt sich etwa die Vertragsbeziehung zwischen Krankenkasse und Versicherten auch genau umgekehrt auffassen, dass nämlich die Versicherung der Prinzipal ist und die Versicherten die Agenten sind. In Abbildung 1 werden die grundlegenden Asymmetrien verdeutlicht. Sie zeigt zudem auch die wesentlichen Leistungs- bzw. Geldströme zwischen den Parteien.



### 5.1 Entscheidung unter Unsicherheit

Die beteiligten Parteien treffen ihre Entscheidungen zwangsläufig in einer unsicheren Umwelt. Dabei versteht man unter Unsicherheit eine Situation, in der die Wahrscheinlichkeit, mit der ein Ereignis eintritt, vollständig oder bis zu einem gewissen Grade unbekannt ist.<sup>47</sup> So kennt die Krankenversicherung nicht die Erkrankungswahrscheinlichkeit eines Versicherten oder der Arzt kann nicht mit Sicherheit den Erfolg einer Behandlung bestimmen. Auch für die bereits erwähnten unterschiedlichen Vergütungssysteme gibt es offenbar unterschiedliche Präferenzen der Ärzte, die wesentlich von der Risikostruktur der Entscheidungsträger abhängen. Wie bereits Pauly (1980) zeigte, würden risikoaverse Ärzte bspw. eher die Einzelleistungsvergütung wählen als andere, weniger risikoaverse Ärzte. Was aber versteht man unter „weniger risikoavers“? Da im folgenden Kapitel verschiedene Annahmen bezüglich des Risikoverhaltens der Parteien getroffen werden, erklärt dieser Unterabschnitt die grundlegenden Begriffe und geht insbesondere auf die Modellierung des Risikoverhaltens der Ärzte ein.<sup>48</sup>

Es wird ein Entscheider betrachtet, der sich verschiedenen Umweltzuständen gegenüber sieht. Die einzelnen Zustände kann er durch seine Handlungen ( $x$ ) verändern, indem er die Konsequenzen dieser Handlungen miteinander vergleicht. Da er sich in einer unsicheren Umwelt bewegt, muss er die unterschiedlichen Konsequenzen mit Wahrscheinlichkeiten belegen, um sie vergleichen zu können. Eine Möglichkeit ist die sogenannte Erwartungsregel, nach der er für alle Handlungen die mit den Wahrscheinlichkeiten gewichtete Summe der Konsequenzen bildet, d.h. der erwartete Nutzen eines Individuums lässt sich als Summe der bewerteten Konsequenzen abbilden.<sup>49</sup> Ein bekanntes Beispiel für diese Regel ist die von Neumann-Morgenstern-Nutzenfunktion, die für  $s$  Zustände folgendes Aussehen hat:

$$U(x) = \sum \pi_s v_{x,s},$$

wobei  $\pi_s$  die Wahrscheinlichkeit für das Auftreten des Zustandes  $s$  und  $v_{x,s}$  die vom Individuum bewertete Konsequenz einer Handlung im Zustand  $s$  ist. Derartige Nutzenfunktionen sind linear in den Wahrscheinlichkeiten und separierbar in  $x$ .

---

<sup>47</sup> Vgl. Sinn (1989, S. 17).

<sup>48</sup> Allgemeine Darstellungen zu diesem Thema finden sich etwa im Kapitel 14 in Deaton und Muelbauer (1980) oder in Sinn (1989).

<sup>49</sup> Auf die wichtige Unterscheidung zwischen Nutzenfunktion über die Konsequenzen und Nutzenfunktion über die Handlungen wird hier nicht explizit eingegangen. Vgl. hierzu etwa Hirshleifer und Riley (1979).



Mit Hilfe solcher Nutzenfunktionen lassen sich relativ mühelos viele mögliche Risikostrukturen abbilden. Wird eine lineare Funktion für  $v$  angenommen, so ergibt sich der Nutzen für ein risikoneutrales Individuum (bspw. für  $s = 2$ :  $U(x) = \pi_1 x_1 + \pi_2 x_2$ ); wird dagegen ein konkaver Verlauf von  $U(x)$  angenommen, so entspricht dies dem Verhalten eines risikoaversen Individuums ( $U' \geq 0$ ,  $U'' < 0$ )<sup>50</sup>.

Dabei stellt  $U'' < 0$  die notwendige und hinreichende Bedingung für die Risikoaversion dar. Will man allerdings auch die Stärke der Risikoaversion untersuchen, so reicht die alleinige Betrachtung der zweiten Ableitung nicht aus, da das Maß sensitiv auf eine Transformation der Nutzenfunktion reagiert. Das Ausmaß der Risikoaversion lässt sich vielmehr mittels der Maße von Arrow (1971) und Pratt (1964) messen,<sup>51</sup> die die zweite Ableitung durch die Division mit der ersten Ableitung der Nutzenfunktion normieren. Dabei wird zwischen absoluter und relativer Risikoaversion unterschieden:

Absolute Risikoaversion: 
$$R_A = -\frac{U''(x)}{U'(x)} \text{ und}$$

Relative Risikoaversion: 
$$R_R = -\frac{xU''(x)}{U'(x)}.$$

Häufig werden diese Maße benutzt, um Aussagen über die Stärke der Risikoaversion treffen zu können, ohne explizite Funktionen formulieren zu müssen. So lassen sich mittels der Maße Funktionsklassen definieren, die ein bestimmtes Risikoverhalten abbilden. Eine wichtige Klasse stellt dabei diejenige der konstanten absoluten Risikoaversion dar.<sup>52</sup> Um eine intuitive Interpretation der Maße geben zu können, wird kurz auf ihre Ableitung am Beispiel der absoluten Risikoaversion eingegangen.

Betrachtet wird ein Individuum mit einem Einkommen  $Y$  und der Nutzenfunktion  $U(Y)$ . Es wird angenommen, dass das Individuum eine Investition tätigen kann, die ihm entweder einen Gewinn oder einen Verlust von  $b$  einbringt. Die Wahrscheinlichkeit eines Gewinns betrage  $\pi$ . Bezeichnet nun  $\pi(Y, b)$  gerade diejenige Wahrscheinlichkeit, bei der das Individuum zwischen

---

<sup>50</sup>  $U', U''$  bezeichnen die erste bzw. zweite Ableitung der Nutzenfunktion.

<sup>51</sup> Vgl. Deaton und Muelbauer (1980).

<sup>52</sup> In diese Klasse fallen alle Nutzenfunktionen, bei denen entsprechend des Maßes der absoluten Risikoaversion der Quotient aus zweiter und erster Ableitung eine Konstante ist.

einem sicheren Einkommen und einer unsicheren Investition – die beide zu einem gleich hohen erwarteten Einkommen führen würden – indifferent ist, so kann gezeigt werden,<sup>53</sup> dass folgende Beziehung gilt:

$$\pi(Y, b) \approx \frac{1}{2} + \frac{b}{4} R_A.$$

Damit nimmt diese „Indifferenzwahrscheinlichkeit“ mit der Risikoaversion zu. Mit anderen Worten: Einem stark risikoaversen Individuum müsste man ein größeres Sicherheitsäquivalent zahlen, damit es die gleiche unsichere Investition tätigt wie ein risikofreudigeres Individuum.

Eine anders geartete Beurteilung risikoaversen Verhaltens wird durch die Berücksichtigung der zeitlichen Dimension unsicherer Entscheidungen erreicht. So kann es von Interesse sein, wie ein risikoaverses Individuum zwischen heutigem Konsum und zukünftigem unsicheren Konsum entscheidet. Ein bekanntes Maß für diese Beurteilung ist die sogenannte „Vorsicht“ oder „Besonnenheit“.<sup>54</sup> Hierbei geht es darum, Aussagen über die Voraussicht des Agenten zu treffen. Es sei deshalb angenommen, dass sein Einkommen in jeder Periode bestimmten Schwankungen unterliegt. Betrachtet man bspw. zwei Perioden, so lässt sich das Einkommen in Periode 2 als Summe des Erwartungswertes des Einkommens ( $\bar{Y}$ ) und einer Abweichung von diesem Erwartungswert ( $\tilde{Y}$ ) darstellen:  $Y = \bar{Y} + \tilde{Y}$ . Dabei gilt  $E(\tilde{Y}) = 0$ . Das Individuum kann sein Einkommen in der ersten Periode auf den heutigen Konsum ( $k$ ) und die Ersparnis ( $s$ ) aufteilen. Das Entscheidungsproblem für das Individuum besteht in der optimalen Wahl von  $s$ , um seinen erwarteten Nutzen über die zwei Perioden zu maximieren:

$$\max_s E \left\{ U(Y_0 - s) + \beta U(s + \bar{Y} + \tilde{Y}) \right\},$$

wobei  $Y_0$  das verfügbare Einkommen der Periode 1 bezeichnet und  $\beta$  den Diskontierungsfaktor für den Konsum in der zweiten Periode kennzeichnet. Um die unterschiedlichen Auswirkungen des unsicheren Einkommensteils in der zweiten Periode beschreiben zu können, bezeichne im Folgenden  $s_1$  die Ersparnis bei Unsicherheit (also  $\tilde{Y} \neq 0$ ) und  $s_2$  die Ersparnis bei sicherem Einkommen ( $\tilde{Y} = 0$ ).

---

<sup>53</sup> Vgl. bspw. Zweifel und Eisen (2001).

<sup>54</sup> Das Konzept der „prudence“ wurde zuerst von Kimball (1990) vorgestellt. Die nachfolgende Darstellung basiert auf Danthine und Donaldson (2002).

Die Bedingung erster Ordnung für das Problem lautet in diesen Fällen:

$$\begin{aligned} \text{(i)} \quad & U'(Y_0 - s_1) = \beta E\{U'(s_1 + \bar{Y} + \tilde{Y})\} \\ \text{(ii)} \quad & U'(Y_0 - s_2) = \beta U'(s_2 + \bar{Y}). \end{aligned}$$

Beim Konzept der „Besonnenheit“ wird nun gefragt, wie viel das Individuum zu zahlen bereit wäre, um die Unsicherheit im Einkommen in der zweiten Periode zu ignorieren. Oder anders ausgedrückt: Welche Kompensation müsste man dem Individuum zahlen, damit seine Konsum- und Sparsentscheidung in beiden Fällen gleich bleiben würde? Kimball (1990) definiert zu diesem Zweck die zwei folgenden Maße und nannte sie absolute und relative Prudence:

$$\begin{aligned} P_A(c) &= -\frac{U'''(c)}{U''(c)} \text{ und} \\ P_R(c) &= -\frac{cU'''(c)}{U''(c)}. \end{aligned}$$

Ähnlich dem Sicherheitsäquivalent bei der Risikoaversion lässt sich hier eine „vorbeugende Kompensationsprämie“ definieren  $[\Psi = \Psi(\bar{Y}, \tilde{Y}, s)]$ , die den zusätzlichen Betrag angibt, den man dem Individuum geben müsste, damit die Lösungen zu Problem (i) und zu Problem (ii) vollständig identisch sind. Kimball (1990) zeigt, dass

$$\text{(I)} \quad \Psi(\bar{Y}, \tilde{Y}, s) \approx 1/2\sigma_{\tilde{Y}}^2 P_a(s + \bar{Y})$$

und dass für zwei unterschiedliche Nutzenfunktionen, für die Folgendes gilt:

$$\begin{aligned} & P_a^1(s) < P_a^2(s), \quad \forall s \text{ dann auch} \\ \text{(II)} \quad & \Psi_2(\bar{Y}, \tilde{Y}, s) > \Psi_1(\bar{Y}, \tilde{Y}, s), \quad \forall \bar{Y}, \tilde{Y}, s \text{ gelten muss.} \end{aligned}$$

Nach (I) ist also die Kompensationszahlung dem Produkt aus Prudence-Index und Varianz des unsicheren Einkommensanteils  $(\sigma_{\tilde{Y}}^2)$  direkt proportional. Das Resultat (II) ist intuitiv zugänglich: Je besonnener ein Individuum ist, um so höher wird die Kompensationszahlung dafür ausfallen, dass das Individuum auch unter Unsicherheit genau den gleichen Betrag in die Periode 2 transferiert wie bei Sicherheit.

### 5.2 Die Prinzipal-Agent-Theorie (Grundlagen)

Man unterscheidet im Allgemeinen die Modelle nach der Art der Informationsverteilung zwischen Prinzipal und Agent. Handelt es sich um Informationen bezüglich der *Handlungen* des Agenten spricht man von Modellen mit verborgener Handlung, betreffen die Informationen dagegen bestimmte *Eigenschaften* des Agenten, die dem Prinzipal verborgen bleiben, so handelt es sich um Modelle mit verborgener Information. Des Weiteren können die Informationsprobleme auch beidseitig auftreten.<sup>55</sup> Typische Beispiele für verborgene Handlungen und Informationen finden sich auf dem Arbeitsmarkt. So kann der Arbeitgeber den Aufwand seiner Angestellten nicht genau beobachten (verborgene Handlung) oder er kann die Produktivität eines Arbeiters nicht richtig einschätzen (verborgene Information). Wie die Beispiele zeigen, lassen sich beide Probleme oftmals nicht eindeutig voneinander trennen (aufgrund der geringen Produktivität betreibt ein Arbeiter wenig Aufwand).<sup>56</sup> Probleme bei Franchise-Verträgen lassen sich als Beispiel für beidseitige Informationsasymmetrien anführen.<sup>57</sup>

In der Gesundheitsökonomie können derartige Modelle auf mehreren Gebieten eingesetzt werden. Adverse Selektion auf Versicherungsmärkten lässt sich bspw. als P-A-Modell mit verborgener Information beschreiben und Moral-Hazard-Probleme in Modellen mit verborgener Handlung. Dabei beschreibt das Phänomen der adversen Selektion auf Versicherungsmärkten die Beobachtung, dass gute Risiken versuchen, für sie vorteilhafte Verträge abzuschließen, der Versicherer aber nicht in der Lage ist, gute und schlechte Risiken eindeutig zu erkennen.<sup>58</sup> Im Falle der Versicherung bezieht sich dagegen das Problem des Moral Hazards auf die Beobachtung, dass Versicherte mit einer umfangreichen Leistungsdeckung, eine höhere Leistungsanspruchnahme aufweisen, als wenn sie keine Versicherungsdeckung hätten.<sup>59</sup> Doppelseitige Informationsprobleme lassen sich ebenfalls im Gesundheitssektor ausmachen – so z.B. in der Arzt-Patienten-Beziehung.

---

<sup>55</sup> Eine grundlegende vertragstheoretische Diskussion dieser Modelle findet sich bspw. in Schweizer (1999) oder auch in Salanié (1997).

<sup>56</sup> Vgl. auch Milgrom und Roberts (1992, S. 169) für ein weiteres Beispiel. Zur Darstellungsmöglichkeit von Fällen als PA-Problem mit verborgener Handlung wie auch als PA-Problem mit verborgener Information vgl. Milgrom (1987).

<sup>57</sup> Vgl. Bhattacharyya und Lafontaine (1995).

<sup>58</sup> Vgl. Arrow (1963).

<sup>59</sup> Vgl. hierzu auch die Ausführungen in Kapitel 3. Das Problem des Moral Hazards wird darüber hinaus in unterschiedlichen Facetten im Verlauf dieser Arbeit noch eine wichtige Rolle spielen – nicht zuletzt bei der empirischen Analyse.

### 5.2.1 Grundmodell bei einseitig verborgener Handlung

Ausgangspunkt der Analyse ist ein gewinnmaximierender, risikoneutraler Prinzipal, der einen Agenten beauftragt, eine bestimmte Aufgabe zu erledigen.<sup>60</sup> Der Prinzipal steht im Allgemeinen vor dem Problem, den Agenten für seine Anstrengungen in geeigneter Weise zu vergüten. Probleme treten insofern auf, da die Interessen des Agenten möglicherweise denen des Prinzipals entgegengesetzt sind und der Prinzipal nicht in der Lage ist, die Anstrengungen, die der Agent zur Erledigung der Aufgabe unternimmt, direkt zu beobachten. Maximiert der Agent seinen Nutzen (Gewinn), so wird er bestrebt sein, eine möglichst hohe Vergütung zu bekommen. Eine hohe Vergütung des Agenten schmälert aber den Gewinn des Prinzipals.

Es gilt bei der Lösung dieses Problems einen Vertrag zu finden, der im Optimum dem Prinzipal durch die Wahl einer optimalen Vergütung den maximalen Gewinn sichert. Der Gewinn des Prinzipals sei  $\Pi$  und die Anstrengung des Agenten zur Erledigung einer bestimmten Aufgabe  $e \in E$  sowie die Vergütung des Agenten  $w$ . Da der Prinzipal die tatsächliche Anstrengung des Agenten nicht beobachten kann, stellt der mit den Anstrengungen verbundene Gewinn  $\Pi$  eine Zufallsgröße dar. Die Zielfunktion des Prinzipals lautet damit:

$$E[\Pi(e) - w(\Pi(e))]. \quad (5.1)$$

Mit anderen Worten hängt der erwartete Nettogewinn des Prinzipals von der gewählten Vergütung  $w$  und dem Aufwand des Agenten  $e$  ab.

Der Agent maximiert seinen Nutzen ( $U$ ) in Abhängigkeit seines Lohns  $w$  und den zu leistenden Anstrengungen  $e$ . Der erwartete Nutzen des Agenten lässt sich in allgemeiner Form daher schreiben als:

$$EU(w(\Pi(e)), e). \quad (5.2)$$

Gesucht wird ein Vertrag, der den erwarteten Nettogewinn des Prinzipals maximiert. Ein solcher Vertrag muss gewährleisten, dass der Agent einen Anreiz besitzt, sich an diesem Vertrag zu beteiligen. D.h. der erwartete Nutzen aus dem Vertrag bei allen möglichen Anstrengungsniveaus muss mindestens so groß wie bei einer alternativen Verwendung seiner Anstrengungen sein, die ihm ein Nutzenniveau von  $U_0$  verschaffen würde:

$$EU(w(\Pi(e)), e) \geq U_0. \quad (5.3)$$

---

<sup>60</sup> Die Darstellung folgt in weiten Teilen der Darstellung bei Tirole (1995).

Gleichung (5.3) beschreibt somit die Bedingung, die erfüllt sein muss, damit sich der Agent an diesem Vertrag beteiligt (*Beteiligungsbedingung*). Darüber hinaus will der Prinzipal ein bestimmtes Anstrengungsniveau  $e$  implementieren. Daher muss die Vergütungsstruktur anreizverträglich für den Agenten sein, d.h. das Anstrengungsniveau  $e$  muss für den Agenten aus allen möglichen Anstrengungsniveaus  $e \in E$  seinen erwarteten Nutzen maximieren:

$$e \in \arg \max_{\tilde{e} \in E} EU(w(\Pi(e)), \tilde{e}). \quad (5.4)$$

Gleichung (5.4) gibt damit die *Anreizverträglichkeitsbedingung* für den Agenten an. Die Aufgabe des Prinzipals besteht nun darin, eine Vergütung  $w^*$  zu wählen – und damit ein optimales Anstrengungsniveau  $e^*$  festzulegen –, sodass sein erwarteter Nettogewinn maximal wird:

$$\max E[\Pi(e) - w(\Pi(e))] \quad (5.5)$$

unter den Nebenbedingungen (5.3) und (5.4).

Die Lösung dieses Problems führt zu einem Second-Best-Optimum. Die Referenzlösung (First-Best) würde man erhalten, wenn das Anstrengungsniveau  $e$  beobachtbar wäre.<sup>61</sup>

Im Allgemeinen stellt die Lösung dieses Problems eine relativ komplexe Aufgabe dar. Daher wird im Folgenden die Nutzenfunktion des Agenten weiter spezifiziert. Es wird angenommen, dass der Agent risikoavers sei, sodass seine Nutzenfunktion  $U$  über  $w$  konkav ist. Weiterhin wird angenommen, dass man die Nutzenfunktion nach dem Nutzen im Einkommen  $u(w)$  und dem Nutzen der Anstrengungen  $\phi(e)$  zerlegen kann, d.h.  $U(w, e) = u(w) - \phi(e)$ . Für  $\phi(e)$  gelte, dass sie eine steigende Funktion von  $e$  sei ( $\phi'(e) > 0$ ).

Des Weiteren wird unterstellt, dass die Anstrengungen des Agenten in einem bestimmten Intervall liegen  $e \in [\underline{e}, \bar{e}]$ . Dabei bezeichne die Dichtefunktion  $f(\Pi, e)$  die Wahrscheinlichkeit, dass das Ergebnis  $\Pi$  bei einer bestimmten Anstrengung  $e$  des Agenten erreicht wird. Es soll gelten, dass die Wahrscheinlichkeit eines besseren Ergebnisses mit dem Anstrengungsniveau zunimmt. Dies entspricht der Beziehung der stochastischen Dominanz erster Ordnung über  $\Pi$ :

$$e_1 > e_2 \Rightarrow F(\Pi, e_1) < F(\Pi, e_2). \quad (5.6)$$


---

<sup>61</sup> Vgl. Tirole (1995).

<sup>62</sup> Vgl. hierzu auch Tirole (1995) S. 124.

Die Lösung des Problems aus Sicht des Prinzipals kann über zwei Stufen erfolgen.<sup>63</sup> Im ersten Schritt wird nach dem kostenminimalen Vertrag gesucht:

$$\begin{aligned} \min_w \int_{\Pi} w(\Pi) f(\Pi, e) d\Pi \\ \text{u.d.Nb.} \quad \int_{\Pi} (u(w(\Pi)) f(\Pi, e) - \phi(e)) d\Pi \geq U_0 \\ e \in \arg \max_{\tilde{e} \in E} \int_{\Pi} (u(\Pi(x)) f(\Pi, \tilde{e}) - \phi(\tilde{e})) d\Pi. \end{aligned} \quad (5.6)$$

Dies führt zu den Minimalkosten für bestimmte Anstrengungsniveaus:  $c(e)$ . Im zweiten Schritt wird der erwartete Profit des Prinzipals über alle implementierbaren Aufwandsniveaus maximiert:

$$\max_e s(e) \equiv \int_{\Pi} (\Pi f(\Pi, e) - c(e)) d\Pi, \quad (5.7)$$

wobei  $s(e)$  den Nettogewinn des Prinzipals bei kostenminimaler Vergütung unter Beachtung von Anreizverträglichkeit und Beteiligungsbereitschaft des Agenten bezeichnet.

Für die Existenz einer eindeutigen Lösung dieses Second-Best-Problems müssen verschiedene Annahmen erfüllt sein, die in den folgenden Punkten zusammengefasst sind.

- (i)  $u' > 0, \quad u'' < 0, \quad \phi_e > 0 \quad \forall e \in E \wedge \forall w \in W$
- (ii)  $E$  ist eine nicht leere Menge.
- (iii) Die Anzahl der Ergebnisse  $\Pi$  ist endlich.
- (iv)  $f(\Pi, e) > 0 \quad \forall \Pi, e$

Sind diese Annahmen erfüllt, so existiert ein optimales Vergütungsschema  $w^*$  und ein optimales Aufwandsniveau  $e^*$ .<sup>64</sup> Im Folgenden soll nur die explizite Lösung für  $w$  angegeben werden.<sup>65</sup>

Die Bedingung erster Ordnung zur Lösung des Problems (5.6) ist gegeben durch:

$$\frac{1}{u'(w(\Pi))} = \lambda_1 + \lambda_2 \frac{f_e(\Pi, e)}{f(\Pi, e)}, \quad (5.8)$$

wobei  $\lambda_1$  der mit der Bedingungsbedingung verbundene Lagrangemultiplikator und  $\lambda_2$  der mit der Anreizverträglichkeitsbedingung verbundene Lagrangemultiplikator ist. Ohne diese Multipli

---

<sup>63</sup> Die allgemeine Formulierung dieses Lösungsansatzes geht auf Grossman und Hart (1983) zurück. Vgl. für die hier gewählte Darstellung auch Thiele und Wambach (1999).

<sup>64</sup> Für einen Beweis vgl. Grossman und Hart (1983).

<sup>65</sup> Für die Herleitung des optimalen Anstrengungsniveaus vgl. wieder Grossman und Hart (1983).

katoren explizit zu bestimmen, wissen wir, dass diese in jedem Falle positiv sein müssen, da sonst eine oder beide Bedingungen für die Teilnahme des Agenten am Vertrag nicht erfüllt sein würden.  $u'$  ist die erste Ableitung der Nutzenfunktion nach der Vergütung und  $f_e$  die erste Ableitung der Dichtefunktion nach den Anstrengungen.

Aus der Bedingung erster Ordnung kann der Vertrag mit den optimalen Löhnen bestimmt werden. Invertierbarkeit von  $U_w$  vorausgesetzt, lässt sich dieser Vertrag schreiben als:

$$w^*(\Pi) = u'(U_w(\Pi))^{-1} \left( \lambda_1 + \lambda_2 \frac{f_e}{f} \right). \quad (5.9)$$

Nach (5.8) bzw. (5.9) umfassen anreizkompatible Vergütungsschemata positive „Löhne“  $w$ , die mit einem besseren Ergebnis steigen sollten. Dies folgt aus der Annahme, dass das Wahrscheinlichkeitsverhältnis  $f_e/f$  in  $\Pi$  steigen soll.<sup>66</sup> Zudem zeigt (5.9), dass  $w$  im Fall asymmetrischer Informationen höher sein muss als im Falle symmetrischer Informationsverteilung, da der (positive) Lagrangemultiplikator  $\lambda_2$  bei vollständiger Information wegfällt. Die Vergütung in der first-best-Welt wäre also in diesem Fall (risikoneutraler Prinzipal und risikoaverser Agent) konstant und nicht vom geleisteten Aufwand abhängig. Allerdings würde die Höhe des Fixbetrages vom Grad der Risikoaversion des Agenten abhängen.<sup>67</sup>

Neben der Bedingung erster Ordnung muss die optimale Lösung auch die Bedingung zweiter Ordnung des Agenten erfüllen. Wie Grossman und Hart (1983) sowie andere gezeigt haben, ist es dafür hinreichend, wenn das Wahrscheinlichkeitsverhältnis  $(f_e/f)$  eine zunehmende Funktion von  $\Pi$  und die Verteilungsfunktion konvex ist ( $F_{ee} \geq 0$ ).<sup>68</sup>

### 5.2.2 Grundmodell bei zweiseitig verborgener Handlung

Wie bereits erwähnt, kann das Problem der verborgenen Handlungen auch auf beiden Seiten auftreten (Double Moral Hazard).<sup>69</sup> In diesem Fall wäre der Prinzipal nicht nur passiv am Produktionsprozess beteiligt, sondern hätte selbst Einfluss auf das Ergebnis. Es bezeichne jetzt  $e_p$  die Anstrengungen des Prinzipals und  $e_A$  die des Agenten, um ein bestimmtes Ergebnis  $\Pi$

<sup>66</sup> Vgl. hierzu auch Milgrom (1981).

<sup>67</sup> Ein Beispiel kann das verdeutlichen: Sei die Nutzenfunktion des Agenten mit  $\ln(w)$  gegeben, so würde sich der Fixbetrag ergeben zu  $w = \lambda_1$ .

<sup>68</sup> Grossman und Hart (1983) beweisen in ihrem Artikel die obigen Aussagen. Vgl. auch Tirole (1995), S. 127.

<sup>69</sup> Vgl. hierzu und im Folgenden Kim und Wang (1998).



zu erreichen. Somit lässt sich eine gemeinsame Funktion der Anstrengungen von Prinzipal und Agent  $\theta(e_p, e_A)$  formulieren, für die folgende Annahmen gelten sollen  $\theta_{e_p} > 0$ ,  $\theta_{e_A} > 0$ ,  $\theta_{e_p e_p} < 0$ ,  $\theta_{e_A e_A} < 0$  und  $\theta_{e_p e_A} \geq 0$ . Die Wahrscheinlichkeit, ein bestimmtes Ergebnis zu erreichen, hängt demnach von den gemeinsamen Anstrengungen von Prinzipal und Agenten ab ( $f[\Pi, \theta(e_p, e_A)]$ ). Die Kosten der Anstrengungen seien für die beiden Parteien gegeben durch  $\phi(e_p)$  und  $\phi(e_A)$  mit jeweils positiver erster und zweiter Ableitung. Damit verändern sich die Zielfunktion und die Nebenbedingungen. Die Anreizkompatibilität erfordert jetzt auch für den Prinzipal, dass seine gewählten Anstrengungen den Nutzen (hier Gewinn) maximieren müssen:

$$\max_{e_p} \int [\Pi - w(\Pi)] f[\Pi, \theta(e_p, e_A)] - \phi(e_p). \quad (5.10)$$

Die Bedingung für den Agenten verändert sich entsprechend. Der Prinzipal muss nun wieder versuchen, kostenoptimale Verträge zu finden, die sowohl die Anreizkompatibilitätsbedingungen als auch die Beteiligungsbedingung des Agenten beachten.

Dieses Problem lässt sich dann entsprechend Gleichung (5.6) wie folgt schreiben:

$$\begin{aligned} \min_{w(\Pi)} \int w(\Pi) f[\Pi, \theta(e_p, e_A)] d\Pi + \phi(e_p) \\ \text{u.d.N.} \quad \int u(w(\Pi)) f[\Pi, \theta(e_p, e_A)] d\Pi - \phi(e_A) \geq U_0 \\ \theta_{e_A} \int u(w(\Pi)) f_{\theta}[\Pi, \theta(e_p, e_A)] d\Pi - \phi_{e_A}(e_A) \\ \theta_{e_p} \int [\Pi - w(\Pi)] f_{\theta}[\Pi, \theta(e_p, e_A)] d\Pi - \phi_{e_p}(e_p). \end{aligned} \quad (5.11)$$

Daraus lässt sich analog zur Ableitung bei einseitig verborgener Information der optimale Vertrag  $w^*(\Pi)$  bei zweiseitiger verborgener Information ableiten. Invertierbarkeit vorausgesetzt, berechnet sich dieser zu:

$$w^*(\Pi) = u'(w(\Pi))^{-1} \frac{\lambda_1 + \lambda_2 \theta_{e_A} (f_{\theta} / f)}{1 + \lambda_3 \theta_{e_p} (f_{\theta} / f)}. \quad (5.12)$$

Dabei stellt  $\lambda_3$  nun den Lagrangemultiplikator dar, der mit der Anreizbedingung des Prinzipalens verbunden ist.

Gleichung (5.12) ist eine Verallgemeinerung von Gleichung (5.9). Denn würde man in Gleichung (5.12)  $\lambda_3 = 0$  setzen, käme man wiederum auf den Ausdruck in Gleichung (5.9). Wenn der Prinzipal selbst Einfluss auf den Produktionsprozess nehmen kann und er seine eigenen

Bemühungen in seinem Nutzenkalkül beachtet, so wird dies Auswirkungen auf die Vergütung des Agenten haben.

### 5.3 Der Konflikt zwischen Arzt und Patient

Der Konflikt zwischen Arzt und Patient entspricht im Wesentlichen der Ausgestaltung des Grundproblems bei einseitig verborgener Information bzw. Handlung. Der Patient als Auftraggeber beauftragt den Arzt eine Behandlung durchzuführen, um den vom Patienten gewünschten Gesundheitszustand wiederherzustellen. Allerdings kann der Patient weder die Handlungen des Arztes genau beobachten, noch ist er in der Lage den Arzt selbst und somit seine Leistungsfähigkeit exakt zu bestimmen. Für den Arzt sei angenommen, dass er aufgrund seiner Fähigkeiten die Schwere der Krankheit genau erkennen und sie dann auch adäquat behandeln kann. Allerdings muss er nicht notwendigerweise diese Erkenntnisse seinem Prinzipal, dem Patienten, in vollem Umfang enthüllen.

Für die Darstellung der Arzt-Patienten-Beziehung muss das Grundmodell nur unwesentlich abgeändert werden. Als Erstes wird jetzt angenommen, dass auch der Prinzipal risikoavers sei und seinen Erwartungsnutzen maximiert. Von der Existenz einer Krankenversicherung wird zunächst abgesehen. Der Nutzen des Patienten soll von seinem Einkommen ( $y$ ), welches er vollständig für Konsumzwecke ausgibt, und dem Gesundheitszustand ( $b$ ) abhängen. Der Patient ist in der Lage, den Nutzen aller möglichen Gesundheitszustände in Abhängigkeit von der geleisteten Arbeit des Arztes monetär zu bewerten ( $b \rightarrow H(e)$ ). Allerdings kann er die Anstrengungen des Arztes nicht genau einschätzen, sodass die Beziehung zwischen ärztlichen Anstrengungen und Gesundheitszustand nicht deterministisch ist. Dieser Zusammenhang lässt sich vielmehr durch die Dichtefunktion  $f(H, e)$  beschreiben. Sie misst, wie sich die Anstrengungen des Arztes auf den Gesundheitszustand auswirken. Im Falle einer Behandlung erhält der Arzt einen Beitrag ( $w$ ). Dieser Beitrag hängt von den ärztlichen Anstrengungen zur Gesundung des Patienten ab,  $w(H(e))$ . Der erwartete Nutzen eines Patienten kann somit durch folgende Nutzenfunktion beschrieben werden:

$$U^P = \int_H \left[ u^P(H - w(H), y) f(H, e) \right] dH. \quad (5.13)$$

Der erwartete Nutzen des Patienten entspricht also der (wahrscheinlichkeits-) gewichteten Summe der Einzelnutzenwerte in allen möglichen Gesundheitszuständen.

Entsprechend lässt sich auch die Nutzenfunktion des Arztes formulieren, dessen einzige Einkommensquelle die Vergütung durch den Patienten sein soll:

$$U^A = \int_H [u^A(w(H)) - \phi(e)] f(H, e) dH, \quad (5.14)$$

d.h., dass der erwartete Nutzen des Arztes in diesem einfachen Modell einzig vom Nettolohn (Vergütung minus Anstrengungen) abhängt.

Das Problem lässt sich analog zum Grundproblem lösen, sodass sich die Bedingung erster Ordnung unter Beachtung der neuen Nutzenfunktion des Prinzipals entsprechend der Lösung von (5.6) sofort hinschreiben lässt:

$$\frac{u_w^P}{u_w^A} = \lambda_1 + \lambda_2 \frac{f_e}{f}. \quad (5.15)$$

Die linke Seite von (5.15) beschreibt das Grenznutzenverhältnis zwischen Arzt und Patient bei optimaler Vergütung. Die rechte Seite gibt hingegen die Bestimmungsgründe für diese optimale Vergütung an. Sie hat zwei Teile, wobei ein Teil der Vergütung ein Fixbetrag sein sollte, der den Arzt zur Annahme des Vertrages bewegen soll. Diese Interpretation folgt aus der Herkunft des Lagrange-Multiplikators  $\lambda_1$ . Er ist mit der Teilnahmebedingung verbunden und gibt den marginalen Effekt einer Änderung dieser Bedingung auf die Zielfunktion an. Je größer also  $\lambda_1$ , desto wichtiger ist die Teilnahme des Arztes am Vertrag. Insbesondere würde das vor allem dann der Fall sein, wenn ein Nichtzustandekommen des Vertrags erhebliche gesundheitliche Konsequenzen für den Patienten hätte (Zweifel 1994, S. 15).

Der zweite Lagrange-Multiplikator  $\lambda_2$  ist mit der Anreizverträglichkeitsbedingung verknüpft. Er wird umso größer sein, je mehr der Patient höhere Anstrengungen des Arztes wünscht. Von besonderer Bedeutung ist der Term  $f_e$ , der misst, wie stark sich die Bemühungen des Arztes auf die Wahrscheinlichkeit eines günstigen Ergebnisses für den Patienten auswirken. Je stärker die Bemühungen des Arztes auf das Ergebnis Einfluss haben, desto höher müsste auch das Honorar für den Arzt ausfallen. Allerdings ist der ärztliche Beitrag durch den Patienten nur schwer zu ermitteln. Denn der Patient ist selten in der Lage, seinen Gesundheitszustand vor wie

auch nach der Behandlung exakt einzuschätzen. Hinzu käme noch, dass er auch den ärztlichen Anteil am Heilungserfolg abschätzen müsste.<sup>70</sup>

Im allgemeinen Fall ohne Versicherung wählt der Patient damit eine Vergütung des Arztes, die zum einen aus einem Fix-Betrag und zum anderen aus einem leistungsabhängigen und/oder gesundheitszustandsabhängigen Betrag besteht. Die funktionale Form der Vergütungsfunktion hängt dabei im entscheidenden Maße vom Grad der Risikoaversion beider Parteien ab.

Erfüllt  $w(e^*; H)$  die obige Bedingung, so ist der Vertrag sowohl anreizverträglich aus Sicht des Arztes, als auch kostenoptimal aus Sicht des Patienten.

Wie würde der Vertrag aussehen, wenn wir Moral Hazard auch beim Prinzipalen zulassen würden? Wobei Moral Hazard jetzt eine für den Arzt verborgene Handlung des Patienten meint, die die Gesundheit des Patienten beeinflussen kann. Solche Handlungen können bspw. vorbeugende Maßnahmen sein, die eine schnelle Gesundung befördern können. Die für den Arzt nicht beobachtbaren Anstrengungen des Patienten lassen sich in gleicher Weise wie im Grundmodell mit zweiseitiger verborgener Handlung modellieren (vgl. Gleichungen (5.10) - (5.12)), nur dass wir hier wieder Risikoaversion auch des Prinzipalen unterstellen.

Benutzen wir die Spezifikationen für die Arzt-Patientenbeziehung, kann die Bedingung erster Ordnung dieses modifizierten Maximierungsproblems für die optimale Vergütung gemäss Gleichung (5.12) wie folgt geschrieben werden:

$$\frac{u_w^P}{u_w^A} = \frac{\lambda_1 + \lambda_2 \theta_A \left( \frac{f_\theta}{f} \right)}{1 + \lambda_3 \theta_P \left( \frac{f_\theta}{f} \right)}. \quad (5.16)$$

Würde der Patient seine eigenen Anstrengungen in seinem Nutzenkalkül mitberücksichtigen, so hätte das auch Auswirkungen auf die Honorierung des Arztes. Nach (5.16) würde eine Erhöhung der Compliance des Patienten den Wert des Quotienten auf der rechten Seite der Gleichung verringern. Unter sonst gleichen Umständen müsste dann auch das Grenznutzenverhältnis sinken, was nur durch eine Verringerung der Vergütung  $w$  erfüllt werden kann. D.h., dass eine bessere Compliance des Patienten nicht nur zu einer verbesserten Gesundheit führt,

---

<sup>70</sup> Vgl. hierzu bspw. auch Schneider (2002), der das Modell von Zweifel (1994) ausführlich darstellt.

sondern auch den Einkommensnutzen – über die geringere Vergütung für ärztliche Leistungen – erhöhen kann.

Aus Sicht des Arztes führt die bessere Compliance des Patienten (unter sonst gleichen Bedingungen) damit zu Einkommenseinbußen. Ist der Arzt hingegen in der Lage, bei gegebener Vergütungsfunktion seine Anstrengungen über das optimale Maß hinaus auszudehnen, könnte es sein, dass er den Compliance-Effekt auffangen kann. Je stärker die Vergütung auf die Anstrengungen des Arztes bezogen ist, desto stärker dürfte die Versuchung des Arztes sein, sich in diesem Sinne zu verhalten.

Ein interessanter Aspekt soll abschließend kurz diskutiert werden. Hierbei geht es um den Zusammenhang von gegenwärtiger Vergütung und dem Vermögen eines Agenten, da vermutet werden kann, dass Agenten in unterschiedlichen Vermögenssituationen auch unterschiedlich auf bestehende Vergütungsregeln reagieren. In einer Erweiterung des Standardmodells (5.1) - (5.7) zeigen Thiele und Wambach (1999), wie das Anfangsvermögen  $a$  des Agenten im Grundmodell mit berücksichtigt werden kann. Damit besitzt die Nutzenfunktion des Agenten jetzt mit dem Vermögen ein weiteres Argument. Im Endeffekt sind somit auch die optimale Vergütungsfunktion  $c(a, e)$  und die Gewinnfunktion  $s(a, e)$  zusätzlich vom Vermögen des Agenten abhängig. Thiele und Wambach können unter bestimmten Annahmen zeigen, dass die erwarteten Auszahlungen des Prinzipals negativ mit dem Vermögen des Agenten korreliert sind,

$$\left( \frac{\partial c(a, e)}{\partial a} > 0 \right).$$

Eine höhere Vermögensausstattung des Agenten erhöht die Kosten der Implementierung eines Vertrags und schmälert somit den Gewinn des Prinzipals.

Eine wichtige Annahme hierfür ist neben der Separierbarkeit der Nutzenfunktion vor allem die, dass der Agent relativ „unbesonnen“ ist. Insbesondere unbesonnen im Vergleich zur Risikoaversion. Technisch ausgedrückt bedeutet dies, dass die absolute „Besonnenheit“ kleiner sein muss als das Dreifache der absoluten Risikoaversion. Diese Bedingung ist bei Nutzenfunktionen mit konstanter absoluter Risikoaversion oder aber bei Nutzenfunktionen mit relativer Risikoaversion von mindestens 0.5 erfüllt.<sup>71</sup>

---

<sup>71</sup> Ein Beispiel für eine Nutzenfunktion mit konstanter absoluter Risikoaversion ist gegeben durch  $U(x) = -\exp(-k x)$ , wobei  $k$  eine Konstante ist. Eine Nutzenfunktion mit relativer Risikoaversion von

Mit anderen Worten hängt demnach die Wirkung eines bestimmten Vertrages zwischen Prinzipal und Agent auf das Ergebnis entscheidend von der Risikoaversion und dem Vermögen des Agenten ab. Ist – wie im angenommenen Fall – die Besonnenheit wesentlich kleiner als die Risikoaversion, dann steigen für den Prinzipal mit zunehmendem „Reichtum“ des Agenten die Kosten der Implementierung eines bestimmten Anstrengungsniveaus.

Gerade bei der Vergütung von Ärzten kann obiges Ergebnis wichtig sein, denn über die Lebenszeit betrachtet, dürfte das Einkommen von Ärzten sehr stark variieren. Am Anfang steht eine hohe finanzielle Belastung (Ausbildungskosten, Kosten für Praxisausstattung oder für die Praxisübernahme). Diese hohen Kosten müssen in der Regel durch Kredite finanziert werden. Gleichzeitig muss ein junger Arzt viel stärker um einen eigenen Patientenstamm kämpfen als Ärzte, die schon jahrelang erfolgreich praktizieren. Dies gilt grundsätzlich auch, wenn ein junger Arzt eine bestehende Praxis übernimmt, denn durch den Arztwechsel wird sicherlich die Wechselbereitschaft der Patienten steigen. Zudem muss er in der Regel seinen Vorgänger für die Vorteile eines bestehenden Patientenstamms finanziell kompensieren. Ältere Ärzte sehen sich dagegen einer anderen Einkommenssituation gegenüber. Sie haben ihre Kredite abgezahlt und verfügen in der Regel über einen gesicherten Patientenstamm. Somit nimmt auch die Bedeutung des Einkommens aus der Arztpraxis relativ zu anderen Einkommensarten ab, sodass gleiche Vergütungsregeln andere Anreize setzen können als bei jüngeren Ärzten.

### 5.4 Der Konflikt zwischen Krankenkasse und Arzt

Die Krankenkasse als zusätzlicher Sachwalter der Patienten ist bestrebt, die Patienteninteressen gegenüber den Ärzten durchzusetzen. Wie im allgemeinen Modell der Arzt-Patienten-Beziehung deutlich wurde, hängt die optimale Vergütung der Ärzte entscheidend davon ab, den Beitrag des Arztes zum Behandlungserfolg einzuschätzen. Da eine solche Einschätzung dem einzelnen Patienten in der Regel nicht möglich ist, besteht die Hoffnung, dass die Krankenkasse oder ein anderer zusätzlicher Sachwalter die Informationsasymmetrien verringern kann.<sup>72</sup>

Durch die bestehenden Krankenversicherungsregelungen greift die Krankenkasse damit als dritter Akteur in entscheidendem Maße in die Produktion von Gesundheitsleistungen ein. Die

---

mindestens 0.5 lässt sich bspw. durch die folgende Nutzenfunktion ausdrücken:  $U(x) = x^p$  mit  $p \leq 0.5$ . Vgl. zu Risikoaversion und Besonnenheit auch die Ausführungen im Abschnitt 5.1.

<sup>72</sup> In den USA übernehmen bspw. die Arbeitgeber häufig die Aufgabe eines zusätzlichen Sachwalters.

Krankenkasse versichert ihr Versichertenkollektiv gegen die hohen Kosten möglicher Erkrankungen. Als Sachwalter ihrer Versicherten trägt sie Verantwortung dafür, dass ihre Versicherten mit der optimalen Menge an Gesundheitsleistungen in der von den Versicherten gewünschten Qualität versorgt werden. Andererseits übernimmt vom Standpunkt der Krankenkasse – oder aber auch eines sozialen Planers, der die gesamtwirtschaftliche Effizienz im Auge hat – die Rolle des Agenten jetzt der Patient. Die Krankenkasse als Vertreter des gesamten Versichertenkollektivs ist in diesem Fall der Prinzipal, der die Aufgabe hat, alle Versicherte optimal zu versorgen. Hieraus ergibt sich das schon erwähnte Problem einer Informationsasymmetrie auf der Nachfrageseite. Damit spielt im Verhältnis zwischen Krankenkasse und Versichertem die nachfrage-seitige Moral-Hazard-Problematik eine entscheidende Rolle. Im Verhältnis zwischen Krankenkasse und Arzt besitzen die Anreizmöglichkeiten gegenüber den Ärzten bezüglich der optimalen Versorgung der Patienten eine besondere Bedeutung. Die Krankenkasse als Vertreter der Patienten kauft Leistungen bei den Ärzten ein. Allerdings entscheidet in der Regel der Patient über den Leistungsumfang – die Kasse hingegen finanziert diese Leistungen.<sup>73</sup> Die entscheidende Frage in der Arzt-Krankenkassen-Beziehung ist daher die nach der richtigen Vergütung der Ärzte.

Nach den Ergebnissen im letzten Abschnitt würde der Patient eine Vergütung seines Arztes wählen, die aus zwei Komponenten bestünde: einem Pauschalanteil und einem „Leistungs“- bzw. Aufwandsanteil. Die Zusammensetzung der beiden Teile hinge im starken Maße von der Schwere der Krankheit des Patienten ab. Was kann die Krankenkasse aus dieser Übung lernen? Zunächst einmal nicht sehr viel, denn sie ist nicht in der Lage, den Gesundheitszustand des Patienten zu kennen, sodass sie überhaupt nicht im Stande wäre, auf Grundlage des Gesundheitszustands des Patienten eine Vergütung der Ärzte festzulegen. Während der Patient zumindest noch eine Ahnung von der Schwere seiner Krankheit hat, ist die Krankenkasse voll und ganz auf die Information Dritter angewiesen. D.h. sie muss sich mit dem begnügen, was sich beobachten lässt: den gestellten Diagnosen, den Verschreibungen von Medikamenten, den Überweisungen in Krankenhäuser oder zu Spezialisten sowie den direkten Behandlungen der

---

<sup>73</sup> Dies trifft vor allem für die Einzelleistungsvergütung zu, wenn keine weiteren Ausgabenbegrenzungen vereinbart werden. Auf der Nachfrageseite soll – wie erwähnt – durch Kostenbeteiligungen der Patienten einer nichtoptimalen Leistungsausweitung entgegengewirkt werden.

Patienten. Voraussetzung hierfür ist natürlich, dass die Krankenkasse in der Tat solche Informationen auch bekommt.<sup>74</sup>

Das folgende Modell von Selden (1990) zeigt, wie eine optimale Vergütung von Leistungserbringern aussehen müsste, wenn die Krankenkasse die Interessen ihrer Patienten gegenüber den Ärzten optimal vertritt. Hierfür wird ein Versicherter betrachtet, der vor Eintritt einer Krankheit eine Krankenkasse wählt.<sup>75</sup> Durch die Kasse wird ihm *ex ante* ein Vertrag angeboten, der sowohl die Versicherung gegen Krankheiten spezifiziert als auch die Vergütung des Leistungserbringers. Der Erwartungsnutzen eines potenziellen Patienten hängt vom Konsum ( $k$ ) und der Gesundheit ab. Der Nutzen der Gesundheit ist jetzt allerdings eine Funktion vom so genannten erreichbaren Gesundheitszustand  $G$ , welcher zum einen von den ärztlichen Anstrengungen ( $e$ ) beeinflusst wird als auch vom Gesundheitszustand *vor* der ärztlichen Behandlung ( $H$ ). Der Gesundheitszustand ist bei Vertragsabschluss nicht bekannt und wird daher mit der Dichtefunktion  $f(H)$  bewertet. Während  $H$  im vorangegangenen Kapitel den Gesundheitszustand *nach* der Behandlung gemessen hatte, misst  $H$  nun also den *ex ante* Gesundheitszustand des Patienten.

Zunächst wird als Referenzfall eine Situation untersucht, in welcher der Versicherte entsprechend seines Gesundheitszustands den Behandlungsaufwand selbst festlegt. In diesem Fall spielt demnach die Entscheidung des Arztes keine Rolle. Der Erwartungsnutzen des Versicherten lässt sich schreiben als

$$U^P = \int_H u^P(G(e, H), k(H)) f(H) dH \quad (5.17)$$

mit  $k(H) = y - P(H)$ , wobei  $y$  das exogene Einkommen des Versicherten bezeichnet und  $P$  die Versicherungsprämie.

In einer erstbesten Welt würde der Versicherte seinen Nutzen maximieren, indem er die Anstrengungen des Arztes selbst wählt und sein Einkommen entsprechend auf medizinische und andere Verwendungen aufteilt. Diese Entscheidungen muss der Versicherte vor Eintritt einer Krankheit treffen. Er besitzt somit zum Zeitpunkt der Entscheidung einen Gesundheitszustand  $H$ .

---

<sup>74</sup> Vgl. zur Datenlage in der Schweiz im Allgemeinen und zur Datengrundlage dieser Untersuchung im Besonderen Kapitel 7.

<sup>75</sup> Vgl. zu diesem Modell auch Breyer und Zweifel (1999, 8. Kapitel).



Sein Maximierungsproblem kann daher geschrieben werden als:

$$\begin{aligned} \max_{e, k, P} \int_H u^P [G(e(H), H), k(H)] f(H) dH \\ \text{u.d.N. } y - P(H) - k(H) \geq 0 \quad \forall H \\ \int_H [P(H) - w e(H)] f(H) dH \geq 0, \end{aligned} \quad (5.18)$$

wobei die zweite Nebenbedingung die Restriktion für die Krankenkasse darstellt, mit  $w$  dem Preis für die ärztliche Leistung.

Die Bedingungen erster Ordnung lauten somit:

$$\begin{aligned} \frac{\partial u^P}{\partial G} \frac{\partial G}{\partial e} &= \lambda^* w \\ \frac{\partial u^P}{\partial k} &= \gamma^* \\ \gamma^* &= \lambda^*. \end{aligned} \quad (5.19)$$

Dabei bezeichnet  $\gamma$  den Lagrange-Multiplikator der ersten Nebenbedingung und  $\lambda$  denjenigen der zweiten Nebenbedingung. Aufgrund der dritten notwendigen Bedingung lässt sich  $\lambda$  auch als Grenznutzen des Einkommens auffassen. Gemäß (5.19) wird der Konsument also einen Vertrag wählen, sodass einerseits der Grenznutzen des Konsums mit dem Grenznutzen des Einkommens im Einklang steht (zweite Bedingung in (5.19)) und andererseits die mit dem Grenznutzen des Einkommens gewichtete Vergütung gerade dem Produkt aus Grenznutzen der Gesundheit und der Veränderung der Gesundheit durch den medizinischen Aufwand entspricht (erste Bedingung in (5.19)). Diese Bedingung drückt somit die Beziehung zwischen dem Einfluss des medizinischen Aufwands auf den Nutzen des Patienten und den Kosten dieses Aufwands aus. Aufgrund der positiven Ableitungen auf der linken Seite von Bedingung 1 in (5.19) sollten höhere Anstrengungen des Arztes auch höher entlohnt werden.

Allerdings ist der Versicherte in einer unsicheren Umwelt nicht in der Lage, einen solchen Vertrag zu wählen: Er kann den Behandlungsaufwand  $e$  nicht *ex ante* festlegen und auf dieser Entscheidung einen Vertrag gründen. Vielmehr obliegt es der Krankenkasse, durch einen geeigneten Vertrag (Vergütung) mit dem Leistungserbringer, die „richtige“ medizinische Behandlung zu veranlassen. Für den Versicherten bedeutet dies, dass er jetzt eine Prämie in Abhängigkeit der vom Arzt gewählten Anstrengungen zahlen muss:  $P = P(e(H))$ .

Die Nutzenfunktion des Versicherten bekommt daher folgendes Aussehen:

$$U^P(G(e, H), y - P(e)). \quad (5.20)$$

Es sei des Weiteren angenommen, dass die Krankenkasse bestrebt ist, Verluste zu vermeiden. Zur Vergütung der Ärzte für ihre geleisteten Anstrengungen ( $e$ ) stehen der Krankenkasse zwei unterschiedliche Vergütungsformen zur Verfügung: einerseits eine Pauschale pro Patient ( $a$ ) und andererseits die Vergütung der Einzelleistungen über eine Vergütungsfunktion  $EZ(e)$  in Abhängigkeit der Anstrengungen. Der Nettoertrag des Arztes ist damit:

$$\Pi(e) = a + EZ(e) - we, \quad (5.21)$$

wobei  $w$  den Preis für die Anstrengungen des Arztes bezeichnet. Nimmt man einen Markt mit vollständiger Konkurrenz für die Versicherungsleistungen an, so muss für die Versicherung gelten, dass die erwarteten Einnahmen gerade die erwarteten Kosten decken, d.h. die Prämie berechnet sich zu:

$$P(e) = \int_H [a + EZ(e(H))] f(H) dH. \quad (5.22)$$

Da durch die Einschaltung der Krankenkasse die Vergütung des Arztes nicht mehr direkt mit dem Gesundheitszustand des Versicherten verbunden ist (vgl. die Nutzenfunktionen (5.23) mit (5.14)), wird über die Gestalt der Nutzenfunktion des Arztes die nach wie vor bestehende Sachwalterbeziehung zwischen Arzt und Patient berücksichtigt. Es wird angenommen, dass ein höherer Nutzen aus einer besseren Gesundheit einen positiven Effekt auf das erwartete Nutzenniveau des Arztes hat. Der Arzt maximiert demnach seinen erwarteten Nutzen unter Berücksichtigung der Nettoertragsrestriktion (5.21) sowie unter Berücksichtigung des Patientennutzens. Der Einfachheit halber wird von einem risikoneutralen Agenten ausgegangen:

$$U^A = \Pi(e) + \alpha U^P(G(e, H), k(e)). \quad (5.23)$$

Der Parameter  $\alpha$  lässt sich dabei als Grad der Sachwalterrolle des Arztes für den Patienten auffassen. Ein perfekter Sachwalter der Interessen des Patienten hätte damit ein  $\alpha = 1$ .

Aus den bisherigen Bausteinen lässt sich nun eine optimale Vergütung herleiten. Nach wie vor geht es darum, den Nutzen des Versicherten zu maximieren. Durch die Beteiligung der Krankenkasse ändern sich jetzt aber die Nebenbedingungen. So muss zum einen sichergestellt sein, dass die Krankenkasse mindestens auf ihre Kosten kommt und zum anderen müssen wieder Anreizbedingung und Teilnahmebedingung des Arztes erfüllt sein.

Aus Sicht der Krankenkasse, die den Nutzen ihrer Versicherten maximieren will, ergibt sich somit folgendes Maximierungsproblem:

$$\begin{aligned}
 & \max_{\Pi(\cdot), P} \int_H^P (G(e, H), y - P(e)) f(H) dH \\
 & \text{u.d.N.} \quad \int_H [a + EZ(e(H)) - P(e(H))] f(H) dH = 0 \\
 & e^* = \arg \max_H \int \Pi(e) + \alpha U^P(G(e, H), y - P(e)) f(H) dH \\
 & \int_H \Pi(e(H)) f(H) dH = 0.
 \end{aligned} \tag{5.24}$$

Bei der Lösung des Maximierungsproblems ist zu beachten, dass der Beitragssatz  $P$  bei der Bestimmung des optimalen Anstrengungsniveaus in der Anreizverträglichkeitsbedingung für den Arzt ein Datum ist (vgl. (5.22)).

Die Lösung dieses Maximierungsproblems hinsichtlich der optimalen Vergütungsfunktion  $EZ(e)$  führt unter Verwendung der Ergebnisse aus (5.19) zu folgender Optimalitätsbedingung:

$$\frac{\partial EZ(e)^*}{\partial e} - w + \alpha \lambda w = 0. \tag{5.25}$$

Durch Integration von (5.25) erhält man schließlich die optimale Vergütungsfunktion:

$$EZ^*(e) = w e (1 - \alpha \lambda). \tag{5.26}$$

Die optimale Pauschale ergibt sich dann direkt durch die Verbindung der Ertragsgleichung des Arztes mit der Teilnahmebedingung des Arztes:

$$a^* = \int_H [w e^*(H) - EZ^*(e^*(H))] f(H) dH. \tag{5.27}$$

Nach (5.26) kann nur eine gemischte Vergütung optimal sein, wenn von einer berufsethischen Einstellung des Arztes ausgegangen wird ( $\alpha > 0$ ). Der Anteil der Kostenerstattung sollte dabei mit einer stärkeren Ausprägung seiner Ethik abnehmen.

Um die Höhe der Kostenerstattung zu bestimmen, müsste der Versicherer allerdings in der Lage sein, sowohl  $\alpha$  als auch  $\lambda$  zu bestimmen. Das bedeutet einerseits, dass der Versicherer die *Einstellung* zum Behandlungserfolg der Leistungsanbieter beurteilen muss. Andererseits müsste er auch den Grenznutzenbeitrag von Gesundheit für den Versicherten und die Behandlungswirkung auf die Gesundheit (die Grenzproduktivität ärztlicher Leistung) bestimmen können (vgl.

(5.19)). Diese Grenzproduktivität ärztlicher Leistung ist daher mit  $f_e$  in Gleichung (5.15) im Grundmodell der Arzt-Patienten-Beziehung vergleichbar.

### 5.5 Der Arzt als Doppelagent

Bisher wurden die Aufgaben des Arztes als Agent der Krankenkassen und der Patienten mehr oder weniger getrennt voneinander betrachtet. Da zwischen diesen Beziehungen Interdependenzen bestehen, soll in diesem Abschnitt versucht werden, die Sachwalterrolle des Arztes für den Patienten wie auch seine Beziehung zur Krankenkasse gemeinsam darzustellen.

Betrachtet man den Arzt sowohl als Agenten der Patienten als auch der Krankenkassen, so lassen sich die bestehenden Vergütungssysteme als Ausdruck verschieden starker Ausprägungen der jeweiligen Sachwalterrolle auffassen. Nach dieser Lesart führt die Einzelleistungsvergütung dazu, dass der Arzt einem perfekten Agenten für die Patienten gleicht. Da der Arzt alle erbrachten Leistungen vergütet bekommt, hat er keinen Anreiz, dem Patienten irgendwelche Wünsche abzuschlagen. Bei Kopfpauschalen lässt sich genau das Gegenteil aussagen: Da der Arzt einen festen Betrag für jeden Patienten erhält, wird er nur das Allernotwendigste zur Gesundheit des Patienten unternehmen und kann in diesem Fall als perfekter Agent für die Krankenkasse bezeichnet werden. In diesem Sinne lässt sich die Rolle des Arztes als Vertreter des Patienten und als Vertreter der Krankenkasse als „Doppelagent“ bezeichnen.<sup>76</sup>

Zur genaueren Untersuchung dieser Wechselwirkungen wird ein Modell betrachtet, welches die unterschiedlichen Beziehungen zwischen den Akteuren abbilden soll.<sup>77</sup> Das Modell lässt sich als Spielart des im Abschnitt 5.3 dargestellten allgemeinen Modells der Arzt-Patienten-Beziehung auffassen – jetzt unter Berücksichtigung der Krankenkasse. Allerdings verzichtet es auf die explizite Ausgestaltung einer Nutzenfunktion für den Arzt und somit auch auf die explizite Herleitung einer Vergütungsfunktion wie dies im Abschnitt 5.4 geschah. Vielmehr wird hier untersucht, welche Auswirkungen Verhaltensänderungen des Arztes – der wiederum auf das Vergütungsschema der Krankenkasse reagiert – auf das Nutzenniveau des Patienten haben.

Zunächst soll das Grundmodell am Beispiel vollständiger Versicherungsdeckung betrachtet werden. Dieser Fall dient gleichzeitig als Referenzlösung.

---

<sup>76</sup> Vgl. Blomquist (1991).

<sup>77</sup> Vgl. zu diesem Modell Blomquist (1991).

**a) Vollständige Information**

Der Erwartungsnutzen des Versicherten sei wiederum separierbar. Speziell wird angenommen, dass sich die Nutzenfunktion aus einer Gesundheitsfunktion  $b_i$  und dem Nutzen aus einem Konsumgut  $u(k_i)$  additiv zusammensetzt. Es gibt  $i = 0, 1, \dots, n$  Gesundheitszustände, wobei jeder Zustand mit einer Wahrscheinlichkeit  $\pi_i$  eintreten kann. Der Erwartungswert des Nutzens lässt sich demnach schreiben als:

$$E = \sum_{i=0}^n \pi_i (b_i + u(k_i)). \quad (5.28)$$

Wir nehmen an, dass der Nutzen aus dem Konsumgut mit dem Konsum einer zusätzlichen Einheit steigt, der Anstieg mit zunehmendem Konsum aber immer geringer ausfallen soll.<sup>78</sup> Der Wert der Gesundheitsfunktion wird durch folgende Gleichungen bestimmt:

$$\begin{aligned} b_0 &= \bar{b} \\ b_i &= b(x_i - \theta_i) < \bar{b} \quad i = 1, \dots, n. \end{aligned} \quad (5.29)$$

Dabei ist  $\theta_i$  ein Parameter, der mit dem  $i$ -ten Gesundheitszustand verknüpft ist und  $x_i$  bezeichnet den Umfang der im Gesundheitszustand  $i$  konsumierten Leistung. Wobei  $i = 0$  denjenigen Zustand kennzeichnet, in dem der Versicherte den besten Gesundheitszustand hat und demzufolge keine medizinischen Leistungen benötigt ( $x_0 = 0$ ). Es gelte vereinfachend, dass im Krankheitsfall die Differenz zwischen einem bestimmten Gesundheitszustand  $i$  und dem nächst besseren Gesundheitszustand eine positive Konstante sei:

$$\theta_i - \theta_{i-1} = \delta > 0 \quad i = 2, \dots, n. \quad (5.30)$$

Für die Gesundheitsfunktion soll das Gleiche wie für den Konsumnutzen gelten (steigender Nutzen und abnehmender Grenznutzen).<sup>79</sup> D.h. für die Gesundheitsfunktion, dass es – in Abhängigkeit vom Gesundheitszustand – für das Individuum besser ist, wenn er mehr medizinische Leistungen konsumiert. Dabei soll gelten, dass ein schlechterer Gesundheitszustand auch mit einer stärkeren Nachfrage nach medizinischen Leistungen verbunden ist. Es wird angenommen, dass das Individuum die Form seiner Gesundheitsfunktion kennt. Entsprechend dieser Annahmen weiß der Patient, dass unterschiedliche Gesundheitszustände mit unterschied

---

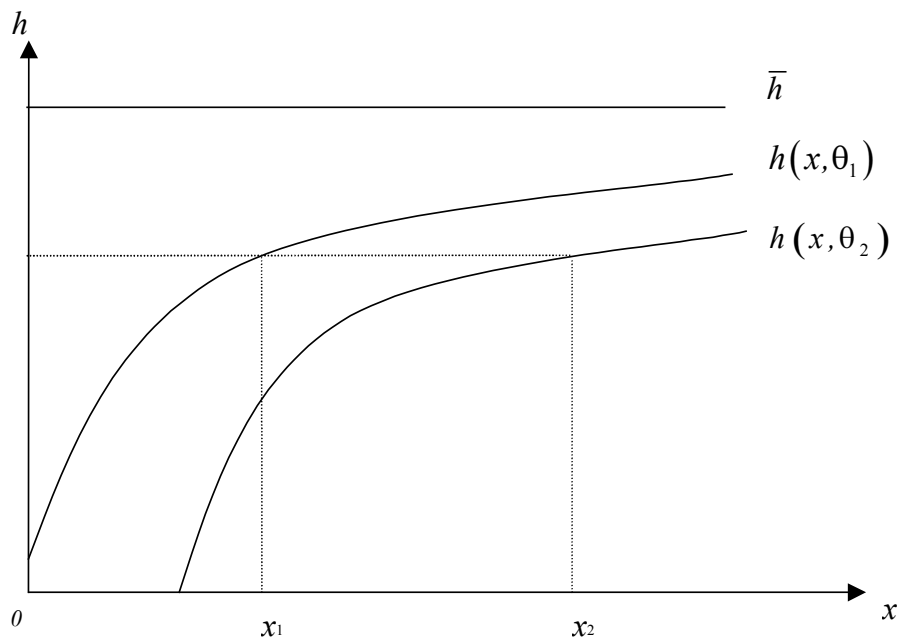
<sup>78</sup> Das heißt:  $u' > 0$  und  $u'' \leq 0$ , wobei die hochgestellten Kommata wiederum die entsprechenden Ableitungen der Nutzenfunktion bezeichnen.

<sup>79</sup> Also auch wieder  $b' > 0$  und  $b'' \leq 0$ .

lichen Nutzenniveaus verbunden sind und dass er in jedem Krankheitszustand den Nutzen durch den Konsum weiterer Leistungen steigern kann, wenn auch mit abnehmender Rate.

Für eine grafische Repräsentation soll aus Vereinfachungsgründen von einer Gesundheitsfunktion mit nur drei Zuständen ausgegangen werden (vollständig gesund, krank, schwerkrank), sodass  $n = 2$  angenommen wird. Diese Gesundheitsfunktion wird in Abbildung 2 dargestellt.

Abbildung 2 Gesundheitsfunktion mit 3 Gesundheitszuständen



In Abbildung 2 ist zu erkennen, dass ein Patient bei gleicher Inanspruchnahme von Gesundheitsleistungen in Abhängigkeit seines aktuellen Gesundheitszustandes unterschiedliche (Gesundheits)-Nutzenniveaus erreicht. Außerdem erkennt man, dass bei einem schlechteren Gesundheitszustand mehr Leistungen nachgefragt werden, um ein bestimmtes Nutzenniveau zu erreichen. Der beste Gesundheitszustand ist demgegenüber unabhängig von der Gesundheitsleistung, da der Versicherte keinerlei Leistungen nachfragen muss.

Im Falle vollständiger Information kann der tatsächliche Gesundheitszustand des Patienten kostenlos vom Patienten selbst, vom Arzt und von der Krankenkasse beobachtet werden. Existiert auf dem Versicherungsmarkt vollkommener Wettbewerb und werden den Versicherten Verträge mit fairen Prämien<sup>80</sup> angeboten, so geht es für den Patienten darum, ein optimales

---

<sup>80</sup> Unter einer fairen Prämie versteht man eine Prämie ohne Kostenaufschlag, vgl. bspw. Breyer et al. (2003).

Bündel an medizinischen Leistungen in Abhängigkeit des jeweiligen Gesundheitszustandes zu wählen. Es gilt also, einen Vertrag mit einer Prämie  $P$  zu finden, der den Versicherern einen Gewinn von Null sichert,<sup>81</sup> sowie ein Leistungspaket  $x_i = x(\theta_i)$  umfasst, welches den Erwartungsnutzen des Versicherten maximiert. Der Vertrag muss demnach folgendes Maximierungsproblem lösen:

$$\begin{aligned} \max_{P, \{x_i\}} E^0 &= \sum_{i=0}^n [\pi_i b_i(x_i - \theta_i)] + u(k - P) \\ \text{u.d.N. } P - \sum_{i=0}^n \pi_i x_i &= 0. \end{aligned} \tag{5.31}$$

Die Bedingungen erster Ordnung lauten dann:

$$\begin{aligned} x_0 &= 0 \\ b'(x_i - \theta_i) - u'(k - P) &= 0. \end{aligned} \tag{5.32}$$

Das Individuum trifft in jedem Gesundheitszustand eine Abwägung zwischen Gesundheitsnutzen und Konsumnutzen. Im Optimum wird der Versicherte gerade so viel Gesundheitsleistungen konsumieren, dass der zustandsabhängige Grenznutzen dem Grenznutzen aus dem Konsum entspricht. Die Lösung bei vollständiger Information zeigt, dass ein Individuum in dieser Situation immer vollständige Versicherung bevorzugt. In (5.32) wird das durch den in jedem Gesundheitszustand gleichen Grenznutzen im Konsum deutlich.<sup>82</sup>

### b) Einseitige Asymmetrie

Der Arzt sei jetzt ein perfekter Agent gegenüber dem Patienten, jedoch kein perfekter gegenüber den Krankenkassen. Ein solches Szenario entspricht der reinen Einzelleistungsvergütung, weil der Arzt in diesem System jede Leistung erstattet bekommt und somit bestrebt sein wird, jeden Wunsch des Patienten zu erfüllen. Nimmt man ferner an, dass der Patient nur weiß, ob er krank ist oder nicht, aber die Schwere der Erkrankung nicht kennt, so wird es im Interesse des Patienten sein, wenn der Arzt die Schwere der Krankheit „überschätzt“. In einem solchen System wäre dann  $x$  nach oben nicht begrenzt und es würde zu keinem Versicherungsvertrag kommen können. Um die Wahl von  $x_i$  zu begrenzen, ist es daher von Versicherungsseite

---

<sup>81</sup> Diese Bedingung bedeutet, dass auf dem Versicherungsmarkt vollständige Konkurrenz herrscht.

<sup>82</sup> D.h., dass er in jedem Fall eine zusätzliche Einheit Konsum realisieren will, egal wie schwer die Krankheit ist. Dies ist nur bei einer Vollversicherung möglich. Denn im Falle einer teilweisen Versicherung (bspw. aufgrund von Kostenbeteiligungen) wäre er im Krankheitsfall an den Kosten beteiligt und müsste somit seinen Konsum einschränken.

notwendig, eine Selbstbeteiligung ( $\sigma$ ) einzuführen. Dabei wird aber weiterhin angenommen, dass der Arzt ein perfekter Agent des Patienten ist, d.h., dass er dem Patienten genau so viele Leistungen anbietet, wie dieser von ihm erwartet. Aufgrund der Selbstbeteiligung des Patienten wird im Idealfall die Höhe dieser Leistung gerade derjenigen Leistung entsprechen, die der tatsächliche Gesundheitszustand erfordert. M.a.W. entspricht die Selbstbeteiligung genau dem *Moral-Hazard*-Effekt.

Die Selbstbeteiligung verändert über die Budgetrestriktion ( $k_i = y - P - \sigma x_i$ )<sup>83</sup> das Maximierungsproblem des Versicherten:

$$\max_{\{x_i\}} E^1 = \sum_{i=0}^n [\pi_i b_i(x_i - \theta_i) + u(k_i)]. \quad (5.33)$$

Die Bedingungen erster Ordnung lauten nun:

$$\begin{aligned} x_0 &= 0 \\ b'(x_i - \phi_i) - \sigma u'(k_i) &= 0 \quad i = 1, \dots, n. \end{aligned} \quad (5.34)$$

Die dazugehörige Prämie lässt sich dann bei gegebenem  $\sigma$  und optimaler Nachfrage in einem Wettbewerbsmarkt bestimmen zu:

$$P = (1 - \sigma) \sum_{i=0}^n \pi_i x_i(P, \sigma). \quad (5.35)$$

Die Bedingungen in (5.34) zeigen, dass im Falle einer exogenen Selbstbeteiligung der Haushalt keine vollständige Versicherungsdeckung mehr wünscht, da der Grenznutzen zustandsabhängig ist.<sup>84</sup>

Wenn der Arzt als perfekter Agent für den Patienten agiert, ist zunächst nicht mit einer dem Gesundheitszustand des Patienten angemessenen Versorgung zu rechnen. Die Anreize einer Vollversicherung des Patienten (*Moral Hazard*) und das Verhalten des Arztes als perfekter Agent seines voll versicherten Patienten würden vielmehr zu einer ungebremsten Leistungsausdehnung führen. Erst eine nachfrageseitige Kostenbeteiligung kann unter den Bedingungen des Einzelleistungsregimes eine angemessene (d.h. nach oben begrenzten) Inanspruchnahme von Gesundheitsleistungen gewährleisten.

---

<sup>83</sup> Dabei bezeichnet  $y$  das Einkommen des Haushaltes.

<sup>84</sup> Es lässt sich mit dem Modell darüber hinaus auch zeigen, dass ein positives  $\sigma$  zwischen 0 und 1 für die Versicherten optimal ist. Vgl. Appendix in Blomquist (1991).



**c) Doppelte Asymmetrien**

Der Arzt sei jetzt kein perfekter Agent mehr für den Patienten, sodass er dem Patienten nicht mehr unbedingt die Wahrheit über  $\theta$  enthüllt. Im Modell lässt sich dieser Fall durch die Annahme modellieren, dass der Arzt bei seinem Patienten immer den schlechtesten Gesundheitszustand diagnostiziert, sodass immer die maximale Leistung abgerechnet wird.<sup>85</sup> Diese Handlungsweise besitzt für ihn zwei Vorteile. Einerseits sichert sie ihm den maximal möglichen Verdienst und andererseits schützt sie ihn am besten vor möglichen Anschuldigungen wegen Kunstfehlern. Für den Sachverhalt der Überversorgung von Patienten kann es demnach zwei unterschiedliche Erklärungsvarianten geben. Einmal ist es die sogenannte angebotsinduzierte Nachfrage und das andere Mal ist es die Angst vor Haftungsansprüchen. Blomquist spezifiziert hierfür eine bestimmte Haftungsregel. Sei die *ex post* beobachtbare Gesundheit des Patienten  $\hat{b}$  eine Zufallsvariable mit Mittelwert  $b(x_i - \theta_i)$ , so wird der Arzt haftbar, wenn folgende Beziehung gilt:

$$b(x_i - \hat{\theta}) - \hat{b} > \gamma > 0, \tag{5.36}$$

wobei  $\hat{\theta}$  den Wert des Gesundheitszustandes angibt, den der Arzt dem Patienten mitteilt. Mit anderen Worten muss der Arzt versuchen, die Differenz zwischen dem diagnostizierten Gesundheitszustand und dem erwarteten Gesundheitszustand nach der Behandlung zu minimieren, um nicht haftbar gemacht zu werden. Daher ist es für den Arzt die beste Strategie, dem Patienten immer  $\hat{\theta} = \theta_n$  zu signalisieren, d.h. dem Patienten in jedem Falle den schlechtesten Gesundheitszustand „einzureden“. Glaubt der Patient dem Arzt dieses Signal, so wird er entsprechend in jedem Zustand eine medizinische Leistung im Umfang von  $x_n$  wählen. In diesem Fall ändert sich das Maximierungsproblem für den Haushalt nicht. Die Lösung des Problems entspricht der  $n$ -ten Lösung von (5.34) und das dazugehörige Nutzenniveau sei mit  $E^2$  bezeichnet. Allerdings muss jetzt für die Berechnung der optimalen Prämie die Beschränkung

$$P = (1 - \sigma)(1 - \pi_0)x_n \tag{5.37}$$

berücksichtigt werden.

---

<sup>85</sup> Es sei unterstellt, dass das ethische Verhalten des Arztes ihm nicht erlaubt, bewusst gesundheitsschädigende Maßnahmen durchzuführen.

Da die Wahl von  $x_n$  nur in einem Fall (im  $n$ -ten Zustand) optimal für den Patienten sein kann, könnte der Patient in jedem anderen Zustand die Nachfrage nach medizinischen Leistungen senken. Daher ist klar, dass

$$E^1 > E^2. \quad (5.38)$$

Wenn der Arzt die Leistungswünsche seiner Patienten im vollen Umfang erfüllt (perfekter Sachwalter für den Patienten), ist der erwartete Nutzen der Patienten größer als in Situationen mit einer nicht perfekten Sachwalterrolle des Arztes, in der der Arzt dem Patienten immer den schlechtesten Gesundheitszustand signalisiert und der Patient seinem Arzt dieses Signal glaubt. Letzterer Umstand lässt sich auch als adverse Compliance interpretieren, denn eine tatkräftige Unterstützung des Heilungsprozesses durch den Patienten setzt auch voraus, dass der Patient erkennt, wenn ihm eine nutzlose Behandlungen angeboten wird.

Obiges Ergebnis wurde unter Berücksichtigung des Moral-Hazard-Effektes auf der Nachfrageseite abgeleitet. Es zeigt sich damit, dass unter einem Einzelleistungsregime bei einer nicht-perfekten Arzt-Patienten-Beziehung und einer absoluten „Hörigkeit“ des Patienten (adverse Compliance) ein niedrigeres Nutzenniveau der Patienten erreicht wird als in Situationen, in denen der Arzt als perfekter Sachwalter seiner Patienten fungiert – unabhängig von der Compliance der Patienten.

Besitzt der Patient dagegen zumindest Kenntnis von der Wahrscheinlichkeitsverteilung von  $\theta$  und glaubt er nicht im vollen Umfang der Einschätzung des Arztes, so schätzt er seinen Gesundheitszustand selbst ein. Da diese Schätzung auf der Verteilung aller möglichen Zustände beruht, wird diese Schätzung einen bestimmten festen Wert annehmen, bspw. den Erwartungswert von  $x$  über alle Gesundheitszustände. Der Arzt als nicht-perfekter Sachwalter wird zwar weiterhin versuchen, die maximale Leistungsmenge dem Patienten anzubieten, kann dies aber nur bis zu einer vom Patienten gesetzten Grenze tun. Damit verändert sich das Maximierungsproblem des Patienten zu:

$$\max_{\{x\}} E^3 = \pi_0 [\bar{b} + u(y - P)] + \sum_{i=1}^n \pi_i [b(x - \theta_i) + u(y - P - \sigma x)]. \quad (5.39)$$

In diesem Fall wird er also in jedem Krankheitszustand den gleichen Umfang an medizinischen Leistungen ( $x$  statt  $x_i$ ) nachfragen.

Die einzige Bedingung erster Ordnung lautet nun:

$$\left[ \sum_{i=1}^n \pi_i (b'(x - \theta_i)) \right] - \sigma (1 - \pi_0) u'(y - P - \sigma x) = 0. \quad (5.40)$$

Die relevante Beschränkung für die optimale Politik der Krankenkasse ist damit:

$$P = (1 - \pi_0)(1 - \sigma)x, \quad (5.41)$$

d.h. die Prämie entspricht jetzt dem Produkt aus der Wahrscheinlichkeit zu erkranken und der von der Versicherung zu übernehmenden (Durchschnitts)-Leistung.

Setzt man den erreichbaren Vertrag unter  $E^2$  in die Bedingungen erster Ordnung von (5.40) ein, so wird die linke Seite der Gleichung negativ.<sup>86</sup> Damit kann der Patient in jedem Zustand seinen Nutzen erhöhen, wenn er die Inanspruchnahme der Leistungen senkt. Es gilt also:

$$E^3 > E^2. \quad (5.42)$$

Der erwartete Nutzen des Patienten ist bei einer nicht perfekten Sachwalterrolle des Arztes dann höher, wenn er auf seine eigene Abschätzung von  $\theta$  vertraut und in jedem Gesundheitszustand die gleiche Menge an Gesundheitsleistungen nachfragt (in Abhängigkeit seiner Schätzung von  $\theta$ ).

Ferner kann auch gezeigt werden, dass unter bestimmten Annahmen<sup>87</sup> weiterhin gilt:

$$E^1 > E^3. \quad (5.43)$$

Der erwartete Nutzen des Patienten ist also dann am größten, wenn die Informationsasymmetrien zwischen Arzt und Patienten am schwächsten ausgeprägt sind ( $E^1 > E^3 > E^2$ ). Welche Auswirkungen der Abbau von diesen Asymmetrien zwischen Krankenkasse und Arzt auf den Nutzen der Patienten haben, wird im folgenden Punkt diskutiert.

#### **d) Verträge mit asymmetrischer Information**

Es sei jetzt angenommen, dass die Ärzte perfekte Sachwalter der Krankenkasse sind. Dabei wird nicht auf die genaue Ausgestaltung der Vergütungsregel eingegangen, sondern lediglich angenommen, dass diese optimal im Sinne der Krankenkasse sei. Im einfachsten Fall könnte das eine Kopfpauschale sein. Außerdem wird von jeglichen Haftungsregeln abgesehen. In diesem Fall

---

<sup>86</sup> Dies setzt voraus, dass die Abschätzung von  $x$  in einem vernünftigen Bereich liegt. Insbesondere muss gelten, dass  $x < x_n$ .

<sup>87</sup> Vgl. ausführlich Blomquist (1991).

wird der Arzt dem Patienten immer den bestmöglichen Gesundheitszustand im Krankheitsfall  $\theta_1$  signalisieren, da mit diesem Zustand der geringste Umfang medizinischer Leistungen im Falle der Krankheit verbunden wäre. Glaubt der Patient dem Arzt, so wird er tatsächlich  $x_1$  konsumieren; glaubt er ihm nicht, wird er in jedem Zustand einen anderen, aber ebenfalls festen Wert  $x$  wählen. Dies führt wieder zu den Bedingungen erster Ordnung aus (5.40), wenn davon ausgegangen werden kann, dass beide  $x$  identisch sind.

Oftmals wird in MCOs von den Versicherten keine Selbstbeteiligung verlangt. Um die Wirkung der Selbstbeteiligung zu studieren, untersucht Blomquist (1991) das untergeordnete Problem der Nutzenmaximierung von (5.39) unter Berücksichtigung von (5.41) durch die optimale Wahl von  $P$  und  $\sigma$ .

Die Maximierung der mit diesem Maximierungsproblem verbundenen Lagrangefunktion führt zu:

$$\begin{aligned} \pi_0 u'(y - P) + (1 - \pi_0) u'(y - P - \sigma x) - \lambda &= 0 \\ x u'(y - P - \sigma x) - \lambda x &= 0. \end{aligned} \tag{5.44}$$

Da diese Gleichungen nur erfüllt sein können, wenn  $\sigma = 0$  gilt, erfährt der Versicherte mit einem derartigen MC-Vertrag einen höheren Nutzen als bei einer herkömmlichen Versicherung mit Selbstbeteiligung. Die wichtige Erkenntnis daraus ist, dass sich der Versicherte durch die Wahl eines MC-Vertrags verbessern kann, obwohl das Informationsproblem zwischen Patient und Arzt nicht gelöst werden konnte (Moral Hazard).

## 5.6 Der Arzt als Experte

Der Arzt ist aufgrund seiner Expertenstellung in der Lage, seinen Patienten Behandlungen zu empfehlen, deren Kosten den Nutzen übersteigen. Der Hauptverband der Schweizer Krankenkversicherer (KSK) geht davon aus, dass ca. 10-15% der Einnahmen von Ärzten und Spitälern auf eine betrügerische Abrechnungspolitik zurückgehen.<sup>88</sup> Ein solches Verhalten des Experten wird hauptsächlich durch die Angst um die eigene Reputation begrenzt. Denn obwohl der Patient aufgrund der beschriebenen Informationsasymmetrien die Behandlungsempfehlungen seines Arztes nicht richtig einschätzen kann, wird er *ex post* das Ergebnis der Behandlung mit dem versprochenen Heilungserfolg vergleichen können. Stellt der Patient eine erhebliche

---

<sup>88</sup> Vgl. Emons (2001, S. 378), der eine Meldung der Sonntagszeitung vom 13.12.1998 wiedergibt.

Diskrepanz zwischen beiden fest, wird er dies dem Arzt anlasten. Somit wäre die Reputation des Arztes beschädigt, was auch erhebliche Auswirkungen auf zukünftige Behandlungen hätte. Denn mit einem beschädigten Ansehen wird es dem Arzt kaum gelingen, neue Patienten anzuziehen bzw. bisherige Patienten zu halten. Der Arzt wird demnach versuchen, diesen negativen Effekt auf seine Reputation zu begrenzen. Insgesamt stellt sich für den Experten die Frage nach dem optimalen Ausgleich zwischen Angebotsinduzierung (und damit Gewinnsteigerung) und negativem Reputationseffekt.

In einem Modell von Wolinsky (1993) kann das opportunistische Verhalten der Leistungserbringer durch zwei Mechanismen eine Begrenzung erfahren.<sup>89</sup> Neben dem Marktverhalten (Suche) der Patienten, die auf einem Wettbewerbsmarkt denjenigen Leistungsanbieter auswählen werden, der ihren Präferenzen am besten entspricht, kann es auch durch einen Reputationseffekt begrenzt werden.

In diesem Modell werden Patienten betrachtet, die zwar wissen, dass sie ein Leiden haben – und deshalb den Arzt aufsuchen –, aber nicht wissen, wie schwer ihre Krankheit ist. Die Ärzte können die Krankheiten hingegen diagnostizieren und mit zwei unterschiedlichen Strategien behandeln: entweder mit einer umfangreichen Behandlung, deren Kosten mit  $H$  bezeichnet werden, oder einer weniger umfangreichen Behandlung mit Kosten  $L$ . Beide Behandlungen besitzen entsprechende Preise, wobei die umfangreiche Behandlung für den Arzt profitabler sein soll ( $p_H > p_L$ ). Durch die asymmetrische Informationsverteilung entstehen dem Arzt damit Möglichkeiten für opportunistisches Verhalten, denn er kann dem Patienten im Falle einer weniger schweren Erkrankung eine Behandlung  $H$  verordnen.

Die Ärzte entscheiden simultan über ihr Behandlungsrepertoire (Angebot nur von  $L$ , nur von  $H$  oder von beiden), ihre Preise und ihre Empfehlung. Die Empfehlungsstrategie besteht entweder im Ratschlag der richtigen Behandlung bei schweren wie bei leichten Fällen oder in der Empfehlung einer teureren Behandlung im Falle eines leichten Falles.

Wolinsky betrachtet zunächst das Modell ohne Reputation. In diesem Modell existiert ein mögliches Marktgleichgewicht bei dem angenommen wird, dass die Suchkosten der Patienten nicht sehr hoch sind.<sup>90</sup> In diesem Fall kommt es zur Spezialisierung der Experten, d.h. die einen Experten werden nur die weniger schwerwiegende Krankheit behandeln und die anderen

---

<sup>89</sup> Vgl. hierzu auch Gaynor (1994).

<sup>90</sup> Es existieren in diesem Modell weitere mögliche Marktgleichgewichte, die hier nicht behandelt werden. Vgl. dazu Wolinsky (1993).

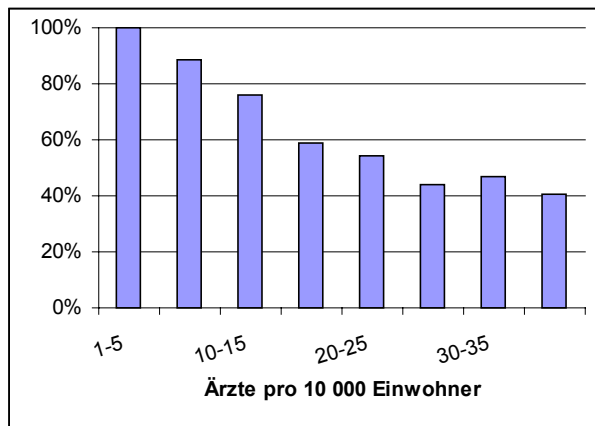
Experten sich auf das schwerwiegendere Problem spezialisieren. Die Spezialisierung führt auch zu einem Auseinanderfallen von Diagnose und Behandlung. Denn der Patient wird, da er nicht weiß, wie schwer seine Erkrankung ist, zunächst einen Experten aufsuchen, um herauszufinden, ob es sich um eine leichte oder schwere Erkrankung handelt. Wolinsky kann zeigen, dass der Patient im Falle geringer Suchkosten zunächst einen Experten wählt, der ausschließlich die leichte Behandlung vornimmt. Dieser Arzt wird dann wahrheitsgemäß den Patienten im Falle einer schweren Erkrankung abweisen, da er sonst im Falle einer falschen Behandlung in der Zukunft Patienten verlieren würde.<sup>91</sup> Vorausgesetzt, der Patient glaubt der Empfehlung des ersten Arztes, wählt er dann auf der zweiten Stufe einen Experten, der ausschließlich schwerwiegende Fälle behandelt. Die Entscheidung für einen solchen Experten folgt geradewegs aus der ersten Entscheidung, denn er hätte auch auf der ersten Stufe schon einen Experten wählen können, der sowohl in der Lage ist, die leichten wie auch die schweren Fälle zu behandeln.

Eine wichtige Implikation des Modells kann damit wie folgt zusammengefasst werden: Sind die Suchkosten für die Patienten gering, z.B. aufgrund einer hohen Wettbewerbsintensität zwischen den Ärzten, so kommt es zu einer Spezialisierung und einer Aufspaltung von Diagnose und Behandlung. Die Spezialisierung schränkt damit die Möglichkeiten der Leistungserbringer ein, Nachfrage zu induzieren. Letztendlich wird die theoretische Schlussfolgerung einer Spezialisierung auch durch die empirische Beobachtung gestützt. Denn nichts anderes stellt die starke Abgrenzung von Allgemeinmedizinern und Spezialisten dar. Man kann feststellen, dass diese Abgrenzung in Gebieten mit einem starken Konkurrenzdruck (hohe Ärztedichte) stärker ausgeprägt ist als in Gebieten mit einer geringeren Arztdichte. Ein Blick auf Schweizer Daten bestätigt dies: Mit zunehmender Arztdichte nimmt der Anteil der Allgemeinmediziner an der Gesamtzahl der Ärzte ab. Andererseits steigt der Anteil der Spezialisten mit zunehmender Arztdichte (vgl. Abbildung 3).

---

<sup>91</sup> Man beachte, dass in diesem Szenario alle Patienten zunächst zum Experten für die leichteren Fälle gehen.

Abbildung 3 Anteil der Allgemeinmediziner an allen Ärzten im Verhältnis zur Arztdichte (Zürich und Genf 1997)



Betrachtet man in diesem Lichte die Stärkung der Hausärzte in Hausarztmodellen, so lassen sich derartige Modelle als bewusste Trennung von Diagnose und Behandlung bei schweren Fällen auffassen. Denn einerseits gilt der Hausarzt sicher als Spezialist für leichtere Erkrankungen, die er diagnostizieren und behandeln kann. Andererseits müssen alle Patienten zunächst bei allen Beschwerden ihren Hausarzt aufsuchen und werden, wenn der Hausarzt aufgrund der Schwere oder der Art der Erkrankung sie selber nicht behandeln kann, zu einem anderen Leistungserbringer überwiesen. Allerdings sieht man auch, dass in einem solchen System die Anreize für die Ärzte richtig gesetzt sein müssen, damit sie tatsächlich die schweren Fälle zu einem Spezialisten überweisen. Die Einzelleistungsvergütung dürfte daher gerade in Hausarztmodellen kontraproduktiv wirken. Da die Patienten fest im Hausarztssystem eingeschrieben sind, können sich die Wirkungen der Einzelleistungsvergütung und des gesicherten Patientenstammes gegenseitig verstärken und auf das Überweisungsverhalten negativ auswirken.

Ein wesentlicher Nachteil des Suchkostenmodells besteht darin, dass die Patienten im Voraus wissen, wie hoch die Kosten der Behandlungsalternativen sind. Dies wurde durch die Annahme ermöglicht, dass die Ärzte gleichzeitig über die Preise und Behandlungsalternativen entscheiden mussten und die Patienten aufgrund dieser Angaben sich entscheiden konnten. Eine derartige Annahme ist im Medizinbereich eher unrealistisch. Geht es um die Reparatur eines mechanischen Gerätes, so ist es vergleichsweise einfach, im Voraus zu bestimmen, was die Reparatur kostet. In der Arzt-Patienten-Beziehung besteht dagegen eine starke Unsicherheit über die notwendige medizinische Versorgung, auch nachdem die (richtige) Diagnose gestellt wurde.

Diesen Aspekt berücksichtigt das Reputationsmodell von Wolinsky. In einem leicht modifizierten Modell wird jetzt angenommen, dass die Patienten exakt zwei Perioden leben. In jeder Periode haben sie eine Krankheit, wissen allerdings im Voraus nicht, von welchem Schweregrad diese Krankheit ist. Unter diesen Annahmen ist es für einen Patienten in der zweiten Periode optimal, zu dem Arzt zu gehen, bei dem er schon in der ersten Periode gewesen ist, wenn der Arzt den Patienten beim ersten Mal nicht abgelehnt, sondern behandelt hatte. Umgekehrt besteht die optimale Strategie des Patienten nach einer Ablehnung durch den Arzt der ersten Periode im Wechsel des Arztes in der zweiten Periode.

Demnach kann es in einer Arzt-Patienten-Beziehung mit der Möglichkeit wiederholter Kontakte zu einer Reduktion der bestehenden Informationsasymmetrien kommen, da diese wiederkehrenden Kontakte den Arzt zwingen, dem Patienten die Wahrheit zu sagen.

Dieses Ergebnis wird auch in einem Prinzipal-Agent-Modell sich wiederholender Begegnungen zwischen Arzt und Patient bestätigt. Jelovac (2001) zeigt in ihrem Modell, in dem es die Möglichkeit einer zweiten Konsultation gibt, dass durch diese kleine Dynamik die Informationsasymmetrien zum Teil verschwinden. Jelovac weist damit auch nach, dass die Behauptung, eine starke Konzentration auf angebotsseitige Kostenbeteiligungen würde zu einer Unterversorgung der Bevölkerung führen, in diesem Kontext widerlegt werden kann. D.h. mit anderen Worten, dass eine angebotsseitige Kostenbeteiligung in Hausarztnetzen oder HMOs mit einer höheren Wahrscheinlichkeit nicht zu Qualitätseinbußen in der Behandlung der Patienten führt, als in einer konventionellen Versorgung, die gleichfalls angebotsseitige Kostenbeteiligungselemente beinhaltet. Der Grund hierfür liegt darin, dass in Hausarztnetzen bzw. HMOs nicht nur eine angebotsseitige Kostenbeteiligung angewendet wird, sondern der Arzt gleichzeitig als Gatekeeper den Behandlungsverlauf des Patienten kontrolliert. Damit wird deutlich, dass es nicht darauf ankommt, MC-Instrumente um jeden Preis anzuwenden, sondern vielmehr um den richtigen Mix der Instrumente.

### **5.7 Zusammenfassung und Ausblick**

Dieses Kapitel betrachtete das Verhalten der Ärzte und der Patienten in informationsökonomischen Modellen. Dabei wurde eine normative Sicht eingenommen und nach einer optimalen Vergütungsstruktur für Allgemeinmediziner gesucht, wenn die Akteure bestimmte Verhaltensannahmen erfüllen.



Aufgrund der bestehenden Informationsasymmetrien wird im Allgemeinen ein zwei-geteilter Tarif gewählt. Der erste Teil des Tarifs stellt einen Fix-Betrag dar, der zweite Teil ist eine Funktion der Anstrengungen des Agenten.

Allerdings ist diese Aufteilungsregel nicht einfach zu implementieren. Da die Kostenaufteilung zwischen Prinzipal und Agent bei einem derartigen Tarif durch das Grenznutzenverhältnis der beiden Parteien bestimmt wird und der Grenznutzen der Parteien vom Ergebnis abhängt, müsste ein solcher Tarif ergebnisabhängig gestaltet werden. Im Gesundheitsbereich sollte das Ergebnis eine Verbesserung der Gesundheit sein. Damit hängt eine optimale Gestaltung des Vergütungsvertrags unter asymmetrischer Informationsverteilung entscheidend von der Messbarkeit des Gesundheitszustandes ab. Im Allgemeinen dürfte die adäquate Messung des Gesundheitszustandes ein ernstes Problem darstellen. In der Praxis werden Vergütungsverträge daher oft allein auf die manifestierten Anstrengungen des Arztes konditioniert.

Sowohl das Standardmodell als auch die vorgestellten Modelle asymmetrischer Informationsverteilung unter Berücksichtigung der Krankenversicherung betrachten jeweils nur einen Prinzipalen und einen Agenten. Es stellt sich daher die Frage, wie die Ergebnisse aussehen, wenn es entweder mehrere Prinzipale oder mehrere Agenten gibt. So kann man sich einen Arzt (Agenten) vorstellen, der in der Lage ist, exklusive Verträge mit verschiedenen Versicherungen (Prinzipalen) zu schließen. Andererseits kann die Krankenkasse auch mehrere Verträge mit Leistungserbringern abschließen.<sup>92</sup> Letztgenannte Möglichkeit bedeutet für die Krankenkasse, dass ihr nun mehrere Beobachtungen (statt einer) zur Verfügung stehen, um einen anreizkompatiblen Vertrag zu gestalten. Denn sie kann auf der Grundlage der Arbeitsergebnisse mehrerer Ärzte wesentlich einfacher Schlussfolgerungen bezüglich des tatsächlichen Anstrengungsniveaus der Mediziner ziehen. Damit können im Rahmen von *Multiple-Agent-Beziehungen* die Informationsasymmetrien zwischen Krankenkasse und Ärzten zumindest zum Teil überwunden werden.<sup>93</sup>

Des Weiteren wurde in den vorgestellten Modellen nur der Fall betrachtet, dass der Arzt eine (einheitliche) Aufgabe zu erledigen hat. In der Praxis muss der Arzt dagegen mehrere und vor allem unterschiedliche Aufgaben durchführen. *Multi-Task-Modelle* erweitern daher das Grundmodell bezüglich der beschriebenen Aufgabe. Multitasking kann sich auf die zusätzlichen

---

<sup>92</sup> Dieser Fall wird sicherlich in Zukunft eine große Rolle spielen, da sowohl in der Schweiz als auch in Deutschland über mehr Wettbewerbsselemente auf der Angebotsseite diskutiert wird. Ein Wettbewerbsselement wäre dabei bspw. die freie Wahl der Leistungserbringer durch die Versicherungen.

<sup>93</sup> Vgl. Karmann (1994).

Aufgaben beziehen, die der Arzt über die rein medizinischen Aufgaben hinaus in einer modernen Arztpraxis zu erledigen hat, wie bspw. organisatorische, administrative oder Managerfunktionen. Andererseits besteht die medizinische Behandlung selbst wiederum aus mehreren Aufgaben. Als Beispiel hierfür kann ein Arzt angesehen werden, der für den Behandlungserfolg seines Patienten verschiedene Aufgaben zu erfüllen hat (Diagnose stellen, Medikamente verschreiben, Hausbesuche durchführen, Überweisung stellen u.v.a. mehr). Die Aufgabe für den Prinzipalen (in diesem Fall die Krankenkasse) besteht in einer solchen Situation darin, das Vergütungssystem so auszugestalten, dass die verschiedenen Leistungen vom Agenten im gewünschten Umfang angeboten werden. Ist dies – wie in der Realität häufig – nicht der Fall, kommt es zu einer Über- oder Unterversorgung der Patienten, je nachdem wie das Verhältnis zwischen Aufwand und Ertrag für jede einzelne Leistung aussieht. Im Extremfall kann es möglich sein, dass der Agent nur eine Aufgabe aus einer Menge von Aufgaben erfüllt. Dies wird insbesondere dann der Fall sein, wenn der Ertrag aus der Vergütung dieser Aufgabe weit größer als die Kosten des Aufwands ist und wenn der Nettoertrag bei Erfüllung der einen Aufgabe auch noch größer als der Nettoertrag bei Erfüllung aller Aufgaben ist. Die in Abschnitt (4.3) beschriebenen Bemühungen um eine Revision des Arzttarifs in der Schweiz lassen sich in diesem Licht auch als einen Versuch betrachten, die Anreizstrukturen innerhalb der Arztvergütung in geeigneter Weise umzugestalten.<sup>94</sup>

Schließlich kann im Multitasking auch die Aufgabe gesehen werden, es als Doppelagent zwei Herren gleichzeitig recht zu machen; also die Aufgabe, sowohl die primären Interessen der Patienten zu vertreten (adäquate Behandlung des einzelnen Patienten), als auch im Interesse der Versichertengemeinschaft zu handeln (kostengünstigste Behandlung aller Patienten).

Eine weitere Forschungsrichtung beschäftigt sich mit dem unterschiedlichen Einfluss der Anstrengungen von Patient und Arzt auf das Ergebnis und wie die Vergütung darauf reagieren sollte. In einem Modell mit zweiseitigem Moral-Hazard mit einer produktionstheoretischen Erweiterung untersuchen Leonard und Zivin (2000) den Zusammenhang zwischen Anstrengungen des Arztes und des Patienten auf der einen und dem Ergebnis auf der anderen Seite. Sind die Anstrengungen des Arztes zu den Anstrengungen des Patienten stark komplementär oder sind beide Anstrengungsniveaus sehr wichtig für das Ergebnis, dann ist eine ergebnisbasierte Vergütung besser als eine anstrengungsbasierte. Vergütungssysteme, die auf den Anstrengungen

---

<sup>94</sup> Vgl. zur Ausgestaltung von Verträgen bei mehreren unterschiedlichen Aufgaben Karmann (1994) und Gaynor (1994).

des Arztes basieren, sind dagegen dann der ergebnisbasierten vorzuziehen, wenn entweder die Anstrengungen des Patienten oder die Anstrengungen des Arztes einen hohen Einfluss auf das Ergebnis haben, aber nicht beide. Aufgrund der Ergebnisse dieses Modells mit beidseitigem Moral-Hazard würde man bei einem chirurgischen Eingriff sicher eine aufwandsbezogene Vergütung vorziehen, weil die Anstrengungen des Arztes sehr entscheidend für das Ergebnis sind. Andererseits sollte die Vergütung der meisten Behandlungsroutinen in einer Allgemeinpraxis eher auf das Ergebnis der Behandlung konditioniert sein, da hier die Anstrengungsniveaus von Arzt und Patient einen entscheidenden Einfluss auf den Behandlungserfolg haben dürften.

Ein Problem bei der Anwendung von Prinzipal-Agent-Modellen auf die Arzt-Patient-Beziehung ist schließlich immer im Auge zu behalten. Entgegen der Standardannahme der P-A-Modelle wird bei Ärzten oftmals unterstellt, dass sie den Nutzen der Patienten in ihren Nutzenüberlegungen berücksichtigen (ethisches Verhalten). Damit sind aber die Nutzenfunktionen von Patient und Arzt nicht unabhängig voneinander, sodass man im Grunde genommen wissen müsste, was in den Nutzenfunktionen von Ärzten und Patienten enthalten ist.<sup>95</sup>

Einen Versuch, die Präferenzen von Patienten zu evaluieren, stellt bspw. die Arbeit von Scott und Vick (1999) dar. Sie befragen in einer Studie Patienten in Aberdeen. In ihrer Untersuchung stellen sie vor allem fest, dass dem Gespräch mit dem Arzt offenbar von den Patienten ein sehr hoher Stellenwert beigemessen wird. Dabei sei nicht die Quantität der erhaltenen Information, sondern die Qualität der Information besonders wichtig.

Über die Frage, was tatsächlich in der Nutzenfunktion des Arztes enthalten ist, kann trefflich gestritten werden.<sup>96</sup> Die Literatur zu den sogenannten Credence Goods – wie auch das Beispiel in Abschnitt 5.6 – bieten Argumente für die Aufnahme des Patientennutzens in die Nutzenfunktion des Arztes. In diesen Modellen geht es um die Begrenzung des eigennützigen Motivs des Arztes. In diesem Sinne sind auch die Modelle zur angebotsinduzierten Nachfrage zu verstehen. Denn sie zeigen, wie ein uneigennütziges Motiv des Arztes einer uneingeschränkten Ausnutzung der Expertenstellung entgegenwirkt. Auch das Modell in Abschnitt 5.4 verwendet eine derartige Berücksichtigung des Patientennutzens in der Nutzenfunktion, um so die Sachwalter-Beziehung zwischen Arzt und Patient abbilden zu können. Dies entspricht aber in

---

<sup>95</sup> Vgl. Mooney und Ryan (1993) oder Zweifel und Eisen (2001).

<sup>96</sup> Eine Diskussion möglicher Argumente in der Nutzenfunktion des Arztes findet sich u.a. bei Kristiansen (1995).

keiner Weise der Standardtheorie, bei der Agent und Prinzipal unabhängig voneinander ihren Nutzen maximieren.<sup>97</sup>

### 6 Die Arztpraxis und Wettbewerb

Dieses Kapitel betrachtet die Beziehung zwischen Arzt und Patient ohne die explizite Modellierung von Informationsasymmetrien. Entscheidend für die folgenden Betrachtungen sind vielmehr Annahmen über die Wettbewerbssituation. Zunächst wird untersucht, wie sich eine Arztfirma unter den Annahmen des vollkommenen Wettbewerbs bzw. des Monopols verhält. Abschnitt 6.1 stellt beide Marktformen kurz vor und untersucht wie der Arzt seine Arbeitszeit in die Firma einbringt. Im Abschnitt 6.2 werden die extremen Annahmen des Monopols und der vollständigen Konkurrenz aufgegeben. Vielmehr wird nun das Verhalten der Ärzte unter der realistischeren Annahme modelliert, dass sie über begrenzte Marktmacht verfügen.

Entsprechend der Untersuchung im 5. Kapitel besteht ein wesentliches Element einer anreizkompatiblen Vergütung in der teilweisen Kostenübernahme durch die Ärzte. Die Stärke dieser Kostenübernahme kommt in der Höhe des fixen Anteils der Vergütung zum Ausdruck. Der zweite wichtige Aspekt einer anreizkompatiblen Vergütung muss in seinem Anreiz gesehen werden, die für den Patienten bestmögliche Behandlung zu gewährleisten. Dieser Anreiz nimmt mit einer stärkeren Bezugnahme der Vergütung auf die Anstrengungen des Arztes zu. Abschnitt 6.3 beschäftigt sich daher mit der angebotsseitigen Kostenbeteiligung sowie mit der Qualität der medizinischen Versorgung unter der Annahme der monopolistischen Konkurrenz. Darüber hinaus wird in diesem Abschnitt auch diskutiert, welche Auswirkungen der Wettbewerb auf das Verhalten der Ärzte besitzt.

Die meisten Ärzte in Hausarztmodellen behandeln nicht nur MC-Versicherte, sondern auch Patienten aus dem konventionellen Bereich. Im Abschnitt 6.4 wird deshalb untersucht, welche Auswirkung ein hoher Managed-Care-Anteil innerhalb einer Arztpraxis auf die Behandlung sowohl der MC-Patienten als auch der konventionell versicherten Patienten besitzt.

#### 6.1 Produktion in einer Arztpraxis

Der Output einer Arztpraxis ist die Gesundheit der Patienten. Gesundheit als Output kann aber nicht gemessen werden, sodass man sich bei ihrer Messung mit Approximationen begnügen

---

<sup>97</sup> Vgl. wieder Mooney und Ryan (1993).

muss. Mögliche Kandidaten sind beobachtbare Outputmaße, die eng mit der Gesundheit korreliert sind. Hierzu zählen die angebotenen Konsultationen oder die erbrachte Menge an medizinischen Leistungen, da im Allgemeinen davon ausgegangen werden kann, dass mit mehr Konsultationen oder einer höheren Menge an medizinischen Leistungen ein besserer Gesundheitszustand erreicht werden kann.

Zur Produktion werden verschiedene Inputs eingesetzt. Wichtige Inputfaktoren sind zunächst die benötigten Arbeitskräfte. Darüber hinaus stellen auch die Anstrengungen des Arztes zum Behandlungserfolg sowie das Wissen und die Fertigkeiten wichtige Inputfaktoren des Arztes dar. Außerdem werden auch technische Inputfaktoren zur Produktion eingesetzt wie etwa Röntengeräte oder Praxislabors.

### 6.1.1 Vollständige Konkurrenz und Monopol

Ausgangspunkt bei der Betrachtung der Arztpraxis ist eine Firma, die auf einem wettbewerblich organisierten Markt als Preisnehmer agiert und ihren Gewinn maximiert. Das zu produzierende Gut sei die Konsultation ( $Kons$ ) mit dem Preis  $p$ . Als Arbeits-Input steht die eigene Arbeitszeit ( $t^A$ ) sowie die Arbeitszeit anderer Arbeitskräfte wie Praxisschwestern, Sekretärinnen usw. zur Verfügung ( $t^P$ ). Der Kapital-Input sei allgemein mit  $K$  bezeichnet und umfasst sowohl die technische Ausstattung der Praxis als auch die laufenden Kosten, wie etwa die Miete für die Praxisräume. Die insgesamt  $i$  Inputfaktoren werden in einem Vektor  $z = \{z_1, \dots, z_i\}$  zusammengefasst. Die Inputpreise können vom Arzt nicht beeinflusst werden und seien gegeben durch den Vektor  $p_z = \{p_{z_1}, \dots, p_{z_i}\}$ .<sup>98</sup> Maximiert die Arztpraxis unter diesen Bedingungen und unter den gewöhnlichen Annahmen für die Produktionsfunktion  $Kons = f(z)$ <sup>99</sup> ihren Gewinn ( $\Pi$ ),

$$\Pi = p \text{Kons} - p_z z, \quad (6.1)$$

erhält man die Bedingungen erster Ordnung für dieses Problem:

$$\frac{\partial f(z)}{\partial z} = \frac{p_z}{p}, \quad (6.2)$$

---

<sup>98</sup> Um die Notation zu vereinfachen, werden  $p$  und  $z$  im Folgenden als einfache Variablen behandelt.

<sup>99</sup> Die Produktionsfunktion sei stetig, differenzierbar und strikt konkav über alle Inputfaktoren.

d.h. der Arzt wird im Optimum so viele Konsultationen anbieten wollen, sodass die Grenzprodukte der Inputfaktoren gerade ihrem Inputpreis entsprechen.<sup>100</sup> Damit ist der Preis einer Konsultation gleich den Grenzkosten ( $GK$ ) ihrer Bereitstellung ( $p = GK$ ).

Mit Hilfe der Inputpreise ausgedrückt, heißt das:

$$GK = \frac{p_z}{f'(z)} \quad (6.3)$$

Im Monopolfall kann der alleinige Anbieter in einem Markt den Preis hingegen beeinflussen und so seinen Gewinn steigern. Dabei wird der Preis durch die Wahl des eigenen Produktionsumfanges und somit auch durch die Wahl der Inputfaktoren indirekt beeinflusst. Die Gewinnfunktion der Arztpraxis hat in diesem Fall folgendes Aussehen:

$$\Pi = p(f(z))f(z) - p_z z. \quad (6.4)$$

Die Bedingung erster Ordnung für ein Gewinnmaximum lautet dann:

$$\frac{\partial f(z)}{\partial z} = \frac{p_z}{p} \left( \frac{|\varepsilon|}{|\varepsilon| - 1} \right) \quad (6.5)$$

wobei  $\varepsilon$  die Preiselastizität der Nachfrage bezeichnet. Sie gibt an, um wie viel Prozent sich die nachgefragte Menge verändert, wenn sich der Preis für das nachgefragte Gut um 1 Prozent ändert. Da die Elastizität in der Regel negativ ist (im Fall einer Nachfragekurve mit negativer Steigung), reicht es aus, sich den Absolutbetrag dieser Größe anzusehen. (6.5) macht deutlich, dass ein Monopolist nie im Bereich einer unelastischen Nachfrage ( $|\varepsilon| < 1$ ) produzieren würde, da dies ihm auf keinen Fall ein Gewinnmaximum verschaffen könnte (die rechte Seite von (6.5) wäre negativ).<sup>103</sup>

---

<sup>100</sup> D.h. beispielsweise, dass eine zusätzliche Konsultation, die durch den zusätzlichen Einsatz eines Inputfaktors zum Marktpreis angeboten werden könnte, im Optimum gerade dem Inputpreis dieses Faktors entsprechen muss.

<sup>101</sup> Mit  $f'(z) = \frac{\partial f(z)}{\partial z}$ .

<sup>102</sup> Mit  $\varepsilon = \frac{\partial \text{Kons}}{\partial p} \cdot \frac{p}{\text{Kons}}$ .

<sup>103</sup> Vgl. Varian (1992).

Setzt man (6.3) in (6.5) ein, so lässt sich die Optimalbedingung des Monopolisten auch schreiben als:

$$p = MC \left( \frac{|\varepsilon|}{|\varepsilon| - 1} \right). \quad (6.6)$$

Im Allgemeinen wird also der Preis die Grenzkosten der Produktion übersteigen. Insbesondere wird der Preis mit einer stärkeren Neigung der Nachfragekurve steigen, während umgekehrt bei einer Steigung von Null (der Arzt sieht sich einer horizontalen Nachfragekurve und damit einer unendlich elastischen Nachfrage gegenüber) das Gleichgewicht bei vollkommenem Wettbewerb erreicht wird ( $p = MC$ ).

Innerhalb der obligatorischen Krankenversicherung der Schweiz sind die Preise für den Arzt ein Datum. Sie werden in Verhandlungen zwischen den Krankenkassen und den Ärzteverbänden festgelegt.<sup>104</sup> Trotzdem besitzt der Arzt verschiedene Möglichkeiten, eine monopolähnliche Stellung auf dem Markt zu realisieren. Zum einen sind aufgrund der Besonderheiten des Marktes die verhandelten Preise für Arztdienstleistungen in der Regel keine marktgerechten Preise. Außerdem sichern Zugangsbeschränkungen, die vornehmlich durch Ausbildungsbeschränkungen realisiert werden, den etablierten Anbietern Monopolrenten. Die besonderen Eigenschaften des Gutes Gesundheit spielen in diesem Zusammenhang ebenfalls eine wichtige Rolle. Denn eine einmal verkaufte Dienstleistung wie etwa eine Diagnose oder eine Behandlung kann nicht weiterverkauft oder gehandelt werden.<sup>105</sup> Schließlich verschafft auch das Nachfrageverhalten der Versicherten den Ärzten eine monopolähnliche Stellung. Da die Patienten sich aufgrund der Versicherungsdeckung nicht dem tatsächlichen Preis der in Anspruch genommenen Leistung gegenübersehen, spielt der Preis als Koordinierungsmechanismus von Angebot und Nachfrage nur eine untergeordnete Rolle. Vielmehr wird der Patient bis zu einem bestimmten Grade – aber unabhängig vom Preis der Leistung – seinem Arzt die Treue halten. Schließlich können auch die im 5. Kapitel dargestellten Informationsasymmetrien dem Arzt zu einer monopolähnlichen Stellung gegenüber seinen Patienten verhelfen. Somit entsteht ein Markt, auf dem der Anbieter in Bezug auf seinen eigenen Patientenstamm eine Monopolstellung inne hat, während er auf Marktebene mit anderen Ärzten um neue Patienten konkurriert.

---

<sup>104</sup> Zum neuen Arzttarif TARMED, der sich gegenwärtig (2002) in der Erprobungsphase befindet und ab dem Jahr 2004 eingeführt werden soll, vergleiche die Ausführungen im Abschnitt 4.3.

<sup>105</sup> Vgl. Gaynor (1994).

### **6.1.2 Input-Monopol bei vertikaler Integration – die Arbeitszeit des Arztes**

Der wesentliche Inputfaktor bei der Produktion in einer Arztpraxis ist die Arbeitszeit des Arztes. Die Wichtigkeit dieses Inputfaktors ergibt sich aus der Doppelrolle, welche die Arbeitszeit des Arztes in der Zielfunktion spielt. Als Unternehmer versucht der Arzt die Arbeitszeit bei einem möglichst geringen Preis vermehrt einzusetzen. Als nutzenmaximierendes Individuum ist er dagegen bestrebt, seine Freizeit zu optimieren. Ein daraus sich ergebendes bekanntes Phänomen ist, dass Ärzte – wie viele andere Selbstständige auch – scheinbar „zu viel“ der eigenen Arbeitszeit in ihre Firma stecken.

Das führt zur Frage: Wie setzt ein Arzt seine Arbeitszeit ein? Einige Studien<sup>106</sup> belegen, dass Ärzte offenbar weniger Hilfspersonal einstellen als sie eigentlich sollten, sofern sie ihren Nutzen, der vom Einkommen und der Freizeit abhängt, maximieren. Es scheint, dass die Ärzte nicht die kostenminimale Inputkombination verwenden. Warum aber verhalten sie sich so? Eine mögliche Erklärung kann in den Substitutionsbeziehungen der Inputfaktoren und einer gleichzeitigen Monopolstellung des Arztes bei einem Inputfaktor liegen.

Der Arzt als Anbieter von Arbeit hat auf dem Input-Markt eine monopolartige Stellung und verkauft dem Arzt als Betreiber der Arztpraxis den Inputfaktor Arbeit. Die Arztpraxis weiß natürlich, dass in jedem Fall der Marktpreis den gemeinsamen Gesamtgewinn maximiert, weil er in Verbindung mit der tatsächlich geleisteten Arbeitszeit die Minimalkostenkombination darstellt. Die Arztpraxis als Firma verlangt aber nach außen den höheren Monopolpreis. Dies kann sie, weil der „öffentliche“ Preis des Inputfaktors auch für alle anderen Firmen der Branche derselbe ist, sodass die anderen Firmen ebenfalls einen Monopolpreis für das Endprodukt verlangen müssen.

Die Beobachtung, dass freipraktizierende Ärzte überdurchschnittlich lang arbeiten, obwohl sie auch weitere Hilfskräfte einstellen könnten, kann durch die oben beschriebene Monopolstellung des Arztes auf dem Inputmarkt erklärt werden. Innerhalb der Organisation rechnet er mit dem Grenzkostenpreis seiner Arbeit, da dies zu einer kostenminimalen Produktion führt. Die Grenzkosten liegen annahmegemäß unterhalb des Monopolpreises, sodass das Arbeitsangebot bei Grenzkosten höher als zum Monopolpreis ist.

Allerdings gibt es auch andere Erklärungen für das beobachtete Phänomen. So kann es sein, dass der Arzt seine eigene Arbeitszeit bei der Behandlung der Patienten als sehr viel wichtiger

---

<sup>106</sup> S. z.B. Reinhardt (1975). Vgl. hierzu und im Folgenden Phelps, Ch. (1997, S. 185-191).



für die Patienten einschätzt als die Arbeitszeit seiner Hilfskräfte. Für eine derartige Abweichung von der Maximierung einer Nutzenfunktion mit Einkommen und Freizeit wird wiederum häufig die Bindung der Ärzte an den Eid des Hippocrates angeführt, der von jedem Arzt u.a. verlangt, dass er das Bestmögliche für die Patienten unternimmt. Das Modell in den folgenden Abschnitten berücksichtigt diese Sichtweise, indem eine zu starke Mengenausweitung zu einem Nutzenverlust für den Arzt führt.

### 6.2 Monopolistische Konkurrenz

Der Markt für Arztdienstleistungen kann und wird in der Literatur auch oft als monopolistischer Wettbewerbsmarkt (monopolistic competition) bezeichnet.<sup>107</sup> Ein solcher Markt zeichnet sich dadurch aus, dass ein jeder der relativ vielen Anbieter ein differenziertes Produkt verkaufen kann. Im Falle des Gesundheitsmarktes kommt es zur Produktdifferenzierung durch die Bevorzugung eines bestimmten Arztes durch den Patienten aufgrund tatsächlicher oder vermuteter Unterschiede in Verfügbarkeit und Qualität der Dienstleistung (Dranove und Satterthwaite 2000). Als Konsequenz kann er entweder den Preis erhöhen oder die Qualität senken, ohne dass er alle seine Patienten verlieren würde. Mit anderen Worten sieht sich jeder Anbieter einer fallenden Nachfragekurve gegenüber (McGuire 2000). Intuitiv kann man drei Ursachen dafür angeben (Feldstein 1999, Dranove und Satterthwaite 2000):

- a) Konsumenten betrachten oft den Kauf von Gesundheitsleistungen als eine notwendige Sache. Deshalb reagiert die nachgefragte Menge nicht auf den Preis, den sie bezahlen müssen.
- b) Konsumenten müssen bei der Wahl eines Anbieters auf dem Gesundheitsmarkt sehr viele nichtmonetäre Attribute bewerten. Daher muss schon sehr viel Negatives passieren, bis ein Patient seinem Arzt den Rücken kehrt.
- c) Aufgrund der Versicherung tragen die Konsumenten nicht die vollständige Last des Preisdifferentials zwischen einem sehr teuren Anbieter und anderen Anbietern. Dadurch wird die Wichtigkeit des Preises relativ zu anderen Bestimmungsgründen der Anbieterwahl reduziert.

Darüber hinaus hängt die Marktmacht des Einzelnen entscheidend von der Gesamtzahl der Anbieter ab. Allgemein würde man erwarten, dass die Marktmacht eines Anbieters dort am

---

<sup>107</sup> Vgl. bspw. Dranove und Satterthwaite (2000).

größten ist, wo er die geringste Konkurrenz zu befürchten hat. Denn dann kann er in einer monopolähnlichen Stellung den Patienten mehr Leistungen anbieten, als diese bei vollständiger Information nachfragen würden. Nimmt der Wettbewerb im Markt zu, erhöht sich die Wahrscheinlichkeit, dass unzufriedene Patienten einen anderen Arzt suchen. Damit wird die Möglichkeit einer zu starken Mengenausweitung durch den zunehmenden Wettbewerb begrenzt.

Andererseits gibt es auch einen Erklärungsansatz, der zu genau entgegengesetzten Ergebnissen führt. Dieser Ansatz geht von der Vorstellung aus, dass die Höhe der Suchkosten entscheidend für die Wechselbereitschaft ist. Unter Suchkosten werden dabei nicht in erster Linie die Kosten für die Überwindung von Entfernungen verstanden, sondern die direkten Kosten bei der Auswahl eines Anbieters. Gibt es nur ein paar Anbieter, so sind die Informationskosten für wechselwillige Konsumenten gering. Versteht man die Suchkosten in diesem Sinn, so stehen sie in einem umgekehrt proportionalen Zusammenhang sowohl zur Entfernung des Patienten vom Arzt als auch zur Arztdichte: Die Suchkosten für die Patienten sind dort hoch, wo die Entfernungen der Patienten zum Arzt gering sind, d.h. die Arztdichte hoch ist (vgl. hierzu Frech III (1996), S. 180 ff.).

Das hier vorgestellte monopolistische Konkurrenzmodell geht in seinen Grundzügen auf ein Modell von T. G. McGuire (McGuire 2000) zurück. Es wird zunächst davon ausgegangen, dass der Arzt sowohl über den Preis seiner Dienstleistung als auch über die Menge selbst entscheiden kann. Im weiteren Verlauf wird dann der Fall betrachtet, bei dem der Arzt nur über die Wahl der Menge seinen Gewinn maximieren kann, nicht aber über den Preis.

Die Nachfrage sei jetzt durch die Nutzenfunktion der Patienten spezifiziert. Diese Vorgehensweise erlaubt es, auch Preis-Mengenkombinationen zu betrachten, die nicht auf der Nachfragekurve liegen.<sup>108</sup>

Es bezeichne  $U(x)$  den Nutzen des Patienten aus einer Behandlung ( $x$ ) und  $NU(x) = U(x) - px$  den Nettonutzen (Nutzen minus Behandlungskosten) zum Preis der Behandlung  $p$ . Es gilt  $U'(x) > 0$  und  $U''(x) < 0$ .

---

<sup>108</sup> Eine derartige Möglichkeit ergibt sich direkt aus der Eigenschaft dieser Dienstleistung der Nicht-Wiederverkäuflichkeit.

Der Patient wird genau dann einen bestimmten Arzt aufsuchen, wenn sein Nettonutzen aus dem Arztbesuch mindestens so groß wie der Nettonutzen beim Aufsuchen eines anderen Arztes ist:  $(NU(x) \geq NU^0)$ .

Auf einem Wettbewerbsmarkt, auf dem der Arzt seine Dienstleistung zu Grenzkosten verkauft, ist der Nettonutzen am größten. Das bedeutet aber auch, dass der Patient durch einen Arztwechsel in diesem Fall keine Nutzensteigerung erfahren kann.

Für einen gewinnmaximierenden Arzt mit Marktmacht ergibt sich folgendes Maximierungsproblem:

$$\begin{aligned} \text{Max } \Pi &= p x - p_x x \\ \text{u.d.N. } U(x) - p x - NU^0 &\geq 0, \end{aligned} \tag{6.7}$$

mit  $p_x$  den Grenzkosten der Behandlung. Dabei spiegelt die Nebenbedingung die Wettbewerbssituation für den einzelnen Arzt wider: Er muss einem Patienten mindestens ein Nutzenniveau  $NU^0$  zusichern, damit dieser ihn wählt. Wäre der Arzt in einem derartigen Markt Monopolist, würde der alternative Nettonutzen der Patienten  $NU^0 = 0$  betragen, da die Patienten keine Alternative zu dem einen Arzt hätten. Das andere Extrem des vollkommenen Wettbewerbsmarktes ergibt sich, wenn das Aufsuchen eines anderen Arztes mit keinen zusätzlichen Kosten für den Patienten verbunden wäre  $(NU(x) = NU^*)$ .

Die Referenzlösung in einem derartigen Modell der monopolistischen Konkurrenz erhält man durch die Maximierung obiger Funktion über den Preis und die Menge, den Entscheidungsvariablen des Arztes, und unter Beachtung der Nettonutzennebenbedingung des Patienten. Wir erhalten 3 Bedingungen:<sup>109</sup>

$$\begin{aligned} x - \lambda x &= 0 \\ p - p_x + \lambda(U'(x) - p) &= 0 \\ U(x) - p x - NU^0 &= 0. \end{aligned} \tag{6.8}$$

Aus der ersten Bedingung ergibt sich  $\lambda = 1$ , was den Umstand widerspiegelt, dass der Arzt in der Lage ist, jeden zusätzlichen Nutzengewinn oberhalb von  $NU^0$  für sich zu verbuchen.<sup>110</sup> Aus der zweiten Bedingung erkennt man dann, dass die Grenzkosten gleich dem Grenznutzen

---

<sup>109</sup> Hier – wie bei den folgenden Modellen – wird davon ausgegangen, dass auch die Bedingungen zweiter Ordnung für ein Gewinnmaximum erfüllt sind.

gesetzt werden ( $p_x = U'(x)$ ), was zu einem effizienten Mengenangebot führt ( $x^*$ ). Aus der dritten Bedingung ergibt sich schließlich  $p$ :

$$p = \frac{U(x^*) - NU^0}{x^*}.$$

Um zu sehen, dass der vom Anbieter geforderte Preis über dem Grenzkostenpreis liegt, kann die vom Arzt gewählte optimale Preis-Mengen-Kombination ( $p, x^*$ ) verwendet werden, um den Gewinn zu berechnen:

$$(p - p_{x^*})x^* = NU(x^*) - NU^0. \quad (6.9)$$

M.a.W. wählt ein Monopolist die Preis-Mengenkombination so, dass er jeglichen Surplus oberhalb von  $NU^0$  für sich verbuchen kann. Der Unterschied zum reinen Monopolmarkt besteht darin, dass der „Monopolist“ auf einem Konkurrenzmarkt die effiziente Menge anbietet. Allerdings bietet er diese Menge zu einem höheren als dem effizienten Preis an ( $p > p_x$ ). Der Monopolist auf einem reinen Monopolmarkt bietet hingegen eine geringere Menge zu einem höheren Preis an – verglichen zur Situation vollständiger Konkurrenz (vgl. auch Abschnitt 6.1).

Eine grafische Lösung dieses Problems ist in Abbildung 4 dargestellt. Das Gleichgewicht unter Wettbewerb wird durch die Menge  $x^*$  und dem Preis  $p = p_x$  gekennzeichnet. Es führt zu einem Nettonutzen für den Patienten in Höhe von  $NU(x^*)$ . Die Fläche des großen Dreiecks unterhalb der Grenznutzenkurve bis zur Grenzkostenkurve  $p_x$  entspricht gerade diesem Nettonutzen (Fläche ABC). Der Gewinn des Arztes bei vollkommenem Wettbewerb ist Null. Besitzt der Arzt dagegen Marktmacht kann er den Preis erhöhen, ohne dass sich die nachgefragte Menge verringert. Durch den Preisanstieg verringert sich der Nettonutzen des Patienten. Der Nutzen des Patienten wird zwar weiterhin durch die konsumierte Menge  $x^*$  bestimmt und entspricht der Fläche  $ABx^*0$ . Allerdings sind jetzt die Kosten für ihn höher, da der Arzt einen höheren als den Grenzkostenpreis verlangt. Die Kosten für den Patienten entsprechen der Fläche  $DFx^*0$ , sodass sein Nettonutzen nur noch  $AED-EFB$  beträgt. Der Gewinn des Arztes ist jetzt positiv und entspricht vollständig der Nutzeneinbuße des Patienten ( $DFBC$ ). Der Arzt kann aber den Preis – und damit seinen Gewinn – nur solange erhöhen, bis er dem Patienten

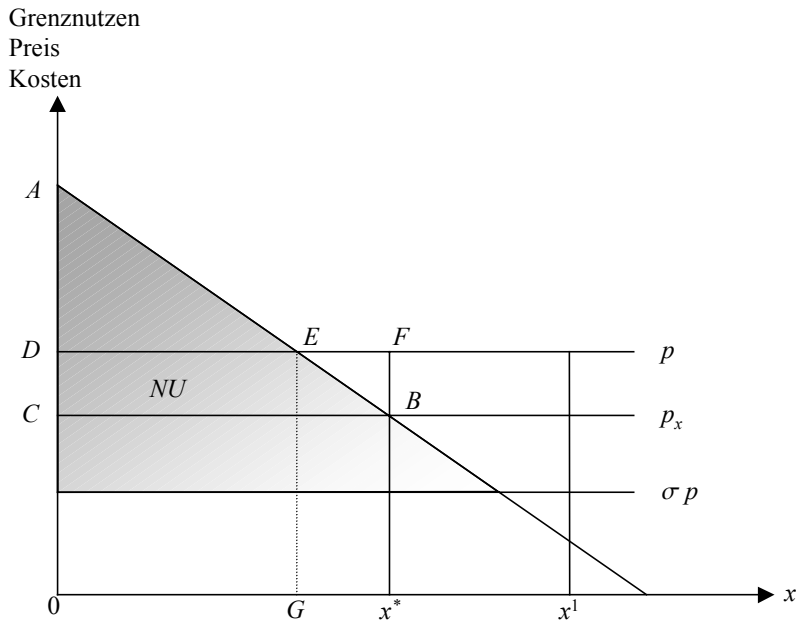
---

<sup>110</sup>  $\lambda$  ist ein Lagrangemultiplikator, der mit der Nebenbedingung der Maximierung verbunden ist.

gerade noch den kleinsten Nettonutzen garantiert, sodass der Patient ihm die Treue hält:

$$NU(x) \geq NU^0.$$

Abbildung 4 McGuire-Modell: Monopolistische Konkurrenz



Der Markt für Arztdienstleistungen in der Schweiz zeichnet sich dadurch aus, dass der Preis für den einzelnen Arzt ein Datum ist. In diesem Fall erfolgt die Maximierung nur durch die Wahl der Menge. Die Bedingungen erster Ordnung lauten dann:

$$\begin{aligned} p - p_x + \lambda(U'(x) - p) &= 0 \\ U(x) - p x - NU^0 &= 0. \end{aligned} \tag{6.10}$$

Diese beiden Bedingungen bestimmen  $x$  und  $\lambda$  eindeutig, wobei  $\lambda$  jetzt nicht notwendigerweise eins sein muss. Über die zweite Bedingung in (6.10) kann  $x$  bestimmt werden. Formt man die zweite Bedingung in (6.10) in ähnlicher Weise wie in (6.8) um, erhält man wiederum Gleichung (6.9). Da der Preis jetzt aber fest vorgegeben ist, wird der Arzt versuchen, die Menge so lange zu erhöhen, wie er noch einen positiven Gewinn aus dem Verkauf der zusätzlichen Einheit erzielen kann. Darüber hinaus lässt sich über diese Bedingung untersuchen, wie sich der Arzt bei einer Preiserhöhung bzw. Preissenkung verhalten wird. Über das Totale Differential erhält man:

$$\frac{dx}{dp} = \frac{-x}{p - U'(x)} < 0. \tag{6.11}$$

D.h. wenn in einem solchen Fall die Preise gesenkt werden, reagiert der Arzt mit einer Mengenausweitung, um die Gewinnschmälerung aufzufangen. Dies gilt so lange wie der fixe Preis oberhalb der Grenzkosten liegt und der Nettonutzen der Patienten nicht unter eine bestimmte Schwelle fällt, die den Patienten veranlassen könnte, den Arzt zu wechseln. Ein derartiges Verhalten wird in der Literatur als angebotsinduzierte Nachfrage (AIN) bezeichnet (vgl. bspw. Fuchs (1974) oder Pauly (1980)). Interessant ist in diesem Zusammenhang, dass Bedingung (6.11) aus einem sehr einfachen Modell ohne explizite Formulierung einer Nachfrageinduzierung oder einer Zieleinkommensmotivation der Ärzte hergeleitet wurde, wie sie in der einschlägigen Literatur oft angenommen wird. Danach weitet der Arzt die Nachfrage zur Erhöhung seines Gewinnes über die vom Patienten normalerweise gewünschte Nachfrage aus. Eine „normalerweise gewünschte Nachfrage“ ist schwierig zu bestimmen. Ganz allgemein kann man sagen, dass diese Nachfrage diejenige Leistungsmenge ist, die ein Arzt mit dem gleichen Informationsstand wie der Patient letzterem anbieten würde. Allerdings ist es aufgrund der verschiedenen Probleme in der Arzt-Patienten-Beziehung nicht einfach, eine solche optimale Leistungsmenge zu bestimmen. Dabei lassen sich die angesprochenen Schwierigkeiten sowohl durch die beschriebenen Informationsasymmetrien im 5. Kapitel, als auch durch die in diesem Kapitel dargelegten Gründe (fehlende Konsumentensouveränität und die Versicherungsdeckung der Patienten) beschreiben. Damit scheinen in diesem einfachen Beispiel schon diejenigen Probleme auf, die uns später bei der empirischen Analyse des Arztverhaltens noch beschäftigen werden. Denn von Angebotsinduzierung kann nur gesprochen werden, wenn der Ausgangspunkt der Betrachtung eine Situation ist, wie wir sie durch die Menge  $x^*$  in Abbildung 4 gekennzeichnet haben. Wäre die Ausgangssituation dagegen durch einen Nachfrageüberhang, also  $x < x^*$  gegeben, würde eine Preissenkung zu einem Abbau des Überhanges führen. Empirisch ist es in der Regel schwierig, zwischen beiden Situationen zu unterscheiden und klare Rückschlüsse auf mögliche Angebotsinduzierung zu ziehen.

Des Weiteren lässt Gleichung (6.11) deutlich erkennen, dass Ärzte auch ohne die Möglichkeit der Preissetzung noch Marktmacht besitzen. Damit zeigt sich die vorher schon prosaisch geäußerte Auffassung der Marktmacht von Ärzten bestätigt.

Bisher wurde nur die Preissetzung eines Dritten bezüglich des Leistungserbringers berücksichtigt. Im Folgenden soll dies auf die Nachfrageseite ausgeweitet werden. In der Regel ist dieser Dritte eine Krankenkasse. Die Versicherungsdeckung hat zur Folge, dass der Patient sich zum Zeitpunkt der Behandlung nur einem Teil der Behandlungskosten gegenüber sieht. Der Kosten

anteil des Patienten sei wieder mit  $\sigma$  bezeichnet ( $0 < \sigma < 1$ ). Der Preis  $p$ , den der Arzt für seine Leistung erhält, liegt oberhalb der Grenzkosten und hängt nicht von der Selbstbeteiligung  $\sigma$  ab. Der Nettonutzen des Patienten wird hingegen von  $\sigma p$  mitbestimmt. Aufgrund der Veränderung des Nettonutzens des Patienten verändert sich das Maximierungsproblem des Arztes in (6.7), sodass sich auch die Bedingungen erster Ordnung ändern (wieder bei feststehendem Preis):

$$\begin{aligned} p - p_x + \lambda(U'(x) - \sigma p) &= 0 \\ U(x) - \sigma p x - NU^0 &= 0. \end{aligned} \tag{6.12}$$

Aus der zweiten Gleichung in (6.12) lässt sich wiederum  $x$  bestimmen. Analog zu (6.9) kann sie umgeformt werden und man erhält mit  $x^1$  die Lösung mit Versicherung:

$$(p - p_{x^1}) \sigma x^1 = NU(x^1) - NU^0. \tag{6.13}$$

Auch hier wird der Arzt versuchen die Leistungsmenge auszuweiten, da er mit jeder verkauften Einheit seinen Gewinn steigern kann. Begrenzt wird sein opportunistisches Verhalten durch die Nettonutzennebenbedingung des Patienten. Aus der ersten Bedingung in (6.12) ist zudem zu sehen, dass so lange  $p > p_x$  gilt, auch  $\sigma p > U'(x)$  gelten muss, da  $\lambda$  positiv ist. Im Vergleich mit der Lösung ohne Versicherung ergibt sich, dass  $x^1 > x^*$  gilt, da zusätzlich zur Marktmacht der Anbieter auf der Nachfrageseite das Moral-Hazard-Problem hinzutritt. Die Versicherungslösung ist ebenfalls in Abbildung 4 dargestellt. Sie macht deutlich, dass sich damit auch das Problem einer Untersuchung der angebotsinduzierten Nachfrage verschärft. Denn ein Teil der (zu hohen) beobachteten Nachfrage beruht eben nicht direkt auf der Angebotsinduzierung seitens der Ärzte, sondern auf der nachfrageseitigen (zu hohen) Inanspruchnahme der Konsumenten. Allerdings ist auch klar, dass Ärzte die zusätzlichen Wünsche ihrer Patienten gern erfüllen, sodass eine eindeutige Trennung zwischen Angebotsinduzierung und Moral Hazard sehr schwer ist.

Moral Hazard auf der Nachfrageseite wird durch die Selbstbeteiligung der Versicherten begrenzt. Ob eine angebotsseitige Kostenbeteiligung genau wie die Kostenbeteiligung auf der Nachfrageseite zu einer Begrenzung der Leistungsmenge führen kann und welche Probleme damit verbunden sind, wird im folgenden Abschnitt diskutiert.

### 6.3 Angebotsseitige Kostenbeteiligung, Qualität und Wettbewerb um Patienten

Bisher war die Vergütung durch  $p \cdot x$  vorgegeben. Im Folgenden soll genauer untersucht werden, wie unterschiedliche Vergütungsformen das Verhalten des Arztes beeinflussen können und wie sich die Wahl der Vergütungsform auf die Qualität der Versorgung und den Wettbewerb um die Patienten auswirken.<sup>111</sup>

Nach wie vor werden – im Unterschied zu den Modellen im 5. Kapitel – vollständig informierte Parteien vorausgesetzt. Aus der Sicht des Patienten lassen sich verschiedene Faktoren unterscheiden, die den Behandlungserfolg und somit den Nutzen erhöhen sollen. Zunächst wird weiterhin die Menge der Leistungen betrachtet, die bspw. nach der Zahl und der Länge von Konsultationen gemessen werden kann. Die Leistung hat einen Preis pro Einheit, sodass sie als Basis für ein Vergütungssystem dienen kann. Davon unterschieden werden Faktoren, die nicht oder nur mit sehr großem Aufwand gemessen werden können. Diese Faktoren sollen unter einem Faktor *Anstrengung* subsumiert werden. Darunter fallen zum einen die Zeit, die ein Arzt für eine Konsultation oder Prozedur aufwendet und zum anderen auch solche Faktoren wie bspw. Fleiß, Vorsicht und Aufmerksamkeit. Während die zuletzt genannten Faktoren, die im Faktor Anstrengung zusammengefasst werden, eher qualitativer Natur sind, umfassen die vertraglich verifizierbaren Faktoren eher quantitative Einflüsse.<sup>112</sup>

Die Anstrengung des Arztes sei wieder mit  $e$  bezeichnet. Diese hat einen positiven Einfluss auf den Behandlungserfolg und somit auch auf den Patientennutzen. Gleiches gilt für die Menge der medizinischen Leistungen  $x$ . Der Nutzen des Patienten aus der Behandlung kann daher geschrieben werden als  $U(x, e)$  mit den Ableitungen  $U_x > 0, U_{x,x} < 0$  sowie  $U_e > 0, U_{e,e} < 0$ . Außerdem wird angenommen, dass die Nutzenfunktion in ihren Argumenten separierbar sei, sodass die Kreuzableitung der Nutzenfunktion Null beträgt ( $U_{x,e} = 0$ ). Die Anstrengung ist für den Arzt ein Kostenfaktor. Dieser Umstand soll durch die Kostenfunktion der Anstrengung  $c(e)$  ausgedrückt werden mit  $c_e > 0$  und  $c_{e,e} > 0$ . Im Unterschied zum Grenznutzen der Anstrengungen für den Patienten steigen die Kosten des Arztes mit den Anstrengungen an.

---

<sup>111</sup> Vgl. wiederum McGuire (2000).

<sup>112</sup> Bei der Einbeziehung der Arbeitszeit in die Qualitätsmerkmale wird von gegenwärtigen Vergütungsformen ausgegangen. Diese nehmen in der Regel nicht auf die Zeit Bezug, obwohl sie in gewissem Maße verifizierbar wäre. Vgl. zum Einsatz der Arbeitszeit des Arztes auch 6.1.2.



Nachfolgend sollen nun zwei bereits bekannte Vergütungssysteme der angebotsseitigen Kostenbeteiligung innerhalb dieses Modells betrachtet werden: die prospektive Vergütung und die Kostenerstattung. Dabei sei auch ein gemischtes Vergütungssystem nicht ausgeschlossen, sodass die Vergütungsregel pro Patient allgemein lautet:

$$a + p_s x, \text{ mit } a \geq 0, c > p_s \geq 0,$$

wobei  $a$  die Pauschale pro Patient bezeichnet und  $p_s$  die Zahlung pro Mengeneinheit. Bei einem reinen Kostenerstattungssystem wäre demzufolge  $a = 0$  und  $p_s = c$ , während eine reine Pauschalvergütung  $p_s = 0$  und  $a > 0$  setzen würde. Die Bedingung  $c > p_s$  drückt die angebotsseitige Kostenbeteiligung in einem gemischten System aus. Eine reine Einzelleistungsvergütung – wie sie im vorherigen Abschnitt angenommen wurde – erhielte man demnach bei  $a = 0$  und  $p_s > c$ .

Im Folgenden wird auch der Wettbewerb um die Patienten berücksichtigt, indem die Anzahl der Patienten direkt in die Gewinnfunktion eingeht. Dabei soll die Patientenanzahl der einzelnen Arztpraxis positiv vom Nettonutzen der Patienten abhängen ( $n(NU)$  mit  $n' > 0$ , wobei  $n' = \frac{\partial n}{\partial NU}$  gilt). Für den Arzt bedeuten mehr Patienten c.p. einen höheren Gewinn. Geht man von identischen Patienten aus, so lässt sich die Gewinnfunktion als Produkt aus der Anzahl an Patienten und dem Ertrag pro Patient schreiben:

$$\Pi = n(NU) \left[ a + (p_s - c(e))x \right], \quad (6.14)$$

mit  $NU = U(x, e) - p_d x$ . Wobei  $p_d$  den Preis pro Einheit für die Patienten bezeichnet, der exogen von der Krankenkasse vorgegeben wird.

Durch die Wahl von  $x$  und  $e$  maximiert der Arzt seinen Gewinn. Die dazugehörigen Bedingungen erster Ordnung lauten:

$$n'(U_x - p_d) \left[ a - (p_s - c)x \right] + n(p_s - c) = 0 \quad (6.15)$$

$$n'U_e \left[ a + (p_s - c)x \right] - n c_e x = 0. \quad (6.16)$$

Eine Umstellung ergibt:

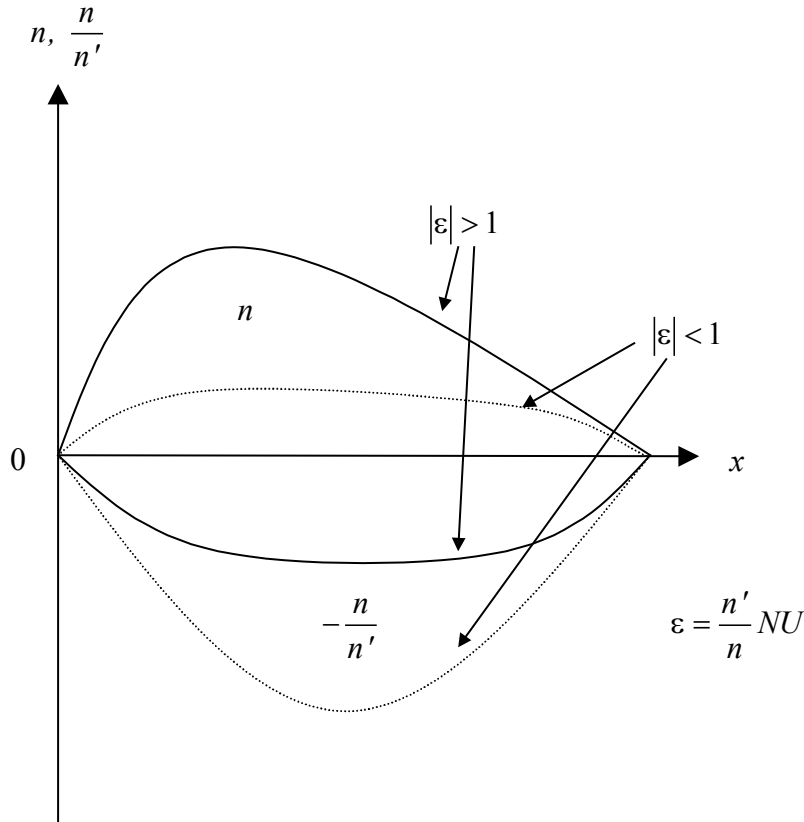
$$(U_x - p_d) \frac{a + (p_s - c)x}{p_s - c} = -\frac{n}{n'} \quad (6.17)$$

$$U_e \left( \frac{a}{x} + (p_s - c) \right) = \frac{n}{n'} c_e. \quad (6.18)$$

Die Bedingungen (6.17) und (6.18) beschreiben die Wahl von  $x$  und  $e$  durch den Arzt in Abhängigkeit der Vergütung.

Zunächst soll die Mengensentscheidung genauer betrachtet werden. Der erste Term auf der linken Seite von (6.17) zeigt die Veränderung des marginalen Nettonutzens der Patienten. Versucht der Arzt, die Leistungsmenge über die vom Patienten gewünschte Menge auszudehnen, wird der Ausdruck negativ ( $U_x$  fällt unter  $p_d$ ). Dieser Ausdruck wird mit dem Term der Vergütungsregel multipliziert und muss im Optimum dem negativen Verhältnis aus bisherigem Patientenstamm und der Veränderung des Patientenstammes entsprechen. In der Regel wird die Funktion der Patientenzahl in Abhängigkeit des Nettonutzens des Patienten zunächst steigen, um ab einem bestimmten Punkt zu fallen. Die Krümmung der Funktion  $n/n'$  hängt entscheidend von der Marktnachfrageelastizität ab: Mit zunehmender Elastizität (= zunehmende Konkurrenz auf dem Markt) wird die Kurve flacher. In Abbildung 5 sind zwei solche Nachfragekurven mit konstanter Elastizität (einmal  $< 1$  und einmal  $> 1$ ) in Abhängigkeit der Menge  $x$  dargestellt:

Abbildung 5 McGuire-Modell:  
Anzahl Patienten und Veränderung der Anzahl bei unterschiedlichen  
Elastizitäten in Abhängigkeit der Leistungsmenge pro Patient



Nun sollen die einzelnen Vergütungsalternativen bezüglich der Mengenentscheidung genauer betrachtet werden. Der erste Fall ergibt sich, wenn  $a = 0$  gesetzt wird und die Vergütung somit keine prospektive Zahlung beinhaltet. In diesem Fall muss der marginale Nettonutzen eines Patienten genau der (negativen) Veränderung des Patientenstammes entsprechen, weil der Vergütungsterm gleich eins ist:

$$(U_x - p_d)x = -\frac{n}{n'}. \quad (6.19)$$

Da die rechte Seite von (6.19) negativ ist, muss auch die linke Seite negativ sein. Dies kann aber nur erfüllt sein, wenn die Grenzkosten des Patienten höher sind als der Grenznutzen. Somit ist eine überoptimale Mengensetzung durch den Arzt gegeben. Eine unbegrenzte Mengenausweitung pro Patient wird allerdings durch die dämpfende Wirkung der Marktnachfrage erreicht. Zunächst nimmt die Attraktivität des Arztes für weitere Patienten mit einem höheren Nettonutzen pro Patient – und damit einer höheren Menge pro Patient – zwar zu, ab einem bestimmten

Punkt wird die steigende Aggressivität in der Verschreibungs- und Behandlungspraxis des Arztes aber zu immer weniger Patienten führen.

Da der Vergütungsterm gleich eins ist, spielt die Unterscheidung zwischen Kostenerstattung und Einzelleistungsvergütung hier keine Rolle. Interessant ist ferner, dass für die Mengensetzung des Arztes das Preisdifferential offenbar uninteressant ist. Er setzt die Menge einzig und allein nach den Patientenwünschen – entsprechend des Grenznutzen und der Grenzkosten des Patienten – und unter Beachtung der Marktnachfrage fest.

Bei einer rein prospektiven Vergütung ( $p_s = 0$ ) reduziert sich (6.17) zu:

$$(U_x - p_d) \begin{pmatrix} \frac{a-c}{x} \\ -c \end{pmatrix} = -\frac{n}{n'} \quad (6.20)$$

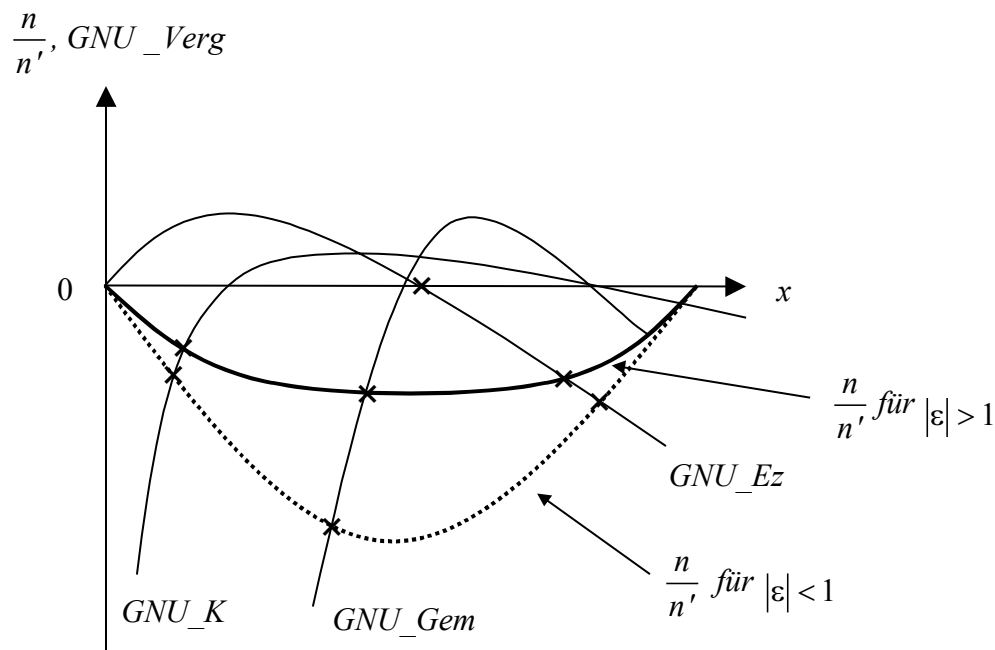
So lange der Durchschnittsertrag die Grenzkosten übersteigt, ist der Vergütungsterm negativ. Der marginale Nettonutzen muss in diesem Falle positiv sein, da die rechte Seite weiterhin negativ ist. M.a.W. muss der Grenznutzen den Nachfragepreis übersteigen. Damit besitzt der Arzt keinen Anreiz mehr, die Menge über das vom Patienten gewünschte Niveau hinaus auszudehnen. Vielmehr wird der Arzt seinen Gewinn erhöhen, wenn er den Patienten weniger Leistungen anbietet; zum einen wegen dem steigenden Durchschnittsertrag und zum anderen wegen der Zunahme des marginalen Nettonutzens. Allerdings kann der Arzt die Mengenreduktion nicht beliebig fortsetzen, da ihm ansonsten sämtliche Patienten abwandern würden.

Die gemischte Vergütung kann die Fehlsteuerungen der Einzelleistungsvergütung und der Kopfpauschalen vermeiden. Denn ihr stehen nun zwei Steuerungselemente zur Verfügung (Kostenerstattungsanteil zur Mengenausweitung und prospektiver Anteil zur Mengenbegrenzung). Im Idealfall (optimale Wahl von  $a$  und  $p_s$ ) ist es möglich, dass eine gemischte Vergütung die Versorgung der Patienten mit der optimalen Leistungsmenge, d.h. einer Leistungsmenge, die der Bedingung  $U_x = p_d$  genügt, sicherstellen kann.

In Abbildung 6 sind die beschriebenen Lösungen für zwei unterschiedliche Marktnachfragen dargestellt. Die äußere Kurve repräsentiert wieder die „unelastische“ Nachfrage. Die schwächer gezeichneten Kurven stellen das Produkt aus Grenznutzenfunktion eines einzelnen Patienten und der Vergütung des Arztes dar ( $GNU_K$ ,  $GNU_{Ez}$  und  $GNU_{Gem}$ ). An den Schnittstellen dieser Funktionen mit den Marktnachfragen finden sich daher die optimalen Mengen unter den

jeweiligen Vergütungssystemen. Die aus Patientensicht optimale Menge ergibt sich aus der Bedingung:  $U_x - p_d = 0$ . Da bei der Einzelleistungsvergütung der Vergütungsterm gleich eins ist, entspricht die Kurve  $GNU_{Ez}$  genau der Nettogrenznutzenkurve des Patienten. Der Schnittpunkt dieser Kurve mit der x-Achse gibt somit die aus Patientensicht optimale Leistungsmenge an.

Abbildung 6 McGuire-Modell: Mengenlösungen



Die Auswirkungen der Anstrengungen auf den Gewinn des Arztes können mit Hilfe von Gleichung (6.18) studiert werden. Die Bedingung beschreibt den grundlegenden Zielkonflikt des Arztes bei der Entscheidung, ob sich zusätzliche Anstrengungen lohnen: Einerseits erhöhen sich die Kosten für den Arzt, andererseits erhöht die Anstrengung die angebotene Qualität und somit auch die Attraktivität des Arztes für die Patienten. Allerdings nehmen die Kosten mit steigenden Anstrengungen weiter zu, während der Grenznutzen des Patienten mit steigenden Anstrengungen sinkt.

Bezüglich der Einzelleistungsvergütung zeigt Bedingung (6.18), dass der Arzt bei hohen Gewinnaussichten die entsprechenden Anstrengungen tätigen wird, während er bei geringeren Gewinnaussichten, seine Anstrengungen reduziert. Das heißt aber in einem solchen System, dass die Gebühr pro Einzelleistung höher als die Grenzkosten ihrer Bereitstellung sein muss, um die gewünschten Anstrengungen zu veranlassen. Bei der rein prospektiven Vergütung gilt analoges,

nur dass hier die Kopfpauschale pro Patient erhöht werden müsste, um höhere Anstrengungen bzw. eine höhere Qualität zu induzieren. Beide Vergütungsregeln generieren bei  $R/x = p_s$  das gleich hohe Anstrengungsniveau. Entsprechend gilt bei  $p_s > R/x$ , dass unter der Einzelleistungsvergütung der Arzt höhere Anstrengungen betreibt. Dies dürfte vor allem bei Patienten der Fall sein, bei denen umfangreiche Behandlungen erforderlich sind. Bei der Betrachtung der gemischten Vergütung lässt sich der letzte Punkt nochmals verdeutlichen: Mit einer höheren Vergütung wird der Arzt größere Anstrengungen bei der Behandlung seiner Patienten unternehmen. Allerdings hat der Regulierer nun zwei Parameter zur Hand, um die Anstrengungen des Arztes (zum Beispiel die Qualität der Behandlung) zu steuern:

$$\frac{de}{dp_s} = \frac{U_e}{\frac{n}{n'}c_{e,e} + c_e U_e - \left(\frac{a}{x} + (p_s - c)\right)U_{e,e}} > \frac{de}{da} = \frac{U_e/x}{\frac{n}{n'}c_{e,e} + c_e U_e - \left(\frac{a}{x} + (p_s - c)\right)U_{e,e}} > 0. \quad (6.21)$$

Bei gegebener Nachfrage und bei gegebener Leistungsmenge pro Patient wird der Arzt durch eine Erhöhung der Einzelleistungsvergütung stärker zu einer Steigerung seiner Anstrengungen animiert als bei einer vergleichbaren Erhöhung der Pauschale.

Nicht nur die Vergütung hat Einfluss auf die Anstrengung des Arztes, sondern auch die Wettbewerbssituation in der Region.<sup>113</sup> Der Einfluss des Wettbewerbs auf die Anstrengungen kann durch Differentiation von (6.18) bestimmt werden:

$$\frac{de}{d(n/n')} = \frac{c_e}{\left(\frac{a}{x} + p_s - c\right)U_{e,e} - c_e U_e - \frac{n}{n'}c_{e,e}} < 0. \quad (6.22)$$

Der Ausdruck ist wegen der negativen zweiten Ableitung von  $U$  negativ – vorausgesetzt, der Vergütungsterm ist positiv. Wie wir gesehen haben, ist die Funktion  $n/n'$  auf einem Konkurrenzmarkt eher klein, während sie bei geringer Markt Elastizität groß wird. Damit erhöht nach (6.22) ein hoher Wettbewerbsdruck die Anstrengungen der Anbieter. Bezieht man diese Aussage wiederum auf die Qualität der ärztlichen Versorgung, so bedeutet sie eine qualitativ bessere Versorgung der Patienten auf einem Wettbewerbsmarkt. Man erkennt auch die unterschiedlichen Auswirkungen des Wettbewerbs bei unterschiedlicher Vergütung. Während der Vergütungsterm unter einem Einzelleistungsregime immer positiv ist, kann er bei einer reinen Kopfpauschale negativ werden. Für Patienten mit schwereren Erkrankungen kann das zur Folge

---

<sup>113</sup> Vgl. Eggleston (2001).

haben, dass sie bei steigendem Wettbewerbsdruck auf die Anbieter weniger Qualität angeboten bekommen im Vergleich zu profitableren Patienten. Die Qualität der Versorgung letzterer Patienten erhöht sich vielmehr.

Um schließlich den Einfluss des Wettbewerbs auf die Aufwandsveränderung des Arztes aufgrund einer Vergütungserhöhung zu bestimmen, differenzieren wir den ersten Ausdruck in (6.21) nach  $n/n'$  und erhalten:

$$\frac{d(d e / d p_s)}{d(n / n')} = - \frac{c_{e,e} U_e}{\left( \frac{n}{n'} c_{e,e} - U_{e,e} \left( \frac{R}{x} + p_s - c \right) \right)^2} < 0. \quad (6.23)$$

Nach (6.23) nimmt mit steigendem Wettbewerbsdruck auch die Sensitivität der Anbieter auf Preisveränderungen zu. Das hat zur Folge, dass die Anstrengungsveränderung aufgrund einer Vergütungsveränderung in einem Markt mit hohem Wettbewerbsdruck größer ist als in einem Markt mit weniger Wettbewerbsdruck.

#### 6.4 Auswirkungen unterschiedlicher Patienten einer Praxis auf Anstrengungen des Arztes

Freipraktizierende Ärzte unterscheiden sich hinsichtlich der Zusammensetzung ihres Patientenstammes. So gibt es Ärzte, die ausschließlich Patienten mit herkömmlichen Versicherungsverträgen behandeln und solche, die sowohl Patienten aus dem konventionellen Sektor und Patienten aus Hausarztmodellen versorgen. Die einzelnen Verträge sind in der Regel mit unterschiedlichen Vergütungssystemen verbunden. Während in den letzten Abschnitten untersucht wurde, welche Auswirkungen derartige Verträge in ihrer reinen Form (ein Arzt behandelt entweder einen MC-Patienten oder einen konventionellen Patienten) auf die Anstrengungen des Arztes ausüben, sollen jetzt die Wechselwirkungen verschiedener Verträge der Patienten auf die Anstrengungen des Arztes untersucht werden.

Glied und Zivin (2002)<sup>114</sup> untersuchen die Wechselwirkungen zwischen dem Anteil an Managed-Care-Versicherten und der Behandlungsintensität für unterschiedliche Patientengruppen. Dieses Modell dient als Ausgangspunkt der folgenden Analyse.

---

<sup>114</sup> Vgl. ebenfalls Glied und Zivin (2000).

Es wird angenommen, dass sich die Ärzte zwei unterschiedlichen Gruppen von Patienten gegenübersehen: einer Gruppe von Patienten mit herkömmlicher Versicherung (mit  $F$  bezeichnet) und Patienten mit MC-Verträgen, wie sie Versicherte in HAM darstellen (mit  $H$  bezeichnet). Der Versichertenanteil einer Praxis, der zu solchen Hausarztmodellen gehört, soll mit  $\alpha$  bezeichnet werden, während  $\alpha_j$  den Anteil von MC-Versicherten im Markt  $j$  repräsentiert. Es bezeichne  $x_i$  die Produktionsfunktion für die Leistung pro Patient und  $p_i$  den entsprechenden Preis pro Leistung ( $i = F, H$ ), wobei angenommen wird, dass  $p_H$  eine Funktion von  $\alpha$  ist ( $p_H = p_H(\alpha)$ ) und  $p_F > p_H$  gelte. Die Modellierung schließt auch den Fall der Kopfpauschalvergütung bei MC-Verträgen mit ein. Denn indirekt hängt das Einkommen des Arztes bei einer Kopfpauschalvergütung auch von den Leistungen ab. So muss der Arzt ein bestimmtes Niveau der ärztlichen Versorgung sicherstellen, um keine Patienten zu verlieren. Es entsteht so ein impliziter Preis für jede Leistung.

Die ärztlichen Anstrengungen seien wiederum mit  $e_i$  bezeichnet und die Kosten mit  $c_i$ . Während die angebotene Menge an Leistungen für HAM-Versicherte außer vom gewählten Anstrengungsniveau auch vom Anteil der HAM-Versicherten auf dem Markt abhängt ( $x_H = x_H(e_H, \alpha_j)$ ), wird die produzierte Menge für die konventionell Versicherten in dieser Arztpraxis von den Anstrengungen für diese Patienten und dem Anteil der in dieser Praxis behandelten HAM-Patienten beeinflusst ( $x_F = x_F(e_F, \alpha)$ ). Die Kosten hängen ausschließlich von den ärztlichen Anstrengungen ab. Es wird also vereinfachend angenommen, dass die Kosten für die beiden Patientengruppen nicht vom Anteil der in Hausarztnetzen versicherten Patienten abhängt.<sup>115</sup> Der Arzt maximiert seinen Gewinn, indem er die entsprechenden Anstrengungsniveaus für beide Patientengruppen optimal wählt:

$$\max_{e_F, e_H} \alpha p_H(\alpha) x_H(e_H, \alpha_j) + (1 - \alpha) p_F x_F(e_F, \alpha) - \alpha c_H(e_H) - (1 - \alpha) c_F(e_F). \quad (6.24)$$

Für den Produktionsprozess wird angenommen, dass die Leistung mit zunehmenden Anstrengungen steigt, allerdings mit abnehmenden Raten. Die Kosten der Arztproduktion steigen hingegen mit zunehmenden Anstrengungen auch immer stärker an.<sup>116</sup>

---

<sup>115</sup> Die hier gewählte Formulierung des Produktionsprozesses im MC-Bereich unterscheidet sich damit vom Modell in Glied und Zivin (2000) in diesem Punkt:  $c_i(e_i)$  statt  $c_i(e_i, \alpha)$ .

<sup>116</sup> Formal heißt das:  $x' > 0, x'' < 0$  und  $c' > 0, c'' > 0$ .



Die Bedingungen erster Ordnung für dieses Maximierungsproblem sind damit gegeben durch:

$$(1-\alpha) p_F \frac{\partial x_F}{\partial e_F} - (1-\alpha) \frac{\partial c_F}{\partial e_F} = 0 \quad (6.25)$$

$$\alpha p_H \frac{\partial x_H}{\partial e_H} - \alpha \frac{\partial c_H}{\partial e_H} = 0. \quad (6.26)$$

Diese Bedingungen drücken aus, dass im Optimum der Grenznutzen und die Grenzkosten der ärztlichen Anstrengungen gleich sind.<sup>117</sup>

Im Folgenden soll nun untersucht werden, wie eine Veränderung des Managed-Care-Anteils innerhalb der Praxis sich auf das Anstrengungsniveau für HAM-Patienten und für konventionelle Patienten auswirkt. In wie weit eine Veränderung des Managed-Care-Anteils auf Marktebene die Anstrengungsniveaus für die beiden Patientengruppen beeinflusst, ist ebenfalls Gegenstand der Untersuchung. Zu diesem Zweck werden die Bedingungen erster Ordnung einer Inspektion bezüglich einer Veränderung von  $\alpha$  bzw.  $\alpha_j$  unterzogen.

#### 6.4.1 Aufwand und Managed-Care-Anteil in der Praxis

Zunächst betrachten wir die Veränderung des Managed-Care-Patientenstammes auf Praxisebene. Durch totale Differentiation von (6.25) und (6.26) lassen sich die uns interessierenden Differentialquotienten berechnen:

$$\frac{d e_F}{d \alpha} = \frac{\left( p_F \frac{\partial x_F}{\partial e_F} - \frac{\partial c_F}{\partial e_F} \right) - (1-\alpha) p_F \frac{\partial^2 x_F}{\partial e_F \partial \alpha}}{(1-\alpha) \left[ \frac{\partial^2 x_F}{\partial^2 e_F} - \frac{\partial^2 c_F}{\partial^2 e_F} \right]} \quad (6.27)$$

$$\frac{d e_H}{d \alpha} = \frac{\left( -p_b \frac{\partial x_H}{\partial e_H} + \frac{\partial c_H}{\partial e_H} \right) - \alpha \frac{\partial p_H}{\partial \alpha} \frac{\partial x_H}{\partial e_H}}{\alpha \left[ p_b \frac{\partial^2 x_H}{\partial^2 e_H} - \frac{\partial^2 c_H}{\partial^2 e_H} \right]}. \quad (6.28)$$

Den ersten Summanden im Zähler bezeichnen wir als Grenzgewinn und den zweiten Summanden als Intensität der Behandlung. Der Grenzgewinn beschreibt, wie sich der Gewinn verändert, wenn sich die Anstrengungen in der relevanten Population verändern. Die Intensität gibt hingegen an, wie sich zusätzliche Umsätze einer Anstrengungserhöhung verändern, wenn sich

---

<sup>117</sup> Die hinreichenden Bedingungen zweiter Ordnung sind aufgrund der Annahmen bezüglich der involvierten Funktionen erfüllt.

die Patientenzusammensetzung auf Praxisebene verändert. Beide Effekte zusammen ergeben die totale Anstrengungsänderung, die durch eine Änderung der Zusammensetzung des Patientenstammes einer Arztpraxis hervorgerufen wird.<sup>118</sup>

Da die Nenner jeweils eindeutig negativ sind, werden die Vorzeichen beider Ausdrücke durch die jeweiligen Zähler bestimmt. Ohne weitere Annahmen sind keine eindeutigen Aussagen über die Vorzeichen möglich. Daher werden zwei Fälle unterschieden. Für beide wird vereinfachend angenommen, dass der Grenzgewinn für konventionell versicherte Patienten wie auch für MC-Patienten positiv (oder an der Grenze Null) sei.

**Fall 1:** Der Anteil der MC-Patienten übt keinerlei Einfluss auf die Produktion der Dienstleistung aus. Hingegen wird von einem Preiseffekt des MC-Anteils ausgegangen, demnach ein höherer MC-Anteil in der Praxis den Preis erhöht, weil sich aufgrund des höheren MC-Anteils die Verhandlungsposition des MC-Arztes gegenüber der Krankenkasse verbessert:

$$\frac{\partial^2 x_F}{\partial e_F \partial \alpha} = 0 \quad \text{und} \quad \frac{\partial p_H}{\partial \alpha} > 0.$$

Damit hat die Intensität der Behandlung keinen Einfluss auf die Anstrengungen im konventionellen Bereich. In dem Maße, wie der Praxisanteil von MC-Patienten zunimmt, wird der Arzt seinen Aufwand für konventionelle Patienten senken und seinen Aufwand für MC-Versicherte erhöhen, da er so seinen Gesamtgewinn erhöhen kann.

$$\frac{de_F}{d\alpha} \leq 0 \quad \text{und} \tag{6.29}$$

$$\frac{de_H}{d\alpha} \geq 0. \tag{6.30}$$

Geht man davon aus, dass die Anstrengungen für MC-Patienten ursprünglich geringer waren als für konventionelle Patienten, so führt eine Erhöhung des MC-Anteils in der Praxis zu einer Angleichung der Anstrengungsniveaus für beide Patientengruppen. Die Angleichung erfolgt damit auf einem niedrigeren Anstrengungsniveau als dem ursprünglichen Anstrengungsniveau für die konventionellen Patienten.

**Fall 2:** Die Ärzte verhalten sich entsprechend der These der angebotsinduzierten Nachfrage. Dies sei allerdings auf den konventionellen Bereich beschränkt, da im Managed-Care-Sektor die

---

<sup>118</sup> Glied und Zivin (2000) bezeichnen diese beiden Effekte als *extensive* und *intensive margin*, wobei sie – wie bereits erwähnt – auch die Kostenveränderung durch die Veränderung des MC-Anteils auf Praxisebene betrachten.

Leistungsanspruchnahme der Versicherten stark überwacht wird und der Arzt aufgrund der Kostenbeteiligung außerdem keinen Anreiz besitzt, die Kosten unnötig auszuweiten. Damit ist die Annahme gerechtfertigt, dass eine Erhöhung des Managed-Care-Anteils in einer Praxis nur für die konventionellen Patienten zu einer Erhöhung der angebotenen Leistungsmenge führt als auch die Intensität erhöht. Zusätzlich wird jetzt angenommen, dass der MC-Anteil der Praxis keinen Einfluss auf den MC-Preis hat:

$$\frac{\partial x_F}{\partial \alpha} > 0, \frac{\partial^2 x_F}{\partial e_F \partial \alpha} > 0 \text{ und } \frac{\partial p_H}{\partial \alpha} = 0.$$

In diesem Fall werden sich die Anstrengungen für einen konventionell Versicherten bei einer Erhöhung des MC-Anteils erhöhen oder leicht senken. Im Vergleich zum ersten Fall erhöhen sich jedoch die Anstrengungen für diese Patientengruppe: Entweder fällt die Reduktion der Anstrengungen nicht so stark aus, oder aber eine Erhöhung des Managed-Care-Anteils führt absolut zu einer Erhöhung der Anstrengungen. Entsprechend der unterstellten Annahmen bezüglich des Preiseffekts sinken die Anstrengungen für die Managed-Care-Versicherten im Verhältnis zum Anstrengungsniveau des ersten Falls. Gleichwohl ist dieses Anstrengungsniveau noch positiv:

$$\frac{de_F}{d\alpha_1} < \frac{de_F}{d\alpha_2} \tag{6.31}$$

$$\frac{de_H}{d\alpha_1} > \frac{de_H}{d\alpha_2} > 0. \tag{6.32}$$

Eine Veränderung der Zusammensetzung des Patientenstamms einer Praxis besitzt einen zweideutigen Effekt auf das Anstrengungsniveau für konventionell versicherte Patienten: Aufgrund eines steigenden Anteils von MC-Patienten sinkt einerseits zwar der Gewinn für diese Patientengruppe, andererseits veranlasst dieser steigende MC-Anteil den Arzt, die Intensität der Behandlung für diese Patientengruppe zu erhöhen, um den Gewinnrückgang aufzufangen.

#### 6.4.2 Anstrengungen und Managed-Care-Konzentration auf Marktebene

Für den Einfluss der Managed-Care-Konzentration im Markt auf die Anstrengungsniveaus der beiden Patientengruppen erhalten wir folgende beiden Ausdrücke:

$$\frac{de_f}{d\alpha_j} = 0 \text{ und} \quad (6.33)$$

$$\frac{de_H}{d\alpha_j} = \frac{-p_H \frac{\partial^2 x_H}{\partial e_H \partial \alpha_j}}{p_H \frac{\partial^2 x_H}{\partial e_H^2} - \frac{\partial^2 c_H}{\partial e_H^2}}. \quad (6.34)$$

Nach (6.33) führt eine Veränderung des Marktanteils von MC-Patienten zu keiner Veränderung der Anstrengungen für Versicherte im konventionellen Bereich. Die Wirkungen einer Marktanteilsveränderung auf die Anstrengungen für MC-Patienten in der Praxis sind dagegen wieder nicht eindeutig bestimmbar. Aufgrund der getroffenen Annahmen ist lediglich das Vorzeichen des Nenners eindeutig bestimmt (negativ). Daher unterscheiden wir wieder zwei mögliche Fälle:

**Fall 1:** Arztpraxen unterscheiden sich nicht unwesentlich in ihrem Praxisstil voneinander. So wird vermutet, dass Ärzte, die in MC-Modellen arbeiten, einen Praxisstil pflegen, bei dem sie weniger Leistungen verursachen als andere Ärzte. Allerdings müssen sie ihren eigenen Praxisstil auch immer mit demjenigen anderer Ärzte im Markt vergleichen. Je größer der Marktanteil von Managed Care in einer bestimmten Region ist, desto eher kann sich auch ein „Managed-Care-Praxisstil“ durchsetzen. Anders ausgedrückt bedeutet dies: Je geringer der Marktanteil von Managed Care ist, desto größer wird die Anpassung des eigenen Praxisstils an den konventionellen Praxisstil ausfallen. Für das hier betrachtete Modell heißt das, dass die Ärzte bei einem steigenden Managed-Care-Anteil besser in der Lage sind, die angebotene Leistungsmenge für die Managed-Care-Patienten zu reduzieren:

$$\frac{\partial x_H}{\partial \alpha_j} < 0 \text{ und } \frac{\partial^2 x_H}{\partial e_H \partial \alpha_j} < 0.$$

Die Inspektion von (6.34) führt unter dieser Annahme daher zu dem Ergebnis:

$$\frac{de_H}{d\alpha_j} < 0. \quad (6.35)$$

MC-Ärzten fällt es damit bei steigendem Managed-Care-Anteil in einer Region immer leichter, ihre Anstrengungen für MC-Patienten zu reduzieren, da die Gefahr, dass sie unzufriedene MC-

Patienten verlieren, immer geringer wird. Da sich gleichzeitig das Anstrengungsniveau für konventionell Versicherte nicht verändert, steigt mit zunehmendem MC-Anteil im relevanten Markt die Diskrepanz in den Anstrengungsniveaus für beide Patientengruppen.

**Fall 2:** Als einen speziellen Praxisstil lässt sich das Selektionsverhalten von Ärzten auffassen. Damit ist gemeint, dass Ärzte versuchen, ganz gezielt möglichst gesunde Versicherte als MC-Patienten zu bekommen. Bei einem steigenden MC-Anteil im Markt bedeutet dieses Verhalten aber, dass sie gezwungen sein werden, auch weniger gesunde Patienten aufzunehmen, weil die gesündesten Patienten schon in MC-Praxen behandelt werden. Damit erhöht sich mit zunehmendem MC-Anteil im Markt die Leistungsmenge wie auch die Intensität der angebotenen Leistung:

$$\frac{\partial x_H}{\partial \alpha_j} > 0, \quad \frac{\partial^2 x_H}{\partial e_H \partial \alpha_j} > 0.$$

Die Inspektion von (6.34) zeigt schließlich, dass die Anstrengungen für MC-Patienten bei einem steigenden MC-Anteil auf Marktebene zunehmen:

$$\frac{\partial e_H}{\partial \alpha_j} > 0. \tag{6.36}$$

Damit würde eine Zunahme des MC-Anteils im Markt zu einer Angleichung des Behandlungsaufwands für MC-Patienten an den Behandlungsaufwand für konventionell Versicherte innerhalb der Praxis führen.

Allerdings dürfte dieses Ergebnis auch von der absoluten Höhe des MC-Marktanteils abhängen. So kann man vermuten, dass es den Ärzten gerade bei einem sehr geringen Marktanteil gelingen kann, besonders gute Risiken aus dem konventionellen Versichertenkollektiv herauszulösen. Damit würde zunächst mit zunehmendem MC-Marktanteil der Aufwand für die MC-Patienten sinken. Erst ab einer bestimmten Schwelle des MC-Anteils käme es nach dieser Sicht zu dem eingangs beschriebenen Effekt, welcher die Ärzte zwingt, auch weniger gute Risiken als MC-Patienten aufzunehmen.

### Wechselwirkungen mit MC-Praxisanteil

Bisher wurde davon ausgegangen, dass die beiden Effekte – MC-Praxisanteil und MC-Konzentration auf dem relevanten Markt – unabhängig voneinander seien. Natürlich ist dies nur eine modellhafte Vereinfachung. Denn der Praxisanteil von MC-Versicherten wird in der

Realität stark von den Marktverhältnissen mitbestimmt: So kann man davon ausgehen, dass Ärzte in Regionen mit einer hohen MC-Konzentration auch einen hohen Anteil an MC-Patienten besitzen – zumindest einen höheren als Ärzte in Regionen mit einem vergleichsweise niedrigen MC-Anteil. Man kann mit anderen Worten von einer hohen positiven Korrelation zwischen MC-Praxisanteil und MC-Konzentration ausgehen.

Bezieht man diese Korrelation in die Betrachtung ein, ergeben sich interessante Wechselwirkungen. So kann ein selektives Verhalten der Ärzte auf Märkten mit einem hohen MC-Anteil den Effekt der Anstrengungserhöhung für die MC-Patienten in der eigenen Praxis verstärken. Steht für den Arzt hingegen die Durchsetzung eines effizienten Praxisstils im Vordergrund, wirken die Effekte in unterschiedliche Richtungen: Einerseits senkt eine höhere MC-Konzentration im Markt die Anstrengungen für die MC-Patienten in der Praxis, andererseits führt eine höhere MC-Konzentration zu einem höheren MC-Anteil in der Praxis und somit zu einer Steigerung der Anstrengungen.

Auch unter Berücksichtigung dieser Wechselwirkungen dürfte sich die Richtung der beschriebenen Effekte nicht ändern. D.h., dass Ärzte mit einer gemischten Praxis auf Veränderungen ihres Patientenstammes reagieren: Verändert sich bei der Produktion der Arztdienstleistungen aufgrund einer Veränderung der Zusammensetzung des Patientenstammes selber nichts und verbessert sich gleichzeitig durch die Erhöhung der eingeschriebenen MC-Patienten die Verhandlungsposition der Ärzte, so ist zu erwarten, dass das Anstrengungsniveau für konventionell versicherte Patienten sinkt, während es für MC-Patienten steigt. Geht man hingegen von einem angebotsinduzierenden Verhalten der Ärzte aus, steigen die Anstrengungen für konventionelle Patienten. Die Veränderung des Anstrengungsniveaus für MC-Patienten würde ebenfalls positiv sein, gleichzeitig aber auch von der MC-Konzentration auf Marktebene abhängen.

### **6.5 Zusammenfassung**

In diesem Kapitel wurde das Verhalten der Ärzte unter unterschiedlichen Wettbewerbsbedingungen untersucht. Wie jede andere Firma auch, versucht die Arztpraxis durch ihre Tätigkeit ein Einkommen zu erzielen. Mit diesem Bestreben steht der einzelne Arzt mit anderen Ärzten in Konkurrenz. Während auf einem Markt mit vollständiger Konkurrenz der Arzt seine Leistungen zu Grenzkosten anbieten würde, wären die Kosten für die Patienten auf einem Monopolmarkt erheblich höher: Entsprechend der Theorie würde der Arzt hier einen Preis in Höhe seines

Grenzertrags verlangen, was gleichzeitig dazu führen würde, dass viele Konsumenten diesen Preis nicht zahlen und keine medizinischen Leistungen in Anspruch nehmen könnten.

Nicht nur aufgrund der räumlichen Marktabgrenzung (wenige Anbieter und viele Nachfrager auf einem Markt) lässt sich der Markt für Arztdienstleistungen als monopolistische Konkurrenz beschreiben. Weitere Merkmale, die für eine derartige Marktform sprechen, liegen in den Eigenschaften des zu verhandelnden Gutes und den Präferenzen der Nachfrager begründet: einerseits Nichtwiederverkäuflichkeit von Arztdienstleistungen, andererseits heterogene Präferenzen der Nachfrager.

Diese Eigenschaften des Marktes führen dazu, dass der Arzt Leistungen auf einem Niveau anbieten kann, die nicht notwendigerweise den Präferenzen der Patienten entsprechen. Die Versicherungsdeckung der Patienten erhöht dabei im Allgemeinen die Differenz zwischen angebotener und optimaler Leistungsmenge. Durch die Wahl einer geeigneten Vergütung kann ein Regulierer das Verhalten der Ärzte allerdings steuern. In Abhängigkeit der gewählten Vergütung (Kopfpauschale oder Einzelleistungsvergütung) werden dabei die angebotenen Leistungsmengen stark differieren: Während der Arzt bei der Einzelleistungsvergütung einen Anreiz besitzt, die Leistungsmenge über die vom Patienten gewünschte Menge auszuweiten, wird es unter Kopfpauschalen zu einem geringeren Angebot kommen. Unter diesen Bedingungen sollte ein Vergütungsvertrag der Ärzte aus einem Pauschalanteil und einem leistungsabhängigen Anteil bestehen. Durch die optimale Wahl der Pro-Kopf-Vergütung und des Leistungsanteils kann der Regulierer die Leistungsanbieter zur optimalen Mengensetzung veranlassen.

Steigender Wettbewerb begrenzt die Marktmacht des einzelnen Arztes: Bei Einzelleistungsvergütung begrenzt er die Mengenausweitung – unter Pauschalvergütung gibt er Anreize, mehr Leistungen anzubieten.

Beide Vergütungssysteme bieten Anreize einer qualitativ guten Behandlung. Die Qualität der Behandlung nimmt bei einer Erhöhung der Kopfpauschale oder der Einzelleistungsvergütung zu. Allerdings wird im Allgemeinen die Qualität unter einer reinen Kopfpauschalen-Vergütung geringer ausfallen als unter einer reinen Einzelleistungsvergütung. Schließlich hat auch der Wettbewerb einen positiven Einfluss auf die angebotene Qualität. Wobei zu berücksichtigen ist, dass unter einer reinen Kopfpauschalenvergütung sich die Qualitätsverbesserung nur auf die profitabelsten Patienten beziehen kann.

Die Untersuchung eines gemischten Patientenstammes einer Arztpraxis zeigt, dass es Spill-over-Effekte zu Gunsten der konventionell versicherten Patienten geben kann. Die Richtung dieser Wirkung hängt dabei entscheidend von den Annahmen bezüglich des Arztverhaltens ab. Würde man dem Arzt angebotsinduzierendes Verhalten unterstellen, wäre eine Erhöhung der Anstrengungen vornehmlich für konventionelle Patienten zu erwarten. Verändert der Arzt aufgrund eines höheren MC-Anteils seine Produktionsweise nicht, sondern richtet seine Anstrengungen auf die Aushandlung höherer Preise für seine MC-Patienten, ist mit einer Verringerung der Anstrengungen für konventionelle und einer Erhöhung der Anstrengungen für MC-Patienten zu rechnen.

Im Unterschied zu den Untersuchungen im Kapitel 5 wurden die Ergebnisse in diesem Kapitel ohne die explizite Berücksichtigung der Informationsasymmetrien zwischen den Parteien hergeleitet. Vielmehr spielte die Wettbewerbssituation der Ärzte hier eine wichtige Rolle. Eigenschaften des Marktes wiederum waren allerdings nicht Gegenstand der Betrachtungen im Kapitel 5. Trotz dieser unterschiedlichen Annahmen sind die Ergebnisse beider Kapitel eng miteinander verbunden.



### III EMPIRISCHE UNTERSUCHUNG

#### 7 Untersuchungsgebiet und Daten

Am Anfang der empirischen Untersuchung steht die Charakterisierung des Untersuchungsgebietes. Dabei sind neben den allgemeinen Merkmalen der betrachteten Kantone Zürich und Genf die Preise für die ärztlichen Leistungen sowie die konkreten Vergütungsverträge von besonderem Interesse. Abschnitt 7.1 stellt die Vergütungsverträge zwischen Ärzteschaft und Krankenversicherungen in den beiden Kantonen vor.

Nachdem im ersten Teil der Arbeit die unterschiedlichen Formen des ärztlichen Angebots relativ allgemein diskutiert worden sind, sollen nun die zu untersuchenden Modelle (HAM und HMOs) etwas genauer vorgestellt werden. Dazu wird auf die jeweilige Geschichte der Netze und die konkrete vertragliche Ausgestaltung zwischen dem Versicherer und den Arztnetzen eingegangen (Abschnitt 7.2). Abschnitt 7.3 dokumentiert dann die konkrete Auswahl der Ärzte, die in die Untersuchungen aufgenommen worden sind. Im Abschnitt 7.4 erfolgt überblicksartig die Vorstellung der in der empirischen Untersuchung verwendeten Daten sowie die Erläuterung der Variablenkonstruktion.

Die Daten stammen aus sehr unterschiedlichen Quellen. Leistungsdaten aus dem konventionellen Bereich und die Daten der Versicherten wurden von der Schweizer Krankenkasse CSS zur Verfügung gestellt. Die CSS ist die zweitgrößte Krankenkasse in der Schweiz mit einem Marktanteil von 15.7%.<sup>119</sup> Im Kanton Zürich liegt sie mit einem Marktanteil von 10.6%<sup>120</sup> etwas unter ihrem durchschnittlichen Marktanteil. Mit 13.4%<sup>121</sup> liegt der Marktanteil in Genf ebenfalls im durchschnittlichen Bereich. Die zweite Datenquelle stellt die amtliche KSK-Statistik über die Schweizer Ärzte dar. Ihr wurden die wesentlichen Eigenschaften der Ärzte entnommen. Als dritte Datenquelle konnte die MHCS-Datenbank verwendet werden, in der die Leistungsdaten der CSS-Versicherten für die Hausarzt- und HMO-Modelle abgelegt sind.<sup>122</sup> Um die Leistungen

---

<sup>119</sup> Die Zahlen spiegeln den Stand Ende 1997 wider (CSS-Versicherung 1998).

<sup>120</sup> Vgl. FN. 119.

<sup>121</sup> Vgl. FN. 119.

<sup>122</sup> Genauer gesprochen, ist die MHCS-Datenbank eine Datenbank in der Daten verschiedener Versicherer von verschiedenen Managed Care Netzen gespeichert sind.

der Ärzte im Untersuchungssample für andere Versicherte berücksichtigen zu können, wurde ebenfalls die Leistungsstatistik der abgerechneten Leistungen für die Jahre 1997 bis 1999 verwendet.<sup>123</sup> Die Einbeziehung dieser Statistik erlaubt es gleichzeitig, die Ärzte auf Marktebene miteinander zu vergleichen. Schließlich wurden verschiedene amtliche Statistiken einbezogen, um auf Gemeindeebene entsprechende Einflussgrößen (z.B. Bevölkerung und Fläche) zu berücksichtigen. Gerade die unterschiedliche Herkunft der Daten erschwert den Umgang und die Verknüpfung dieser Daten. Nicht nur etwaige technische Probleme sind davon betroffen, sondern vielmehr auch Probleme bei der Definition und Abgrenzung des Inhalts bestimmter Variablen.

Ärzte behandeln nicht ausschließlich Patienten einer Krankenversicherung. Daher wird im Abschnitt 7.5 untersucht, ob die verwendeten Daten der CSS-Versicherung für die untersuchten Ärzte repräsentativ sind.

#### **7.1 Vergütungsverträge in Zürich und Genf**

Die Vergütungsverträge zwischen den Ärzten und den Krankenkassen für ambulante Leistungen werden zentral auf Kantonsebene durch die lokalen Ärztesellschaften und den Verbänden der Krankenkassen ausgehandelt. In den Verträgen sind neben den allgemeinen Pflichten der Vertragsparteien und ihrer Mitglieder sowie dem Vorgehen bei eventuellen Streitigkeiten vor allem alle Tarifpositionen und ihre Verrechenbarkeit untereinander geregelt.

Die zwei folgenden Abschnitte beschreiben kurz die relevanten Verträge für Zürich und Genf getrennt; der dritte Unterabschnitt versucht, beide Tarifvereinbarungen für ausgewählte Leistungen gegenüberzustellen und mögliche Anreizwirkungen abzuleiten.

##### **7.1.1 Tarifvertrag Zürich**

Der in den Jahren 1997-1999 gültige Tarifvertrag basiert auf einem Grundvertrag aus dem Jahre 1977. Er wurde in den folgenden Jahren immer wieder angepasst und war im Untersuchungszeitraum in der Fassung vom Januar 1995 gültig.

Der Vertrag wurde zwischen der Ärztesgesellschaft des Kantons Zürich und dem Verband der Krankenkassen im Kanton Zürich abgeschlossen. Er ist in einen allgemeinen und einen Teil mit

---

<sup>123</sup> Diese Statistik umfasst alle Abrechnungen der in der Grundversorgung tätigen Ärzte und wird vom KSK erstellt. Die Gesamtstatistik wird dann den einzelnen Krankenversicherungen wieder zur Verfügung gestellt. Für die Bereitstellung dieser Statistik danke ich der CSS.

den einzelnen Tarifpositionen untergliedert. Im ersten Teil wird der Geltungsbereich festgelegt, sind die allgemeinen Pflichten der Vertragsparteien und ihrer Mitglieder geregelt, werden allgemeine Ausführungen zum Tarif gemacht und Fragen von Beanstandungen, Beschwerden und Streitigkeiten erörtert. Hinsichtlich möglicher auftretender Beanstandungen legt der Vertrag zunächst den Verhandlungsparteien formlose gegenseitige Verständigungen nahe. Bei Problemen zwischen dem einzelnen Arzt und der Krankenkasse wird letztere in der Regel vom kasseneigenen Vertrauensarzt vertreten. Kommt es zu keiner gütlichen Einigung, kann der Beschwerdeführer bei der so genannten Blauen Kommission Beschwerde einreichen. Die Zusammensetzung dieser Kommission wird durch die Ärztesgesellschaft in einem separaten Reglement festgelegt.<sup>124</sup> Darüber hinaus gelten für die sogenannte Pauschalbeanstandung, d.h. eine allgemein unwirtschaftliche Behandlung durch den Arzt, Sonderregelungen. Neben vielen speziellen Voraussetzungen für ein Beanstandungsverfahren ist vor allem interessant, dass eine derartige Beanstandung nur vom Kassenverband und nicht durch eine einzelne Kasse initiiert werden darf und dass vor der Einleitung des Verfahrens in jedem Fall der Vertrauensarzt des Beschwerdeführers persönlich mit dem betreffenden Arzt gesprochen haben muss.

Im zweiten Teil sind die einzelnen Taxen für die unterschiedlichen Verrichtungen aufgelistet. Die Taxen lassen sich in zwei Gruppen einteilen: einer Art Grundgebühr für die Inanspruchnahme des Arztes und Gebühren für alle zusätzlich erbrachten Leistungen. Grundgebühren werden für Konsultationen, telefonische Beratungen und Hausbesuche erhoben, wobei jede der drei Dienstleistungen wiederum nach drei Kategorien differenziert wird (normaler Arztbesuch, Notfall und Arztbesuch in der Nacht). Verschiedene Zuschläge für erhöhten Zeitaufwand können diese Grundgebühren ergänzen. Außerdem wird für die erste Konsultation innerhalb von 90 Tagen ein Zuschlag verrechnet.

#### **7.1.2 Tarifvertrag Genf**

Der für den Untersuchungszeitraum gültige Tarifvertrag für ambulante medizinische Leistungen im Kanton Genf stammt aus dem Jahr 1981. Auch hier ist der Vertrag in zwei Teile geteilt: Der erste Teil enthält generelle Bestimmungen und Definitionen von bestimmten Leistungen. Der zweite umfasst die Auflistung der Taxen für die einzelnen Leistungen.

Die im allgemeinen Teil aufgelisteten Definitionen umfassen Erklärungen von Konsultationen, Hausbesuchen und Notfällen. So wird bspw. eine Konsultation als eine persönliche Unterre

---

<sup>124</sup> Vgl. hierzu auch die Statuten der Ärztesgesellschaft Zürich (AGZ 1999).

dung des Arztes mit dem Patienten im Sprechzimmer von 20-minütiger Dauer definiert. Darüber hinaus sind wie im Kanton Zürich auch hier im allgemeinen Teil Regelungen hinsichtlich Beschwerden und Schlichtung enthalten. Streitigkeiten im Rahmen der Anwendung des Tarifs werden durch eine Schlichtungskommission gelöst, die paritätisch durch Repräsentanten der Genfer Ärzte und Vertretern der Krankenkassen besetzt ist. Sollte keine Einigung erzielt werden können, kommen Bundesgesetze zur Anwendung. Der Punktwert ist in diesem Vertrag mit 0.50 SFr. festgeschrieben.

Die Auflistung der Taxen erfolgt ebenfalls wieder in zwei Teilen. Zunächst werden die allgemeinen Leistungstaxen aufgelistet. Diese umfassen die Taxen für die Konsultationen, die Hausbesuche, die Notfälle, Konsultationen mit Kollegen, das Anfertigen von Gutachten sowie für Entfernungsentgelte. Im zweiten Teil sind schließlich die Leistungen in Untergruppen aufgelistet, die (im Allgemeinen) zusätzlich zu den Grundleistungen verrechnet werden können.

#### **7.1.3 Vergleich der Tarife in Zürich und Genf**

Beide Vertragswerke weisen grundsätzlich die gleiche Struktur auf: In einem allgemeinen Teil werden prinzipielle Regelungen zwischen den Vertragspartnern geregelt, im Leistungsteil die einzelnen Tarifpositionen abgehandelt. Dennoch zeigen sich im Vergleich interessante Unterschiede, sowohl im allgemeinen Teil als auch in der Leistungstaxierung.

Während der Vertrag von Zürich relativ ausführlich die Regelungen bei Beschwerden eines Vertragspartners beschreibt, stellt der Vertrag in Genf auf Regelungen des Bundesrechts ab. Wesentlich interessanter ist allerdings die unterschiedliche Besetzung der Schlichtungskommission für Streitigkeiten bei der Auslegung von Tarifen oder Unklarheiten bezüglich einzelner Rechnungen. Während in Zürich die Verantwortung für die Besetzung der „Blauen Kommission“ bei der Ärztesgesellschaft liegt, wird im Vertrag von Genf ausdrücklich die Notwendigkeit einer paritätischen Besetzung hervorgehoben. Außerdem sind im allgemeinen Teil des Genfer Vertrags Definitionen der Grundleistungen sowie Anforderungen an die Verrechnung eines Extra-Zeitaufwands enthalten, die im Zürcher Vertrag nicht explizit aufgeführt sind. Schließlich wird im Genfer Vertrag auch der Punktwert der Leistungen festgeschrieben, während er im Zürcher Vertrag weiterer Verhandlungen – die dann auch beschrieben werden – bedarf.

Der eigentliche Tarifteil der Verträge unterscheidet sich einmal in verschiedenen Besonderheiten hinsichtlich des Aufbaus der Tarifstruktur und den Punktwerten. Um die Preise besser vergleichen zu können, werden im Folgenden die einzelnen Punktwerte gleich in Frankenbeträge

### III Empirische Untersuchung

umgerechnet. Wie bereits erwähnt, betrug der Punktwert in Genf 0.5 SFr. In Zürich betrug er im Untersuchungszeitraum dagegen 0.8 SFr. Die folgende Tabelle (Tabelle 1) zeigt, dass die Preise für verschiedene Leistungen stark differieren. Betrachtet man zunächst nur bestimmte Grundleistungen (Konsultationen und Hausbesuche) so fällt auf, dass im Durchschnitt die Gebührensätze in Zürich höher liegen als in Genf. Allerdings können die Ärzte für diese Leistungen verschiedene Zuschläge verrechnen. Zeitaufschläge fallen dann an, wenn die Konsultation oder der Besuch länger als 20 Minuten (Genf) bzw. 30 Minuten (Zürich) dauern. Allerdings sind diese Zuschläge nur bis zu einem bestimmten Umfang anrechenbar (vgl. FN 2 und 3 der Tabelle 1).

Tabelle 1 Grundleistungen und Preise in Zürich und Genf (1997-1999) in SFr.

Pos.	Bezeichnung	Zürich			Genf		
		Normal	Notfall	Nacht	Normal	Notfall	Nacht
<b>Grundleistungen</b>							
1	Konsultation	18.4	28.8	64	24	42,5	42,5
2	Hausbesuche	80 <sup>1</sup>	136 <sup>1</sup>	192 <sup>1</sup>	55	89,5	124
3	Zeitzuschlag	24 <sup>2</sup>	36 <sup>2</sup>	48 <sup>2</sup>	36.5 <sup>3</sup>	48.5 <sup>3</sup>	60.5 <sup>3</sup>
4	Erstkonsultationszuschlag	12 <sup>4</sup>	12 <sup>4</sup>	12 <sup>4</sup>	/	/	/
5	Telefonische Beratung	9.6		24	18 <sup>5</sup>		24
6	Kilometergeld	/	/	/	4	5,5	7
<b>Zuschläge f. Anamnese u. Status</b>							
		Groß	Klein	Groß	Klein		
7	Neurologischer Status	120	40	122.5	30.5		
8	ORL Status	40	40	30.5	12		
9	Orthopädischer Status	60	40	48.5	18		
10	Gynäkologischer Status	20	20	30.5	12		

<sup>1</sup> Bei gleichzeitiger Behandlung mehrerer Mitglieder eines Einzel- oder Kollektivhaushaltes wird nur einmal eine Besuchs-Pos. gerechnet; im Übrigen einfache Konsultation oder – wenn begründet – Notfall- oder Nachtkonsultation.

<sup>2</sup> Über eine halbe Stunde, pro angefangene Viertelstunde, soweit keine anderen honorierten Extraleistungen diesen Zeitaufwand bedingen.

<sup>3</sup> Zeitaufwand über 20 Minuten, maximal aber für zusätzlich 30 Minuten. Kann nur 5 Mal pro Krankenschein verrechnet werden.

<sup>4</sup> Zuschlag für erste Konsultation oder Hausbesuch innerhalb von 90 Tagen.

<sup>5</sup> Es wird zusätzlich noch eine kurze telefonische Beratung unterschieden (9.5 SFr.).

Aufgrund der unterschiedlichen Zusammensetzung der ausgewählten Tarifpositionen für Grundleistungen in Tabelle 1 lässt sich kein eindeutiges Bild hinsichtlich der Frage gewinnen, ob eine bestimmte Position in dem einen oder anderen Kanton höher vergütet wird oder nicht. Immerhin erkennt man, dass eine gewöhnliche Konsultation ohne Zeitzuschlag in Zürich billiger ist als in Genf (ca. 24%); eine Nachkonsultation ist in Zürich dagegen um ca. 50% teurer. Die Vergütung für einen normalen Hausbesuch ist dann in beiden Kantonen gleich, wenn der

Arzt in Genf mindestens 6 km für den Hausbesuch zurücklegen muss. Durch den Erstkonsultationszuschlag wird in Zürich ein zusätzlicher Anreiz geschaffen, möglichst viele Patienten zu behandeln. Alles in allem scheinen die Grundleistungen in Zürich daher etwas höher vergütet als in Genf, wobei aber die unterschiedliche Verrechenbarkeit einzelner Tarifpositionen in den Kantonen beachtet werden muss. Die Zuschläge für Anamnese und Status fallen in Zürich für den kleinen Status eindeutig höher aus als in Genf; der Preis für den großen Status ist dagegen in beiden Kantonen im Durchschnitt gleich.

#### 7.2 Die untersuchten HAM und HMOs in der Schweiz

Seit den ersten Gründungen von Hausarztmodellen und HMOs in der Schweiz ist die Zahl der Netze rasant angestiegen. Im Jahre 1999 waren in der gesamten Schweiz 3,800 Ärzte in Hausarztmodellen tätig. Sie versorgten 350,000 Versicherte von mehr als 12 Versicherern in mehr als 50 Hausarztnetzen.<sup>125</sup> Gegenwärtig ist eine Stagnation eingetreten. Wie im ersten Teil dieser Arbeit dargelegt, kann es dafür vielfältige Gründe geben (vgl. hierzu Kapitel 3 und 4).

Im Kanton Zürich gab es im Jahre 1999 10 Hausarztnetze mit 51,717 Versicherten und 485 Ärzten.<sup>126</sup> Damit versorgte jeder Arzt durchschnittlich 107 HAM-Versicherte. Der Schweizer Durchschnitt dieser Verhältniszahl beträgt 102. Der Versichertenanteil der HAM betrug 4.35%. In den 6 existierenden HMOs (Stand: 1999) versorgten 25.3 Ärzte 21,976 Versicherte, was einem Durchschnitt von rund 869 Versicherten pro Arzt entspricht. Außer diesen 6 HMOs gibt es in Zürich-Stadt und in Winterthur noch jeweils eine HMO der Swica<sup>127</sup>.

Im Kanton Genf existierten im Jahr 1999 3 Hausarztnetze mit insgesamt 15,890 Versicherten und 624 Ärzten, d.h. also mit nur durchschnittlich 25 Versicherten pro Arzt. Außer einer Swica-HMO, für die keine Daten verfügbar sind, gibt es in Genf keine weitere HMO. Damit ist die Verbreitung von Managed Care in Genf ungefähr vergleichbar mit Zürich (3.9% in Genf und 4.35% in Zürich).

##### 7.2.1 *HAM Wintimed*

Das Hausarztmodell Wintimed ist das älteste Hausarztnetz der Schweiz. Es wurde 1994 als ein vom BSV bewilligtes Pilotprojekt in Winterthur gegründet. In diesem Netz beteiligen sich

---

<sup>125</sup> Vgl. Schweizerische Ärztezeitung (1999).

<sup>126</sup> Vgl. hierzu und im Folgenden Bundesamt für Sozialversicherung (2000b).

<sup>127</sup> Diese HMO fehlt in den Statistiken.

derzeit 25 Ärzte, die 7870 Versicherte von 3 verschiedenen Versicherern versorgen.<sup>128</sup> Damit kommen auf einen Hausarzt 314 eingeschriebene Versicherte, was deutlich über dem Zürcher wie auch dem Schweizer Durchschnitt liegt.

Laut Statuten des Arztnetzes<sup>129</sup> müssen die beteiligten Ärzte Facharztstitel in der Allgemeinmedizin, der Inneren Medizin oder der Pädiatrie besitzen und über Praxiserfahrung verfügen sowie in der Stadt Winterthur tätig sein. Die Verträge zwischen Versicherungsträgern und Wintimed regeln weitere Details. In dem hier betrachteten Vertrag wird ein FMH-Titel oder eine Berufserfahrung von mindestens 6 Jahren von den Ärzten verlangt. Diejenigen Ärzte, die einen Vertrag mit der Arcovita – der Managed-Care-Organisation des Versicherungsträgers – abschließen, verpflichten sich, an Qualitätszirkeln teilzunehmen, die durch den Hausarztverein selbst angeboten werden. Die Vergütung der Ärzte erfolgte anfänglich ausschließlich über Einzelleistungsentgelte, die dem kantonalen Einzelleistungstarif entsprachen. Allerdings wird für jede Rechnungsperiode ein medizinisches Budget berechnet, welches die Leistungen pro Altersklasse berücksichtigt sowie ein Einsparungsziel formuliert. Dabei wird jeder einzelne Arzt am finanziellen Risiko über eine Gewinn- und Verlustbeteiligung bis zu 5000 SFr. pro Jahr beteiligt.<sup>130</sup> Diese Regelung galt auch noch im Untersuchungszeitraum für das CSS-Versichertenkollektiv. Für die Zukunft war vorgesehen (ab 2000), eine echte Kopfpauschalvergütung mit einer Absicherung gegen hohe Risiken einzuführen. Das Vorhaben stieß allerdings seitens der Ärzte zunächst auf Ablehnung. Wobei die Ablehnung nicht mit einem erhöhten Einkommensrisiko erklärt, sondern mit angeblichen Berechnungsfehlern und der Komplexität der Formel argumentiert wurde. Insofern hätte man die 5000 SFr. als eine Approximation der durchschnittlichen Risikobereitschaft der in diesem Netz beteiligten Ärzte auffassen können. Allerdings unterzeichneten die Ärzte von Wintimed im Jahr 2002 zu fast den gleichen Bedingungen wie 2000 einen Kopfpauschalvertrag mit der CSS.<sup>131</sup>

Auch zum Thema Risikoselektion bietet der betrachtete Vertrag interessante Einblicke. In einem Artikel, in dem es um die Rechte und Pflichten der Ärzte geht, wird explizit darauf hingewiesen, dass die Wintimed-Ärzte nicht darauf verpflichtet werden können, einen Hausarztversicherten anzunehmen, der sich bei ihm einschreiben will. Darüber hinaus kann der Wintimed-Arzt eine

---

<sup>128</sup> Stand: 1.1.1999. Vgl. Forum Managed Care (2000b).

<sup>129</sup> Vgl. Wintimed (2000).

<sup>130</sup> Vgl. Schweizerische Ärztezeitung (1999).

<sup>131</sup> Vgl. zur Capitationberechnung Beck (2001).

weitere Behandlung von Versicherten ablehnen, wenn das Vertrauensverhältnis zwischen Arzt und Patient schwer gestört ist, worunter auch die Nichtbezahlung der ärztlichen Honorarrechnung fällt. Nur in Notfällen ist der Wintimed-Arzt auch weiterhin zur Behandlung verpflichtet.<sup>132</sup> Diese Regelungen erlauben es den Ärzten, im Prinzip Risikoselektion zu betreiben. Auch bei den weitergehenden empirischen Untersuchungen wird es von Interesse sein, ob es möglich ist, die *ex post* bekundete starke Risikoaversion der Ärzte und die „vertraglich genehmigte“ Risikoselektion zu erkennen und zu separieren.<sup>133</sup>

#### **7.2.2 HAM Winterthur Andelfingen**

Anfang 1996 wurde der Verein „Hausärzte der Bezirke Winterthur und Andelfingen“ gegründet, aus dem 1997 das Hausarztnetz hervorging. Ende 1998 arbeiteten 104 Ärzte für das Netz und versorgten insgesamt 34,035 Versicherte von 4 unterschiedlichen Versicherungsträgern - unter anderem der CSS. Damit musste jeder Hausarzt durchschnittlich 327 HAM-Patienten versorgen. Die hier berücksichtigte Stichprobe umfasst insgesamt 98 Ärzte und 1871 CSS-Versicherte.<sup>134</sup>

Die Vergütung besitzt eine Erfolgs- bzw. Verlustbeteiligung. Der Erfolg wird am Budget des Netzes gemessen, welches sich aus den gegenwärtigen Leistungen pro Versichertem und der Definition eines Einsparungsziels ergibt. Allerdings wurde im ersten Betriebsjahr noch keine Erfolgs- bzw. Verlustbeteiligung berechnet.<sup>135</sup> Ansonsten gelten die gleichen vertraglichen Bestimmungen wie bei dem HAM Wintimed. Insbesondere gilt auch bei diesem Netz das zuvor Gesagte zur Zukunft der Zusammenarbeit zwischen Hausarztnetz und der CSS-Versicherung: Der Versuch, eine pauschalierte Vergütung durch die CSS-Versicherung ab dem Jahr 2000 einzuführen, scheiterte auch bei diesem Netz, sodass ab dem Jahr 2000 nur noch Versicherte von drei anderen Versicherungen sich in diesem Netz befinden. Für die Versicherten der CSS bedeutete das entweder ein Zurück in den konventionellen Sektor oder – falls sie im Hausarztmodell weiter eingeschrieben sein wollten – einen Wechsel der Versicherung.<sup>136</sup>

---

<sup>132</sup> Außerdem sind die Hausärzte beim Ausscheiden eines Kollegen aus dem Hausarztnetz verpflichtet, dessen Patientenstamm zu übernehmen – vorausgesetzt die Patienten wünschen dies. Vgl. auch FN 136.

<sup>133</sup> *Ex post* bekundete Risikoaversion bezieht sich auf den Umstand, dass die Ärzte dieses Netzes den erwähnten Capitation-Vertrag mit teilweiser Risikoübernahme durch die Krankenkasse zunächst nicht akzeptierten.

<sup>134</sup> Diese Angaben beziehen sich auf das Jahr 1998.

<sup>135</sup> Diese Angaben beruhen auf den Angaben, die in den entsprechenden Verträgen zwischen der Arcovita AG und den Hausärzten stehen.

<sup>136</sup> Weitere Untersuchungen könnten dieses „natürliche Experiment“ ausnutzen, um neue Erkenntnisse bezüglich Moral Hazard und der Selbstselektion der Versicherten zu bekommen.



Beide Netze versorgten Ende 1998 41,905 Versicherte, was einem Anteil an der Gesamtbevölkerung in den beiden Bezirken von 16% entspricht.<sup>137</sup> Der Marktanteil der beiden betrachteten Hausarztmodelle ist damit höher als der Zürcher Durchschnitt.

#### 7.2.3 *HAM Genf*

Das Genfer Hausarztmodell hat sich aus einer HMO entwickelt, in der zunächst ausschließlich Studenten und Angestellte der Universität Genf versorgt wurden. Die aus einer Kollektivversicherung hervorgegangene HMO bestand seit 1992. Im Jahre 1994 wurde dann die Mitgliederbeschränkung für Universitätsangehörige aufgehoben und es wurden weitere Ärzte engagiert. Seit diesem Zeitpunkt lässt sich die einstige HMO als HAM auffassen. Obwohl nur ein Teil der ehemaligen Kollektivversicherten in die neue HMO mitgegangen sind, ist die Risikostruktur der HMO (und anfänglich sicherlich auch des HAMs) günstiger als im Genfer Durchschnitt gewesen. Die Tatsache, dass ein Großteil der Versicherten Studenten waren, legt einen solchen Schluss nahe. Außerdem wird diese Sicht durch eine Studie (Perneger et al. 1995) bestätigt. Demnach lagen die Kosten in den letzten 12 Monaten vor der Umwandlung der Kollektivversicherung in eine HMO für die Versicherten, die dann der HMO nicht beigetreten sind, um 39% höher als bei den Versicherten, die sich mit dem Transfer einverstanden erklärten.<sup>138</sup> Ein solches natürliches Experiment bestätigt zudem, dass es bei der Wahl unterschiedlicher Versicherungsverträge (hier HMO- und Kollektivvertrag) zur Selbstselektion kommt.

Hinsichtlich Risikoselektion innerhalb des HAMs lassen sich keine expliziten Rückschlüsse ziehen, da ein Vertrag zwischen Versicherung und HAM nicht vorliegt. Gleiches gilt auch im Hinblick auf die Risikoaversion der Ärzte. Allerdings ist bekannt (CSS-Angaben), dass im Untersuchungszeitraum ähnliche Regelungen galten wie in den beiden HAM im Kanton Zürich. Im Unterschied zu den beiden Zürcher Hausarztmodellen schloss das HAM in Genf aber für das Jahr 2000 einen neuen Vergütungsvertrag mit der CSS ab, der vollständig auf einer Capitationvergütung mit einer Rückversicherung für sehr teure Fälle beruht. Dies lässt immerhin *ex post* den Schluss zu, dass die Ärzte in diesem HAM eher risikofreudigere Individuen sind als ihre Kollegen im Kanton Zürich.<sup>139</sup> Aus dem Selbstverständnis des Netzes ist zudem erkennbar, dass

---

<sup>137</sup> Vgl. Forum Managed Care (2000b).

<sup>138</sup> Perneger et al. (1995), S. 13.

<sup>139</sup> Natürlich erscheint ein derartiger Rückschluss an dieser Stelle ziemlich gewagt, da es auch andere Gründe für das unterschiedliche Verhalten der Ärzte im Genfer HAM geben kann. So könnte bspw. eine sehr günstige Struktur

die Befürwortung einer Pauschalvergütung durch die Ärzteschaft eine grundsätzliche Auseinandersetzung der Ärzte mit ihrem Beruf verlangt. In einem kürzlich erschienenen Artikel verweisen die Verantwortlichen des Netzes darauf, dass eine pauschalierte Vergütung es ihnen erlaube, ihren Beruf „mit mehr Freiheit, mehr Intelligenz und mehr Eleganz“ auszuüben.<sup>140</sup> Außerdem betonen die Autoren im Hinblick auf Anstrengungen zur Kostensenkung, dass diese nur Erfolg haben können, wenn Ärzte und Patienten an einem Strang ziehen, d.h., dass nicht nur die Ärzte sich in Qualitätszirkeln weiterbilden, sondern auch, dass die Patienten am Wissen der Ärzte teilhaben müssen, um Verantwortung übernehmen zu können und um sich an den therapeutischen Entscheidungen zu beteiligen.<sup>141</sup>

Das HAM hat sich seit seiner Gründung ständig vergrößert. Während 1992, also im Jahr der Gründung des Gesundheitsplans, sich 11 Ärzte beteiligten, die 3613 Versicherte versorgten, stieg die Zahl der Ärzte Ende 1998 auf 44 an, die nun 10 338 Versicherte von drei verschiedenen Versicherern medizinisch versorgen. Das entspricht durchschnittlich 235 Versicherten pro Arzt. Damit liegt diese Zahl nicht nur deutlich über der Vergleichszahl für den Kanton Genf (25 Versicherte pro Arzt), sondern auch noch über dem Schweizer Durchschnitt von ca. 100 Versicherten pro Arzt.

In der vorliegenden Stichprobe befinden sich 43 Ärzte, die pro Jahr ca. 4771 CSS-Patienten zu versorgen haben.<sup>142</sup> Außer dem betrachteten Hausarztnetz existieren seit geraumer Zeit noch zwei weitere Netze, die jedoch wegen der zu geringen Zahl an Versicherten (ca. 20 pro Arzt) zur Zeit keine große Bedeutung besitzen. Bezogen auf die Gesamtbevölkerung des Kantons Genf waren Ende 1998 2.6% der Versicherten in einem Hausarztmodell versichert.<sup>143</sup>

#### **7.2.4 HMO Zürich**

Die Arcovita-HMO<sup>144</sup> in Zürich wurde 1996 gegründet. In ihr arbeiteten im Untersuchungszeitraum drei Ärzte, die Angestellte der CSS sind. Sie erhalten ein festes Gehalt pro Monat sowie leistungsabhängig einen zusätzlichen Bonus bei besonderem Erfolg der HMO. Aufgrund des

---

des Versichertenbestandes Ursache für die Entscheidung sein. Die folgende empirische Analyse soll auch darüber Aufschluss geben.

<sup>140</sup> Vgl. Schaller und Raetzo (2002, S. 20).

<sup>141</sup> Vgl. ebenfalls Schaller und Raetzo (2002).

<sup>142</sup> Vgl. auch die Übersicht in Tabelle 2 im Abschnitt 7.4.4.

<sup>143</sup> Eigene Berechnung.

<sup>144</sup> Die Arcovita, die Managed-Care-Organisation der CSS, wurde 1999 umbenannt und heißt seitdem Medinet AG.

Angestelltenverhältnisses kommen sie außerdem in den Genuss weiterer Ver- und Begünstigungen laut Anstellungsbedingungen der CSS-Versicherung wie bspw. einem zugesicherten Jahresurlaub – und damit verbunden, einer Vertretung für die Urlaubszeit –, Haushaltszulagen für Verheiratete sowie Kinderzulagen gemäss gesetzlichen Vorschriften. Hinzu kommen noch verschiedene Regelungen, welche die zusätzliche Altersversorgung und Vergünstigungen bei der Krankenversicherung betreffen. Insgesamt scheint es ein Vergütungspaket zu sein, welches wohl eher dem stark risikoscheuen Arzt einen Anreiz gibt, bei dieser Art HMO mitzuarbeiten

Die drei beschäftigten Ärzte versorgten Ende 1998 2750 Versicherte. In der hier betrachteten Stichprobe sind 1509 CSS-Versicherte einbezogen (ebenfalls Stand Ende 1998). Die Leistungen wurden unter Mithilfe von drei Praxisassistenten und einer Gesundheitsschwester erbracht. Die Lohnsumme betrug 1998 für alle Beschäftigten 507,000 Franken (Angaben laut CSS). Dies bedeutet für einen angestellten Arzt in der HMO ein geschätztes Bruttoeinkommen von knapp 90,000 SFr. im Jahr.

Neben dieser Medinet-HMO existieren in der Stadt Zürich noch 3 SanaCare HMOs, eine HMO der Swica sowie eine ärzte-eigene HMO (Medix). Eine weitere ärzte-eigene HMO hat am 1.1.2000 in Zürich-Hirslanden eröffnet. Insgesamt sind somit derzeit 27 Ärzte in der Stadt Zürich in HMOs beschäftigt.<sup>145</sup> Diese 27 Ärzte versorgen insgesamt 20,200 Versicherte, was einem Anteil an der Gesamtbevölkerung von knapp 6%<sup>146</sup> entspricht.

### 7.3 Arztselektion

Ziel der Untersuchung ist es, einen Vergleich zwischen verschiedenen Formen der Gesundheitsversorgung und -finanzierung vorzunehmen. Insbesondere soll das Verhalten der Leistungserbringer in diesem Prozess beleuchtet werden. Speziell wird hierbei der Fokus auf diejenigen Ärzte gelegt, die im Allgemeinen am Anfang jeder Inanspruchnahme von medizinischen Leistungen stehen. Allerdings lässt sich dieser Kreis von Leistungserbringern nicht eindeutig eingrenzen, da im konventionellen Bereich die Versicherten das Recht auf freie Arztwahl besitzen. Somit reicht es nicht aus, nur die Ärzte zu untersuchen, die den Facharztstitel Allgemeinarzt tragen. Immerhin kann aber davon ausgegangen werden, dass Chirurgen oder Augenärzte in der Regel nicht zu den ersten Ansprechpartnern eines Patienten zählen. Trotzdem

---

<sup>145</sup> Ohne die im Jahr 2000 neugegründeten HMO. Vgl. auch Forum Managed Care (2000a).

<sup>146</sup> Eigene Berechnung auf Grundlage der Bevölkerungszahlen für Zürich (Stadt) 1998.

sei darauf hingewiesen, dass nur ein Teil der von den Versicherten in Anspruch genommenen Leistungen durch die hier gewählte Modellierung abgebildet werden kann, da – wie erwähnt – im konventionellen Bereich prinzipiell freie Arztwahl besteht. Um diesem Umstand in beschränktem Maße Rechnung zu tragen, wird eine separate Nachfrageschätzung durchgeführt, bei der die gesamten Leistungen eines Versicherten berücksichtigt werden. Die Gesamtnachfrage wird dann in geeigneter Weise in das Gesamtmodell übernommen bzw. zur Konstruktion von Instrumentalvariablen verwendet. Eine nähere Erklärung dieser Vorgehensweise erfolgt bei der Vorstellung der einzelnen Modelle.

Im Folgenden wird konkret der Selektionsprozess der zu untersuchenden Ärzte und der von ihnen erbrachten Leistungen beschrieben. Ziel des Prozesses ist es, eine hohe Vergleichbarkeit zwischen den Ärzten zu erreichen. Aus diesem Grund berücksichtigt die Untersuchungen nur Ärzte, von denen angenommen werden kann, dass sie ein Großteil ihres Einkommens durch die ärztliche Tätigkeit erzielen. Dies trifft bspw. nicht auf Ärzte zu, die bereits pensioniert sind oder kurz vor der Pensionierung stehen und nur noch wenige Patienten im Jahr behandeln (z.B. Freunde oder Bekannte). Daher werden in der Regel nur solche Ärzte berücksichtigt, die in einem Jahr einen Umsatz von mindestens 100,000 SFr. hatten.<sup>147</sup>

Es muss ebenfalls berücksichtigt werden, dass die Ärzte sich einem Patientenstamm gegenübersehen, der sehr heterogen ist und daher die erbrachten Leistungen nur schwer miteinander vergleichbar sind. Für den Vergleich von Hausarztmodellen und konventionellem Sektor kommt noch hinzu, dass die Patienten eines Hausarztnetzes in gewisser Weise homogener als im konventionellen Bereich sind, da sich einerseits nur Versicherte einer bestimmten Region in ein solches Netz einschreiben dürfen (vgl. Abschnitt 7.2) und andererseits die Selbstselektion der Versicherten dazu führt, dass in HAM in der Regel gesündere Versicherte anzutreffen sind. Obwohl durch die Einbeziehung verschiedener sozioökonomischer Variablen der Versicherten ein Teil dieser Heterogenität der vom Arzt behandelten Patienten in der Untersuchung berücksichtigt werden kann, wird ein Großteil der Heterogenität durch in der Regel nichtbeobachtbare Morbiditätsfaktoren bestimmt, die durch eine Berücksichtigung des Wohnorts der Patienten zumindest zum Teil aufgefangen werden soll. Eine noch genauere Beschreibung jedes Arzt-Patienten-Kontakts kann durch die Betrachtung der Entfernung, die der Patient zu seinem Arzt

---

<sup>147</sup> Der Umsatz der Ärzte ergibt sich aus der Summe der abgerechneten direkten Arztleistungen und dem Umsatz aus der Selbstdispensation.

zurücklegen muss, entstehen. Alle verwendeten Variablen werden im nächsten Abschnitt vorgestellt.

In der Untersuchung werden Ärzte aus den Kantonen Zürich und Genf untersucht, die Versicherte der CSS-Versicherung in den Jahren 1997, 1998 und 1999 behandelt haben. Ferner wird zwischen Ärzten unterschieden, die in der Zürcher HMO, im Hausarztmodell Wintimed in Winterthur oder im HAM Winterthur/Andelfingen und im HAM in Genf arbeiten. Als Vergleichskollektiv auf der Arztseite wurden alle Ärzte ausgewählt, die in den jeweiligen Kantonen als Grundversorger tätig sind. Des Weiteren wurde das Vergleichskollektiv der Ärzte auf die Fachbereiche (Partnerart) eingeschränkt, die typischerweise in den Managed-Care-Organisationen (MCOs) vorkommen, sowie auf diejenigen, die auch von MCO-Versicherten relativ frei gewählt werden können und damit ebenfalls in den Grundversorgungsbereich gehören. Zu dieser Gruppe gehören vor allem die Fachbereiche Allgemeinmedizin, Allgemeine Innere Medizin sowie die Kinderärzte. Darüber hinaus wurden auch Konkordatsnummern<sup>148</sup> der Gemeinschaftspraxen berücksichtigt, da es sich hierbei oft um gemeinschaftlich geführte Arztpraxen von Internisten und Allgemeinmedizinern handelt. Die Einbeziehung von Arztpraxen, die als Gemeinschaftspraxen abrechnen, ist deshalb wichtig, weil einige HMOs der Partnerart nach bei den Allgemeinmedizinern, andere dagegen bei den Gemeinschaftspraxen eingeordnet sind.

Eine weitere Einschränkung des Untersuchungskollektivs ergibt sich aus der unterschiedlichen Verfügbarkeit der Daten. Einerseits werden nur Ärzte einbezogen, von denen bestimmte Eigenschaften bekannt sind (wie Berufserfahrung, Titel und Geschlecht), andererseits kommen nur Ärzte in das Untersuchungskollektiv, die im gesamten Untersuchungszeitraum (also in allen 3 Jahren) tatsächlich auch mindestens eine bestimmte Anzahl von Leistungen abgerechnet haben.

Über den gesamten Zeitraum (1997 – 1999) waren in dem so konstruierten Untersuchungsgebiet insgesamt 1319 Grundversorger tätig – 926 in Zürich und 393 in Genf. Vernachlässigt man die Umsatzbeschränkung von 100,000 SFr., so waren es insgesamt 1726 Ärzte. Tabelle 2 zeigt die Arztstruktur aller Ärzte in den beiden Kantonen nach Arztgruppen aufgeteilt.

---

<sup>148</sup> Konkordatsnummern sind diejenigen Nummern, unter denen die Ärzte beim Verband der Schweizer Krankenkversicherer (KSK) angemeldet sind.

Tabelle 2 Anzahl Ärzte nach Kanton und Partnerart (volles Sample)

Kanton	Allgemein	Innere	Kinder	Gemeinschaft
Genf	211	243	75	14
Zürich	791	288	97	7

#### 7.4 Datenbeschreibung, Variablenkonstruktion und erste deskriptive Statistiken

Alle in der Untersuchung verwendeten Leistungsdaten der CSS-Versicherung wurden nach dem Behandlungsdatum erhoben, d.h. nicht – wie im betriebswirtschaftlichen Sinne üblich – nach dem Abrechnungsdatum. Dieser Vorgehensweise liegt die Sichtweise zu Grunde, dass das Interesse einer solchen Untersuchung dem tatsächlichen Leistungsgeschehen in einer bestimmten Periode gilt und nicht der buchhalterischen Frage, *wann* diese Leistungen bei der Versicherung vom Arzt und/oder dem Patienten abgerechnet worden sind. Zwischen Behandlung und Abrechnung kann eine große Zeitspanne liegen. Das Problematische daran ist nicht die Zeitspanne an sich, sondern der Umstand, dass es sich offensichtlich um Zufallsschwankungen handelt (vgl. hierzu Beck 2000). Schwierigkeiten können immer dann entstehen, wenn die nach Behandlungsbeginndatum erhobenen Daten mit Daten einer höheren Ebene verglichen werden müssen, die nach Abrechnungsdatum erfasst worden sind. Dies ist bspw. beim Vergleich von CSS-Leistungsdaten mit den Leistungsdaten aller Krankenversicherer der Fall.

Die verwendeten Variablen lassen sich in drei große Gruppen einteilen. Als erstes sind diejenigen Variablen zu erwähnen, welche auf Arzzebene gemessen werden. Hierunter fallen sowohl Informationen über die Ärzte selbst und Informationen über die Arztpraxen als auch die Leistungserbringung der Ärzte. Die zweite Gruppe umfasst Variablen, die Eigenschaften der Patienten beschreiben. Hierhin gehören zunächst ebenfalls soziodemografische Angaben. Dazu kommen verschiedene Angaben zum Versicherungsverhältnis sowie Morbiditätsfaktoren. Die dritte Variablengruppe schließlich beschreibt die Marktstruktur im Untersuchungsgebiet.

##### 7.4.1 Arzt – Variablen

Jeder Arzt wird durch soziodemografische Angaben sowie die Charakteristika seiner Arztpraxis beschrieben. Alle Eigenschaften des Arztes bzw. der Arztpraxis entnehmen wir der KSK-Statistik der Schweizer Ärzte. Der Hauptuntersuchungsgegenstand ist allerdings die Leistungstätigkeit des Arztes in seiner Praxis sowie die Art seiner Leistungserbringung.

### Eigenschaften des Arztes

Die wichtigsten soziodemografischen Variablen des Arztes sind Alter, Geschlecht, Berufserfahrung und Qualifikation.

Die Erfahrung des Arztes im Beruf wird durch die Variable *EINTRITT* in den Beruf abgebildet, der Qualifikationsgrad dagegen durch zwei Variablen (*DOKTOR* und *ARZT*). Erstere umfasst Ärzte mit höherem akademischen Abschluss, *ARZT* hingegen Ärzte mit Diplom. Beiden Variablen nehmen entsprechend des jeweiligen Qualifikationsgrad den Wert 1 an. Arztgemeinschaften, insbesondere HMOs, erhalten für beide Variablen den Wert 0 und stellen somit die Referenzkategorie dar. Arztgemeinschaften lassen sich auch im Zusammenspiel mit den Variablen für das Geschlecht der Ärzte identifizieren, denn für Arztgemeinschaften bezeichnen wir in diesem Fall das Geschlecht mit *NEUTRAL* (ansonsten *MANN* oder *FRAU*). Die folgende Tabelle 3 beschreibt die 1726 Ärzte über die Dummy-Variablen für das Geschlecht und den Qualifikationsgrad.

Tabelle 3 Codierung von Geschlecht und Qualifikationsgrad der Ärzte

NEUTRAL	MANN	FRAU	DOKTOR	ARZT	Anzahl
0	0	1	0	1	31
0	0	1	1	0	373
0	1	0	0	1	55
0	1	0	1	0	1241
1	0	0	0	0	26

Die Variable *EINTRITT* gibt an, ab wann ein Arzt mit einer eigenständigen Nummer beim KSK gemeldet ist. Sie dürfte sehr stark mit dem Aufnahmezeitpunkt der eigenen Praxistätigkeit korreliert sein und wird daher als Berufserfahrung der Ärzte interpretiert. In Fällen, wo der Arzt einen Wohnortwechsel vollzogen hat oder aber anderweitig die Nummer wechselte, trifft der Zusammenhang allerdings nicht mehr zu. Hinzu kommt, dass das Minimum von *EINTRITT* das Jahr 1977 ist. Das heißt dass die maximale Berufserfahrung der Ärzte als „mindestens“ interpretiert werden muss, da ältere Ärzte sicherlich auch schon vor 1977 in einer eigenständigen Praxis gearbeitet haben. Eine Analyse dieser Fälle wie auch der Wechsler ist allerdings im Rahmen unserer Untersuchung nicht möglich, sodass die beschriebenen Interpretationsmöglichkeiten im Hinterkopf bleiben sollten.

Für das Alter liegen für ca. 1/3 der Ärzte keine Angaben vor. Daher wurde auf eine Einbeziehung des Alters verzichtet. Immerhin ist es indirekt in der Variablen *EINTRITT* mit enthalten.

Denn das Alter ist für diejenigen Ärzte mit Altersangabe signifikant positiv mit der Variablen *EINTRITT* korreliert (Korrelation = 0.59).

#### **Eigenschaften der Arztpraxis**

Hinsichtlich der Praxisausstattung ist zunächst bekannt, ob es sich um eine Allgemeinarztpraxis (*ALLGEMEIN*), eine Kinderarztpraxis (*PÄDIATRIE*), eine Praxis der allgemeinen inneren Medizin (*INNERE*) oder eine Gemeinschaftspraxis (*GEM\_KSK*) handelt. Unter einer Gemeinschaftspraxis versteht man entsprechend der Systematik des KSK mehrere Ärzte, die nach außen unter einer Arztnummer auftreten, so z.B. HMOs. Wie bereits erwähnt, ist diese Systematik nicht eindeutig, da einige HMOs auch bei den Allgemeinmedizinern eingeordnet sind. Neben dieser unscharfen Einteilung gibt es auch Ärzte, die zwar eigenständig abrechnen, aber trotzdem in einer Gemeinschaftspraxis tätig sind. Für diese Ärzte wurde eine weitere Variable *GEM* berechnet, die den Wert 1 annimmt, wenn der Arzt mit weiteren Ärzten in den selben Praxisräumen arbeitet. *GEM\_ANZ* gibt an, wie viele Ärzte gemeinsam in der Praxis arbeiten. Eine Arztpraxis lässt sich weiterhin durch den Grad der technischen Ausstattung beschreiben. In unserer Untersuchung wird der Ausstattungsgrad durch die Variable *ROENTGEN* charakterisiert. Sie zeigt an, ob eine Arztpraxis mit einem Röntgenapparat ausgerüstet ist. Außerdem wird die Art der Praxisführung auch durch den Ort der Praxis mitbestimmt. Damit können wir Unterschiede in der Praxisführung durch die Berücksichtigung des Standortes der Praxis und dessen Besonderheiten einfangen. Auf diese Eigenschaften wird bei der Vorstellung der Gemeindecharakteristika zurückgekommen (vgl. Abschnitt 7.4.3). Schließlich wissen wir, ob die Ärzte in einem von der CSS-Versicherung mitbetriebenen Hausarztnetz oder in einer HMO arbeiten. Auch diese Eigenschaften lassen sich durch zwei Dummies einfangen.

In Tabelle 4 sind alle Eigenschaften mit ihren Mittelwerten bzw. Anteilswerten für das Jahr 1998 aufgeführt.



Tabelle 4 Arzt- und Praxiseigenschaften

Variable	Mittelwert / Anteil	Variable	Mittelwert / Anteil
NEUTRAL	1.51%	ALLGEMEIN	58.05%
MANN	75.09%	INNERE	30.76%
FRAU	23.41%	PÄDIATRIE	9.97%
DOKTOR	93.51%	GEM_KSK	1.22%
ARZT	4.98%	GEM	23.93%
EINTRITT	1985.93	GEM_ANZ	0.71
		HMO	0.64%
		HAM	8.34%
		ZÜRICH	68.50%
		ROENTGEN	62.69%

#### Leistungen der Ärzte

Die erbrachten Leistungen der Ärzte lassen sich in bestimmte Kategorien gliedern. Durch die Zerlegung der abgerechneten Leistungen soll ein möglichst genaues Bild der Leistungstätigkeit einer Praxis entstehen. So kann man zwischen Leistungen unterscheiden, die zu einem Einkommen für den Arzt führten (*INT\_LST*) oder aber veranlasst worden sind (*EXT\_LST*). Zum anderen lassen sich die Leistungen nach ihrer Art voneinander abgrenzen, d.h. ambulante Leistungen, Medikamente, stationäre Leistungen usw. Außerdem werden die Leistungen nach der Zugehörigkeit des Arztes zu einem Hausarztnetz differenziert.

Der Untersuchung liegen keine Informationen bezüglich der gestellten Diagnosen zu Grunde. Vielmehr kann nur auf die oben dargelegte Einteilung der erbrachten Leistungen nach ihrer Art zurückgegriffen werden. Obwohl die fehlenden Diagnosen eine Einschränkung darstellen, ist es dennoch möglich, auch unter diesen Umständen bestimmte Einsichten in das Verhalten der behandelnden Ärzte zu bekommen. Denn aufgrund der relativ großen Anzahl der betrachteten Ärzte, dürften Unterschiede im Patientenstamm der einzelnen Ärzte im Durchschnitt nicht besonders stark ins Gewicht fallen. Außerdem wird für Besonderheiten der Arztpraxen (vgl. Tabelle 4) und ebenfalls für bestimmte Merkmale der Patienten (vgl. folgenden Abschnitt) kontrolliert. Darüber hinaus erlaubt es die konkrete Betrachtung der Arzt-Patienten-Beziehung, weitere Merkmale der Arztpraxen zu berücksichtigen (vgl. Abschnitt 9.4.3).

In Tabelle 5 sind die Leistungsvariablen – mit den entsprechenden deskriptiven Statistiken – aufgelistet. Die internen Leistungen umfassen in der Regel nur die direkten Arztkosten, also die Vergütung des Arztes für die abgerechneten Leistungen. Bei Ärzten mit einer Praxisapotheke (Selbstdispensation) sind in den internen Leistungen auch die Leistungen durch den Medika

### III Empirische Untersuchung

menten- bzw. Hilfsmittelverkauf enthalten. Die internen Leistungen lassen sich daher als Umsatz der Arztpraxis interpretieren.

Bei den externen Leistungen, also den vom Arzt veranlassten Leistungen, stellen die Medikamente und die ambulanten Leistungen die größten Kostenblöcke dar. In den externen ambulanten Leistungen wurden dabei sehr unterschiedliche Leistungen zusammengefasst. Sie umfassen vor allem die Kosten für Chiropraktoren, Ergotherapeuten, Logopäden und auch bestimmte Naturheilverfahren.

Die in Tabelle 5 dargestellten Leistungen sind in der Regel nicht die Gesamtleistungen der beteiligten Ärzte, da die meisten Ärzte sicherlich auch Patienten anderer Krankenversicherungen behandeln. Ob die hier verwendeten Behandlungsdaten der Ärzte für diese tatsächlich typisch sind oder ob das behandelte Versicherungskollektiv für die untersuchten Ärzte eher untypisch ist, kann durch einen Vergleich der gesamten Leistungen der Ärzte (pro Patient) mit den Pro-Patient-Leistungen aus Tabelle 5 erfolgen. Ein solcher Vergleich wird im Abschnitt 7.5 durchgeführt.

Tabelle 5 Leistungen der Ärzte <sup>a</sup>

Jahr	OKP			HAM / HMO		
	1997	1998	1999	1997	1998	1999
Ärzte	1726	1726	1726	138	145	145
Patienten <sup>b</sup>	80.618	76.284	81.413	40.254	40.179	39.910
Konsultation	4.03	3.71	3.75	2.360	3.498	2.584
Direkte Arztleistung	432.309	394.783	437.219	256.721	348.126	276.218
Medikamente (Selbstdispensation)	66.729	67.268	81.931	25.110	30.401	42.356
Umsatz_intern	499.038	462.051	519.150	324.445	422.204	361.068
Konsultation	0.724	0.697	0.761	2.502	4.114	3.660
Medikamente (Apotheke)	211.251	223.323	243.626	180.828	291.753	255.206
Ambulant	31.138	33.324	39.272	390.704	470.019	486.418
Labor	41.606	42.983	41.958	66.462	82.471	66.122
Physio	44.858	45.104	41.300	70.841	58.231	52.365
Prävention	0.201	0.186	0.174	0.964	0.871	0.828
Umsatz_extern	335.277	352.693	367.306	708.267	903.321	860.918
Gesamtleistung <sup>c</sup>	2778	3026	3036	1033	1325	1221.97

<sup>a</sup> Leistungen, Konsultationen und Besuche pro Patient, Arzt und Jahr

<sup>b</sup> Pro Arzt und Jahr

<sup>c</sup> Pro Patient und Jahr

Die Leistungen der Ärzte in Tabelle 5 wurden nach der Zugehörigkeit der Patienten zu einem Hausarztnetz gegliedert. Alle Ärzte, die Hausarztversicherte behandelten, haben auch im konventionellen Bereich Patienten versorgt. Bei einem Vergleich der rohen Zahlen für die

unterschiedliche Zugehörigkeit der Patienten erkennt man deutliche Unterschiede in der Zusammensetzung und auch der Höhe der Leistungen pro Patient. Die Versicherten der HAM suchen im Durchschnitt seltener einen Arzt auf als die Versicherten im konventionellen Bereich und der Umsatz pro Arzt und Patient ist in HAM deutlich geringer. Interessant ist ferner, dass diejenigen Ärzte, die HAM-Versicherte behandeln, im Durchschnitt die Hälfte ihres Patientenstammes aus den Hausarztnetzen rekrutieren. Dass die externen Umsätze für die HAM-Versicherten pro Arzt um rund das 10-fache höher sind als für konventionell Versicherte, ist nicht verwunderlich, da der Definition nach in den Hausarztnetzen alle veranlassten Leistungen von den Hausärzten ausgehen. Im konventionellen Bereich muss man dagegen beachten, dass freie Arztwahl herrscht, sodass die Versicherten einerseits mehrere Hausärzte nutzen können und zum anderen auch gleich direkt einen Spezialisten aufsuchen oder in das Spital gehen können. In der letzten Zeile ist daher zum Vergleich die durchschnittliche Gesamtleistungen der Patienten in den unterschiedlichen Netzen angegeben: Während im Hausarztnetz außer den Spitalkosten sämtliche von den Patienten verursachten Leistungen erfasst sind, machen die in unserer Untersuchung berücksichtigten Leistungen im konventionellen Sektor nur ca. 1/3 der Gesamtleistungen aus.

#### **7.4.2 Patient – Variablen**

Auch die Patienten der Ärzte werden durch ihre soziodemografischen Eigenschaften beschrieben. Hinzu kommen hier Variablen, die verschiedene Eigenschaften der Versicherungsverträge widerspiegeln sowie Variablen, die Morbiditätsfaktoren der Versicherten auffangen sollen. Ein besonderes Merkmal stellt dabei die Variable *SPT1* dar, die angibt, ob ein Versicherter im Vorjahr im Spital war oder nicht.

Aus dem Versicherungsvertrag sind die Länge der Versicherungszeit ( $M$ ) und – bei vorzeitiger Vertragsbeendigung ( $M < 12$ ) – der Grund des Ausscheidens bekannt. Ein wichtiger Ausscheidungsgrund ist der Tod eines Versicherten. Denn wie verschiedene neuere Untersuchungen gezeigt haben, stellt der nahende Tod eine entscheidende Determinante für extrem hohe Gesundheitskosten dar, sodass diese Variable auch einen hohen Erklärungsgehalt für die Höhe der in Anspruch genommenen Arztleistung haben dürfte.<sup>149</sup> Eine weitere Variable, die aus dem Versicherungsvertrag entnommen werden kann, ist *UNFALL*. Sie gibt an, ob der betroffene Versicherte gleichzeitig bei der CSS eine Kranken- und Unfallversicherung besitzt. Der Versi

---

<sup>149</sup> Vgl. zu dem Zusammenhang zwischen Gesundheitskosten, Alter und Todesnähe bspw. Zweifel et al. (1999).

cherte kann die Unfalldeckung in der obligatorischen Krankenversicherung ausschließen, wenn er bereits über den Arbeitgeber gegen Unfall versichert ist. Deshalb sind Personen mit Unfallabschluss entweder Selbständige oder aber Rentner bzw. andere Nichterwerbstätige. Eine weitere Variable des Versicherungsvertrages ist *FR*, die Franchise. Dieser wählbare absolute Selbstbehalt kann 4 Stufen annehmen.<sup>150</sup> Da im Sample auch HMO-Versicherte enthalten sind, die keine Franchise haben, wurde die Variable hier mit 5 Stufen modelliert, 0 für ein HMO-Mitglied und 1-5 für die anderen Franchisen.

Tabelle 6 Merkmale der Versicherten

Merkmal	Beschreibung	1997	1998	1999
Merkmale Vertrag und Patient				
N	Anzahl Versicherte	168,809	168,552	163,035
END	Versicherung im lfd. Jahr beendet (in %)	6.68%	6.38%	9.45%
TOD	Endgrund der Versicherung aufgrund des Todes (in %)	0.78%	0.77%	0.78%
M	Anzahl der Versicherungsmonate (1-12)	11.59	11.59	11.58
UNFALL	Wenn Kranken- und Unfallversicherung bei der CSS abgeschlossen sind = 1 (in %)	69.74%	69.13%	69.49%
FR	Franchisestufe (0 = HMO, 1-5 entsprechend der Franchisestufen)	1.47	1.54	1.59
SEX	Mann = 1 (in %)	46.26%	46.31%	45.71%
ALTER	Alter des Versicherten	39.54	39.35	38.73
PR	Gewährte Prämienverbilligungen	179.59	244.07	372.54
Leistungsmerkmale Patient				
SPT1	1 wenn Patient im Vorjahr mind. 1 Tag im Spital war (in %)	9.27%	9.33%	8.70%
LST_GES	Gesamte in Anspruch genommene Leistung pro Jahr in Franken	2,122.81	2,258.25	2,381.99

Wichtige Informationen bezüglich des Vermögensstatus liefert die Variable *PR*. Sie gibt den jährlichen Prämienzuschuss des Staates zur Prämie des jeweiligen Versicherten an. Prämienzuschüsse werden Versicherten gewährt, für die die Aufbringung der Prämie eine zu große wirtschaftliche Last darstellt. Die Prämienverbilligungen werden gemeinsam von Bund und Kanton gezahlt. In Zürich erhielten in den Jahren 1996 bis 1998 durchschnittlich 23.5% der Versicherten eine Prämienverbilligung, in Genf sogar 33%. Die durchschnittliche jährliche Prämienverbilligung betrug in Zürich im genannten Zeitraum 830 SFr., in Genf dagegen 1,306 SFr.<sup>151</sup> Damit liegen sowohl die Bezügerquote als auch die Höhe der ausgezahlten Prämienver

<sup>150</sup> Vgl. auch die Ausführungen zu den allgemeinen Regelungen der Schweizer Krankenversicherung im Anhang.

<sup>151</sup> Vgl. zu diesen Zahlen Bundesamt für Statistik (2000).

billigungen in unserem Sample deutlich unter den entsprechenden Vergleichswerten für die jeweiligen Kantone.

#### **7.4.3 Marktstruktur – Variablen**

Die letzte Gruppe umfasst Variablen, die Besonderheiten der Marktstruktur beschreiben, denen sich die Ärzte in den einzelnen Gebieten gegenüber sehen. Als Bezugsgröße wurde die Gemeinde gewählt. Auf dieser Ebene ist es gerade noch möglich, die relevanten Daten zu erfassen.

Theoretisch lässt sich ein Markt durch die Menge aller Käufer und Verkäufer definieren, die durch ihr Zusammenspiel den Preis des Gutes auf diesem Markt bestimmen (vgl. bspw. Stigler und Sherwin 1985). Eine Möglichkeit, empirisch Märkte zu definieren, ist die Identifizierung von geografischen Regionen, in die relativ wenige Konsumenten von außen hineinkommen und aus denen relativ wenige Konsumenten hinausgehen. Auf den Märkten für Arztdienstleistungen würde sich bspw. die Abgrenzung der regionalen Märkte voneinander durch die Einteilung nach der Entfernung, die ein potentieller Patient einer Region zum nächsten Arzt zurücklegen müsste, sowie den geografischen Gegebenheiten anbieten. In der vorliegenden Arbeit wird, wie gesagt, ein regionaler Markt durch die Gemeinde bestimmt. Einer derartig einfachen Marktabgrenzung liegt der Gedanke zugrunde, dass Patienten zum nächsten Anbieter gehen. Gerade bei der hier zu untersuchenden Arztgruppe ist diese Annahme nicht unrealistisch. Die durchschnittliche Entfernung der Wohnortgemeinde eines Patienten zur Gemeinde der Arztpraxis wird allerdings später noch eine Rolle spielen, wenn es darum geht, die konkrete Leistungserbringung des Arztes bei bestimmten Patienten abzubilden.

Insgesamt gibt es in Zürich 171 und in Genf 45 Gemeinden. Von diesen 216 Gemeinden werden in der Regel allerdings nur diejenigen Gemeinden betrachtet, in denen Ärzte im Untersuchungszeitraum tätig waren. Das waren 1999 168 Gemeinden.<sup>152</sup>

Neben wichtigen Merkmalen, welche die Bevölkerungsstruktur in den Gemeinden abbilden (*BEV* – Gesamtbevölkerung und jeweils als Anteil an der Gesamtbevölkerung, *GEST\_ANT* – Anteil der Gestorbenen, *GEB\_ANT* – Anteil der Geborenen, *MANN\_ANT* – Anteil der Männer und *AUSL\_ANT* – Anteil der Ausländer an der Gesamtbevölkerung), konnte auch die Größe der Gemeinde (*HA*) einbezogen werden. Darüber hinaus gelang es ebenfalls, Informationen über die durchschnittliche Einkommenssituation der Gemeinden im Untersuchungszeit

---

<sup>152</sup> Bei der Schätzung der Arztdichte in den Gemeinden werden alle Gemeinden mit einbezogen (Vgl. Abschnitt 9.1).

raum zu berücksichtigen. Als Approximation für das Einkommen der Bevölkerung kann der Steuerertrag pro Person (*ST\_P\_K*) aufgefasst werden. Diese Größe umfasst alle Steuereinnahmen der Gemeinde pro Kopf der Wohnbevölkerung.<sup>153</sup>

Die Anzahl der in einer Gemeinde tätigen Ärzte wird hauptsächlich durch die Arztdichte (*ARZTD*) gemessen. Sie wird pro 10,000 Einwohner angegeben.

Die letzte Variable auf Marktebene approximiert die durchschnittlichen Gesundheitsausgaben einer Gemeinde. Diese Variable konnte aus einer Untersuchung der ETH Zürich übernommen werden, die für das KSK mit Hilfe der Leistungsdaten aller großen Krankenversicherer für jede Gemeinde in der Schweiz die durchschnittlichen Gesundheitskosten berechnet.<sup>154</sup> Der verwendete Durchschnittswert basiert auf den Jahren 1997 und 1998 und passt somit recht gut in die vorliegende Untersuchung.

Tabelle 7 Marktstrukturvariablen

Merkmal	Beschreibung	1997	1998	1999
N	Anzahl Gemeinden mit Arzt	169	168	168
BEV	Gesamtbevölkerung	9095	9163	9210
GEST_ANT	Anteil Verstorbener	0.007	0.0068	0.0069
GEB_ANT	Anteil Geborener .	0.011	0.011	0.011
AUSL_ANT	Anteil Ausländer	0.178	0.178	0.179
MANN_ANT	Anteil Männer	0.494	0.494	0.494
HA	Fläche der Gemeinde in ha	961*	964*	964*
ST_P_P	Steuerertrag pro Person	1412.1**	1403.9**	1403.9**
ARZTD	Anzahl Ärzte pro 10000 Einwohner	14.63	14.85	14.47
DK	Durchschnittskosten pro Monat	143.5	143.2	143.2

Erklärungen:

\* Gemeindestand von 1998

\*\* Werte für die Steuerjahre 1997/98

#### 7.4.4 *Versichertenstruktur in Genf und Zürich im Vergleich mit den Hausarztkollektiven*

Tabelle 8 gibt einen Überblick der Unterschiede zwischen den Versicherten in den Hausarztnetzen und den Versicherten, die in der normalen OKP versichert sind. Neben den bereits vorgestellten Merkmalen werden auch die Anzahl Versicherter und die Bruttoleistung pro Versichertenmonat für die einzelnen Netze und ihre Vergleichskollektive betrachtet.

<sup>153</sup> In der Schweiz liegt die Steuerhoheit bei den Kantonen und Gemeinden. Die Gemeinden können ausgehend vom Kantonssteuersatz einen eigenen höheren oder niedrigeren Steuerfuß festlegen.

<sup>154</sup> Ziel dieser Berechnungen war und ist die einheitliche Einteilung der Gemeinden in Risikozonen, nach denen die Krankenversicherer einheitlich ihre Prämien berechnen sollen. Zur Zeit der Untersuchung teilte jeder Versicherer die Kantone noch individuell in Prämienregionen ein. Ab 1.1.2004 gibt es einheitlich festgelegte Zonen für jeden Kanton. Die neue Einteilung bezieht sich auf die oben erwähnte Untersuchung der ETH Zürich.

### III Empirische Untersuchung

Tabelle 8 Versicherte OKP / HAM und HMO 1997 – 1999

Netz	Jahr	Versicherte	Gest	Mann	Alter	Spital-Vorjahr	Brutto-Lst	Präm-verb	Pat-Anteil
HMO-Zürich	1997	1,460	0.27%	52.67%	39.98	3.15%	83.16	358.11	75.34%
HMO-Zürich	1998	1,411	0.28%	52.02%	40.70	1.77%	91.76	419.24	71.79%
HMO-Zürich	1999	1,441	0.76%	52.39%	40.93	0.49%	102.62	415.11	77.10%
HAM-Wi-An	1997	1,219	0.41%	44.54%	37.35	7.22%	118.90	135.77	66.12%
HAM-Wi-An	1998	1,625	0.31%	44.49%	36.85	4.62%	122.26	191.45	75.63%
HAM-Wi-An	1999	1,828	0.49%	46.28%	36.03	1.04%	88.43	187.02	67.94%
HAM-Winti	1997	376	0.00%	43.88%	40.18	6.12%	118.50	114.27	63.03%
HAM-Winti	1998	484	0.41%	44.21%	40.22	5.37%	98.39	179.38	69.01%
HAM-Winti	1999	498	0.80%	45.38%	40.70	1.00%	89.36	186.27	67.47%
HAM-Genf	1997	4,958	0.18%	50.04%	30.47	1.65%	115.22	279.78	81.14%
HAM-Genf	1998	4,607	0.13%	49.47%	31.34	0.87%	125.77	392.43	78.73%
HAM-Genf	1999	4,648	0.30%	49.83%	32.27	0.56%	139.86	572.74	80.21%
Vergleichskollektiv	Jahr	Versicherte	Gest	Mann	Alter	Spital-Vorjahr	Brutto-Lst	Präm-verb	Pat-Anteil
Zürich-Stadt	1997	28,459	1.46%	42.64%	46.37	12.56%	247.25	352.36	79.60%
Zürich-Stadt	1998	27,461	1.57%	42.74%	46.47	13.57%	273.16	400.53	80.08%
Zürich-Stadt	1999	25,976	1.49%	42.21%	46.02	12.35%	282.95	393.35	80.98%
Zürich-Land	1997	86,550	0.54%	48.16%	37.35	8.11%	129.52	106.89	74.75%
Zürich-Land	1998	90,300	0.49%	48.16%	36.99	8.16%	138.12	137.07	75.95%
Zürich-Land	1999	86,031	0.48%	47.34%	35.75	7.15%	141.49	157.67	76.39%
Genf	1997	45,822	0.90%	44.39%	40.43	10.54%	260.65	194.78	83.78%
Genf	1998	42,700	0.95%	44.27%	40.64	10.47%	280.45	350.60	84.08%
Genf	1999	42,648	1.02%	43.86%	41.02	11.17%	305.98	771.37	85.52%

Tabelle 8 zeigt die Besonderheit des HAMs Genf. Hier sind die Versicherten nicht nur im Schnitt 10 Jahre jünger als im Vergleichskollektiv, sondern es sind auch viel mehr Männer vertreten (durchschnittlich 6% mehr als im konventionellen Bereich). So sind die enormen Kostenunterschiede wie auch die deutlich geringeren Spitalisierungs- und Mortalitätsraten nicht weiter verwunderlich. Ein Teil dieser „verzerrten“ Struktur ist aber auf die Gründungsumstände dieses Hausarztmodells zurückzuführen (ursprünglich HMO für Studenten).<sup>155</sup> In der HMO in Zürich sieht es ähnlich aus. Hier liegt das Durchschnittsalter zwischen 6 und 7 Jahren unter dem Durchschnitt des Vergleichskollektivs und der Anteil der versicherten Männer liegt um 10%

<sup>155</sup> Vgl. auch die Ausführungen zur Geschichte dieses HAM in Abschnitt 7.2.3.

über dem Durchschnitt der Vergleichsgruppe. Auch die Mortalitätsraten und die Zahlen zur Spitalinanspruchnahme im Vorjahr deuten auf ein viel gesünderes Versichertenkollektiv in der HMO hin. Dementsprechend liegen die verursachten Kosten in der HMO auch deutlich unter den Kosten des Vergleichskollektivs. Bei den beiden Hausarztmodellen im Kanton Zürich ergibt sich ein differenzierteres Bild. Teilweise kehren sich sogar die Verhältnisse. So ist das Durchschnittsalter im HAM Wintimed bis zu 5 Jahre höher als in der Vergleichsgruppe. Aber auch hier verursachen die Versicherten geringere Kosten als im Vergleichskollektiv. Es fällt bei den Hausarztnetzen in Zürich besonders auf, dass die Kosten pro Monat und Versicherten von 1997 bis 1999 kontinuierlich gefallen, während sowohl in den anderen MCOs (Zürcher HMO und HAM in Genf) als auch im konventionellen Sektor die Kosten von Jahr zu Jahr gestiegen sind.

#### **7.5 Repräsentativität der verwendeten Daten**

Die verwendeten Daten erscheinen in Bezug auf das Untersuchungsgebiet repräsentativ. Wie dargestellt, wurde nicht nur auf die Homogenität der betrachteten Ärzte geachtet, sondern auch berücksichtigt, dass die zu vergleichenden Ärzte sich einer möglichst homogenen Nachfrage gegenübersehen (Versicherte aus den gleichen Kantonen unter Berücksichtigung der Wohngemeinde und der Entfernung zum Arzt). Ob der Patientenstamm der Ärzte in der Stichprobe dem durchschnittlichen Patientenstamm eines jeden Arztes entspricht, kann durch einen Vergleich der an CSS-Versicherten erbrachten Leistungen mit der gesamten Leistungserbringung des jeweiligen Arztes untersucht werden.

Dafür wurden Allgemeinärzte, Kinderärzte und Ärzte der allgemeinen inneren Medizin sowie die Gemeinschaftspraxen herangezogen, die im Jahr 1999 mindestens 20 CSS-Patienten im konventionellen Bereich behandelt haben und einen Gesamtumsatz von mindestens 100,000 SFr. erzielten ( $N = 1,342$ ). Im Durchschnitt deckt unsere Stichprobe mit den CSS-Daten rund 1/8 des Patientenstammes der hier betrachteten Arztpraxen ab.

Die durchschnittlichen Gesamtleistungen pro Patient liegen in allen betrachteten Kategorien etwas höher als die Leistungen pro Patient für die CSS-Versicherten (vgl. Tabelle 9). Die Mittelwerte für die veranlassten Leistungen (Apotheke, Labor und Physiotherapie) sind statistisch nicht signifikant unterschiedlich. Dagegen weisen die direkten Leistungen der Arztpraxen



signifikante Unterschiede auf. Dies gilt auch für die durchgeführten Konsultationen und die Hausbesuche.

Tabelle 9 Vergleich Gesamtleistung – Leistung CSS (1999)<sup>a</sup>

Ärzte = 1,342	CSS		KSK	
	Mittelwert	Standardabw.	Mittelwert	Standardabw.
<b>Ambulant</b>	405.88**	297.39	442.04**	312.39
<b>Medizin v. Arzt</b>	96.05**	126.96	108.93**	135.83
<b>Medizin v. Apotheke</b>	247.85	351.84	262.36	347.27
<b>Labor</b>	42.64	66.77	45.87	64.92
<b>Physio</b>	42.85	47.13	48.76	45.23
<b>Konsultationen</b>	3.70**	1.78	4.22**	1.78
<b>Besuche</b>	0.06**	0.13	0.10**	0.14
<b>Patienten</b>	100.26	76.01	883.04	597.37

<sup>a</sup> Leistungen, Konsultationen und Besuche als Angaben pro Patient

\*\* p < 0.05

Die signifikanten Unterschiede könnten auf einen durchschnittlich besseren Gesundheitszustand der CSS-Versicherten hinweisen. Andererseits kann eventuell auch die Unvollständigkeit der CSS-Daten oder aber die hohe Zufallskomponente in den KSK-Daten ausschlaggebend für die Unterschiede sein.<sup>156</sup> Allerdings lässt sich darüber kein abschließendes Urteil fällen. Wichtig ist jedoch, dass die Tätigkeit der Ärzte in der Stichprobe sich mit der Tätigkeit der betrachteten Ärzte bei allen Patienten im Groben vergleichen lässt; dass sich also die in den Untersuchungen betrachteten Patienten nicht vollständig vom durchschnittlichen Patientenstamm der Arztpraxis unterscheidet.

## 8 Methodologische Probleme der empirischen Analyse

Die in der empirischen Analyse betrachteten Datenmodelle stützen sich naturgemäß auf die erzielten Ergebnisse aus den theoretischen Abhandlungen in den Kapiteln 5 und 6. Allerdings sind auch die empirischen Gegebenheiten, wie sie in Kapitel 7 beschrieben worden sind, zu berücksichtigen. Daraus ergeben sich prinzipielle Schwierigkeiten. Denn zum einen müssen die theoretischen Ableitungen in testbare Hypothesen „übersetzt“ werden und zum anderen fehlen häufig die entsprechenden Daten, um bestimmte Hypothesen tatsächlich testen zu können.

---

<sup>156</sup> An dieser Stelle sei wiederum an den Umstand erinnert, dass die Versicherungsdaten der CSS nach Behandlungsbeginn gegliedert sind, während die Daten des KSK dem Abrechnungsdatum zugeordnet sind. Vgl. zu dieser Thematik wieder Beck (2001).

Abschnitt 8.1 widmet sich daher den grundlegenden Schwierigkeiten der empirischen Analyse. Neben allgemeinen Problemen beim Übergang von der Theorie zur Empirie müssen verschiedene Schwierigkeiten in der empirischen Analyse selbst beachtet werden. Abschnitt 8.2 betrachtet deshalb ökonomische Probleme, die bei allen nachfolgenden Schätzungen von Bedeutung sind.

#### **8.1      Prinzipielles zum Übergang von den theoretischen Modellen zu den empirischen Schätzungen**

In der theoretischen Analyse wurde herausgearbeitet, dass das Verhalten des einzelnen Arztes entscheidend von den verschiedenen Beziehungen zu den anderen an der Gesundheitsproduktion beteiligten Parteien abhängt. Diese Beziehungen sind durch asymmetrische Informationsverteilungen gekennzeichnet, die es der jeweils informierteren Partei erlauben, auf Kosten der uninformatierteren Partei, einen Vorteil zu erzielen. So können Ärzte möglicherweise zu viele Leistungen erbringen oder Patienten mit einer nicht ausreichenden Qualität versorgen. Patienten wiederum können aufgrund der Versicherungsdeckung ihre Nachfrage nach Gesundheitsleistungen über das optimale Maß ausdehnen. Außerdem wurde untersucht, welche Auswirkungen unterschiedliche Marktgegebenheiten auf das Arztverhalten haben.

Konkret konnte im Kapitel 5 gezeigt werden, wie eine Vergütung aussehen müsste, damit der Arzt sich in optimaler Weise für seine Patienten einsetzt und die Vorgaben der Krankenkassen umsetzt. Eine Schlussfolgerung lautete, dass der erwartete Nutzen für die Patienten dann am größten ist, wenn die Informationsasymmetrien zwischen Arzt und Patienten am schwächsten sind. Es zeigte sich, dass die Informationsasymmetrien mit der Einführung von MCOs verringert werden können. Ein wichtiges Instrument stellt dabei eine anreizkompatible Vergütung dar.

Im Kapitel 6 wurde dann das Verhalten von Ärzten mit dem üblichen mikroökonomischen Instrumentarium beleuchtet. Dabei kam die Analyse, die eine Arztpraxis als Firma innerhalb eines Marktes betrachtet, zu den grundsätzlich gleichen Ergebnissen wie die Untersuchung in Kapitel 5. Die wichtigsten Ergebnisse beziehen sich zunächst wieder auf die Gestalt einer anreizkompatiblen Vergütungsfunktion. Zusätzlich wurden die Ursachen für angebotsinduzierendes Verhalten von Ärzten angesprochen sowie die Struktur des Marktes, auf dem Arztleistungen angeboten werden, beleuchtet.

Aus diesen Überlegungen ergeben sich dann auch die ersten testbaren Hypothesen. Zwei mögliche Fragen könnten lauten:

1. Haben sich mit der Einführung von MCOs die Informationsasymmetrien zwischen Arzt und Patient verringert, sodass der Nutzen der Patienten gesteigert werden konnte?
2. Gibt es signifikante Unterschiede bei der Arztvergütung zwischen einer MCO und dem konventionellen Bereich und hat die Vergütung einen Einfluss auf das Verhalten der Ärzte?

Allerdings muss zunächst untersucht werden, ob die gestellten Fragen auch mit dem zur Verfügung stehenden Datenmaterial beantwortet werden können. Genauer gesagt, geht es um zwei Fragen, die bei jeder Hypothesenaufstellung zu berücksichtigen sind. Zum einen muss entschieden werden, ob und wie eine derartige Fragestellung operationalisiert werden kann und zum anderen wie die Operationalisierung der theoretischen Variablen durchgeführt werden soll. Dabei hängt die Operationalisierung der Fragestellung von der grundsätzlichen Verfügbarkeit der Daten bzw. deren Struktur ab. So können Fragen, die Aufschluss über die Wirksamkeit ärztlichen Handelns bei bestimmten Krankheitsbildern geben sollen, natürlich nur mit Hilfe von Diagnosedaten beantwortet werden. Oder aber die Struktur der Daten erlaubt die Beantwortung der gestellten Fragen einfach nicht. Dies ist in ähnlicher Weise bei der oben gestellten zweiten Frage der Fall: Da es – wie wir gesehen haben – bisher kaum (signifikante) Unterschiede in der Arztvergütung gibt, kann auch kein unterschiedlicher Einfluss der Arztvergütung herausgearbeitet werden. Davon vollkommen unabhängig ist die Frage, ob sich das Verhalten des Arztes nach Systemen unterscheidet. Mögliche Verschiedenheiten müssen dabei ohne die Kenntnis der genauen Diagnosen gemessen werden. Wie bei der Vorstellung der Variablen schon dargelegt, können wir zwar erfassen, wo welche Art von Leistung erbracht wurde; welche Diagnose bzw. Erkrankung diese Leistungen verursacht haben, wissen wir dagegen nicht. Trotzdem sollte ein vernünftiger Vergleich aufgrund der sehr großen Anzahl von betrachteten Fällen möglich sein (vgl. auch Abschnitt 7.5).

Die Operationalisierung der theoretischen Variablen dagegen hängt in erster Linie von dem theoretischen Konstrukt und erst in zweiter Linie von der Datenverfügbarkeit ab. So ist z.B. die theoretische Variable *Nutzen des Patienten aus einer Behandlung* tatsächlich nur ein theoretisches Konstrukt, das nicht beobachtet werden kann. Derartige Konstrukte können in ökonometrischen Modellen aber trotzdem operationalisiert werden. Man verwendet dazu so genannte

latente Variablen, die zwar noch einen Bezug zu den theoretischen Konstrukten aufweisen, aber ansonsten den Vorteil besitzen, dass man sie mittels realer Variablen messen kann.

Die angesprochene Problematik tritt sowohl bei der empirischen Umsetzung der Modelle aus Kapitel 5 wie auch der Modelle aus Kapitel 6 auf. Denn wir kennen weder die genaue Struktur der Nutzenfunktionen von Patienten und Ärzten, noch die genaue Struktur der Produktionsfunktion einer Arztpraxis. Trotzdem lassen sich aufgrund der vorangegangenen theoretischen Überlegungen testbare Hypothesen ableiten.

## 8.2 Ökonometrische Probleme und Lösungen

Die ökonometrischen Probleme und die Verfahren zur Lösung dieser Probleme sind vielfältig. Neben dem bereits erwähnten Konzept latenter Variablen, welches in der Mikroökonomie eine sehr breite Verwendung gefunden hat, ist es vor allem die mögliche Endogenität von erklärenden Variablen, die bei fast allen Schätzungen eine Rolle spielt. Hinzu kommen Probleme, die mit den unterschiedlichen Ebenen der Daten verbunden sind. In vielen Fällen untersucht man Daten auf Mikroebene und verbindet sie mit Daten einer höheren Aggregationsebene. Bei derartigen Schätzungen gilt es, die Korrelation innerhalb wie auch zwischen den verschiedenen Ebenen zu berücksichtigen.

### 8.2.1 Modellierung mit Hilfe Latenter Variablen

Häufig können wir nur die Auswirkungen der Entscheidung eines Individuums beobachten, wissen aber über den Entscheidungsprozeß relativ wenig. Als Beispiel lässt sich die Entscheidung eines Versicherten nennen, ob er einen Arzt aufsuchen soll oder nicht. Das Ergebnis der Entscheidung (der Arztbesuch) kann beobachtet werden, die Faktoren, die für die Entscheidung verantwortlich sind, dagegen nicht. Vorrangig dürfte der subjektiv wahrgenommene Gesundheitszustand von Bedeutung sein. Darüber hinaus dürfte aber bspw. auch die Art des Versicherungsvertrags (Kostenbeteiligungsregelung) einen Einfluss ausüben. Das Charakteristische an derartigen Situationen ist, dass man es mit *unbeobachtbaren* Entscheidungsprozessen und *beobachtbaren* Entscheidungen zu tun hat. In vielen Fällen kann man solche Entscheidungssituationen durch so genannte latente Variablen beschreiben.<sup>157</sup> Im dargestellten Fall eines Arztbesuches definiert man den Nutzen des Versicherten aus einem Arztbesuch als eine latente Variable. Es wird angenommen, dass der Versicherte dann einen Arzt aufsuchen wird, wenn der erwartete

---

<sup>157</sup> Vgl. bspw. Verbeek (2000).

Nutzen aus einem Arztbesuch höher eingeschätzt wird, als der Nutzen ohne einen Arztbesuch. Obwohl diese beiden Nutzenwerte nicht beobachtbar sind, kann man doch das Ergebnis des Entscheidungsprozesses beobachten. Die getroffene Entscheidung für oder gegen einen Arztbesuch dient somit als eine Annäherung für die im Hintergrund ablaufenden Entscheidungsprozesse bzw. Nutzenabwägungen.

Formal lässt sich der Sachverhalt am Beispiel der Entscheidung eines Arztbesuches wie folgt illustrieren: Sei  $y_i^*$  der Nutzen eines Versicherten aus den beiden Alternativen Arztbesuch ( $i = 1$ ) oder kein Arztbesuch ( $i = 2$ ).<sup>158</sup> Es wird angenommen, dass er versucht, seinen Nutzen zu maximieren, sodass er diejenige Alternative mit dem höchsten Nutzen wählt. Dabei kann  $y_i^*$  eine beliebige Funktion von Attributen der Alternativen und/oder der Individuen sein. Entscheidend ist nun, dass man trotz des nichtbeobachtbaren Nutzens und des nichtbeobachtbaren funktionalen Zusammenhangs zwischen Nutzen und Merkmalen der Alternativen bzw. der Versicherten, Aussagen über den Zusammenhang zwischen der Entscheidung des Individuums und der dahinterstehenden Nutzenabwägung treffen kann. Denn wenn die Annahme eines Versicherten als Nutzenmaximierer zutreffend ist, gilt:

$$\begin{aligned} y &= 1 && \text{falls } y_1^* = \max(y_1^*, y_2^*) \\ y &= 0 && \text{sonst} \end{aligned},$$

d.h. man beobachtet die Wahl eines Individuums, falls diese Wahl für das Individuum mit dem höchsten Nutzen verbunden ist. Eine solche Formulierung hat den Vorteil, dass eine komplizierte Nutzenabwägung durch die getroffene Entscheidung plausibel abgebildet werden kann. Eine äquivalente Darstellung, die in der Literatur häufig verwendet wird, ist die Darstellung mit Hilfe einer Indexfunktion  $I$ , die für die Entscheidung eines Arztbesuches folgendes Aussehen hat:

$$y = I(y_1^* - y_2^* > 0).$$

Es ist klar, dass die abhängige Variable binär ist und somit Schätzmethoden angewandt werden müssen, die dieser Besonderheit der Ergebnisvariablen Rechnung tragen. Probit- oder Logitmodelle stellen daher in solchen Situationen die Methoden der Wahl dar (vgl. etwa Greene 1997 und Jones 2000).

---

<sup>158</sup> Wobei ein \* im Folgenden eine unbeobachtbare Variable kennzeichnet.

### 8.2.2 Endogenität I

Oftmals arbeiten empirischen Untersuchungen Zusammenhänge zwischen einer zu erklärenden Variablen und verschiedenen unabhängigen Variablen heraus, die dann mit Hinweis auf die mögliche Endogenität einer oder mehrerer erklärender Variablen in Frage gestellt werden.

Die Problematik soll am Beispiel der angebotsinduzierten Nachfrage (AIN) erläutert werden.<sup>159</sup> Im theoretischen Teil wurde schon mehrfach auf dieses Phänomen eingegangen. Man stelle sich eine einfache Nachfragegleichung auf dem Gesundheitsmarkt vor, in der  $Q_N$  die nachgefragte Menge und  $P_N$  sowie  $P_A$  die exogenen Nachfrage- bzw. Angebotspreise darstellen. Die Nachfrage nach Gesundheitsleistungen soll dabei nicht nur vom Preis, dem sich die Patienten gegenübersehen, und verschiedenen anderen Nachfragefaktoren  $N$  abhängen, sondern auch von der Anbieterdichte, die hier durch die Arztdichte ( $AD$ ) gemessen wird. Für die Arztdichte sei nun angenommen, dass sie nicht exogen vorgegeben sei, sondern vielmehr selbst wieder eine Funktion der Nachfrage darstellt.

Die beiden Gleichungen lassen sich dann wie folgt aufschreiben:

$$\begin{aligned} Q_N &= \alpha_0 + \alpha_1 P_N + \alpha_2 N + \alpha_3 AD + u \quad 160 \\ AD &= \beta_0 + \beta_1 P_A + \beta_2 A + \beta_3 Q_N + \varepsilon. \end{aligned} \quad (8.1)$$

Die Arztdichte wird des Weiteren neben dem exogenen Anbieterpreis auch von anderen Angebotsfaktoren ( $A$ ) beeinflusst. Da die Verfügbarkeit der Ärzte in diesem Modell einen direkten Einfluss auf die Nachfrage ausübt, lässt es sich als ein Modell der angebotsinduzierten Nachfrage bezeichnen.

Eine wichtige Annahme, die erfüllt sein muss, damit die Parameter in dieser Form konsistent geschätzt werden können, ist die Annahme der Unkorreliertheit zwischen den erklärenden Variablen und dem Störterm. Eine derartige Korrelation kann u.a. durch Messfehler einer oder mehrerer erklärender Variablen entstehen. Sie kann aber auch – wie in unserem Beispiel – durch die Endogenität einer oder mehrerer erklärender Variablen verursacht werden. Im konkreten Fall heißt das, dass der uns interessierende Parameter  $\alpha_3$  nicht konsistent geschätzt werden kann. Eine Möglichkeit, trotzdem zu konsistenten Schätzungen der interessierenden Parameter

---

<sup>159</sup> Vgl. etwa Labelle et al. (1994), Folland (1997), Getzen (1997) oder aber auch Benstetter (2002).

<sup>160</sup> Die Größen  $u$  und  $\varepsilon$  bezeichnen Störterme, die Zufallseinflüsse und nichtbeobachtbare Faktoren einfangen sollen.  $\alpha_i$  und  $\beta_i$  sind zu schätzende Parameter ( $i = 0, 1, 2, 3$ ).

zu gelangen, besteht in der Schätzung der *reduzierten Form* des Gleichungssystems, in der die endogene Variable alleine durch exogene Faktoren erklärt wird.

Im konkreten Beispiel erhält man durch einfache Umformungen von (8.1) diese *reduzierte Form* des Gleichungssystems:

$$\begin{aligned} Q_n &= \Pi_1 + \Pi_2 P_N + \Pi_3 P_A + \Pi_4 N + \Pi_5 A + w \\ AD &= \Gamma_1 + \Gamma_2 P_N + \Gamma_3 P_A + \Gamma_4 N + \Gamma_5 A + v. \end{aligned} \tag{8.2}$$

Das Gleichungssystem (8.2) kann nun mittels OLS<sup>161</sup> konsistent geschätzt werden. Allerdings interessiert man sich im Allgemeinen nicht hauptsächlich für die Parameter der reduzierten Form, sondern für die *Strukturparameter* aus (8.1). Nur wenn das Gleichungssystem eindeutig identifizierbar ist, können die Parameter aus der reduzierten Form gewonnen werden. Eindeutig identifizierbar wäre das System, wenn die Anzahl der Parameter in der reduzierten Form mit der Anzahl der Parameter in der Strukturform übereinstimmen würde. Im betrachteten Fall ist die Anzahl der Strukturparameter allerdings kleiner als die Anzahl der Parameter in der reduzierten Form, sodass man aus der reduzierten Form keine eindeutigen Werte für die Strukturparameter berechnen kann. Das System ist somit *überidentifiziert*.<sup>162</sup>

Konsistente Schätzungen können in diesem Fall trotzdem durch eine zweistufige Methode erlangt werden. Dazu schätzt man auf der ersten Stufe das Modell in seiner reduzierten Form. Die so erhaltenen Schätzungen für die endogenen Variablen sind unverzerrte und konsistente Schätzungen für diese Variablen, die in der Strukturform des Modells auf der zweiten Stufe an Stelle der originalen Werte der endogenen Variablen verwendet werden, um konsistente Schätzungen der interessierenden Strukturparameter zu erhalten.

Da auch in den hier zu betrachtenden Modellen die Arztdichte als erklärende Variable eine wichtige Rolle spielt, muss die mögliche Endogenität dieser Variablen beachtet werden. Darüber hinaus können auch weitere Variablen endogen sein. Sind entsprechende Instrumente vorhanden, kann man über den dargestellten Ansatz die Endogenität kontrollieren. Als Instrumente bezeichnet man Variablen, die hoch mit der endogenen Variablen korreliert sind, aber nicht mit dem Störterm der interessierenden Strukturgleichung. Im konkreten Fall der Arztdichte müssen demnach Variablen gefunden werden, die zwar in enger Beziehung zur Arztdichte stehen, aber

---

<sup>161</sup> OLS steht für Ordinary Least Square (Methode der kleinsten Quadrate).

<sup>162</sup> Daraus ergibt sich andererseits, dass die Strukturparameter dann nicht identifizierbar sind, wenn die Anzahl der Strukturparameter größer ist als die Anzahl der Parameter des reduzierten Modells.

nicht gleichzeitig die Nachfragegleichung beeinflussen. Wie viele empirische Untersuchungen belegen, ist das keine leichte Aufgabe. So wurden etwa Quadratterme oder höhere Potenzen der Bevölkerung in einer Gemeinde oder weitere Variablen der Angebotsstruktur (die Anzahl an Krankenhausbetten bspw.) als Instrumente vorgeschlagen.<sup>163</sup> Allerdings lassen sich auch immer Gründe gegen solche Instrumente finden, da die Richtigkeit der Wahl statistisch nicht getestet werden kann.<sup>164</sup> Für den Forscher bedeutet dies, dass er die Wahl der Instrumente möglichst gut begründet und sich letztendlich auf seine Intuition verlassen muss.

#### **8.2.3 Endogenität II**

Ähnliche Probleme treten auch im Zusammenhang mit der Evaluierung bestimmter institutioneller Regelungen auf dem Gesundheitsmarkt auf. So ist es bspw. von Interesse, ob die neuen Formen der Krankenversicherung in der Schweiz nachhaltige Wirkungen auf die Ausgaben der obligatorischen Krankenversicherung ausüben. Um derartige Wirkungen bestimmen zu können, reicht es nicht aus, die Ausgaben für bestimmte Subpopulationen (Versicherte mit Managed-Care-Verträgen versus Versicherte ohne Managed-Care-Verträgen) miteinander zu vergleichen. Denn das Interesse bei der Evaluierung der gesundheitspolitischen Maßnahme liegt darin zu zeigen, dass die Maßnahme nachhaltige Auswirkungen auf das Verhalten der von der Maßnahme Betroffenen zeitigt und somit echte Kosteneinsparungen erzeugen kann. Eine bloße Betrachtung der Korrelation zwischen Ausgaben und Beteiligung an der Maßnahme kann diese Erkenntnis nicht liefern. Im Folgenden wird die Problematik etwas näher erläutert.<sup>165</sup>

Es werden die Kostenunterschiede zwischen Managed-Care-Versicherten und Versicherten mit herkömmlichen Versicherungsverträgen betrachtet. Bezeichne  $Y_{1i}$  die potenziellen Leistungen eines Versicherten  $i$ , wenn er einen Managed-Care-Vertrag abschließen würde und  $Y_{0i}$  die potenziellen Leistungen des gleichen Versicherten  $i$ , wenn er keinen MC-Vertrag abgeschlossen hätte. Ob ein Versicherter einen MC-Vertrag hat, wird durch  $D_i = 1$  angezeigt. Im Fall, dass der Versicherte keinen MC-Vertrag gewählt hat, ist dementsprechend  $D_i = 0$ .

---

<sup>163</sup> Vgl. für einen Überblick zur empirischen Literatur der angebotsinduzierten Nachfrage bspw. Evans (1974), Getzen (1997), Folland (1997) oder Breyer et al. (2003).

<sup>164</sup> Wie Dranove und Wehner (1994) gezeigt haben, kann die Verwendung des Instrumentalvariablen-Ansatzes sogar zu vollkommen unplausiblen Ergebnissen führen.

<sup>165</sup> Die hier gewählte Darstellung schöpft aus der umfangreichen Evaluierungsliteratur. Diese hat vor allem bei der Evaluierung auf dem Arbeitsmarkt einen bedeutenden Beitrag für das Verständnis der mit jeder Evaluierung verbundenen Probleme geleistet. Vgl. stellvertretend Schmidt (2000) und Lechner (2002) zu diesem Thema.



Damit kann die beobachtete Leistung für einen Versicherten  $i$  wie folgt geschrieben werden:

$$\begin{aligned} Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} \\ &= Y_{0i} + D_i (Y_{1i} - Y_{0i}). \end{aligned} \tag{8.3}$$

Gleichung (8.3) verdeutlicht das fundamentale Problem jeder Evaluierung: Es ist nicht möglich, für ein und dasselbe Individuum beide Zustände zu beobachten. Daher ist auch der individuelle *kausale Effekt* ( $Y_{1i} - Y_{0i}$ ) schlichtweg nicht bestimmbar.

Ohne weitere Annahmen gilt dies ebenso für den durchschnittlichen Effekt der Wahl eines MC-Vertrages:

$$\alpha = E[Y_{1i} - Y_{0i}], \tag{8.4}$$

der die erwartete Leistungsanspruchnahme mit einem MC-Vertrag mit der erwarteten Leistungsanspruchnahme ohne MC-Vertrag für ein zufällig ausgewähltes Individuum aus der Gesamtpopulation misst; und ebenso gilt dies für den durchschnittlichen Effekt bezogen auf einen zufällig gezogenen Versicherten, der einen MC-Vertrag gewählt hat:

$$\alpha_{MC} = E[Y_{1i} - Y_{0i} | D_i = 1]. \tag{8.5}$$

In der Literatur wird dieser Effekt als *treatment on the treated* bezeichnet. Um die Schwierigkeiten bei der Bestimmung des Erwartungswertes in (8.5) zu verdeutlichen, kann man  $\alpha_{MC}$  auch wie folgt ausdrücken:

$$\begin{aligned} \alpha_{MC} &= E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1] \\ &= E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 0] + \{E[Y_{0i} | D_i = 0] - E[Y_{0i} | D_i = 1]\}. \end{aligned} \tag{8.6}$$

Die ersten beiden Erwartungswerte in (8.6) sind eindeutig identifizierbar. Der Ausdruck in der geschweiften Klammer dagegen kann nicht identifiziert werden, da man die potenziellen Leistungen eines Versicherten mit MC-Vertrag im Falle, dass er keinen Vertrag hat, nicht beobachten kann (zweiter Term in der geschweiften Klammer). Dieser zweite Effekt wird deshalb auch das *Kontrafaktische* genannt (vgl. Schmidt 2000).

Um nun trotzdem  $\alpha_{MC}$  identifizieren zu können, sind weitere Annahmen notwendig. Im einfachsten Fall nimmt man an, dass die Ergebnisse für die Versicherten, die keinen MC-Vertrag haben, unabhängig von  $D$  sind. Dies trifft z.B. in sozialen Experimenten zu, bei denen die

Teilnehmer nicht entscheiden können, ob sie einen MC-Vertrag wählen möchten oder nicht.<sup>166</sup> Formal kann diese Unabhängigkeitsannahme wie folgt ausgedrückt werden:

$$Y_{0i} \perp\!\!\!\perp D_i. \quad (8.7)$$

Unter dieser Annahme ist  $E[Y_{0i} | D_i = 1] = E[Y_{0i} | D_i = 0]$ . Damit kann der Managed-Care-Effekt identifiziert werden, denn in Gleichung (8.6) heben sich die Effekte in der geschweiften Klammer auf.<sup>167</sup>

Allerdings ist Annahme (8.7) sehr stark und wird in der Realität oft nicht erfüllt werden können. Das trifft auch im vorliegenden Beispiel zu, denn die Versicherten wählen den Versicherungsvertrag selbst aus. Damit ist eine Identifizierung des kausalen Effekts nur möglich, wenn zusätzliche beobachtbare Faktoren bereitstehen, die sowohl die Leistungsanspruchnahme als auch die Wahl der Versicherten vollständig abbilden können. Bezeichne  $X$  derartige Faktoren, so lässt sich diese Annahme (auch: bedingte Unabhängigkeitsannahme) formal schreiben als:

$$Y_{0i} \perp\!\!\!\perp D_i | X_i. \quad (8.8)$$

Damit ist  $E[Y_{0i} | X_i, D_i = 1] = E[Y_{0i} | X_i]$ , sodass der kausale Effekt ebenfalls wieder identifizierbar ist. Denn Gleichung (8.6) lässt sich jetzt schreiben zu:

$$\alpha_{MC} = E[Y_{1i} | X_i] - E[Y_{0i} | X_i]. \quad (8.9)$$

D.h. wenn die potenziellen Leistungen eines Versicherten unabhängig von der bedingten Wahl des Versicherten sind (in Abhängigkeit von  $X$ ), kann der kausale Effekt identifiziert werden.

In der Praxis werden verschiedene Methoden angewandt, um den kausalen Effekt zu bestimmen. Neben parametrischen Verfahren der Regressionsanalyse haben sich in den vergangenen Jahren vermehrt nichtparametrische Verfahren durchgesetzt. Während in ökonometrischen Modellen versucht wird, durch Annahmen über den funktionalen Zusammenhang zwischen beobachtbaren Faktoren und der Wahlentscheidung der Versicherten, den kausalen Effekt zu identifizieren, gehen die nichtparametrischen Verfahren den Weg des direkten Vergleiches. Die Idee dabei ist, möglichst homogene Gruppen in  $X$  zu konstruieren, deren Ergebnisse sich nur noch durch den „Behandlungseffekt“ unterscheiden (vgl. bspw. Heckman et al. 1998).

---

<sup>166</sup> Die berühmte RAND-Untersuchung ist ein Beispiel für derartige soziale Experimente. Vgl. Manning et al. (1987).

<sup>167</sup> Vgl. bspw. Angrist (2000).

Die beiden Verfahren unterscheiden sich in den zu Grunde liegenden Annahmen: Während die nichtparametrischen Verfahren über die bedingte Unabhängigkeitsannahme versuchen, den kausalen Effekt zu identifizieren, benötigen die parametrischen Verfahren zur Identifikation neben den Annahmen über den funktionalen Zusammenhang auch explizite Verteilungsannahmen bezüglich der involvierten Störterme.

Damit versuchen die parametrischen Verfahren eine Abschätzung bezüglich des Kontrafaktischen  $E[Y_{0i} | D_i = 1]$  zu machen, um so die Selektionsverzerrung, die durch  $E[Y_{0i} | D_i = 0] - E[Y_{0i} | D_i = 1]$  gegeben ist, bestimmen zu können. Ein derartiger parametrischer Ansatz wird auch weitestgehend in dieser Arbeit verwendet (vgl. die Schätzung der Nachfrage im Abschnitt 9.3 und die Schätzung des Gesamtmodells in Abschnitt 9.4). Prinzipiell sind bei der Schätzung zwei Wege denkbar: Man kann erstens entsprechend der Vorgehensweise im letzten Abschnitt die Endogenität der interessierenden Variablen durch eine Instrumentalvariablenschätzung berücksichtigen. Zweitens kann direkt eine Abschätzung der Selektionsverzerrung vorgenommen und explizit im Modell verwendet werden (*Kontrollfunktionsansatz*). Die bekannteste Anwendung hierfür stellt sicherlich das Selektionsmodell von Heckman dar.<sup>168</sup> Wie Vella und Verbeek (1999) zeigen, liefern beide Methoden unter bestimmten Umständen die gleichen Ergebnisse. Dies hat bedeutende Implikationen für den Instrumentalvariablenschätzer, da gezeigt werden kann, dass wichtige Punkte bei dem Kontrollfunktionsansatz ebenso für den Instrumentalvariablenansatz relevant sind. In der vorliegenden Untersuchung wird weitestgehend auf den Kontrollfunktionsansatz zurückgegriffen. Dieser wird explizit am Beispiel der Nachfrage in Abschnitt 9.3 hergeleitet.

#### **8.2.4 Hierarchische Datenstruktur**

Oftmals werden in empirischen Untersuchungen Daten von unterschiedlichen Aggregations Ebenen verwendet. Dieses Problem tritt auch bei der Untersuchung der Arzt-Patientenbeziehung auf. Denn jeder Arzt behandelt eine bestimmte Anzahl von Patienten, was dazu führt, dass die Beobachtungen für die Patienten nicht unabhängig voneinander sind, sondern auch von Eigenschaften des Arztes abhängen können. Welche Konsequenzen dieses Mikro-Makro-Problem<sup>169</sup> für die Schätzung der interessierenden Effekte hat, soll im Folgenden kurz beleuchtet werden.

---

<sup>168</sup> Vgl. Heckman (1979).

<sup>169</sup> So nannten Snow und Thurston (2000) das beschriebene Problem.

Ein Arzt behandelt im Laufe einer bestimmten Periode mehrere Patienten. Jeder Arzt wird dabei mit  $i$  und jeder Patient mit  $j$  indexiert. Das Ergebnis der Behandlung durch den Arzt ( $y$ ) lässt sich durch verschiedene Variablen  $x$  erklären. Messfehler und fehlende erklärende Faktoren überlagern die Beziehung zwischen  $y$  und  $x$ . Obwohl diese Faktoren unbeobachtbar sind, lassen sich aufgrund der hierarchischen Datenstruktur nicht nur Annahmen über ihre gemeinsame Verteilung treffen. Vielmehr kann man diese Störterme den beiden Modellebenen zuordnen. Das Modell mit zwei Ebenen lässt sich daher wie folgt schreiben:

$$y_{ij} = \beta' x_{ij} + \mu_i + \varepsilon_{ij}, \quad (8.10)$$

wobei  $\mu_i$  den Fehlerterm für Arzt  $i$  und  $\varepsilon_{ij}$  den Fehlerterm für Patient  $j$  bei Arzt  $i$  bezeichnet. D.h. das Behandlungsergebnis von Patient  $j$  bei Arzt  $i$  lässt sich durch gemeinsame Faktoren  $x_{ij}$  erklären; es wird allerdings durch nichtbeobachtbare individuelle Störfaktoren auf Arzteebene und nichtbeobachtbare Faktoren der Patienten innerhalb der Arztpraxis beeinflusst. Nimmt man nun für die beiden Störterme an, dass sie unabhängig und normalverteilt sind, wobei für beide eine konstante Varianz unterstellt wird ( $\sigma_\mu, \sigma_\varepsilon$ ), lässt sich die Gesamtvarianz von  $y_{ij}$  schreiben als:

$$\text{Var}(y_{ij} | \beta' x_{ij}) = \sigma_\mu^2 + \sigma_\varepsilon^2. \quad (8.11)$$

Damit ist also die Varianz nicht über das gesamte Sample identisch, sondern unterscheidet sich in Abhängigkeit der jeweiligen Arztpraxis und der in ihr behandelten Patienten. Daher sind die Fehlerterme für einen bestimmten Arzt  $i$  aufgrund der gemeinsamen Komponente  $\mu_i$  innerhalb einer Arztpraxis miteinander korreliert:

$$\text{Corr}[\varepsilon_{it} + \mu_i, \varepsilon_{is} + \mu_i] \neq 0. \quad (8.12)$$

Für den Korrelationskoeffizienten ergibt sich unter Berücksichtigung der einzelnen Varianzen:

$$\rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}. \quad (8.13)$$

Diese Korrelation ist im klassischen Regressionsmodell Null, da dann die Intra-Gruppen-Varianz  $\sigma_\mu^2$  Null beträgt. Im allgemeinen Fall wird die Korrelation aber von Null verschieden sein, was Konsequenzen für die Schätzung der interessierenden Parameter hat.

In den letzten Jahren wurden in der Gesundheitsökonomie solche Modelle verstärkt eingesetzt. Beispielhaft seien an dieser Stelle die Arbeiten von Scott und Shiell (1997a und 1997b) erwähnt, die den beschriebenen Rahmen verwenden, um den Einfluss des Wettbewerbs und des Vergütungssystems auf das Verhalten von Ärzten zu untersuchen. Dabei berücksichtigen sie insgesamt drei Ebenen der Datenstruktur – Konsultation, Arzt und Region. Die Autoren zeigen, dass eine Nichtberücksichtigung der besonderen Datenstruktur zu einer erheblichen Unterschätzung der Standardfehler führen kann, was zu einer zu optimistischen Beurteilung des Einflusses auf die Ergebnisvariable führt.<sup>170</sup>

Allerdings trifft diese Verzerrung nur auf Effekte zu, die mehrere Ebenen betreffen – wie bspw. der Einfluss der Arztdichte (Marktebene) auf die Nachfrage nach Gesundheitsleistungen eines einzelnen Patienten. Effekte, die dagegen auf ein und derselben Untersuchungsebene wirken, werden hiervon nicht tangiert (vgl. Snow und Thurston 2000).

## 9 Schätzung und Ergebnisse

Das Hauptziel der empirischen Untersuchung besteht in einer Analyse der Wirkung von Managed-Care-Instrumenten auf das Verhalten von Ärzten. In einem Gesamtmodell des ärztlichen Verhaltens werden zu diesem Zweck verschiedene Teilaspekte der ärztlichen Tätigkeit untersucht. So wird das Arztverhalten durch die Betrachtung der Überweisungen zu Fachärzten, die Konsultationen, die veranlassten Gesamtleistungen und die direkten Arztleistungen abgebildet. Um einen möglichen Managed-Care-Effekt identifizieren zu können, betrachten wir den konkreten Arzt-Patienten-Kontakt. Dadurch wird es möglich, zwei unterschiedliche Wirkungsrichtungen des Managed-Care-Effekts zu untersuchen. Einerseits wirkt Managed Care direkt auf die Leistungserbringer, andererseits aber auch auf die Versicherten bzw. die Patienten.

Beide Effekte sind auf der beschriebenen Untersuchungsebene endogen. D.h., dass sowohl die Entscheidung des Versicherten für Managed Care als auch die Beteiligungsentscheidung des Arztes für Managed-Care-Modelle die Leistungshöhe beeinflussen. Nur wenn es gelingt, diese Selektionseffekte herauszufiltern, kann der Managed-Care-Effekt identifiziert werden. In der Schätzung des Gesamtmodells wird der Endogenität des MC-Effekts Rechnung getragen, indem

---

<sup>170</sup> Eine kurze Einführung in die Multilevel-Analyse sowie eine Darstellung ihrer Anwendungsmöglichkeiten in der Gesundheitsökonomie bieten Rice und Jones (1997).

die Vertragswahl der Versicherten und die Beteiligungsentscheidung der Ärzte berücksichtigt wird.

Deshalb ist es zunächst notwendig, Abschätzungen beider Effekte zu erhalten. Die *Vertragswahl* der Versicherten kann innerhalb eines Nachfragemodells berücksichtigt werden (Abschnitt 9.3). Für die Beteiligungsentscheidung des Arztes wird vermutet, dass dafür bestimmte Praxiseigenschaften und die Produktivität der Arztpraxis verantwortlich sind. Eine Approximation dieses Effekts kann durch die Ermittlung der *Effizienz* der Arztpraxen erfolgen (Abschnitt 9.2). Darüber hinaus kann man die einzelne Arzt-Patienten-Beziehung nicht losgelöst von ihren Umweltfaktoren beurteilen, welche insbesondere in der *Marktstruktur* ihren Ausdruck finden. Die Marktstruktur soll daher als erstes etwas näher untersucht werden (Abschnitt 9.1).

Im Kapitel 8 wurde allgemein auf den Übergang von der Theorie zur Empirie eingegangen. Die konkreten empirischen Modelle erfordern aber eine eigenständige theoretische Basis. Auf der Grundlage der Ergebnisse aus den theoretischen Untersuchungen wird für alle vorgestellten Modelle der Übergang explizit beschrieben. Dabei wird im Gesamtmodell direkt auf die abgeleiteten theoretischen Analysen zurückgegriffen. Die Modelle zur Marktstruktur und zur Effizienz basieren hingegen auf eigenständigen theoretischen Fundamenten, die ihre Berechtigung gleichwohl aus den vorangegangenen Überlegungen gewinnen.

#### 9.1 Marktstruktur und Verhalten des Arztes auf Marktebene

Nach den Ergebnissen in Kapitel 6 wird der Markt für Arztdienstleistungen am ehesten durch das Modell der monopolistischen Konkurrenz beschrieben. Diese Vermutung kann auf unterschiedliche Arten getestet werden. So zum Beispiel mit einem von Panzar und Rosse<sup>171</sup> erarbeiteten Test, der die Marktstruktur mit Hilfe der Erlöse und der Inputpreise der einzelnen Firma testet. Sie zeigen durch eine komparativ statische Analyse, dass ihre Teststatistik, nämlich die Summe aller Elastizitäten bezüglich der Erlöse für jede Marktstruktur einen bestimmten Wert annehmen muss.<sup>172</sup> Der vorliegenden Untersuchung stehen aber keine Inputpreise zur Verfügung, sodass ein Test über die Elastizitäten nicht möglich ist. Immerhin sind die Erlöse bzw. Umsätze der Ärzte, sowie die Arztdichte bekannt. Da entsprechend der theoretischen Untersuchung vor allem die Anzahl der Ärzte in einer bestimmten Region ein wichtiger Indikator für die

---

<sup>171</sup> Vgl. Wong (1996).

<sup>172</sup> Vgl. Wong (1996).

monopolistische Konkurrenz ist, soll im Folgenden über diesen Ansatz die Marktstruktur getestet werden.

Um die Arztdichte pro Gemeinde zu erklären, kann man zunächst von einer allgemeinen Formulierung ausgehen:

$$AD = f(Patient, Arzt, Spez, Gem), \quad (9.1)$$

d.h., dass die Arztdichte der Allgemeinmediziner von Eigenschaften der potenziellen Patienten (*Patient*), von bestimmten Eigenschaften der bereits in dieser Region tätigen Ärzte (*Arzt*), von Art und Umfang der verfügbaren Spezialleistungen in der Region (*Spez*) sowie Faktoren, die bestimmte Unterschiede zwischen den einzelnen Gemeinden (*Gem*) berücksichtigen, beeinflusst wird. Die grundlegende Idee dabei ist, dass sowohl nachfrageseitige als auch angebotsseitige Faktoren die Arztdichte beeinflussen.

Wie bei *Brasure et al.* (1999) sei ein regionaler Markt mit einer bestimmten Bevölkerungsgröße angenommen.<sup>173</sup> Zunächst wird die Bevölkerung von einem Arzt versorgt, sodass dieser Arzt anfangs als Monopolist agiert, d.h., dass er Leistungen anbietet, deren Preis über den Grenzkosten liegen (vgl. Abschnitt 6.1). Gleichzeitig sei angenommen, dass der Preis gerade den Durchschnittskosten entspricht, sodass die Durchschnittskostenkurve (DK) die Nachfragekurve (N) gerade am Punkt des Monopolpreises tangiert. In einer solchen Situation entstehen dem Monopolisten keine Gewinne, obwohl der Preis höher als der Wettbewerbspreis ist und die nachgefragte Menge unter der unter Wettbewerbsbedingungen nachgefragten Menge liegt. Aufgrund dieser Null-Gewinne besitzt auch kein weiterer Anbieter den Wunsch, in diesen Markt einzutreten.<sup>174</sup>

Durch die Zunahme der Bevölkerung in der Region verschiebt sich die Nachfragekurve nach außen. Die zusätzliche Nachfrage verschafft dem einzigen Anbieter auf dem Markt einen positiven Gewinn und ist gleichzeitig mit positiven Gewinnaussichten für weitere Anbieter verbunden. Ein zweiter Anbieter wird genau dann in den Markt eintreten, wenn er seine Null-Gewinnsgrenze erreicht hat. Steigt die Nachfrage weiter, nehmen auch die Profite zu und es wird wieder ein Punkt erreicht, wo es für einen weiteren Anbieter von Interesse sein kann, in den Markt einzusteigen, sofern er zumindest seine Kosten durch die unternehmerische Tätigkeit

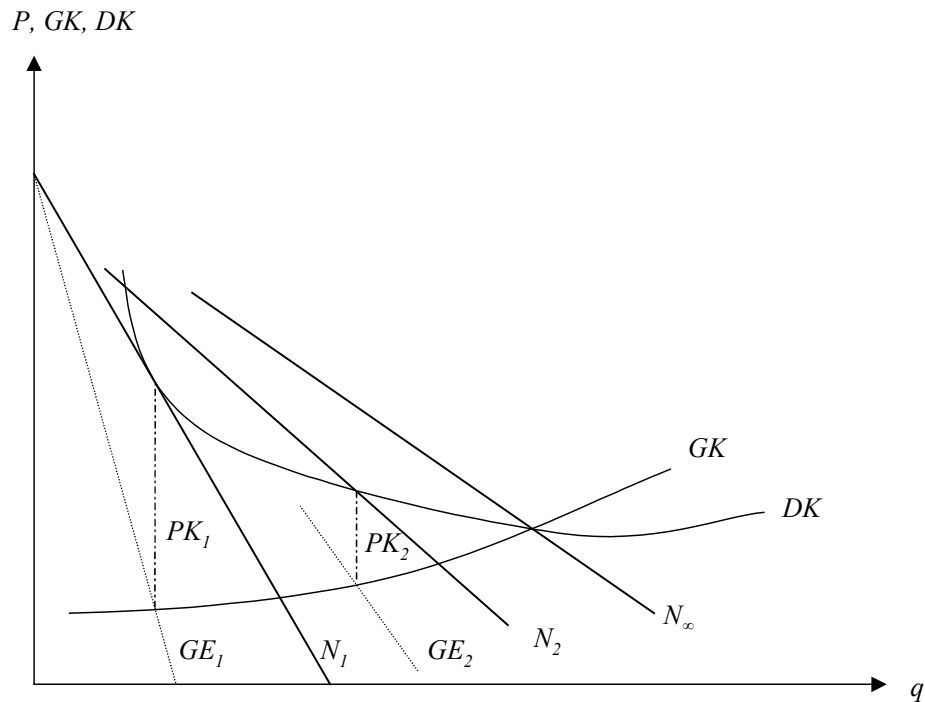
---

<sup>173</sup> *Brasure et al.* (1999) wenden dabei die Idee von *Bresnahan und Reiss* (1991) auf den Arztmarkt an.

<sup>174</sup> Die Null-Gewinn-Bedingung kann daher als Markteintrittsbarriere aufgefasst werden. Die Durchschnittskosten, die zu diesem Null-Gewinn führen, können bspw. aufgrund hoher Ausbildungsinvestitionen entstehen.

decken kann. Mit jedem weiteren Anbieter nimmt der Wettbewerb auf dem Markt zu. Dies kommt in einem fallenden Preis-Grenzkostenverhältnis zum Ausdruck, welches schließlich bei vollkommenem Wettbewerb, d.h. sehr vielen Anbietern, eins wird. Abbildung 7 verdeutlicht den beschriebenen Zusammenhang.

Abbildung 7 Angebot und Nachfrage bei unterschiedlichen Marktgrößen



Quelle: modifiziert nach Brasure et al. (1999)

$P$  = Preis,  $GK$  = Grenzkosten,  $DK$  = Durchschnittskosten,  $N$  = Nachfrage,  $GE$  = Grenzerlös,  $PK$  = Preis-Kosten-Margin,  $q$  = Menge

Daraus kann man nun Rückschlüsse auf die jeweilige zusätzliche Bevölkerung ziehen, welche notwendig ist, damit ein weiterer Anbieter in den regionalen Markt hineingeht. Aufgrund des zunächst fallenden Preis-Grenzkostenverhältnisses wird vermutet, dass das Bevölkerungswachstum mit zunehmender Anbieterzahl zunimmt, wobei die Bevölkerung für die Standortwahl der Ärzte als exogen betrachtet wird. Ist einmal das Wettbewerbsniveau erreicht, kommt es zu neuen Markteintritten bei gleichen Zuwachsraten der Bevölkerung.



### 9.1.1 Ökonometrisches Modell

Um die notwendigen Zuwachsraten der Bevölkerung bestimmen zu können, wird ein geordnetes Probitmodell geschätzt.<sup>175</sup> Mit Hilfe eines solchen Modells lassen sich die Wahrscheinlichkeiten berechnen, mit denen in einer Region, welche durch die Faktoren  $x$  charakterisiert wird, eine bestimmte Anzahl von Ärzten tätig sind.  $N_{it}$  bezeichne die Anzahl Ärzte, die in einer bestimmten Region  $i$  im Jahr  $t$  beobachtet wird. Die Regionen können entsprechend ihrer Ärztezahl in  $k$  Kategorien ( $0 \leq k \leq K$ ) eingeteilt werden.

Bezeichne nun  $y = 0, \dots, K$  diese Kategorien, so lässt sich das Modell allgemein wie folgt schreiben:

$$\begin{aligned} y_{it} &= 0 && \text{wenn } N_{it} = 0 \\ y_{it} &= 1 && \text{wenn } 0 < N_{it} \leq \mu_{1,t} \\ &\vdots && \\ y_{it} &= K && \text{wenn } \mu_{K-1,t} \leq N_{it} \end{aligned} \quad (9.2)$$

wobei  $\mu = [\mu_{1,t}, \dots, \mu_{K-1,t}]$  zu schätzende Parameter darstellen, die mit dem Trennungspunkt der Kategorien verbunden sind und sich für die Kategorien über die Zeit verändern können. Als Haupterklärungsvariable wird die logarithmierte Bevölkerung betrachtet. Da das Hauptinteresse der Untersuchung aber in der Berechnung der notwendigen Bevölkerungsgröße für einen bestimmten Trennungspunkt  $k$  liegt, kann das empirische Modell auch folgendermaßen geschrieben werden:

$$\Pr(y_{it} = k) = \Pr(\mu_{k-1,t} \leq \beta_{\text{Bev}} \ln(\text{Bev}_{it}) + \beta' x_{it} + \varepsilon_{it} < \mu_{k,t}). \quad (9.3)$$

Dabei bezeichnet  $\beta$  den zu schätzenden Parametervektor der erklärenden Faktoren des Wahrscheinlichkeitsmodells und  $\varepsilon$  den Störterm. Über die geschätzten Trennungspunkte lassen sich nun die Nachfrageniveaus bestimmen, die notwendig sind, um genau einen bestimmten Trennungspunkt zu erreichen:

$$\ln(\text{Bev}_{it}) = \frac{\mu_{k-1,t} - \beta' x_{it}}{\beta_{\ln \text{Bev}}}. \quad (9.4)$$

Damit lassen sich mit (9.4) die Bevölkerungsgrößen berechnen, die mindestens erforderlich sind, um ein bestimmtes Angebotsniveau  $k$  zu erreichen.

---

<sup>175</sup> Diese Vorgehensweise wurde von Brassure et al. (1999) übernommen. Eine kritische Würdigung dieses Ansatzes erfolgt am Ende dieses Abschnitts.

Um Gleichung (9.4) lösen zu können, müssen die  $\mu$  s und  $\beta$  s mit dem geordneten Probitmodell geschätzt werden. Für die Wahrscheinlichkeiten ergibt sich mit  $\Phi$  der Verteilungsfunktion der Normalverteilung:<sup>176</sup>

$$\begin{aligned}
 \Pr(y = 0) &= \Phi(-\beta'x), \\
 \Pr(y = 1) &= \Phi(\mu_1 - \beta'x) - \Phi(-\beta'x), \\
 \Pr(y = 2) &= \Phi(\mu_2 - \beta'x) - \Phi(\mu_1 - \beta'x), \\
 &\vdots \\
 \Pr(y = K) &= 1 - \Phi(\mu_{K-1} - \beta'x).
 \end{aligned}
 \tag{9.5}$$

### 9.1.2 Daten

Grundlage der zu erklärenden Variablen ( $y$ ) ist die Anzahl Ärzte (Allgemeinmediziner, Ärzte der inneren Medizin sowie Kinderärzte) in einer Gemeinde. Es gibt relativ viele Gemeinden, in denen überhaupt kein Arzt tätig ist und einige wenige Gemeinden, in denen sehr viele Ärzte arbeiten. Dabei spielen die großen Städte Zürich und Genf, aber auch Winterthur eine besondere Rolle. Insgesamt werden 171 Gemeinden im Kanton Zürich und 45 Gemeinden im Kanton Genf untersucht. Im Unterschied zur Abgrenzung des Untersuchungsgebietes für die CSS-Daten (vgl. Tabelle 7), werden hier jetzt also alle Gemeinden in den beiden Kantonen betrachtet.

Tabelle 10 zeigt die durchschnittlichen Bevölkerungszahlen für Gemeindekategorien mit unterschiedlicher Anzahl Ärzte. Man erkennt, dass die Arztzahlen mit zunehmender Bevölkerung in der Regel deutlich steigen (Ausnahmen bilden die 7. und die 11. Kategorie). Im Lichte der Theorie kann das zunächst sowohl für einen Wettbewerbs- als auch einen Markt mit monopolistischer Konkurrenz sprechen.

---

<sup>176</sup> Wobei die Indizes im Folgenden vernachlässigt werden Vgl. zum geordneten Probitmodell Greene, Kapitel 19.8 (1997).

Tabelle 10 Anzahl Ärzte und Gemeindegröße (1999)

Y	Anzahl Ärzte	Ärzte Max	Ärzte Min	Anzahl Gemeinden	Bevölkerungsdurchschnitt
0	0	0	0	53	825
1	1	1	1	44	1,953
2	2	2	2	29	2,721
3	3	3	3	12	3,515
4	4	4	4	15	5,470
5	5	5	5	5	4,959
6	6	6	6	8	6,532
7	7	7	7	4	5,472
8	9	10	8	10	9,001
9	11.78	13	11	9	11,133
10	15	16	14	7	15,069
11	17.64	19	17	11	13,576
12	25.83	36	20	6	18,738
13	351.33	577	113	3	199,109

Die großen Städte, die in der 13. Kategorie vereint sind, stellen eine eigene Kategorie dar. Im Unterschied zu allen anderen Kategorien grenzt diese Kategorie nicht direkt an die vorgehende Kategorie an. Um konsistente Schätzungen für die anderen Marktgrößen zu erhalten, wurden daher die Städte zunächst aus der Analyse ausgeschlossen.

In Tabelle 11 sind die deskriptiven Statistiken der erklärenden Variablen angegeben. Neben der Haupterklärungsvariablen *LNPOP*, der logarithmierten Bevölkerung, sind dies Variablen, die zum einen die Bevölkerungsstruktur und somit die Nachfragesituation in der Gemeinde berücksichtigen, und zum anderen auch die geografischen Besonderheiten widerspiegeln sollen (Anteil Männer an der Gesamtbevölkerung – *MANN\_ANT*, Anteil Gestorbener – *GEST\_ANT*, Anteil Geborener – *GEB\_ANT*, Anteil Ausländer – *AUS\_ANT* und die Bevölkerungsdichte – *BEVD*). Die robusten Durchschnittskosten pro Gemeinde sollen den durchschnittlichen Gesundheitszustand der Bevölkerung und das Nachfrageverhalten der Bevölkerung approximieren (*DK*). Hinzu kommen noch Variablen, welche die Angebotsituation in der Region ausdrücken. Erstens kann davon ausgegangen werden, dass das Angebot an Spezialmedizin in einer Region einen Einfluss auf die Arztdichte von Allgemeinmedizinern hat (Anzahl von Spezialisten der inneren Medizin – *ANZ\_INN* und Anzahl von Chirurgen in der Gemeinde – *ANZ\_CH*). Zweitens ist zu prüfen, ob die Verfügbarkeit an Krankenhausbetten einen Einfluss auf die Arztdichte hat oder nicht. Letztere Variable wird durch die Anzahl an Spitälern (*ANZ\_SPIT*) angegeben.

Tabelle 11 Deskriptive Statistik der erklärenden Variablen für die Schätzung der Anzahl Ärzte (1999) für  $Y = 0 - 12$

	Mittelwert	Standardfehler	Minimum	Maximum
ZH	0.793	0.406	0	1
LNBEV	7.894	1.076	5.553	10.258
MANN_ANT	0.496	0.014	0.424	0.530
GEST_ANT	0.007	0.003	0.000	0.023
GEB_ANT	0.011	0.003	0.005	0.023
AUS_ANT	0.157	0.094	0.023	0.456
BEVD	7.000	9.246	0.398	61.480
DK	138.967	26.667	98.010	213.32
ANZ_IN	0.216	0.824	0	7
ANZ_CH	0.549	1.763	0	15
ANZ_SPIT	0.117	0.337	0	2
ST_P_K	1337	1030	364	6980

Drittens geht noch die Höhe des Steuerertrags pro Person in jeder Gemeinde ( $ST\_P\_K$ ) in SFr. In die Schätzung ein. Wie Standardfehler, Minimum und Maximum dieser Variablen zeigen, existieren zwischen den Gemeinden erhebliche Unterschiede im Steuerertrag pro Kopf, sodass man auch von erheblichen Einkommensunterschieden der Gemeinden ausgehen kann.

### 9.1.3 Schätzung

Das geordnete Probitmodell wurde mittels Maximum-Likelihood geschätzt, wobei die Panel-Datenstruktur (217 Gemeinden über 3 Jahre) berücksichtigt wurde.

Der positive Koeffizient für  $LNBEV$  erweist sich als hoch signifikant (vgl. Tabelle 12). Die Zusammensetzung der Bevölkerung in der Gemeinde hat ebenfalls einen großen Einfluss auf die Wahrscheinlichkeit, wie viele Ärzte in einer Gemeinde tätig sind (signifikante Koeffizienten von  $MANN\_ANT$ ,  $GEB\_ANT$  und  $GEST\_ANT$ ). Auch die Bevölkerungsdichte hat einen signifikanten Einfluss. Interessant sind ferner die Koeffizienten der angebotsseitigen Erklärungsfaktoren.

Die Koeffizienten in einem geordneten Probitmodell sind vorsichtig zu interpretieren. Da man Wahrscheinlichkeiten betrachtet – die sich über die Kategorien zu eins summieren –, kann ein positiver Koeffizient nicht einfach als ein positiver Effekt auf die Wahrscheinlichkeiten gewertet werden. Der Effekt eines Bevölkerungswachstums auf die Wahrscheinlichkeit, dass  $k$  Ärzte in der Region tätig sind, lässt sich wie folgt berechnen:

$$\frac{\partial \Pr(y = k)}{\partial \ln Bev}$$

Die Wahrscheinlichkeitsänderung bei einem positiven Parameter in der ersten Kategorie ist negativ, während sie in der letzten Kategorie positiv ist. Denn aufgrund von (9.5) gilt für diese beiden Kategorien:

$$\begin{aligned} \frac{\partial \Pr(y=0)}{\partial \ln Bev} &= -\phi(\beta'x) \beta_{\ln Bev} \\ \frac{\partial \Pr(y=13)}{\partial \ln Bev} &= \phi(\mu_{12} - \beta'x) \beta_{\ln Bev}. \end{aligned} \tag{9.6}$$

Während diese beiden Wahrscheinlichkeitsveränderungen aufgrund einer Veränderung in der erklärenden Variablen vom Vorzeichen her eindeutig bestimmt sind, gilt dies nicht mehr uneingeschränkt für die dazwischenliegenden Kategorien. Deren Vorzeichen wird vielmehr auch von der Dichtefunktion selbst mitbestimmt.<sup>178</sup> Konkret können wir also den Einfluss der Bevölkerung auf die Wahrscheinlichkeiten nur für die erste und die letzte Kategorie angeben. Für die erste Kategorie folgt aus dem positiven Koeffizienten der Bevölkerung: Die Wahrscheinlichkeit, dass in einer Gemeinde kein Arzt tätig ist, nimmt mit zunehmender Bevölkerungszahl ab.

Mit Hilfe der Vorzeichen der geschätzten Parameter lassen sich auf die gleiche Weise weitere interessante Ergebnisse ableiten: So gibt es zwischen den Kantonen Genf und Zürich keine signifikanten Unterschiede. Da die Arztzahlen im Kanton Genf wesentlich höher liegen als im Kanton Zürich (8.9 vs. 6.8 Ärzte pro 10,000 Einwohner in der Stichprobe) werden etwaige Unterschiede zwischen den Kantonen durch die anderen erklärende Variablen abgebildet. Weil das neben der Bevölkerungszahl vor allem auch Faktoren der Nachfrage sind (Bevölkerungsdichte, Anteil Geborener, Anteil Gestorbener, Durchschnittskosten der Gemeinde), lassen sich die höheren Arztzahlen in Genf als Ausdruck der unterschiedlichen Präferenzen der Bevölkerung interpretieren. Interessant ist auch, dass mit einem höheren Steuerertrag pro Person die Wahrscheinlichkeit zunimmt, dass in einer Gemeinde überhaupt ein Arzt anzutreffen ist. Schließlich spielen die weiteren Angebotsfaktoren (Anzahl verschiedener Spezialisierungsrichtungen) eine wichtige Rolle: In Gemeinden mit vielen Ärzten befinden sich eher Spitäler und Chirurgen. Für Ärzte der speziellen Inneren Medizin gilt das Gegenteil. Danach nimmt die

---

<sup>177</sup> Mit  $\phi(\cdot)$  der Dichtefunktion der Normalverteilung.

<sup>178</sup> Denn es gilt für die Kategorien  $k = 1, \dots, K-1$ :  $\frac{\partial \Pr(y=k)}{\partial \ln Bev} = [\phi(-\beta'x) - \phi(\mu_{k-1} - \beta'x)] \beta_{\ln Bev}$ . Vgl. Greene (1997).

Wahrscheinlichkeit, dass es in einer Gemeinde keinen Allgemeinarzt gibt, mit steigender Anzahl Ärzte der speziellen inneren Medizin zu. Dies kann damit erklärt werden, dass in diesem Fall Ärzte der speziellen Inneren Medizin Aufgaben von Allgemeinmedizinern übernehmen.

Tabelle 12 Geordnetes Probitmodell für Anzahl Ärzte ( $Y = 0 - 12$ )

VARIABLE	Koeffizient	Standardfehler	VARIABLE	Koeffizient	Standardfehler
ZH	0.145	0.421	DK	0.0008	0.006
LNPOP	2.312 **	0.131	ANZ_IN	-0.299 **	0.111
MANN_ANT	-30.662 **	5.361	ANZ_CH	0.312 **	0.064
GEST_ANT	-55.071 **	20.410	ANZ_SPIT	0.382 *	0.192
GEB_ANT	-1.468	18.896	ST_P_K	0.00012 **	5.26E+05
AUS_ANT	-1.042	1.057			
BEVD	0.058 **	0.009	LOG-L	-750.33	
			LOG-L(0)	-1420.04	

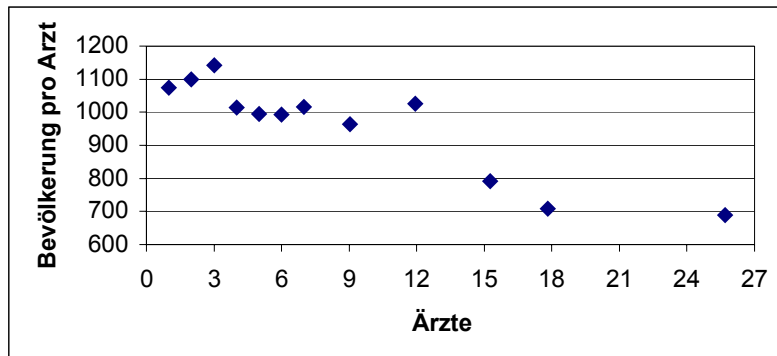
\*)  $p < 0.05$ , \*\*)  $p < 0.01$

Dass die Durchschnittskosten (DK) in den Gemeinden keinen Einfluss auf die Wahrscheinlichkeit einer Arztansiedlung haben, ist zunächst verwunderlich. Bei einer näheren Inspektion der Korrelation zwischen den erklärenden Variablen zeigt sich, dass diese Variable hoch mit anderen Variablen (vor allem nachfrageseitigen Faktoren) korreliert ist. Aus diesem Grund wurde das Modell unter Vernachlässigung der Durchschnittskosten und des Ausländeranteils, der ebenfalls für die hohe Multikollinearität verantwortlich ist, neu geschätzt. Die Neuschätzung ergab bis auf eine Ausnahme keine signifikanten Veränderungen hinsichtlich der Schätzung der anderen Koeffizienten. Die Ausnahme betrifft den Koeffizienten der Kantonsvariable (*ZÜRICH*), die jetzt fast signifikant positiv ist ( $p = 0.065$ ). D.h., dass die Hauptunterschiede zwischen den beiden betrachteten Kantonen durch die unterschiedlich hohen Kosten in den Kantonen widergespiegelt werden.

Entsprechend (9.4) lassen sich nun mit den geschätzten Koeffizienten die uns interessierenden Trennungspunkte der Bevölkerung für die einzelnen Kategorien berechnen. In Abbildung 8 sind die resultierenden Bevölkerungszahlen pro Arzt in Abhängigkeit der Arztlzahl für das Jahr 1999 abgetragen. Danach „lohnt“ es sich für den ersten Arzt in den Markt einzutreten, wenn in der Gemeinde mehr als 1087 Personen leben. Die Grafik zeigt, dass zwischen 1 und 3 Ärzten pro Gemeinde die Bevölkerung pro Arzt zunimmt, danach zwischen 4 und 12 Ärzten ungefähr gleich bleibt und in größeren Gemeinden (Gemeinden mit mehr als 12 Ärzten) kontinuierlich abnimmt, wobei sich diese Abnahme bei 18 Ärzten stark abschwächt. Auch die in der Grafik

nicht berücksichtigten Städte passen in dieses Bild, denn die durchschnittliche Bevölkerung pro Arzt in diesen Gemeinden beträgt 567.

Abbildung 8 Geschätzte Bevölkerung pro Arzt (1999)



Die geschätzten Trennpunkte deuten daher darauf hin, dass in den Gemeinden mit wenigen Ärzten am ehesten von einer Situation mit monopolistischer Konkurrenz gesprochen werden kann, da hier die Bevölkerungszahlen von einer Stufe zur nächsten relativ stark ansteigen. Im mittleren Bereich sprechen die Ergebnisse für einen wettbewerblich organisierten Markt, da in diesem Bereich ein zusätzlicher Arzt bei fast konstantem Bevölkerungswachstum in den Markt eintritt. In den Gemeinden mit mehr als 12 Ärzten führt der Konkurrenzdruck schließlich dazu, dass zusätzliche Ärzte mit einem immer kleiner werdenden Patientenstamm zu kämpfen haben. Ab 18 Ärzten in einer Gemeinde kann hingegen wieder von einem Wettbewerbsmarkt gesprochen werden, da die Bevölkerung pro Arzt ab dieser Ärztezah ungefähr gleich bleibt. Die großen Städte mit durchschnittlich 567 Einwohner pro Arzt passen – wie gesagt – auch sehr gut in dieses Bild, was nicht zuletzt durch die Schätzung des Modells unter Einbeziehung der 13. Kategorie bestätigt wird.

#### 9.1.4 Diskussion und Kritik

Die vorgestellte Schätzung erfolgte auf Gemeindeebene und postuliert auf Marktebene einen engen Zusammenhang zwischen Nachfrage und Angebot. Genauer gesagt, geht das Modell davon aus, dass die Ärzte sich bei ihrer Niederlassungsentscheidung hauptsächlich daran orientieren, wie hoch die Nachfrage in den einzelnen Gemeinden ist. Darüber hinaus dürften aber noch andere Faktoren bei dieser Entscheidung eine Rolle spielen. So können die Präferenzen der Ärzte für eine urbanere Umgebung eventuell erklären, warum sie mit einem viel geringeren Patientenstamm zufrieden sind und somit eventuell auch Einkommenseinbußen in

Kauf nehmen. Andererseits muss ein verstärkter Wettbewerb nicht notwendigerweise zu einer Verringerung des Einkommens für den einzelnen Arzt führen. Denn die Marktmacht des einzelnen Arztes kann gerade durch den Wettbewerb verstärkt werden.

Kritisch anzumerken ist die Marktabgrenzung über die politischen Gemeinden. So sind die über die Gemeinden identifizierten Märkte sicherlich nicht abgeschlossen genug, um der Marktdefinition zu entsprechen (vgl. Abschnitt 7.4.3). Für weitere Untersuchungen für die ganze Schweiz würde sich daher die Definition des Marktes auf einer höheren Aggregationsebene wie den Bezirken oder den sogenannten MS-Regionen anbieten.<sup>179</sup> Eine derartige Abgrenzung konnte in dieser Untersuchung nicht vorgenommen werden, da nur zwei Kantone in der Analyse berücksichtigt worden sind.

Trotz der kritischen Anmerkungen bleibt festzuhalten, dass die Arztzahl in den Gemeinden durch die Nachfrage im Allgemeinen und die Präferenzen der Bevölkerung im Besonderen determiniert wird. Aus diesem Grund scheint die Festlegung einer bestimmten Arztzahl für eine bestimmte Region von außen eher schwierig zu sein. In diesem Sinne steht eine Betrachtung von Verhältniszahlen oder deren Veränderung über die Zeit, wie sie häufig betrieben wird, um bestimmte „wünschenswerte“ Entwicklungen zu beschreiben, im Gegensatz zu der hier vorgenommenen Analyse. Die hypothetische Aussage: „Bis zum Jahre 2020 entsteht ein Mangel von 35,000 Ärzten“ verdeutlicht dies.<sup>180</sup> Vielmehr geht es um die Bestimmung der *optimalen* Anzahl von Ärzten für bestimmte Bevölkerungsgrößen, wobei die unterschiedlichen Präferenzen wie auch die jeweiligen Marktbedingungen in den Regionen zu berücksichtigen sind. Im Kontext der vorangegangenen theoretischen Untersuchungen zur optimalen Arztvergütung bedeutet dies, dass eine Vergütung auch die jeweiligen Marktbedingungen berücksichtigen muss.

## 9.2 Effizienz von Allgemeinartzpraxen

Grundsätzlich hat die Effizienzmessung zum Ziel, diejenigen Produzenten zu finden, welche aus einem gegebenen Input den bestmöglichen Output erzeugen können bzw. zur Produktion eines bestimmten Outputniveaus die bestmögliche Inputkombination verwenden. Die Produktion ist darüber hinaus nur dann effizient, wenn derjenige Inputmix verwendet wird, der die Kosten bei

---

<sup>179</sup> MS bedeutet „mobilité spatiale“ und bezeichnet die regionale Gliederung der Schweiz nach einer bestimmten Typologie, sodass sich in einer MS-Region „ähnliche“ Gemeinden befinden (Hänggeli et al. 2001).

<sup>180</sup> Vgl. zu dieser Problematik auch Getzen (1997).



gegebenen Inputpreisen minimiert, oder wenn der Output so gewählt wird, dass er den Gewinn bei gegebenen Outputpreisen maximiert.<sup>181</sup>

Konkret geht es bei der Betrachtung von Arztpraxen um die Erreichung einer möglichst guten Gesundheit der Patienten (Output) mit möglichst geringen finanziellen Mitteln (minimalen Kosten) und damit um den optimalen Einsatz von knappen Ressourcen. Die Kosten der medizinischen Behandlung sind die selbsterbrachten und veranlassten Leistungen der Arztpraxen, die innerhalb der sozialen Krankenversicherung abgerechnet werden. Dabei umfassen die selbsterbrachten Leistungen die Behandlungsleistungen der Ärzte und direkt vom Arzt verabreichte Medikamente. Die veranlassten Leistungen bestehen aus Verschreibungen, Überweisungen an Spezialärzte, Labortests und Krankenhauseinweisungen. Die Variable Gesundheit lässt sich durch das Outputmaß „Konsultation“ approximieren.<sup>182</sup> Für viele Erkrankungen dürfte ein strikt positiver Zusammenhang zwischen Konsultationen pro Patient und Gesundheit tatsächlich gelten. Allerdings kann der Arzt durch seine Tätigkeit während einer Konsultation weitere Konsultationen vermeiden, indem er bspw. größere Medikamentenpackungen verschreibt oder einen weiteren Arztbesuch des Patienten in Abhängigkeit des Genesungsprozesses stellt. Dadurch ist der Arzt in der Lage, in dem einen oder anderen Fall genauso viel „Gesundheit“ zu erzeugen wie mit mehreren Konsultationen. Außerdem sind die Probleme der Patienten sehr vielfältiger Natur. So gibt es Krankheiten, bei denen wenige Konsultationen ausreichen, um den gewünschten Gesundheitszustand wieder herzustellen. Es gibt aber auch chronische Erkrankungen, die eine kontinuierliche Behandlung erfordern und dann auch nicht einen bestimmten Gesundheitszustand gewährleisten können. In diesem Sinne bedeutet die Approximation der Gesundheit durch die Anzahl Konsultationen pro Patient die Annahme einer sehr guten Sachwalterfunktion des Arztes.

Weitere Outputparameter können daneben Qualitätsmerkmale sein, die in der Realität allerdings nur schwer zu messen sind. Aufgrund der Vorherrschaft der Einzelleistungsvergütung kann allerdings davon ausgegangen werden, dass eine qualitätsgerechte Versorgung der Patienten gewährleistet ist; zumindest in dem Sinne, dass es zu keiner Unterversorgung kommt.<sup>183</sup>

---

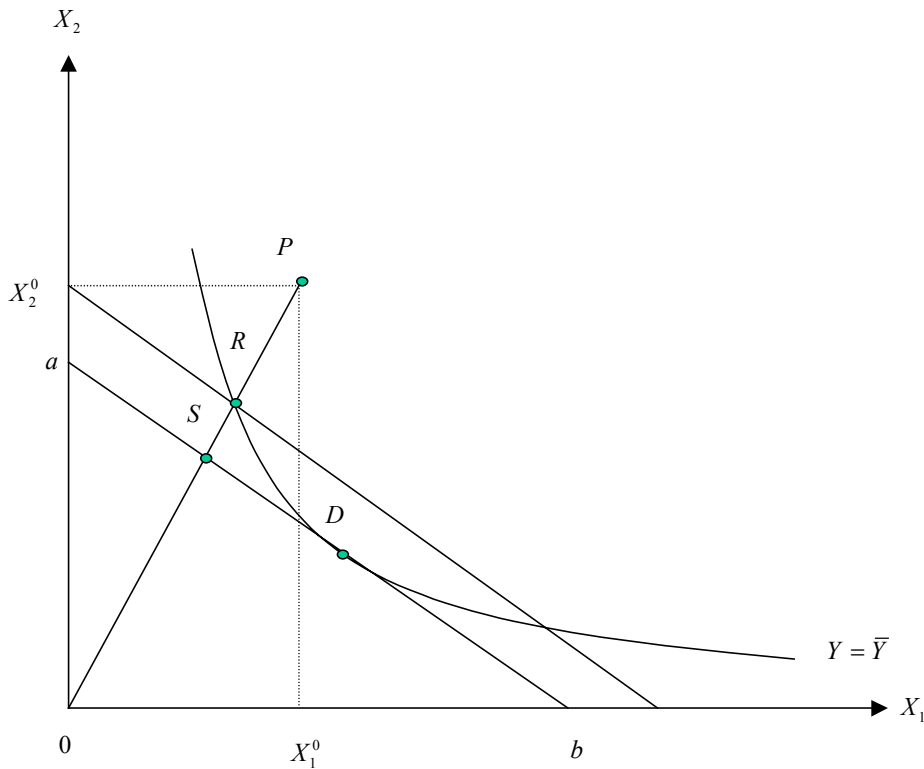
<sup>181</sup> Vgl. Kumbhakar und Lovell (2000).

<sup>182</sup> Gesundheit bedeutet in diesem Zusammenhang immer „erreichbarer Gesundheitszustand“ und nicht perfekte Gesundheit. Vgl. auch die Ausführungen im Abschnitt 6.

<sup>183</sup> Wie in früheren Abschnitten bereits dargelegt, führt die Einzelleistungsvergütung eher zu einem „Zuviel“ an Qualität (vgl. vor allem Abschnitt 6.3).

**9.2.1 Effizienzmessung und stochastic frontier analysis (SFA)**

Grundlage der Effizienzmessung bildet die Arbeit von Farrell (1957), dessen Effizienzmaße sich am besten durch eine Grafik veranschaulichen lassen.<sup>184</sup>



Betrachtet man der Einfachheit halber nur zwei Inputs – etwa Arztleistungen und Medikamente ( $X_1, X_2$ ) – und als einzigen Output die Konsultation ( $Y$ ), so lässt sich für die Produktionstechnologie  $f(X_1, X_2)$ , für die vereinfachend konstante Skalenerträge angenommen werden, eine Isoquante eines bestimmten Produktionsniveaus ( $Y = 1$ ) angeben, welche die effizienten Inputkombinationen zur Produktion von  $Y$  beschreibt. Betrachtet man nun einen beliebigen Punkt  $P$ , in dem mit den beiden Inputs  $X_1^0$  und  $X_2^0$  gerade dieses Outputniveau erzeugt wird, so sieht man, dass diese Inputkombination ineffizient ist, da man zwar den gleichen Inputmix aber von beiden Inputs weniger einsetzen könnte, um das gleiche Outputniveau zu erzeugen. Eine mögliche Inputkombination stellt Punkt  $R$  auf der Isoquante dar.

Das Verhältnis der Strecken  $OP$  und  $OR$  misst daher die sogenannte *Technische Effizienz* ( $TE$ ):

$$TE = OR / OP \quad (0 < TE \leq 1). \quad (9.7)$$

<sup>184</sup> Vgl. Hollingsworth et al. (1999).

$TE$  ist somit für all diejenigen Produzenten 1, die auf der Isoquante liegen (so auch Punkt  $R$ ). Aufgrund dieses Maßes sind alle Punkte auf der Isoquante also gleich effizient, unabhängig davon, welche Inputkombination sie zur Erreichung dieses Zieles verwenden. Die Isoquante der effizienten Input-Output-Kombinationen kann daher auch als Effizienzfront bezeichnet werden.

Die Kosten setzen sich aus den mit Inputpreisen gewichteten Inputs zusammen. Die Gerade  $ab$  stellt die Minimalkostenkombination zu gegebenen Inputpreisen dar, um das Outputniveau  $\bar{Y}$  zu erreichen. Der Produzent im Punkt  $D$  ist *kosteneffizient* ( $KE$ ), da die Gerade der Minimalkostenkombination in diesem Punkt die Isoquante tangiert. Wäre das Inputpreisverhältnis ein anderes –  $X_1$  bspw. relativ teurer als  $X_2$  –, so könnte auch Punkt  $R$  die kosteneffiziente Inputkombination sein. Im dargestellten Fall ist Punkt  $R$  bezüglich der Kosten gegenüber Punkt  $D$  ineffizient, obwohl beide Punkte Inputkombinationen kennzeichnen, die eine technisch effiziente Produktion erlauben. Die Gesamteffizienz berechnet sich schließlich aus dem Produkt der ersten beiden Effizienzmaße und wird als *allokative* Effizienz bezeichnet.

Wir untersuchen hier nur die Technische Effizienz der ärztlichen Produktion. Für das obige Beispiel mit zwei Inputfaktoren wird das Verhältnis von Arztleistungen zu Medikamenten bei effizienter Produktion einer Konsultation im Vergleich zu dem tatsächlich gemessenen Inputverhältnis einer bestimmten Arztpraxis betrachtet.

Die technische Effizienz der Arztpraxen wird durch ein ökonometrisches Verfahren ermittelt, der *stochastic frontier analysis* (SFA).<sup>185</sup> Einen alternativen Ansatz stellt die Data Envelopment Analysis (DEA) dar. Während die ökonometrischen Verfahren parametrische Verfahren sind, ist die DEA ein nichtparametrisches Verfahren.

Mit der SFA wird versucht, die  $TE$  eines jeden Arztes zu bestimmen. Dabei wird angenommen, dass die Abweichungen des einzelnen Arztes von der Effizienzfront zum einen von unbeobachtbaren Faktoren abhängen, jedoch zum anderen auch von Effizienzunterschieden. Der Hauptunterschied zu den nichtparametrischen Verfahren der Effizienzfrontanalyse besteht darin, dass mit Hilfe der vorhandenen Daten eine Effizienzfront geschätzt wird, während die Effizienzfront bei der DEA durch die Lösung eines linearen Maximierungsproblems bestimmt wird.

---

<sup>185</sup> Andere ökonometrische Schätzansätze zur Effizienzmessung werden u.a. in Giuffrida und Gravelle (2001) diskutiert.

Beide Verfahren haben Vor- und Nachteile. So liegt der Hauptvorteil der DEA in der nichtparametrischen Bestimmung der Effizienzfront selbst, da keine Annahmen bezüglich der funktionalen Form der Produktionsfunktion notwendig sind.<sup>186</sup> Diese Annahmen sind für die SFA hingegen sehr wichtig, da sie ein wesentlicher Bestandteil für die Schätzung der Produktionsfunktion sind. Die DEA hat zudem den Vorteil, dass sie auch mit einem komplexen Produktionsumfeld umgehen kann, d.h. es können auch Technologien mit mehreren Outputs betrachtet werden.<sup>187</sup> Hinter beiden Methoden stehen ebenfalls (unterschiedliche) Annahmen, die nicht getestet werden können. So geht die DEA von der Annahme aus, dass es keine Messfehler oder andere zufällige Abweichungen gibt; hingegen müssen in der SFA Verteilungsannahmen bezüglich des Fehlerterms gemacht werden.<sup>188</sup> Aufgrund der stochastischen Formulierung hat die SFA schließlich den Vorteil, dass die Güte des Modells getestet werden kann (Signifikanztest der Parameter und von Restriktionen). Bootstrapverfahren erlauben es mittlerweile auch in der DEA, zumindest Konfidenzintervalle für die Effizienzfront zu erzeugen.<sup>189</sup>

Beide Verfahren werden häufig zur Ermittlung der relativen Effizienz von Leistungserbringern im Gesundheitswesen eingesetzt. Insbesondere gibt es eine breite Literatur zu Effizienzvergleichen von Krankenhäusern.<sup>190</sup> Die Anwendung dieser Methoden auf Vergleiche ambulanter Leistungserbringer bildet allerdings noch die Ausnahme.<sup>191</sup>

Wie bereits erwähnt, wird die Effizienzfront der SFA aus der Schätzung einer Produktionsfunktion gewonnen. Die Abweichung einer einzelnen Beobachtung von dieser Front wird nicht nur als zufällige Abweichung aufgefasst – wie es bei der gewöhnlichen OLS-Schätzung geschehen würde –, sondern in zwei Effekte zerlegt: in eine systematische und eine zufällige Abweichung,

---

<sup>186</sup> Ausgenommen sind hiervon Annahmen bezüglich der Skalenerträge: Eine Annahme über konstante, abnehmende oder steigende Skalenerträge muss explizit gemacht werden.

<sup>187</sup> Vgl. Jacobs (2001).

<sup>188</sup> Vgl. vor allem Newhouse (1994). Im gleichen Heft (*Journal of Health Economics* 13 (1994)) finden sich weitere Beiträge, welche die Vor- und Nachteile der Anwendung des einen oder des anderen Verfahrens im Gesundheitsbereich diskutieren.

<sup>189</sup> Vgl. bspw. Gonzales und Miles (2002).

<sup>190</sup> Hollingsworth et al. (1999) geben einen Überblick über nichtparametrische Verfahren der Effizienzmessung im Krankenhausbereich. Dagegen verwenden Rosko (2001) sowie Li und Rosenman (2001) die SFA für Effizienzvergleiche von Krankenhäusern. Einen Vergleich beider Methoden an Hand eines Datensatzes von englischen Krankenhäusern bietet Jacobs (2001). Sie kommt in ihrem Aufsatz zu dem Schluss, dass die Methoden in Bezug auf die Reihenfolge der ermittelten Effizienzwerte keine systematischen Unterschiede aufweisen.

<sup>191</sup> Giuffrida und Gravelle (2001) wenden sowohl parametrische als auch nichtparametrische Verfahren zur Bestimmung der relativen Effizienz von Gemeinden bei der medizinischen Grundversorgung in England an. Staat (2000) untersucht die Effizienz von Allgemeinmedizinern in Österreich mittels DEA. Pai et al. (2000) bestimmen die Effizienz von Allgemeinmedizinern bei der Behandlung einer einzelnen Krankheit (Sinusitis).

wobei die systematische Abweichung als Maß der technischen Ineffizienz interpretiert wird. Allgemein kann das Produktionsmodell – zunächst ohne die zufällige Abweichung – folgendermaßen ausgedrückt werden:<sup>192</sup>

$$y_i = f(x_i; \beta) TE_i, \quad (9.8)$$

d.h., dass die Produktion  $i$  sich aus der Produktionsgrenze zusammensetzt, die sich aufgrund der unterstellten Technologie ergibt – welche man auch als das maximal zu erreichende Outputniveau auffassen kann – und einer individuellen Abweichung von der Produktionsgrenze.  $TE$  stellt somit die Abweichung der tatsächlichen Produktion von der maximal möglichen Produktion dar:

$$TE_i = \frac{y_i}{f(x_i; \beta)}. \quad (9.9)$$

Der Produktionsprozess wird nun zusätzlich noch von zufälligen Schocks überlagert, die ebenfalls produzentenspezifische Ausprägungen annehmen ( $v_i$ ). Nimmt man speziell einen Produktionsprozess der Cobb-Douglas-Form an, so lässt sich das Modell wie folgt formulieren:

$$y_i = \prod_n x_{ni}^{\beta_n} \exp(v_i) TE_i, \quad (9.10)$$

oder in log-linearer Form:

$$\ln(y_i) = \sum_n \beta_n \ln(x_{ni}) + v_i - u_i, \quad (9.11)$$

mit  $v$  dem zufälligen Störterm, für den angenommen wird, dass er unabhängig von  $u$  und identisch normalverteilt ist, und  $u$ , der nichtnegativen (technischen) Ineffizienzkomponente, aus der  $TE$  berechnet werden kann,  $TE_i = \exp(-u_i)$ . Letztere stellt nun auch eine Zufallsvariable dar, deren Verteilung – unter der Bedingung der Nichtnegativität – noch genauer spezifiziert werden muss.

Die gewählte Formulierung der Produktionsfunktion soll sowohl sinkende als auch steigende Skalenerträge erlauben, sodass nicht notwendiger Weise  $\sum_n \beta_n = 1$  gelten muss. Trotzdem stellt die modifizierte Cobb-Douglas-Produktionsfunktion noch eine relativ starke Restriktion bezüglich der Produktionstechnologie dar. Oft werden daher in ökonometrischen Produktionsanalysen flexiblere funktionale Formen verwendet, die es erlauben, diese Restriktionen zu testen.

---

<sup>192</sup> Vgl. zu der hier gewählten Formulierung des Modells Kumbhakar und Lovell (2000).

Die am häufigsten verwendete Funktion ist die sogenannte Translog-Produktionsfunktion. Sie lässt sich als eine Approximation einer unbekanntenen funktionalen Form über die zweiten Ableitungen einer solchen Funktion interpretieren. Über eine Taylor-Approximation lässt sich daher die verallgemeinerte Cobb-Douglas-Funktion als Translog-Produktionsfunktion folgendermaßen darstellen:<sup>193</sup>

$$\ln(y_i) = \sum_n \beta_n \ln(x_{ni}) + \sum_k \sum_n \beta_{kn} a_{kn} \ln(x_{ki}) \ln(x_{ni}) + v_i - u_i \quad (9.12)$$

$$\forall k \leq n \text{ und mit } a_{kn} = \begin{cases} \frac{1}{2} & \text{wenn } k = n \\ 1 & \text{wenn } k < n \end{cases}.$$

Diese Produktionsfunktion enthält nun als Spezialfälle die allgemeine Cobb-Douglas-Funktion ( $\beta_{11} = \dots = \beta_{kn} = 0, \forall k \leq n$ ) und zusätzlich die restringierte Cobb-Douglas-Funktion mit  $\sum_n \beta_n = 1$ .

Um für alle Beobachtungen die Ineffizienz berechnen zu können, müssen noch Verteilungsannahmen bezüglich der Ineffizienzkomponente gemacht werden. In Frage kommen naturgemäß ausschließlich Verteilungen mit positiven Werten. Häufig verwendete Verteilungen sind hierbei die Halb-Normal-, die zensierte Normal- und die Exponentialverteilung. Die Angemessenheit der gewählten Verteilung kann durch die Signifikanz der Verteilungsparameter überprüft werden.

Für die Schätzung der Produktionsfunktion wurden noch weitere Variablen einbezogen. So gilt es bspw. zu berücksichtigen, dass sich die einzelnen Arztpraxen auch durch bestimmte individuelle Technologieunterschiede unterscheiden. Die Praxisausstattung (Röntgengeräte) gehören genauso dazu wie die sozioökonomischen Charakteristika des einzelnen Arztes bzw. der Arztpraxis. Außerdem dürfte es auch geografische Unterschiede geben, die es zu berücksichtigen gilt. Hier denke man vor allem an die Größe des zu versorgenden Raumes (Gemeinde) oder aber auch an die Wettbewerbsintensität, die eine Rolle spielen dürfte, hier gemessen als Arztdichte und Bevölkerungsdichte.

Da Daten über 3 Jahre zur Verfügung stehen, kann das Effizienzfrontmodell im Panel geschätzt werden. Dabei wird durch die Einbeziehung des Zeitindexes als zusätzlichen Inputfaktor berücksichtigt, dass sich über die 3 Jahre technologische Veränderungen ergeben können.

---

<sup>193</sup> Vgl. Kapitel 6 und Kapitel 7 in Greene (1997).

Insbesondere soll auch der Effizienzterm nicht nur zwischen den einzelnen Arztpraxen variieren dürfen, sondern auch über die Zeit. D.h. es wird explizit zwischen technologischen Veränderungen und technischen Effizienzveränderungen unterschieden.<sup>194</sup>

Hauptaugenmerk wird auf die Erklärung der Effizienzunterschiede gelegt. Entsprechend wollen wir auch untersuchen, inwieweit bestimmte Faktoren Effizienzunterschiede zwischen den Ärzten erklären können. Insbesondere interessiert hierbei, ob Wettbewerbs- und Managed-Care-Effekte für Effizienzunterschiede verantwortlich sind. Entsprechend der theoretischen Untersuchung wird vermutet, dass es gerade diese Effekte sind, die Unterschiede im Verhältnis von eingesetzten Inputs zum gemessenen Output erklären können.

Oft verwendet man hierfür ein zweistufiges Verfahren, bei dem in einer nachgeschalteten Regression die geschätzten Effizienzwerte durch bestimmte Variablen erklärt werden sollen. Allerdings ist eine solche Vorgehensweise inkonsistent mit der Annahme der identischen Verteilung der Effizienzeffekte. Deshalb wird hier ein einstufiges Verfahren gewählt, welches das Problem vermeidet, indem der Erwartungswert der Effizienzeffekte ( $m$ ) in der Effizienzfrontschätzung als Funktion bestimmter arzt spezifischer Variablen und einer Wettbewerbsvariablen ( $z$ ) modelliert wird. In dem  $z$ -Vektor wird auch ein Zeitindex berücksichtigt, um so Effizienzveränderungen über die 3 Jahre bewerten zu können.<sup>195</sup>

Aus den genannten Bestandteilen und Gleichung (9.12) lässt sich nun unter Berücksichtigung des Zeitindex  $t$  die endgültige Schätzgleichung aufstellen:

$$\ln(y_{it}) = \sum_n \beta_n \ln(x_{nit}) + \sum_k \sum_n \beta_{kn} a_{kn} \ln(x_{kit}) \ln(x_{nit}) + \sum_l \delta_l \text{Arzteigenschaften}_{lit} + \sum_m \eta_m \text{Gemeindeeigenschaften}_{mit} + v_{it} - u_{it}, \quad (9.13)$$

wobei  $v_{it}$  identisch verteilte Zufallsvariablen mit  $N(0, \sigma_v^2)$  darstellen, die unabhängig von  $u_{it}$  sind. Die  $u_{it}$  sind dagegen nichtnegative Zufallsvariablen, die den Effizienzeffekt abbilden sollen. Für diese wird angenommen, dass sie unabhängig und mit einer zensierten Normalver

---

<sup>194</sup> Im Kontext der Analyse von Paneldaten stellt dieser Ansatz daher ein *random effects model* dar. Alternativ könnte man auch ein Modell mit fixen Zeiteffekten (*fixed effect model*) verwenden, bei dem keine Unterscheidung zwischen technologischen Veränderungen und Effizienzveränderungen gemacht wird. Vgl. zu Panel-Daten-Modellen i.A. z.B. Greene (1997) und zur Diskussion von unterschiedlichen Panel-Daten-Modellen in der SFA Coelli et al. (1998) und Khumbhakar und Lovell (2000). Zum *random effects model* vgl. auch die Ausführungen in Abschnitt 8.2.4 und die Schätzung des Gesamtmodells (Abschnitt 9.4).

<sup>195</sup> Der gewählte Modellierungsansatz geht auf Battese und Coelli (1995) zurück.

teilung verteilt sind - zens  $N(m_{it}, \sigma_u^2)$ .<sup>196</sup> Dabei lässt sich der Erwartungswert durch die  $\varepsilon$ -Variablen abbilden:

$$m_{it} = \sum_0 \delta_{\sigma \varepsilon_{oit}} . \quad (9.14)$$

Zur Schätzung des simultanen Gleichungssystems (9.13) - (9.14) wird die Maximum-Likelihood-Methode verwendet. Die Likelihoodfunktion wird hierfür durch die Varianzterme des Systems ausgedrückt, für die gilt:  $\sigma^2 = \sigma_v^2 + \sigma_u^2$ . Zusätzlich definiere  $\gamma$  das Verhältnis aus der Varianz des Ineffizienzterms und der Gesamtvarianz:  $\gamma = \frac{\sigma_u^2}{\sigma^2}$ . Für  $\gamma$  muss gelten:  $0 \leq \gamma \leq 1$ . Ein bestimmter Wert für diesen Parameter kann daher ein guter Startwert für die iterative Lösung der LogLikelihood sein.<sup>197</sup> Er dient gleichzeitig dem Test, ob die Annahme eines Ineffizienzterms gerechtfertigt ist ( $\sigma_u^2 = 0 \Rightarrow \gamma = 0$ ).<sup>198</sup>

### 9.2.2 Daten

Für die Untersuchung werden die Abrechnungsdaten des KSK über 3 Jahre (1997 – 1999) verwendet. Ärzte, die nicht in der gesamten Untersuchungszeit Leistungen abgerechnet haben, wurden ausgeschlossen. Darüber hinaus beschränkt die Forderung, dass möglichst nur voll berufstätige Ärzte im Sample vertreten sein sollen, die Anzahl der untersuchten Ärzte weiter. Die Ausschlussgrenze für die Ärzte wurde mit 20,000 SFr. relativ niedrig angesetzt. Natürlich sind Ärzte mit 20,000 SFr. Umsatz nicht unbedingt voll berufstätig. Mit der gewählten Grenze wird aber sichergestellt, dass Ärzte, die nur ein oder zwei Patienten im Jahr behandeln, aus der Untersuchung ausgeschlossen sind. Eine weitere Heraufsetzung der Umsatzgrenze hätte das Untersuchungskollektiv dagegen zu stark eingeschränkt, sodass die 20,000 SFr. einen guten Kompromiss darstellen. Das Untersuchungskollektiv umfasst somit noch 1,531 Ärzte der Fachgebiete Allgemeinmedizin, Pädiatrie und allgemeine innere Medizin, sowie einige KSK-

---

<sup>196</sup> Die zensierte Normalverteilung stellt eine Verallgemeinerung der Halb-Normal-Verteilung dar – letztere ist zensiert bei Null – und trägt somit auch der Kritik bezüglich einer zu starken Einschränkung möglicher funktionaler Formen des stochastischen Effizienzeffektes Rechnung. Natürlich bleibt die grundlegende Kritik an der Notwendigkeit einer Verteilungsannahme trotzdem bestehen (vgl. Coelli et al. 1998).

<sup>197</sup> Vgl. Coelli (1996). Die Log-Likelihood hierzu wird im Appendix in Battese und Coelli (1992) angegeben.

<sup>198</sup> Vgl. Battese und Coelli (1995).



Nummern von Gemeinschaftspraxen, die aber den Allgemeinmedizinerinnen zugeordnet worden sind.<sup>199</sup>

Das Outputmaß ist die Konsultation. Die Möglichkeit einer solchen Modellierung wurde schon an anderer Stelle diskutiert.<sup>200</sup> Dort wurden ebenfalls die wichtigsten Inputfaktoren erwähnt. So wissen wir, dass die Arbeitszeit des Arztes ein ganz wesentlicher Produktionsfaktor ist und dass die Kapitalkosten für die Ausstattung der Arztpraxis eine bedeutende Rolle spielen. Darüber hinaus müssen mögliche Unterschiede zwischen den Arztpraxen, die mit dem Gesundheitszustand der Patienten verbunden sind, beachtet werden. Um auch die Anzahl der von der Arztpraxis versorgten Patienten in der Produktionsfunktion berücksichtigen zu können, wurden der Output wie auch alle Inputs durch die Anzahl Patienten dividiert, sodass die Einheit der folgenden Analyse die Konsultation pro Patient und Jahr ist.

Die mit dem Einsatz der genannten Inputfaktoren verbundenen Kosten sollen die relevanten Inputgrößen approximieren. Sie setzen sich aus den Kosten für direkte Arztleistungen (*ARZT*), den Kosten für Medikamente (vom Arzt selbst oder vom Apotheker *MEDA* und *MED*), den externen Laborkosten (*LAB*), den externen Physiotherapiekosten (*PH*) und den externen Röntgenkosten (*ROE*) zusammen. Entsprechend der Hypothese des Arztes als vollkommener Sachwalter seiner Patienten spiegeln sie den tatsächlichen Aufwand der Arztpraxis unter Berücksichtigung der Patientenheterogenität wider. Auf die Einbeziehung der Spitalkosten musste verzichtet werden, da sie offenbar nicht für alle Ärzte messbar waren.<sup>201</sup>

Mögliche Effizienzunterschiede sollen durch die  $\alpha$ -Variablen erklärt werden. Neben einem Zeitindex für technische Effizienzveränderungen über die Zeit, gehören verschiedene arzt-spezifische Variablen hierher. Uns interessiert hierbei vor allem, ob es Effizienzunterschiede zwischen den Ärzten in den Hausarztmodellen (*HAM\_CSS*) bzw. HMOs (*HMO*) und den traditionellen Ärzten gibt. Allerdings lässt sich die Hausarztnetz-beteiligung nur auf die CSS-Beteiligung beziehen. Es kann nicht ausgeschlossen werden, dass Ärzte, die hier nicht als Mitglieder eines CSS-Hausarzt-netzes auftreten, bei anderen Krankenversicherungen an Hausarzt-netzen beteiligt sind. Immerhin lässt sich der CSS-Hausarzt-netzstatus eines Arztes mit demjenigen eines Mitgliedes des Zürcher Hausarztvereins vergleichen (*HAUS\_ZH*). Die

---

<sup>199</sup> Darüber hinaus beachte man die weiteren allgemeinen Ausschlusskriterien für das Untersuchungskollektiv in Kapitel 7.

<sup>200</sup> Vgl. die entsprechenden Ausführungen in Kapitel 6 und am Anfang dieses Abschnitts.

<sup>201</sup> Vgl. die Ausführungen zur Beschreibung der hier verwendeten Variablen im Abschnitt 7.5.

Variable identifiziert Ärzte, die in diesem nur in der Stadt Zürich ansässigen und auf Initiative der Ärzte entstandenen Vereins tätig sind. Eine weitere Kontrolle des CSS-Effektes wird durch die Einbeziehung der CSS-Beteiligung aller Ärzte erreicht (CSS).

Über die arzt-spezifischen Einflussgrößen hinaus, von denen vermutet wird, dass sie die Effizienzunterschiede erklären können, wollen wir auch untersuchen, ob die Marktmacht einen Einfluss auf die Unterschiede ausübt. Durch die Einbeziehung dreier Dummy-Variablen kann die Wettbewerbssituation der Ärzte aus den Ergebnissen des vorangegangenen Abschnitts abgebildet werden (*WB* und *WB1* für die Gemeinden mit einem Wettbewerbsmarkt und *WB\_M* für Gemeinden mit monopolistischer Konkurrenz, sodass die Referenzkategorie durch die Gemeinden mit 13 - 18 Ärzten gebildet wird).

In Tabelle 13 sind alle deskriptiven Statistiken der in der Untersuchung verwendeten Variablen für das Jahr 1999 zusammengetragen.

Tabelle 13 Effizienz-Schätzung: Deskriptive Statistiken

N = 1531	Mittelwert	Minimum	Maximum		Mittelwert	Minimum	Maximum
KONS	4.56	0.06	35.74	ZH	0.70	0	1
ARZT	497.88	62.24	5390.96	HA	3614.52	129	8780
MEDA	101.99	0	1483.40	MANN_ANT	0.48	0.45	0.52
MED	266.97	0	5778.09	BEV_D	38.56	0.66	108.55
LAB	46.81	0	1465.33	GEST_ANT	0.01	0.001	0.015
PH	47.21	0	425.93	GEB_ANT	0.01	0.006	0.021
ROE	7.92	0	285.61	GEM_ANZ	0.53	0	6
EINTRITT	1985.68	1977	1997	ST_P_K	1154	364	6981
ROENTGEN	0.65	0	1	ARZT_D	14.98	2.44	29.64
MANN	0.76	0	1	WB_M	0.08	0	1
TIT_ARZT	0.05	0	1	WB	0.18	0	1
TIT_DOK	0.93	0	1	WB_1	0.64	0	1
PA100	0.60	0	1	HAM_CSS	0.09	0	1
PA105	0.30	0	1	CSS	0.99	0	1
				HMO1	0.01	0	1
				HAUS_ZH	0.12	0	1

Der durchschnittliche Arzt unseres Samples hat demnach ca. 4.56 Konsultationen pro Patient im Jahr 1999 durchgeführt, dabei einen Umsatz (direkte Arztleistungen und Medikamente vom Arzt) von rund 600 SFr. erwirtschaftet und gleichzeitig Leistungen in Höhe von ca. 370 SFr. bei anderen Leistungserbringern verursacht.

Entsprechend der Produktionsfunktion (9.13) werden nun die Konsultationen pro Patient und die Inputfaktoren pro Patient logarithmiert und durch *Lnvariable* bezeichnet. Da einige Ärzte

nicht alle Inputfaktoren gleichzeitig einsetzen, wurde vor der Logarithmierung sämtlichen Inputs eine Einheit hinzugefügt, sodass der kleinste Wert der logarithmierten Inputs nun 0 beträgt. Schließlich wurden noch die gemischten wie die Quadratterme gebildet. Die Untersuchung der so gebildeten Inputmatrix ergab, dass sie eine hohe Multikollinearität aufweist. Daher wurden die Inputfaktoren physiotherapeutische Leistungen (*PH*) und veranlasste Röntgenleistungen (*ROE*) zu einer neuen Variablen *Out* zusammengefasst.<sup>202</sup>

#### 9.2.3 Schätzung und Diskussion

Die erste Hälfte von Tabelle 14 enthält die Resultate für die Schätzung der Produktionsfunktion.<sup>203</sup> Es zeigt sich zunächst, dass die Spezifikation einer allgemeinen Cobb-Douglas-Produktionsfunktion abgelehnt werden muss. Die Nullhypothese für einen derartigen Funktionstyp (sämtliche Koeffizienten der gemischten und Quadrat-Terme gemeinsam nicht von Null verschieden) wird auf dem 1%-Signifikanzniveau abgelehnt.<sup>204</sup> Vielmehr sind fast alle Quadratterme signifikant größer Null und 7 von 15 Koeffizienten der gemischten Terme signifikant kleiner Null. Letzteres impliziert eine substitutive Beziehung dieser Inputfaktoren. Besonders hervorgehoben werden kann hier die substitutive Beziehung der beiden Inputfaktoren Arztleistung und Verschreibung von Medikamenten (*LALMA* und *LALM*)<sup>205</sup>, die zeigen, dass ein Arzt, der bspw. die Verschreibung von Medikamenten steigert, gleichzeitig weniger Arztleistungen anbietet. Von den 8 nicht signifikanten gemischten Termen der Inputfaktoren ist vor allem das Produkt aus den Inputfaktoren Medikamente vom Arzt und Medikamente aus der Apotheke interessant. Dieses Ergebnis widerspiegelt die Tatsache, dass die Ärzte nicht in der Lage sind, den einen Faktor durch den anderen zu ersetzen. D. h., wenn Ärzte die Möglichkeit besitzen, Medikamente selbst zu verkaufen (Selbstdispensation der Ärzte), so werden sie diese Möglichkeit auch uneingeschränkt wahrnehmen.

---

<sup>202</sup> Für die gemischten Terme der Inputfaktoren wurden der Einfachheit halber Abkürzungen verwendet, die sich im Zusammenspiel mit den Originalvariablen erklären. Eine Zusammenstellung aller Variablenklärungen findet sich außerdem im Anhang (vgl. Anhang I).

<sup>203</sup> Die Schätzungen wurden mit dem Computerprogramm FRONTIER 4.1 von Tim Coelli durchgeführt (vgl. Coelli 1996).

<sup>204</sup> Für diesen und die folgenden Tests wurde ein verallgemeinerter LR-Test durchgeführt (vgl. Coelli et al. 1998):  
$$LR = -2 \left[ \ln(L(H_0)) - \ln(L(H_1)) \right]$$
 mit  $L(H_0)$  und  $L(H_1)$  den Werten der Likelihoodfunktion unter der Nullhypothese bzw. unter der Alternativhypothese. Dieser Test ist unter  $H_0$  asymptotisch  $\chi^2$ -verteilt.

<sup>205</sup> *LALMA* bezeichnet den gemischten Term aus *LARZT* und *LMARZT*; *LALM* ist demzufolge das Produkt aus *LARZT* und *LMED*.

Interessant ist ferner, dass die Summe der Koeffizienten der direkten Effekte kleiner als eins ist (die Hypothese, dass  $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 = 1$ , kann auf dem 1%-Signifikanzniveau abgelehnt werden).<sup>206</sup> Damit erfolgt die Produktion der Konsultation pro Patient mit sinkenden Skalenerträgen, d.h., dass zur Steigerung der Produktion um eine Einheit (eine Konsultation mehr pro Patient) alle Inputfaktoren um mehr als eine Einheit erhöht werden müssen. Interessant ist in diesem Zusammenhang, dass eine höhere Verordnung von Laboruntersuchungen mit geringeren Konsultationen verbunden ist. Dieses Ergebnis kann dadurch erklärt werden, dass viele Ärzte eigene Labore betreiben und damit mehr Konsultationen anbieten als Ärzte, die über keine eigenen Laboreinrichtungen verfügen und die Patienten für Laboruntersuchungen überweisen.

---

<sup>206</sup> Wobei in diesem Test die technologische Veränderung  $\alpha_6$  nicht berücksichtigt worden ist.

### III Empirische Untersuchung

Tabelle 14 Effizienz-Schätzung: Resultate

Variable	Koeffizient	Standardfehler	Variable	Koeffizient	Standardfehler
Konstante	5.746**	0.992	LALAB	7.68E-03	5.28E-03
LARZT	0.274**	0.088	LALMA	-0.026**	0.003
LMARZT	0.161**	0.021	LALM	-0.032**	0.004
LMED	0.180**	0.027	LALOUT	2.44E-03	4.60E-03
LLAB	-0.085**	0.033	LAT	-2.70E-03	6.67E-03
LOUT	0.074**	0.030	LABLMA	-4.22E-03**	1.87E-03
T	0.082	0.046	LABLM	-4.21E-03	2.50E-03
LARZT2	0.034**	0.007	LABLOUT	5.89E-03**	2.73E-03
LMARZT2	6.79E-03**	1.22E-03	LABT	-1.00E-03	3.64E-03
LMED2	0.016**	0.002	LMALM	3.38E-03*	1.72E-03
LLAB2	5.92E-03**	2.05E-03	LMALOUT	-5.59E-03**	1.73E-03
LOUT2	-3.65E-03*	1.91E-03	LMAT	-5.40E-03**	2.02E-03
T2	-0.015**	0.006	LMLOUT	-0.017**	0.002
			LMT	-2.10E-03	2.70E-03
			LOUTT	3.25E-03	3.28E-03
EINTRITT	-0.0036**	0.0005	ZH	0.219**	0.014
ROENTGEN	-0.0220**	0.0067	MANN_ANT	-0.468	0.461
MANN	-0.0382**	0.0068	BEV_D	2.22E-04	1.73E-04
TIT_ARZT	-0.0655**	0.0259	GEST_AN	-3.18E-03	1.67E-03
TIT_DOK	-0.0090	0.0233	ST_P_K	-1.69E-05**	4.19E-06
GEM_ANZ	-0.0069**	0.0025	ARZT_D	-4.97E-05	8.14E-04
PA100	-0.0592**	0.0112			
PA105	-0.1208**	0.0124			
KONSTANT	-1.936**	0.203			
T1	0.018	0.014			
WBM	-0.486**	0.075	HERF	-0.41**	0.0008
WB	0.184**	0.045			
WB1	0.927**	0.085			
HAM_CSS	-1.391**	0.134			
CSS	-0.250**	0.095			
HMO	-0.439**	0.165			
HAUS_ZH	-0.652**	0.066			
Sigma <sup>2</sup>	0.266**	0.020	LR-Test:	517	
GAMMA	0.935**	0.007			

\*\*p< 0.01, \*)p<0.05

Von besonderer Bedeutung sind die Ergebnisse zur Erklärung der Effizienzunterschiede. Zunächst kann überprüft werden, ob Ineffizienz überhaupt auftritt. Ein LR-Test lehnt die Nullhypothese, nach der alle Koeffizienten der Ineffizienz-Gleichung gemeinsam gleich Null sind, ab. Darüber hinaus bestätigt der LR-Test für  $H_0: \gamma = 0$  die gewählte Spezifikation einer

stochastischen Ineffizienzkomponente ( $H_0$  wird abgelehnt). Vielmehr geht ein Großteil der Residualvarianz auf das Konto des Ineffizienzeffektes ( $\gamma$  kann maximal 1 werden im Falle, dass  $\sigma_v^2 = 0$  ist).

Außerdem sind offenbar die gewählten Faktoren eng mit der Effizienz der Ärzte verbunden, da alle Faktoren gemeinsam einen signifikanten Beitrag bei der Erklärung von Effizienzunterschieden der Ärzte liefern können. Interessant sind hierbei vor allem die Wirkungen des Wettbewerbs und von Managed Care. Nach unserer Einteilung der Gemeinden ist die Ineffizienz bei der Bereitstellung von Konsultationen in Gemeinden mit weniger als 4 Ärzten im Verhältnis zu Gemeinden mit mehr als 13 Ärzten und weniger als 18 Ärzte am geringsten. In den Gemeinden, in denen wir aufgrund des räumlichen Wettbewerbs auf einen Konkurrenzmarkt geschlossen hatten, sind die Ineffizienzen dagegen etwas größer. Diejenigen Ärzte in Gemeinden mit sehr vielen Ärzten zeichnen sich schließlich durch eine sehr hohe Ineffizienz aus. Dieses Ergebnis wird auch durch eine alternative Schätzung bestätigt, in der für die drei Wettbewerbs-Dummies der Herfindahl-Index zur Messung der Wettbewerbsintensität verwendet wurde. Der Herfindahl-Index gibt die Konzentration bezogen auf die zu versorgenden Patienten an. Er nimmt den Wert 1 beim Monopol an und geht gegen Null, wenn vollständige Konkurrenz herrscht.<sup>207</sup> Insgesamt deuten die Ergebnisse der Wettbewerbswirkungen auf die Existenz angebotsinduzierter Nachfrage hin. Denn in einem Marktmodell ohne Angebotsinduzierung würde man genau das Gegenteil erwarten: Mit steigendem Wettbewerb wäre auch eine steigende Effizienz der Anbieter verbunden.

Am interessantesten sind die Ergebnisse aber hinsichtlich der Wirkungen von Managed Care auf die Effizienz. Es ist klar erkennbar, dass HAM Ineffizienzen deutlich verringern. Allerdings können in diesem Modell nur HAM betrachtet werden, die durch einen Vertrag mit der Krankenkasse CSS verbunden sind. Daher erfolgt der Vergleich auch zu Ärzten, die eventuell ebenfalls einem HAM angeschlossen sind, deren Mitgliedschaft hier aber nicht beobachtet werden kann. Zumindest kann man aber erkennen, dass die Ärzte in einem CSS-HAM deutlich geringere Ineffizienzen aufweisen, als andere Ärzte, die nur im konventionellen Bereich Leistungen bei der CSS abrechnen. Für die HMOs kann ebenfalls ein negativer Effekt nachgewiesen werden. M.a.W. können die sechs in unserem Sample betrachteten HMOs gleichfalls einen Beitrag zur Verringerung von Ineffizienzen leisten. Dieser Effekt fällt aber wesentlich geringer

---

<sup>207</sup> Vgl. bspw. Tirole (1995).

aus, als für die betrachteten Hausarztnetze. Die Verlässlichkeit dieser Ergebnisse ist beschränkt, da die Rechnungslegung der HMOs in vielen Fällen nur ungenügend erfolgte. Immerhin kann die Effizienzmessung der einzigen HMO, die CSS-Versicherte behandelt hat, durch die Einbeziehung der anderen HMOs besser eingeordnet werden, was für die Untersuchung des Managed-Care-Effekts auf Mikro-Ebene noch von Bedeutung sein wird. Die Mitgliedschaft im Zürcher Hausarztnetz steht in einem signifikant negativen Zusammenhang zur Ineffizienz. Damit arbeiten auch Ärzte in derartigen Hausarztnetzen kostengünstiger als Ärzte im rein konventionellen Bereich. Interessant ist das Ergebnis auch deshalb, weil dieses Hausarztnetz in Zürich-Stadt angesiedelt ist, wo – nach unseren Ergebnissen – aufgrund des scharfen Wettbewerbs zwischen den Ärzten eigentlich am (kosten-) ungünstigsten produziert wird.

#### **9.2.4 Zusammenfassung und Kritik**

Zusammenfassend lässt sich festhalten, dass das Angebot von Ärzten (Konsultationen pro Patient) sehr gut durch eine Trans-Log-Produktionsfunktion spezifiziert werden kann. Dabei stellte sich heraus, dass die Konsultationen pro Patient zum Teil ineffizient produziert werden. Ineffizienz wurde durch die technische Effizienz der Anbieter gemessen, die angibt, wie die verwendeten Inputfaktoren zum geschätzten Outputniveau im Verhältnis stehen. Die eingesetzten Ressourcen zur Produktion einer Konsultation stellen die Inputfaktoren dieser Untersuchung dar. Die Ineffizienz der Anbieter konnte hauptsächlich durch zwei Faktoren erklärt werden. Zum einen führt ein hoher Wettbewerbsdruck zu einem ineffizienten Angebot, was als ein Hinweis auf angebotsinduzierte Nachfrage aufgefasst werden kann. Allerdings muss man die besondere Datenstruktur berücksichtigen und bei der Interpretation der Effekte vorsichtig sein. Zieht man die Unterschätzung des Standardfehlers des interessierenden Effekts mit in Betracht, dürfte der Wettbewerbseffekt trotzdem noch eindeutig negativ sein. Gemessen an der Variablen *HERF* würde selbst eine Vergrößerung des Standardfehlers um den Faktor 250 nichts an der Signifikanz des Koeffizienten ändern.

Zum anderen konnte ein stark negativer Effekt zwischen der Beteiligung an einem HAM oder einer HMO und der Ineffizienz festgestellt werden. Dieses Ergebnis wird bei der Schätzung des Gesamtmodells noch von Bedeutung sein. Aufgrund der hohen Korrelation zwischen MC und der Ineffizienz kann die Effizienzvariable als Instrument bzw. als Kontrollvariable für mögliche Selektionserscheinungen auf der Arztseite dienen. Denn es wurde zwar mit dem hier geschätzten Modell eine Annahme über die Kausalität zwischen Hausarztnetzteilnahme und Effizienz

getroffen – die Ärzte sind effizient, weil sie sich an HAM und HMOs beteiligen –, trotzdem kann die Wirkungsrichtung auch in der entgegengesetzten Richtung verlaufen. Die Kausalitätsproblematik entspricht dabei genau der Frage nach der Selbstselektion der Ärzte, weswegen die Verwendung der Effizienz als Instrument in der Gesamtschätzung effektiv für eine mögliche Selektion kontrollieren kann.

Es sind vor allem zwei Punkte zum vorgestellten Effizienz-Modell kritisch anzumerken, die bei der Beurteilung der Ergebnisse berücksichtigt werden sollten. Erstens verwendet die Untersuchung die Konsultation pro Patient als Outputmaß einer Arztpraxis. Die Produktion einer Konsultation wird hauptsächlich als Funktion der eingesetzten Ressourcen begriffen. Damit unterstellt das Modell gleichzeitig eine hohe (technische) Effizienz derjenigen Ärzte oder Arztpraxen, die viele Konsultationen bei einem gleichzeitig geringen Ressourcenverbrauch anbieten. Daher sind weitergehende Rückschlüsse, die von einer weiter gefassten Definition dieses Outputmaßes ausgehen (Anzahl an Konsultation als Proxie für die Verbesserung der Gesundheit der Patienten) nur sehr bedingt möglich. Der zweite kritische Punkt ist eng mit dem ersten verbunden. Denn geht man gerade davon aus, dass der Gesundheitszustand der Patienten der Fokus jeder nachhaltigen Gesundheitspolitik sein sollte (egal ob man nun den „Staat“ oder die Krankenkassen als Agenten der Patienten betrachtet), so fehlen dem vorgestellten Modell wichtige Indikatoren, die Art und Schwere der Erkrankungen abbilden können. Die Tatsache, dass der Input durch den Aufwand entsprechend der durchschnittlichen Krankheitsschwere der Patienten (d.h. den Kosten der Behandlung) approximiert wurde, kann diese Kritik nur teilweise abschwächen. Insgesamt lassen sich die Ergebnisse nur unter der Annahme interpretieren, dass sich die Patienten der hier betrachteten Ärzte im Durchschnitt nicht durch ihren Gesundheitszustand unterscheiden. Obwohl eine solche Annahme für die meisten der hier betrachteten Ärzte mit der relativen Homogenität der Ärzte (nur Grundversorger) und der großen Patientenzahl eines jeden Arztes gerechtfertigt werden kann, wird sie dann problematisch, wenn man von einer systematischen Selektion von Patienten mit einer guten Gesundheit zu bestimmten Ärzten ausgehen muss. Ein solcher Fall könnte aber gerade für die Patienten von HMOs und HAM zutreffen. Im nächsten Abschnitt wird dieser Punkt eingehender untersucht.



#### 9.3 Nachfrageschätzung

Aufgrund der Tatsache, dass am Anfang eines jeden Arztbesuches die Entscheidung eines potenziellen Patienten steht, ob ein Arztbesuch notwendig ist oder nicht, spielt die Nachfrage nach Arztdienstleistungen für das Angebot eine entscheidende Rolle.

In diesem Abschnitt soll daher ein Modell zur Schätzung der nachfrageseitigen Erklärung der Inanspruchnahme und der Kosten vorgestellt werden. Insbesondere spielt hierbei auch die Frage eine wichtige Rolle, inwieweit die beobachteten geringeren Kosten in HMOs und Hausarztmodellen (vgl. 7.4.4) auf die Selbstselektion der Versicherten zurückzuführen sind. Denn es wird vermutet, dass es gerade diejenigen Versicherten mit einem guten Gesundheitszustand sind, die sich für ein MC-Modell entscheiden. Andererseits kann aber auch die Wahl eines besonderen Versicherungsvertrags durch den Versicherten zu einer Leistungsbegrenzung geführt haben. Letzterer Effekt lässt sich als Managed-Care-Effekt auffassen.

Man muss die möglichen Wirkungen von Managed Care von den Wirkungen nachfrageseitiger Kostenbeteiligungen unterscheiden. Während nachfrageseitige Kostenbeteiligungen, wie Selbstbehalte und prozentuale Kostenbeteiligungen, direkt eine Verhaltensänderung bei den Versicherten bewirken, wirken Managed-Care-Methoden sowohl verhaltensändernd auf der Nachfrage- wie auch auf der Anbieterseite. So wird einerseits die freie Arztwahl der Konsumenten eingeschränkt und andererseits führt die Kostenbeteiligung der Ärzte zu einer höheren finanziellen Verantwortung der Mediziner für die Leistungsanspruchnahme der Patienten.<sup>208</sup>

Die Wirkungen von Selbsthalten werden in der folgenden Untersuchung nur am Rande berücksichtigt. Es ist aber klar, dass sie – aufgrund der Wahlmöglichkeit der Versicherten – ebenfalls von Selbstselektionseffekten mitbestimmt werden. Für die Schweiz gibt es verschiedene neuere Studien, welche die Wirkung von freiwilligen Selbsthalten kontrovers diskutieren.<sup>209</sup> Während Nicolet et al. (2001) und Schellhorn (2001) hauptsächlich Selbstselektion für die geringere Leistungsanspruchnahme auf höheren Selbsthaltsstufen verantwortlich machen, zeigen Werblow und Felder (2003), dass Selbstbehalte trotz vorhandener Selbstselektion der Versicherten auch verhaltensändernd wirken.

---

<sup>208</sup> Gleichzeitig muss aber betont werden, dass damit nur die direkten Effekte angesprochen sind. Unberührt hiervon sind mögliche indirekte Effekte, die bspw. bestimmte Selbstbeteiligungsregelungen auf der Nachfrageseite auf die Leistungserbringer ausüben können.

<sup>209</sup> Für einen Überblick zu den Selbsthaltsregelungen vgl. die Ausführungen zum Schweizer Krankenversicherungsgesetz im Anhang I.

Für die Untersuchung des ärztlichen Verhaltens ist die Frage nach der Selbstselektion der Versicherten deshalb von besonderem Interesse, weil damit ein zusätzlicher Indikator bereitsteht, der den Gesundheitszustand eines Patienten approximieren kann: Gibt es eine spürbare Selektion der guten Risiken, d.h. wählen sie verstärkt MC-Verträge, so zeigt sich darin ein besserer Gesundheitszustand dieser Versicherten. Der Selektionseffekt muss daher vom Managed-Care-Effekt getrennt werden. Wenn die Separierung beider Effekte gelingt, kann auch die Wirkung einer MC-Beteiligung auf das Verhalten der Ärzte besser identifiziert werden.

Aus dem Gesagten wird klar, dass man es hier mit einem Identifikationsproblem zu tun hat, wie es im Abschnitt (8.2.3) beschrieben wurde. Um die durchschnittlichen Effekte im konkreten Kontext identifizieren zu können, wird die Versicherungswahl der Versicherten in einem ökonometrischen Modell explizit berücksichtigt. Die Nachfrage nach Gesundheitsleistungen wird dann in einem zweiten Schritt unter Berücksichtigung der Wahlentscheidung der Versicherten geschätzt. Es ist bei der Nachfrageschätzung außerdem zu berücksichtigen, dass nicht alle Versicherten in einem Jahr Gesundheitsleistungen nachfragen. Daher wird auch für diese Selektionsverzerrung kontrolliert.

### 9.3.1 Modell

Sei  $u_{1i}^*$  der potenzielle Nutzen des Versicherten  $i$ , falls er eine MC-Versicherung wählt;  $u_{0i}^*$  soll dagegen den potenziellen Nutzen des gleichen Versicherten bezeichnen, falls er keinen MC-Vertrag besitzt. Die potenziellen Nutzenabwägungen sind wieder nicht beobachtbar. Beobachten können wir nur die tatsächliche Vertragswahl  $D_i = 1$ , falls der Versicherte  $i$  eine MC-Versicherung gewählt hat und  $D_i = 0$  sonst. Geht man davon aus, dass der Entscheidungsprozess von beobachtbaren Faktoren  $Z_i$  wie auch unbeobachtbaren Faktoren  $\varepsilon_i$  abhängt, so lässt sich die Wahl eines MC-Versicherungsvertrages für einen Versicherten  $i$  wie folgt formulieren:

$$\begin{aligned} u_i^* &= u_{1i}^* - u_{0i}^* \\ u_i^* &= \alpha' Z_i + \varepsilon_i \\ D_i &= 1 \text{ falls } u_i^* > 0, \text{ ansonsten } 0. \end{aligned} \tag{9.15}$$

D.h. der Versicherte vergleicht die möglichen Nutzenniveaus mit und ohne MC-Vertrag. Der nichtbeobachtbare Vergleich kann aber durch beobachtbare Faktoren erklärt werden.<sup>210</sup> Die

---

<sup>210</sup> Dabei bezeichnet  $\alpha$  den zu schätzenden Koeffizientenvektor.

tatsächliche Wahl wird schließlich durch  $D$  angezeigt. Daher kann das Modell in (9.15) durch ein Probit- oder im einfachsten Fall durch ein lineares Wahrscheinlichkeitsmodell geschätzt werden.

Die Nachfrage nach Gesundheitsleistungen  $y_i$  lässt sich bspw. durch einen linearen Zusammenhang ausdrücken. Dabei wird davon ausgegangen, dass die Leistungsanspruchnahme eines Versicherten sowohl von verschiedenen exogenen Faktoren ( $X$ ) als auch der Vertragswahl ( $D$ ) abhängt. Unbeobachtbare Faktoren sind wieder in einem Störterm zusammengefasst ( $\varpi_i$ ).

$$y_i = \beta' X_i + \delta D_i + \varpi_i. \quad (9.16)$$

Aufgrund von (9.15) ist  $D$  in (9.16) endogen. Damit sind die Störterme  $\varepsilon_i$  und  $\varpi_i$  miteinander korreliert. Unter Berücksichtigung dieser Korrelation kann der Erwartungswert der Leistungen für  $D = t$ ,  $t = 0, 1$  folgendermaßen geschrieben werden:

$$E[y_i | D_i = t] = \beta' X_i + \delta D_i + E[\varpi_i | D_i = t]. \quad (9.17)$$

Gleichung (9.17) macht den Unterschied zwischen exogenem und endogenem Einfluss der Versicherungswahl des Versicherten auf die Leistungsanspruchnahme deutlich: Könnte die Wahl als exogen betrachtet werden, wäre der Erwartungswert von  $\varpi$  unabhängig von der Vertragswahl und damit entsprechend der Standardannahmen Null. Bei endogener Wahl hingegen ist dieser Erwartungswert ungleich Null. Um ihn identifizieren zu können, müssen Annahmen bezüglich der gemeinsamen Verteilung der Störterme getroffen werden. Wir gehen im Folgenden von einer bivariaten Normalverteilung für die Störterme aus. Mit den Annahmen bezüglich der Störterme lässt sich der MC-Effekt nun identifizieren. Denn es gilt aufgrund der Korrelation zwischen der Vertragswahl und der Leistungsanspruchnahme für die Erwartungswerte von  $\varpi$  in den beiden Regimes ( $D = 0, 1$ ):

$$\begin{aligned} E[\varpi_i | D_i = 1] &= \rho \sigma_\omega \left( \frac{\phi(\alpha' Z_i)}{\Phi(\alpha' Z_i)} \right) \\ E[\varpi_i | D_i = 0] &= \rho \sigma_\omega \left( \frac{-\phi(\alpha' Z_i)}{1 - \Phi(\alpha' Z_i)} \right). \end{aligned} \quad (9.18)$$

Dabei bezeichnet  $\rho$  die Korrelation zwischen  $\varpi$  und  $\varepsilon$ ,  $\sigma_\omega$  die Varianz von  $\varpi$  sowie  $\phi$  und  $\Phi$  die Dichte bzw. die Verteilungsfunktion der Standardnormalverteilung.<sup>211</sup>

---

<sup>211</sup> Vgl. Greene (1997).

Setzt man nun (9.18) in (9.17) jeweils für  $D = 0, 1$  ein und bildet die Differenz, so erhält man:

$$E[y_i | D_i = 1] - E[y_i | D_i = 0] = \delta + \theta \left( \frac{\phi_i}{\Phi_i (1 - \Phi_i)} \right). \quad (9.19)$$

Damit entspricht  $\delta$  gerade dem durchschnittlichen Behandlungseffekt derjenigen Versicherten, die einen Managed-Care-Vertrag gewählt haben (treatment on the treated). Wir erinnern uns, dass dieser Effekt in Abschnitt (8.2.3) mit  $\alpha_{MC}$  bezeichnet wurde. Ein Vergleich mit Gleichung (8.6) zeigt, dass offenbar gelten muss:

$$E[Y_{0i} | D_i = 0] - E[Y_{0i} | D_i = 1] = \theta \left( \frac{\phi_i}{\Phi_i (1 - \Phi_i)} \right). \quad (9.20)$$

Damit können wir – unter den gestellten Annahmen – mit obiger Schätzstrategie den unbeobachtbaren Effekt identifizieren.

Schreiben wir für die beiden Erwartungswerte in (9.18):

$$E[\varpi_i | D_i] = \theta_1 \lambda_1 (Z_i | D_i)$$

so lässt sich unsere Schätzgleichung in folgender Weise schreiben:

$$E(y_i | D_i) = \beta' X_i + \delta D_i + \theta_1 \lambda_1 (Z_i | D_i). \quad (9.21)$$

Allerdings müssen wir noch beachten, dass die Entscheidung über die Leistungsanspruchnahme selber auch nicht direkt beobachtbar ist. Vielmehr beobachten wir eine Mischung aus zwei Entscheidungen, die nacheinander erfolgen: Zunächst muss der Versicherte entscheiden, ob er einen Arzt aufsuchen soll oder nicht. Falls er zu einem Arzt geht, wird dann die Entscheidung über die Höhe der Gesundheitsleistungen getroffen. Letztere Entscheidung ist dabei nicht allein durch den Patienten bestimmt, sondern auch durch die Entscheidung des Arztes. Die Entscheidung des Versicherten über einen Arztbesuch wird über den unbeobachtbaren Nutzen des Versicherten aus einem Arztbesuch modelliert. Dieser Nutzen lässt sich somit entsprechend der Methodik aus Abschnitt 8.2.1 als latente Variable modellieren.

Bezeichnen wir mit  $a^*$  den Nutzen eines Individuums aus einem Arztbesuch, so beobachten wir nur dann eine positive Leistungsanspruchnahme ( $y$ ), wenn der Nutzen positiv ist. Diese Entscheidung lässt sich mit  $a = 1$  indizieren und kann wiederum mittels eines Probitmodells abgebildet werden. Geht man realistischer Weise davon aus, dass die Entscheidung über einen Arztbesuch und die Leistungsanspruchnahme von den gleichen Faktoren ( $X$ ) sowie ebenfalls

von der Vertragswahl ( $D$ ) beeinflusst wird; und nimmt man vereinfachend an, dass der Nutzen eines Arztbesuches durch die Gleichung  $a_i^* = \gamma' X_i + \kappa D_i + \eta_i$  gegeben sei, lässt sich das Modell hinsichtlich dieses Selektionsmechanismus' in folgender Weise aufschreiben:

$$\begin{aligned} E[y_i | a_i = 1] &= E[y_i, \gamma' X_i + \kappa D_i + \eta > 0] \\ &= \beta' X_i + \delta D_i + E[\varpi_i | \eta_i > -\gamma' X_i - \kappa D_i]. \end{aligned} \quad (9.22)$$

Nimmt man auch hier eine bivariate Normalverteilung für die Störterme an und berücksichtigt entsprechend die Korrelation und die Varianz der Störterme, lässt sich der Ausdruck vereinfachen zu:

$$E[y_i | a_i = 1] = \beta' X_i + \delta D_i + \theta_2 \lambda_2 (\gamma' X_i + \kappa D_i). \quad (9.23)$$

Bringen wir beide Selektionsmechanismen zusammen, so lässt sich der Erwartungswert der Leistungsausgaben auch schreiben zu:

$$E[y_i | D_i = t, a_i = 1] = \beta' X_i + \delta D_i + \theta_1 \lambda_1 (\alpha' Z_i | D_i) + \theta_2 \lambda_2 (\gamma' X_i + \kappa D_i). \quad (9.24)$$

Während der Selektionseffekt hinsichtlich der Wahl eines Managed-Care-Vertrages nicht nur durch die funktionale Form von  $\lambda_1$ , sondern vor allem auch durch die zusätzlichen Variablen in  $Z$ , die nicht in  $X$  enthalten sind, identifiziert werden kann, erfolgt die Identifizierung des Selektionseffektes bezüglich der Inanspruchnahme ausschließlich über die funktionale Form von  $\lambda_2$ , da davon ausgegangen wurde, dass die Entscheidung über die Inanspruchnahme und die Höhe der Inanspruchnahme von den gleichen erklärenden Variablen abhängt.<sup>212</sup>

Gleichung (9.24) kann nun entweder direkt geschätzt werden (sogenannter Kontrollfunktionsansatz) oder mittels Instrumentalvariablenschätzung für die Versicherungswahl. Für den Kontrollfunktionsansatz müssen zunächst  $\lambda_1$  und  $\lambda_2$  geschätzt werden. Allerdings muss beachtet werden, dass die beiden Selektionsmechanismen nicht unabhängig voneinander sind, sodass die Vertragswahl ( $D$ ) in  $\lambda_2$  wieder endogen ist. Wir verwenden im Folgenden deshalb einen gemischten Ansatz, indem wir zunächst die Wahl der Versicherten explizit modellieren (Schätzung von (9.15)). Die erhaltene Schätzung für die Vertragswahl verwenden wir dann als Instrument im Selektionsmodell der Nachfrage (9.23).

---

<sup>212</sup> Vgl. für einen Überblick zu den Selektionsmodellen und den unterschiedlichen Modellierungsmöglichkeiten Maddala (1985).

Insgesamt erhalten wir somit ein 3-stufiges Nachfragemodell in Abhängigkeit der Versicherungswahl (vgl. Tabelle 15). Auf der ersten Stufe wählt der Versicherte einen Vertrag in Abhängigkeit von verschiedenen erklärenden Faktoren. Die Wahrscheinlichkeit, dass ein Versicherter einen MC-Vertrag wählt, kann dabei durch ein einfaches Probit-Modell geschätzt werden. Auf der zweiten Stufe trifft er dann die Entscheidung über den Arztbesuch. Schätzt man die Wahrscheinlichkeit einer positiven Leistungsanspruchnahme bei exogener Vertragswahl, so kann die Schätzung mit einem Probit-Modell geschehen. Allerdings ist die Wahl des Versicherungsvertrages endogen, sodass sie bei der Schätzung auf der zweiten Stufe in geeigneter Weise berücksichtigt werden muss.<sup>213</sup> Auf der dritten Stufe schließlich treffen Patient und Arzt gemeinsam die Entscheidung über die Höhe der notwendigen Leistungsanspruchnahme. Auch hier muss unterschieden werden, ob die Vertragswahl bei der Schätzung der Leistungsanspruchnahme als exogen gegeben angesehen werden kann (OLS-Schätzung) oder ob die Vertragswahl wieder als endogene Variable eingeht.

Tabelle 15 3-stufiges Nachfragemodell

Stufe	Bezeichnung	Ökonometrisches Modell
1. Stufe	Wahl des Vertrages	Probit
2. Stufe	Entscheidung über den Arztbesuch	Binomiales Probit
		Binomiales Probit mit Instrumentalvariablen (GMM)
3. Stufe	Nachfrage nach medizinischen Leistungen	OLS <sup>a)</sup> Schätzung unter Berücksichtigung der 2. Stufe
		2SLS <sup>b)</sup> Schätzung unter Berücksichtigung der 2. Stufe und der 1. Stufe

<sup>a)</sup> OLS: Ordinary Least Square (Kleinst-Quadrateschätzer).

<sup>b)</sup> 2SLS: Two-Stage Least Square (zweistufiger Kleinst-Quadrateschätzer).

Während im beschriebenen Nachfragemodell keine Trennung zwischen der ärztlichen Komponente und der Patientenentscheidung auf der dritten Stufe stattfinden kann, wird eine derartige Unterscheidung einer der Hauptuntersuchungspunkte im Gesamtmodell sein (vgl. Abschnitt 9.4). Wesentlich für die Schätzung der Nachfrage im vorliegenden Modell ist die Berücksichtigung der endogenen Vertragswahl der Versicherten (Managed Care oder konventionelle Versicherung) bei der Nachfrageentscheidung. Unberücksichtigt bleibt – wie erwähnt – die möglicherweise ebenfalls endogene Wahl der Franchise. Dieser Sachverhalt sollte bei der Schätzung und der Interpretation der Ergebnisse berücksichtigt werden.

---

<sup>213</sup> Eine nähere Erklärung der verwendeten Schätzmethode erfolgt bei der Vorstellung der Schätzung.

#### 9.3.2 Daten

Die Daten der Untersuchung wurden zu einem großen Teil bereits im Abschnitt 7.4 vorgestellt. Außer den dort vorgestellten „Original-Variablen“ verwenden wir noch abgeleitete Variablen sowie zusätzliche, die vor allem die Versicherungswahl der Versicherten abbilden sollen. Die zusätzlichen Variablen zeigen, ob ein Versicherter bei der betrachteten Versicherung außer der obligatorischen Grundversicherung auch noch weitere (private) zusätzliche Krankenversicherungen abgeschlossen hat. Für Zusatzversicherungen werden risikoäquivalente Prämien verlangt. Die Entscheidung für eine oder mehrere Zusatzversicherungen gibt Aufschluss über Risikoneigung, Präferenzen für bestimmte Behandlungsmethoden, den (erwarteten) Gesundheitszustand, aber auch über die Einkommenssituation des Versicherten. Sie umfassen bei der hier betrachteten Krankenkasse Versicherungen für Alternativmedizin (*ALT*), für Notfälle (*NOTF*), für die Pflege in Kureinrichtungen (*KURPFL*), für chronische Erkrankungen (*CHRO*), für zahnmedizinische Leistungen (*ZAHNPFL*) sowie mehrere Versicherungen für die stationäre Behandlung. Letztere decken nicht nur die Unterbringung in halbprivaten oder privaten Abteilungen von öffentlichen oder privaten Spitälern in der ganzen Schweiz ab, sondern garantieren auch die freie Wahl des behandelnden Arztes (*SPITAL\_H*, *SPITAL\_P*). Hinzu kommt eine Versicherung, welche die Hotelkosten (Zuschläge für Ein- oder Zweibettzimmer z.B.) eines Aufenthaltes in der allgemeinen Abteilung eines Spitals übernimmt (*SPITAL\_S*). Schließlich lassen sich verschiedene Zusatzversicherungen, bzw. Teile davon, mit der obligatorischen Krankenversicherung zu einer Standardversicherung kombinieren (*STAND*). Für diese Standardversicherung optierten 89.3% Versicherte der vorliegenden Stichprobe.

Das Untersuchungskollektiv wird dadurch begrenzt, dass nur Versicherte betrachtet werden, die über den gesamten Untersuchungszeitraum bei der Krankenversicherung CSS versichert waren. Da die Leistungen der Vorjahre als Indikatoren für den Gesundheitszustand der Patienten verwendet werden, ist eine solche Eingrenzung notwendig. Insgesamt waren damit 122,392 Versicherte im Untersuchungssample.

Die verwendeten erklärenden Variablen mit ihren deskriptiven Statistiken sind in Tabelle 16 aufgelistet.

Tabelle 16 Nachfrageschätzung: Deskriptive Statistik

N=122392	Beschreibung	Mittelwert	Std.Abw.	Minimum	Maximum
<b>Patienten</b>					
ZH	Kanton: ZH = 1	0.694	0.461	0	1
UNFALL	Unfalleinschluss = 1	0.711	0.454	0	1
MANN	Geschlecht: Mann = 1	0.447	0.497	0	1
ALTER	Alter in Jahren	41.590	22.651	2	103
ALTER2	Alter im Quadrat	2242.81	2013.87	4	10609
KIND	Alter 0 – 18: Kind = 1	0.201	0.401	0	1
JUGEND	Alter 19 – 26: Jugend = 1	0.085	0.278	0	1
SPT1	Mindestens 1 Tag im Spital=1	0.098	0.297	0	1
LPR	Ln(Prämienvverbilligung)	2.340	3.240	0	9.33
LNL97	Ln(Leistungen 1997)	5.498	3.079	0	12.09
LNL98	Ln(Leistungen 1998)	5.626	3.062	0	12.09
<b>Gemeinde</b>					
MANN_ANT	Männeranteil an Bevölkerung	0.479	0.014	0.42	0.52
BEV_D	Bevölkerung pro Fläche (ha)	31.035	32.576	0.398	108.55
GEB_BEV	Anteil Geborener an Bev.	111.018	16.693	47.393	227.58
GEST_BEV	Anteil Gestorbener an Bev	83.895	29.428	0	233.39
AUSL_BEV	Anteil Ausländer an Bev.	2652.29	1067.02	231.788	4558.01
ARZTD	Anzahl Ärzte pro 100,000 Bev.	26.789	17.426	0	56.84
LDK	Ln(Durchschnittskosten)	5.08	0.19	4.59	5.40
<b>Zusatzversicherung</b>					
STAND	Kombination versch. Zusatzvers.	0.893	0.309	0	1
ALT	Alternativversicherung	0.699	0.459	0	1
NOTF	Notfallversicherung	0.698	0.459	0	1
KUR_PFL	Kurpflegeversicherung	0.623	0.485	0	1
CHRO	Vers. für chronische Erkrank.	0.183	0.387	0	1
ZAHNPFL	Zahnpflegeversicherung	0.034	0.181	0	1
SPITAL_H	Spitalvers. (Halbprivat)	0.177	0.381	0	1
SPITAL_P	Spitalvers. (Privat)	0.107	0.309	0	1
SPITAL_S	Spitalvers. (Zusatzkosten)	0.027	0.162	0	1

Da Kinder und Jugendliche Versicherungsverträge erhalten, die sowohl in der Prämienhöhe als auch in Bezug auf die Kostenbeteiligungsregelungen sich von denjenigen Verträgen für Erwachsene unterscheiden, wurden außer dem Alter der Versicherten noch zwei Dummy-Variablen (*KIND* und *JUGEND*) für diese Personengruppen eingeführt.

### 9.3.3 Schätzung

Wie bei der Vorstellung des Modells bereits erwähnt, erfolgt die Schätzung in drei Stufen: Auf der ersten Stufe wird die Wahl des Versicherungsvertrags betrachtet. Das Ergebnis der ersten Schätzung dient auf der zweiten und dritten Stufe als Instrument für die auf diesen Stufen endogene, zu erklärende Variable „Vertragswahl“. Da nicht alle Versicherten die gleichen Wahlmöglichkeiten hinsichtlich einer Managed-Care-Beteiligung besitzen, werden die folgenden



Schätzungen für unterschiedliche Teilpopulationen durchgeführt. An den entsprechenden Stellen wird auf das gewählte Kollektiv jeweils eingegangen. Darüber hinaus ist zu beachten, dass bei der Wahlentscheidung bezüglich einer Managed-Care-Beteiligung auch die Franchise-Wahl von Bedeutung ist. Denn genauso wie es bei der Einschreibung in ein Hausarztmodell oder die Beteiligung in einer HMO einen Prämienrabatt gibt, erhalten die Versicherten bei der Wahl eines höheren Selbstbehalts ebenfalls einen Prämienrabatt. In Verbindung mit der Wahl eines MC-Vertrages ergibt sich dabei folgendes Bild: Schreibt sich ein Versicherter in einer HMO ein, so erhält er einen Prämienrabatt von bis zu 40%. Selbstbehalte sowie prozentuale Kostenbeteiligungen der Versicherten sind in HMOs nicht vorgesehen. Schreibt sich ein Versicherter dagegen in ein Hausarztmodell ein, so kann er mit einer Prämienermäßigung von bis zu 20% rechnen. Darüber hinaus gelten die gleichen Selbstbehaltsregelungen wie im konventionellen Sektor.<sup>214</sup>

Zwar kann die Selbstbehaltswahl hier nicht explizit berücksichtigt werden, immerhin kann man aber die gewählte Franchise als exogene Variable bei der Nachfrage berücksichtigen und dann bei der Interpretation der Ergebnisse der möglichen Endogenität Rechnung tragen.

#### **Wahl des Versicherungsvertrages**

Zunächst soll die Wahl des Versicherungsvertrages durch die Versicherten beleuchtet werden. Wir betrachten dabei alle drei möglichen Untersuchungsgebiete. Die erste Schätzung gibt die Vertragswahl unter Berücksichtigung aller Versicherten wieder, unabhängig davon, ob die Versicherten tatsächlich eine Wahlmöglichkeit hinsichtlich eines Hausarztmodells bzw. einer HMO hatten. Die zweite Kolonne der Schätzung gibt die Ergebnisse für diejenigen Versicherten an, die in einer Gemeinde wohnen, in der sie die Wahl zwischen HAM und konventioneller Versorgung innerhalb der Krankenversicherung CSS hatten. Die dritte Kolonne schließlich zeigt Ergebnisse nur für die Stadt Zürich, d.h. für Versicherte, die tatsächlich einen HMO-Vertrag hätten wählen können. Die abhängigen Variablen in den drei Varianten sind: MC = 1, bei Wahl eines Managed-Care-Vertrags (HAM oder HMO), HAM = 1, bei Wahl eines HAM-Vertrags und HMO = 1, bei Wahl eines HMO-Vertrags.

---

<sup>214</sup> Vgl. zu den Selbstbeteiligungsregelungen die Ausführungen zum Schweizer Krankenversicherungsgesetz im Anhang I.

Tabelle 17 Nachfrageschätzung: Vertragswahl der Versicherten 1999

	MC		HAM		HMO	
N	122392		73452		21806	
	Koeff.	Std.fehl.	Koeffizient	Std.fehler	Koeffizient	Std.fehler
<b>Patienten</b>						
ZÜRICH	-0.100**	0.047	-0.589**	0.048		
KONSTANT	-7.417**	0.564	-0.558	0.716	-10.932**	1.273
UNFALL	-0.125**	0.017	-0.117**	0.020	-0.203**	0.039
SEX	-0.035**	0.014	-0.026	0.016	-0.069**	0.032
ALTER	-0.024**	0.002	-0.033**	0.003	-0.057**	0.006
ALTER2	1.74E-04**	2.32E-05	2.71E-04**	2.66E-05	4.77E-04**	5.33E-05
KIND	-0.256**	0.046	-0.329**	0.053	-0.578**	0.105
JUGEND	-0.204**	0.032	-0.219**	0.037	-0.402**	0.072
SPT	-1.001**	0.052	-0.936**	0.057	-2.246**	0.154
LPR	-0.034**	0.002	-3.60E-03**	2.30E-04	-2.50E-04**	2.36E-05
LNL97	-0.023**	0.003	-0.031**	0.003	-0.061**	0.006
LNL98	-0.001	0.003	-0.001	0.003	0.002	0.006
<b>Gemeinde</b>						
MANN_ANT	10.592**	1.046	4.666**	1.343	12.135**	2.417
BEV_D	-0.007**	0.000	-1.08E-03**	4.73E-04	-0.012**	0.001
GEB_BEV	1.22E-03**	4.25E-04	8.62E-04*	4.51E-04	0.002**	0.001
GEST_BEV	4.07E-03**	3.74E-04	-3.81E-03**	4.49E-04	-0.004**	7.86E-04
AUSL_BEV	-1.59E-04**	1.30E-05	-1.57E-04**	1.47E-05	-4.00E-04**	2.71E-05
ARZTD	0.008**	0.001	3.60E-03**	8.51E-04	0.009**	0.002
LDK	0.014**	0.001	-0.004**	0.001	0.035**	0.001
<b>Zusatzversicherung</b>						
STAND	-0.386**	0.022	-0.056**	0.027	-0.249**	0.053
ALT	-0.452**	0.017	-0.396**	0.019	-0.771**	0.037
NOTF	-0.961**	0.025	-0.989**	0.026	-2.428**	0.054
KUR_PFL	1.011**	0.024	1.154**	0.025	2.733**	0.056
CHRO	-0.074**	0.026	-0.212**	0.032	-0.424**	0.068
ZAHNPFL	0.000	0.039	-0.055	0.045	-0.095	0.090
SPITAL_H	-0.120**	0.021	-0.192**	0.024	-0.349**	0.047
SPITAL_P	-0.417**	0.032	-0.437**	0.038	-0.959**	0.082
(Pseudo-) R <sup>2</sup>	0.37		0.41		0.47	

\*\*) p < 0.01, \*) p < 0.05

Nach Tabelle 17 weisen die meisten Koeffizienten bei einer relativ guten Güte des Modells<sup>215</sup> die erwarteten Vorzeichen auf. Da die Leistungen der beiden Vorjahre Indikatoren des Gesundheitszustandes darstellen, bedeuten der signifikant negative Koeffizient von *LNL97* und der nicht signifikante Koeffizient von *LNL98*, dass die Versicherten bei der Wahlentscheidung bezüglich einer MC-Beteiligung eine eher langfristige Sicht einnehmen. Nicht verwunderlich ist ferner, dass die Arztdichte einen positiven Einfluss auf die Wahrscheinlichkeit der Wahl einer MC-Versicherung ausübt. Denn sowohl die Hausarztmodelle als auch die HMOs sprechen ausschließlich Versicherte an, die in Gemeinden mit einer hohen Arztdichte leben.

<sup>215</sup> Wir verwenden eine Formulierung des Pseudo-R<sup>2</sup> nach McKelvey und Zavoina (1975). Vgl. hierzu auch LIMDEP (1998).

Für die Untersuchung ist die Betrachtung der Zusatzversicherungen von besonderer Wichtigkeit. Denn durch ihre Berücksichtigung soll letztendlich die Identifizierung des MC-Effektes erfolgen: Sie sind die einzigen Variablen, die ausschließlich zur Schätzung der Wahlentscheidung der Versicherten herangezogen werden. Außer bei der Zahnpflegeversicherung sind alle anderen Koeffizienten signifikant von Null verschieden. Darüber hinaus ist zu erkennen, dass alle Zusatzversicherungen – bis auf eine Ausnahme – einen negativen Effekt auf die Wahl eines Managed-Care-Vertrages ausüben. Nur die Kurpflegeversicherung hat einen positiven Einfluss auf die Wahrscheinlichkeit, einen MC-Vertrag zu wählen. Die Hypothese, dass die Zusatzversicherungen gemeinsam keinen Einfluss auf die Vertragswahl ausüben, kann dann auch auf hohem Signifikanzniveau abgelehnt werden ( $p = 0$ ).

Zusammenfassend lässt sich über die Wahl besonderer Versicherungsverträge (MC-Verträge) durch die Versicherten festhalten, dass vor allem jüngere Erwachsene, die über ein relativ gutes Einkommen verfügen (d.h. nicht auf Prämienverbilligungen angewiesen sind) und zudem einen guten langfristigen Gesundheitszustand aufweisen, sich für derartige Verträge entscheiden. Abschließend sei auf den interessanten Umstand hingewiesen, dass aufgrund des positiven Koeffizienten von Alter im Quadrat die Wahrscheinlichkeit einer MC-Beteiligung ab dem Rentenalter wieder ansteigt (ab ca. 65).

#### **Nachfrage - Arztbesuch**

In der Regel wird es der Versicherte sein, der entscheidet, ob er einen Arzt aufsucht. Vor allem wird dies für den Beginn einer Behandlungsepisode gelten. In diesem Sinne dürften all diejenigen Versicherten, für die man positive Leistungen beobachten kann, zumindest einmal im Jahr eine solche Entscheidung positiv getroffen haben. Allerdings ist nicht klar zu bestimmen, welche Versicherten tatsächlich einen Arzt aufgesucht haben und welche nicht, weil man in der Regel nur die Leistungen erfasst, die oberhalb der gewählten Franchise anfallen. Da den Patienten die Leistungen von den Leistungserbringern in Rechnung gestellt werden,<sup>216</sup> haben sie grundsätzlich keinen Anreiz, bei den Kassen Kostenerstattung zu beantragen, bevor die aufgelaufenen Beträge nicht den Franchisenbetrag erreichen. Innerhalb des Franchisenbetrages gibt es daher keine gesicherten Angaben über den tatsächlichen Umfang der bezogenen Leistungen. Aus diesem Grund müsste man die gewählten Franchisestufen als natürliche Schwellen für die Trennung

---

<sup>216</sup> Dieser so genannte *tiers garant* wird im ambulanten Bereich durchwegs, im stationären System teilweise angewandt.

zwischen positiver Leistungsanspruchnahme und keiner Leistungsanspruchnahme wählen.<sup>217</sup> Allerdings würde bei einer derartigen Vorgehensweise vernachlässigt, dass die Patienten sich in Abhängigkeit der Franchisestufen unterschiedlichen Preisen gegenübersehen und ein Vergleich von Leistungen auf unterschiedlichen Franchisestufen nicht möglich wäre. Deshalb wird in der folgenden Analyse eine konstante Schwelle gesetzt. Wie andere Studien auch, die explizit die Wirkungen von Wahlfranchisen untersuchen,<sup>218</sup> verwenden wir dafür die dritte Wahlfranchise (600 SFr.). Damit gehen zum einen Leistungen der beiden unteren Franchisestufen eventuell „verloren“, während zum anderen auf den beiden hohen Stufen möglicherweise Leistungen eingehen, die unterhalb des Franchisebetrags liegen.

Auch bei dieser Schätzung wurden wieder alle 3 Varianten unterschiedlicher Untersuchungskollektive berücksichtigt. Im Folgenden wird allerdings nur das Ergebnis für das Gesamtkollektiv wiedergegeben, da sich die Ergebnisse für die Untergruppen nicht wesentlich unterscheiden.

Tabelle 18 Nachfrageschätzung: Arztbesuche

N = 122,392	Probit (Lst > 600)		Probit mit Instrumenten (GMM) (Lst >600)	
	Koeffizient	Standardfehler	Koeffizient	Standardfehler
<b>KONSTANTE</b>	-2.544**	0.096	-2.546**	0.114
<b>ZH</b>	-0.065**	0.025	-0.064**	0.030
<b>UNFALL</b>	0.008	0.010	0.010	0.012
<b>GESCHL</b>	-0.117**	0.009	-0.119**	0.010
<b>ALTER</b>	0.019**	0.002	0.019**	0.002
<b>ALTER2</b>	-2.56E-05*	1.45E-05	-2.28E-05	1.71E-05
<b>KIND</b>	0.023	0.029	0.020	0.035
<b>JUGEND</b>	0.067**	0.020	0.067**	0.024
<b>PR_ERM</b>	5.30E-03**	1.40E-03	3.46E-03**	7.41E-04
<b>LNL97</b>	0.113**	0.002	0.114**	0.002
<b>LNL98</b>	0.160**	0.002	0.162**	0.002
<b>SPT1</b>	-0.033*	0.017	-0.038**	0.018
<b>BEV_D</b>	-4.26E-04	2.77E-04	-3.26E-04	3.31E-04
<b>GEB_BEV</b>	3.53E-04	2.57E-04	3.42E-04	3.03E-04
<b>ARZTD</b>	8.63E-04	4.88E-04	8.25E-04	5.67E-04
<b>DK</b>	3.08E-03**	4.58E-04	2.93E-03**	5.51E-04
<b>MC</b>	-0.237**	0.018	-0.117	0.097
<b>(FR)</b>	(-0.08)**	0.006	(-0.08)**	0.007
<b>(Pseudo-) R<sup>2</sup></b>	0.58		0.58	
<b>Hausman-Test</b>			9.2	

\*\*\*) p < 0.05; \*) p < 0.01

<sup>217</sup> Dies entspricht somit der Sicht der Krankenkasse, für die ebenfalls nur die Kosten oberhalb der Franchisen relevant sind.

<sup>218</sup> Vgl. Nicolet et al. (2001) und Werblow und Felder (2003).

Die rechte Seite von Tabelle 18 enthält das Ergebnis für die Probitschätzung mit Instrumenten. Für diese Schätzung wurde ein verallgemeinerter Momentenschätzer verwendet (GMM), der es bei nichtlinearen Schätzungen erlaubt, Instrumente zu berücksichtigen.<sup>219</sup> Die linke Seite der Tabelle zeigt die Ergebnisse des Probit-Modells, wenn die Entscheidung für eine Managed-Care-Organisation als exogen betrachtet wird. Der Hausmanntest belegt die Endogenität von MC und bestätigt somit auch die Wahl der Instrumentalvariablen.<sup>220</sup>

Zunächst erkennt man, dass die Versicherten in Zürich eine signifikant geringere Wahrscheinlichkeit aufweisen, zum Arzt zu gehen, als die Versicherten in Genf. Dabei bedeutet „weniger häufig zum Arzt gehen“, dass die Wahrscheinlichkeit, Leistungen oberhalb von 600 SFr. zu verursachen, in Zürich um 2.8% geringer ist als in Genf.<sup>221</sup> Für Männer ist diese Wahrscheinlichkeit erwartungsgemäß geringer als für Frauen. Mit dem Alter nimmt die Wahrscheinlichkeit von mehr Arztkontakten zu, wobei aber Jugendliche ebenfalls eine hohe Wahrscheinlichkeit aufweisen. Hohe Prämienverbilligungen erhöhen die Wahrscheinlichkeit eines Arztbesuches genauso wie hohe Leistungen in den Vorjahren. Letzteres Ergebnis unterstreicht die Bedeutung des Gesundheitszustandes bei der Entscheidung für oder gegen einen Arztbesuch. Von den Erklärungsfaktoren auf Marktebene ist nur eine einzige signifikant von Null verschieden: Die Durchschnittskosten der Wohngemeinde des Versicherten haben einen positiven Einfluss auf die Wahrscheinlichkeit, einen Arzt aufzusuchen. Dass die anderen Marktvariablen nicht signifikant von Null verschieden sind, könnte von der teilweise recht hohen Korrelation zwischen diesen Variablen herrühren.

Hinsichtlich des Hauptuntersuchungsgegenstandes zeigt sich, dass die Versicherten in Hausarzt-netzen und HMOs keine geringere Wahrscheinlichkeit aufweisen, Arztbesuche in Anspruch zu nehmen, als die Versicherten in der konventionellen Versorgung. Während im einfachen Probit-

---

<sup>219</sup> Verallgemeinerter Momentenschätzer deshalb, weil er eine ganze Klasse von Spezialfällen enthält. Der bekannteste Spezialfall stellt der OLS-Schätzer dar, denn dieser beruht auf der Momentenrestriktion  $E[x_i, \varepsilon_i] = 0$ , d.h. der Annahme der Unkorreliertheit der erklärenden Variablen und den Störtermen. Bei einer oder mehreren endogenen Variablen ist gerade diese Annahme verletzt und man verwendet einen Instrumentalvariablen-schätzer für den nun die Momentenrestriktion  $E[z_i, \varepsilon_i] = 0$  bei gleichzeitiger Korrelation von  $z$  mit  $x$  gelten muss. Vgl. zum GMM-Ansatz bspw. Greene (1997) oder auch Verbeek (2001).

<sup>220</sup> Vgl. Greene (1997).

<sup>221</sup> Der marginale Effekt einer Variablen auf die Wahrscheinlichkeit ist gegeben durch  $\phi(\beta \cdot x) \beta_k$  mit  $\phi$  der Dichte der Standardnormalverteilung. Für binäre Regressoren gilt diese Ableitung im Allgemeinen nicht, wie aber verschiedentlich gezeigt wurde, kann die angegebene Ableitung als eine gute Annäherung an den tatsächlichen Effekt aufgefasst werden (vgl. Greene 1997). Die marginalen Effekte wurden am Stichproben-Mittelwert evaluiert.

Modell mit exogener Wahlentscheidung die Variable *MC* einen signifikant negativen Einfluss auf die Wahrscheinlichkeit, einen Arzt aufzusuchen, ausübt, verschwindet dieser Effekt, wenn der Endogenität von *MC* Rechnung getragen wird. Die beobachtete geringere Wahrscheinlichkeit von Arztbesuchen im Managed-Care-Bereich ist daher offensichtlich auf den Umstand zurückzuführen, dass ein besonderer Typus Versicherter MC-Versicherungen nachfragt. Dieser besondere Typus wurde hier durch bestimmte, die Vertragswahl beeinflussende Faktoren beschrieben, die das Selektionsverhalten abbilden sollen.

An dem Ergebnis ändert sich auch nichts, wenn zusätzlich die Franchisewahl (*FR*) als exogener Erklärungsfaktor in die Schätzung einbezogen wird. Wie zu erwarten, besitzt die Wahl einer höheren Franchise einen signifikant negativen Effekt auf die Wahrscheinlichkeit, einen Arzt aufzusuchen. Allerdings steht hinter dieser Variablen auch wieder eine Entscheidung des Versicherten, sodass *FR* ebenfalls endogen sein dürfte. Um den kausalen Effekt auch dieser Wahlentscheidung bestimmen zu können, müssten weitere Instrumente gefunden werden, welche die Franchisewahl in geeigneter Weise abbilden können. Das dürfte in der Praxis schwierig sein, da MC- und Franchisewahl ähnliche individuelle Entscheidungsprozesse betreffen.<sup>222</sup>

#### **Nachfrage – Leistungsanspruchnahme**

Die Höhe der in der laufenden Periode beobachteten Leistungen ist schließlich auf der dritten Stufe zu erklären. Da diese Leistungen extrem schief verteilt sind, werden sie in logarithmierter Form in der Schätzung verwendet. Die logarithmierten Leistungen sind annähernd normalverteilt.

Tabelle 19 weist in der rechten Hälfte die korrigierten Resultate mit dem Instrument für MC sowie der entsprechenden Korrektur für Lambda aus; in der linken Hälfte sind hingegen die Resultate unter der Annahme einer exogenen Wahl des Vertrages dargestellt. Auch die Schätzung der Leistungsanspruchnahme wurde wieder für alle drei Untersuchungsgebiete durchgeführt. Da sich die einzelnen Ergebnisse nicht wesentlich voneinander unterscheiden, wird wiederum nur das Ergebnis für die gesamte Stichprobe dargestellt.

---

<sup>222</sup> Vgl. aber auch die Ausführungen zu den Unterschieden von MC-Instrumenten und nachfrageseitigen Kostenbeteiligungen am Anfang dieses Abschnitts.

Tabelle 19 Nachfrageschätzung: Leistungsanspruchnahme

N = 67,898	Probit – OLS		GMM – 2SLS	
	Koeff.	Std.Fehler.	Koeff.	Std.Fehler.
KONSTANTE	-1.559**	0.311	-1.645**	0.120
ZH	-0.048	0.057	-0.049**	0.018
UNFALL	0.074**	0.024	0.077**	0.008
GESCHL	-0.162**	0.020	-0.168**	0.007
ALTER	0.044**	0.004	0.044**	0.001
ALTER2	-1.52E-04**	3.11E-05	-1.48E-04**	9.98E-06
KIND	-0.373**	0.071	-0.371**	0.023
JUGEND	0.019	0.051	0.022	0.016
PR_ERM	8.90E-03**	1.13E-03	8.78E-03**	3.51E-04
LNL97	0.259**	0.007	0.264**	0.003
LNL98	0.421**	0.010	0.430**	0.004
SPT1	0.074**	0.031	0.068**	0.010
BEV_D	-1.59E-03**	5.98E-04	-1.38E-03**	1.96E-04
GEB_BEV	6.43E-04	6.02E-04	6.34E-04**	1.90E-04
ARZTD	2.35E-03**	1.08E-03	2.28E-03**	3.40E-04
DK	7.42E-03**	1.06E-03	7.15E-03**	3.49E-04
MC	-0.477**	0.048	-0.180**	0.069
(FR)	(-0.15**)	(0.01)	(-0.15**)	(0.05)
LAMBDA	3.428**	0.092	3.445**	0.042
R <sup>2</sup>	0.373		0.374	
HAUSMAN-TEST			179	

\*\*p < 0.01, \*) p < 0.05

Im Unterschied zur Schätzung der Wahrscheinlichkeit, überhaupt einen Arzt aufzusuchen, ist der Managed-Care-Effekt jetzt signifikant negativ. D.h. auch wenn der Endogenität der Vertragswahl Rechnung getragen wird, verursachen Managed-Care-Versicherte immer noch deutlich geringere Kosten als vergleichbare Versicherte in der konventionellen Versorgung. Der selektionskorrigierte Koeffizient von MC in Höhe von  $-0.18$  bedeutet, dass die Leistungsanspruchnahme eines durchschnittlichen MC-Versicherten um 16.5% unter den Ausgaben für einen durchschnittlichen Versicherten im konventionellen Bereich liegen. Berücksichtigt man die Selektion der Versicherten dagegen nicht, ist dieser Wert mehr als doppelt so hoch (37.8%).<sup>223</sup> D.h. mit anderen Worten, dass ungefähr die Hälfte der beobachteten geringeren Leistungsanspruchnahme von Versicherten im Managed-Care-Bereich auf die Selektion der guten Risiken zurückzuführen ist, während die andere Hälfte durch die verhaltensändernde Wirkung der MC-Beteiligung zustande kommt.

<sup>223</sup> Da es sich bei dem vorliegenden Modell um ein semilogarithmisches Modell handelt, lassen sich die Koeffizienten der erklärenden Variablen im Allgemeinen als Veränderungsrate der abhängigen Variablen interpretieren. Wie Halvorsen und Palmquist (1980) gezeigt haben, gilt dies nicht für Dummy-Variablen. Vielmehr ist in diesen Fällen eine Korrektur vorzunehmen ( $\tilde{\alpha} = e^{\alpha} - 1$ ).

Die anderen erklärenden Faktoren sind in der Regel signifikant. So ist im Kanton Zürich die Leistungsanspruchnahme durchschnittliche um 4.8% geringer ist als im Kanton Genf. Alter und Geschlecht tragen ebenfalls signifikant zur Erklärung der Leistungsanspruchnahme bei: Die Leistungen nehmen mit dem Alter zu, allerdings mit abnehmenden Raten. Männer sind im Durchschnitt um 15% billiger als Frauen. Die Bedeutung des Gesundheitszustandes für die Nachfrage nach Gesundheitsleistungen der Versicherten wird bestätigt, denn die Leistungsanspruchnahme der Vorjahre ist deutlich signifikant von Null verschieden. In die gleiche Richtung weist die Variable Spital im Vorjahr (*SPT1*): War ein Versicherter im Vorjahr mindestens einen Tag im Spital, so ist im folgenden Jahr eine 7.7% höhere Leistungsanspruchnahme zu erwarten als ohne einen Spitalaufenthalt im Vorjahr.

Interessanterweise sind die Marktvariablen in der Regel signifikant. Vor allem ist jetzt die Arztdichte signifikant positiv. Dies kann man als Anzeichen einer Angebotsinduzierten Nachfrage auffassen (AIN). Allerdings sollte man mit solchen Rückschlüssen vorsichtig sein, da auch eine umgekehrte Kausalität wirken könnte.<sup>224</sup> Dieses Endogenitätsproblem kann behoben werden, wenn entsprechende Instrumente für die Arztdichte gefunden werden (vgl. Abschnitt 8.2.2). Entsprechend der Schätzung in Abschnitt 9.1 lässt sich die Anzahl von Ärzten vor allem durch die Bevölkerung der jeweiligen Gemeinde erklären. Diese ist aber schon Bestandteil der Nachfragegleichung, sodass andere Instrumente gefunden werden müssen. Um die Exogenität der Arztdichte testen zu können, wurden daher die drei Angebotsvariablen Anzahl von Internisten, Chirurgen und Spitälern (vgl. Abschnitt 9.1) als Instrumente verwendet. Die neue Schätzung mit den Instrumenten bestätigt das dargestellte Ergebnis in Tabelle 19. Trotzdem sollte es mit Vorsicht betrachtet werden, da einerseits die Instrumentenauswahl – wie gesagt – schwierig ist und andererseits auch wieder das Problem unterschiedlicher Ebenen auftaucht, welches hier nicht berücksichtigt werden konnte.

#### **9.3.4 Zusammenfassung und Ausblick**

Um die Unterschiede in der Leistungsanspruchnahme von Versicherten im Managed-Care- und im konventionellen Bereich zu untersuchen, wurde ein 3-stufiges Entscheidungsmodell geschätzt. Bei der Schätzung muss berücksichtigt werden, dass die Höhe der Leistungsanspruchnahme selbst von der Wahl des Versicherungsvertrages abhängt. Daher würde eine

---

<sup>224</sup> Unter der AIN-Hypothese würde man sagen, dass die Nachfrage dort hoch ist, wo die Anbieterdichte hoch ist. Allerdings könnte die Arztdichte auch dort am größten sein, wo die Wahrscheinlichkeit, dass viele Patienten einen einzelnen Arzt aufsuchen, vergleichsweise hoch ist.



Nichtberücksichtigung dieser *Selbstselektion* zu verzerrten Schätzungen des MC-Effekts führen. Durch die Modellierung der Wahlentscheidung des Versicherten und die Beachtung der geschätzten Wahl in der Nachfrage soll die Selektionsverzerrung behoben werden. Dabei dienen verschiedene Variablen über den Abschluss weiterer privater Versicherungsverträge der Erklärung der Wahl eines MC-Vertrages. Damit sind die Zusatzversicherungen hauptsächlich für die Identifizierung des MC-Effektes verantwortlich, da andere wichtige Erklärungsfaktoren wie Geschlecht, Alter, Gesundheitszustand sowohl die Wahl des Versicherungsvertrages als auch die Leistungsanspruchnahme beeinflussen.

Wird die Selektionsverzerrung in der beschriebenen Weise behoben, dann unterscheidet sich die Wahrscheinlichkeit, einen Arzt aufzusuchen, bei MC-Versicherten nicht signifikant von derjenigen für konventionell Versicherte. Hingegen weisen die MC-Versicherten ca. 16% geringere Kosten als Versicherte mit herkömmlichen Versicherungsverträgen auf. Aufgrund eines Vergleichs mit einer Schätzung ohne Selektionskorrektur lässt sich die Aufteilung zwischen Managed-Care- und Selektionseffekt bei ca. 50% festmachen. Damit ist ungefähr die Hälfte der beobachteten Einsparung in den MC-Modellen auf die günstigere Risikostruktur der Versicherten und die damit verbundene häufigere Wahl eines MC-Vertrages durch diese Versicherten zurückzuführen. Die andere Hälfte der Reduktion in der Leistungsanspruchnahme kann dagegen einem veränderten, kostenbewussteren Verhalten der Versicherten zugeordnet werden.

Die Möglichkeit einer Angebotsinduzierung durch die Ärzte kann nicht ausgeschlossen werden. Indizien dafür waren einerseits die nicht signifikante Arztdichte bei der Entscheidung des Versicherten, einen Arzt aufzusuchen, und andererseits der signifikante Einfluss der Arztdichte auf die Leistungsanspruchnahme selbst. Während die nicht signifikante Arztdichte bei der Erklärung der Arztbesuche zeigt, dass diese Entscheidung patientengesteuert ist, weist die signifikante Arztdichte bei der Erklärung der Leistungsanspruchnahme in Richtung einer Beeinflussung der Nachfrage durch die Anbieter hin.

Der Managed-Care-Effekt wurde in diesem Abschnitt ausschließlich nachfrageseitig definiert, d.h. die Kosteneinsparungen im MC-Bereich wurden ausschließlich den Verhaltensänderungen der Versicherten zugeschrieben. Im folgenden Gesamtmodell wird dieser globale Managed-Care-Effekt zerlegt. Denn es ist zu vermuten, dass ein Teil der erzielten Einsparungen auf die Tätigkeit der Ärzte zurückzuführen ist.

## 9.4 Gesamtmodell – Angebot

### 9.4.1 Herleitung des Modells (Allgemein)

Kapitel 5 und 6 beschäftigten sich theoretisch mit der Arzt-Patienten-Beziehung. Während im 5. Kapitel die Nutzenfunktion des Arztes im Mittelpunkt stand, wurde im 6. Kapitel die Arztpraxis als Firma betrachtet und somit der Nutzen des Arztes durch den Gewinn aus der ärztlichen Tätigkeit beschrieben. In beiden Abschnitten stand die normative Frage im Vordergrund, wie ein optimales Vergütungssystem unter bestimmten Annahmen aussehen müsste.

Die Ergebnisse im Kapitel 5 zeigen u.a., dass der Nutzen des Arztes ( $U_A$ ) nicht nur durch das Einkommen ( $Y$ ), sondern auch durch den Patientennutzen beeinflusst wird. Letzterer hängt im Zusammenhang mit einem Arztbesuch entscheidend vom Gesundheitszustand des Patienten ab. Daher ist es wichtig, den Gesundheitszustand des Patienten ( $G$ ) in der Nutzenfunktion des Arztes zu berücksichtigen. Darüber hinaus ist bekannt, dass der Patientennutzen auch noch von anderen Faktoren beeinflusst wird. So spielen die Patientenpräferenzen für bestimmte Behandlungsmethoden und Ärzte eine wichtige Rolle. Außerdem kommt offenbar den Suchkosten eine wesentliche Bedeutung zu. Die Patienten sind umso eher bereit, einem ihrer Meinung nach schlechten Arzt den Rücken zu kehren, wenn ihre Suchkosten gering sind. Die mit den Patienten verbundenen Faktoren werden im Folgenden in einem Vektor  $P$  zusammengefasst.

Das Nutzenniveau des Arztes wird nicht ausschließlich durch Einkommensinteressen und die Patientenpräferenzen bestimmt, sondern noch von anderen Faktoren ( $X$ ) beeinflusst. Dazu zählen bspw. die Freizeit oder aber die intellektuelle Zufriedenheit mit der Arbeit.<sup>225</sup> Ein weiterer wichtiger Aspekt ergab sich aus der zusätzlichen Betrachtung des Vermögens ( $W$ ) eines Arztes in Prinzipal-Agent-Modellen. Ärzte reagieren je nach Vermögen unterschiedlich auf Vergütungssysteme. Auch dieser Aspekt soll daher in der empirischen Untersuchung berücksichtigt werden. In allgemeiner Form lässt sich damit die Nutzenfunktion eines Arztes wie folgt aufschreiben:

$$U_A = f(Y, W, X; G, P). \quad (9.25)$$

Im Kapitel 6 wurde hergeleitet, wie sich ein gewinnmaximierender Arzt bei monopolistischer Konkurrenz verhält, wenn er bei einem vorgegebenen Vergütungssystem ausschließlich durch die Mengewahl und die Wahl eines optimalen Anstrengungsniveaus seinen Gewinn verändern

---

<sup>225</sup> Vgl. Kristiansen (1995).

kann. Die mit dem Maximierungsproblem des Arztes verbundenen Bedingungen erster Ordnung seien an dieser Stelle nochmals in Erinnerung gerufen:

$$(U_x - p_d) \frac{a + (p_s - c)x}{p_s - c} = -\frac{n}{n'} \quad (6.17)$$

$$U_e \left( \frac{a}{x} + (p_s - c) \right) = \frac{n}{n'} c_e, \quad (6.18)$$

mit  $x$  der Menge an medizinischen Leistungen,  $e$  den Anstrengungen des Arztes, die notwendig sind, um diese Menge zu produzieren; weiterhin mit den Parametern des Vergütungssystems:  $a$  der Pauschale,  $p_s$  dem (exogenen) Anbieterpreis sowie  $c$  den Grenzkosten der Bereitstellung einer Mengeneinheit und  $c_e$  den Grenzkosten für die Erhöhung der Anstrengungen. Der Preis, dem sich die Patienten gegenübersehen, wird mit  $p_d$  bezeichnet.  $U_x$  und  $U_e$  bezeichnen die Grenznutzen des Patienten bezüglich  $x$  und  $e$ . Die Variable  $n$  gibt die Anzahl Patienten einer Arztpraxis wieder. Sie hängt vom Nettonutzen der Patienten ab. Für ihre erste Ableitung  $n'$  wird ein positives Vorzeichen angenommen, d.h. mit einem höheren Nettonutzen für die Patienten steigt auch die Patientenzahl der Arztpraxis an.

Diese Gleichungen charakterisieren die Faktoren, die den Gewinn einer Arztpraxis maßgeblich beeinflussen. Gleichung (6.17) spiegelt den Zielkonflikt zwischen Mengenerhöhung – und damit Gewinnsteigerung – und den Patienteninteressen wider. Interessant ist hierbei vor allem, dass der Arzt bei der Mengenentscheidung für einen bestimmten Patienten nicht nur den Nettonutzen des einzelnen Patienten im Auge behalten muss, sondern auch die Wirkungen seiner Entscheidung auf seinen gesamten Patientenstamm. Allerdings ist bei einer derartigen Interpretation eine wichtige Annahme des theoretischen Modells zu berücksichtigen: identische Patienten. Diese kritische Annahme wird uns auch bei den folgenden Untersuchungen begleiten.

Die Bedingung (6.18) zeigt, wie der Arzt sein optimales Anstrengungsniveau wählen sollte, wenn er seinen Gewinn maximieren möchte. Auch hierbei geht es um einen Zielkonflikt, nämlich den Konflikt zwischen Gewinnsteigerung aufgrund höherer Anstrengungen einerseits und den zusätzlichen Kosten andererseits. Dabei muss er sowohl die direkten Auswirkungen der Aufwandssteigerung für seine Patienten im Auge behalten, als auch den Einfluss auf seinen Patientenstamm. So erhöht sich bspw. durch eine Qualitätssteigerung der Behandlung nicht nur der Nutzen für den einzelnen Patienten, sondern ebenfalls die Attraktivität der Arztpraxis für potenzielle Patienten.

Berücksichtigt man die wichtigen Faktoren für ein Optimum in der Gewinnfunktion des Arztes selbst, so kann sie geschrieben werden als:

$$\Pi = f(NU, n(p-c)x, \varepsilon), \quad (9.26)$$

d.h. der Gewinn des Arztes wird nicht nur einfach durch den Umsatz und die Kosten der Arztpraxis determiniert ( $n(p-c)x$ ), sondern auch direkt durch den Nettonutzen der Patienten ( $NU$ ) und die Markt elastizität  $\varepsilon$  beeinflusst.

Schließlich konnte gezeigt werden, dass es die Möglichkeit von Spill-over-Effekten (vgl. Abschnitt 6.4) gibt. Bei den Ärzten unserer Stichprobe ist die Wahrscheinlichkeit für derartige Effekte hoch, da die meisten Ärzte, die in Hausarztmodellen arbeiten, auch Patienten aus dem konventionellen Bereich versorgen.

Geht man davon aus, dass das Einkommen des Arztes ausschließlich aus seiner Praxistätigkeit erwächst, kann man nun die Gewinnfunktion des Arztes in die Nutzenfunktion einsetzen und erhält folgende modifizierte – indirekte – Nutzenfunktion:

$$U_A = f(\Pi[NU, x, \varepsilon], W; G, P). \quad (9.27)$$

Der Nutzen des Arztes lässt sich also als eine Funktion des Einkommens und des Vermögens des Arztes auf der einen Seite und des Gesundheitszustandes des Patienten auf der anderen Seite beschreiben. Wobei das laufende Einkommen des Arztes durch den Gewinn der Arztpraxis bestimmt wird.

#### **9.4.2 Herleitung des Modells (im Besonderen)**

Im Folgenden werden einzelne Modelle betrachtet, die entweder Bestandteile obiger Nutzenfunktion oder den Nutzen aus einer bestimmten ärztlichen Handlung untersuchen. So wird in einem ersten Modell das Überweisungsverhalten der Ärzte beleuchtet. Konkret wird der Nutzen eines Arztes untersucht, wenn er einen Patienten zu einem anderen Leistungserbringer überweist. Dagegen werden im zweiten Modell die Konsultationen in einer Arztpraxis analysiert. Die Konsultation betrifft den Nutzen des Arztes in zweierlei Hinsicht: Einerseits ist sie die grundlegende Outputeinheit auf deren Grundlage ein Großteil der Vergütung erfolgt, andererseits ist sie aber auch ein Indikator für den zeitlichen Aufwand, den ein Arzt für seine Patienten aufbringt. Die Schätzungen des Überweisungsverhaltens und der Konsultationen sollen nicht zuletzt Aufschluss über den *Casemix* in einer Arztpraxis geben. Die beiden folgenden Modelle untersu

chen dann schließlich den Gewinn bzw. den Gesamtumsatz des Arztes unter Berücksichtigung des Casemix.

Allgemein lässt sich das ökonometrische Modell mit den unterschiedlichen Ergebnisvariablen entsprechend obiger Herleitung schreiben zu:

$$y_{ijk} = \alpha_0 + \alpha_1 A_i + \alpha_2 P_j + \alpha_3 AP_{ij} + \alpha_4 Gem_k + v_{ijk}, \quad (9.28)$$

wobei  $A$  Eigenschaften eines Arztes  $i$  beschreibt,  $P$  Charakteristika von Patient  $j$  umfasst und  $AP$  die besondere Zusammensetzung des Patientenstamms auf Praxisebene sowie bestimmte Eigenheiten jeder einzelnen Arzt-Patienten-Beziehung widerspiegeln (casemix). In  $AP$  sind damit sowohl Variablen enthalten, die auf Praxisebene gemessen werden (wie etwa das Durchschnittsalter der Patienten einer Praxis), als auch Variablen, die auf der Ebene des einzelnen Arzt-Patienten-Kontaktes gemessen werden (etwa die Anzahl der Leistungsnummern).<sup>226</sup> Die Eigenschaften des Marktes ( $GEM$ ) variieren über die  $k$  betrachteten Gemeinden in der Stichprobe. Entsprechend der Klassifizierung aus Abschnitt 8.2.4 entspricht das beschriebene Modell in (9.28) einem hierarchischen Datenmodell mit drei Ebenen: Arzt, Patient und Markt.

Der uns interessierende Haupteffekt – die Wirkung von Managed Care auf das Arztverhalten – wird auf der Ebene des Arztes bzw. auf der Ebene des Patienten gemessen. Die besondere Datenstruktur kann durch die explizite Berücksichtigung eines arztspezifischen Störterms in sogenannten *random effects models* erreicht werden (vgl. auch Abschnitt 8.2.4). D.h., dass der Störterm in (9.28) in eine Arztkomponente  $\mu_i$  und einen Störterm der Patienten in den Arztpraxen (in den unterschiedlichen Gemeinden) zerlegt wird  $\varepsilon_{ijk}$ . Der möglichen Korrelation zwischen Markteigenschaften und Variablen der Arzt- bzw. Patientenebene soll durch die Berücksichtigung hierfür geeigneter Variablen begegnet werden (vgl. Abschnitt 9.4.3).

Ein weiterer wichtiger Punkt betrifft die mögliche Selektion der Ärzte und Patienten. Im Abschnitt 9.3 wurde gezeigt, wie die Selektion von bestimmten Versicherten in Managed-Care-Verträge bei der Schätzung der Nachfrage berücksichtigt werden kann. Dabei wurde die Versicherungswahl in geeigneter Form in die Nachfrage eingefügt. Das Ergebnis der Wahlentscheidung der Versicherten soll daher auch im Folgenden Verwendung finden. Hierfür wird der im Abschnitt 8.2.3 vorgestellte Kontrollfunktionsansatz verwendet, indem die geschätzte Mills-

Ratio aus der Probit-Schätzung der Wahlentscheidung in der Schätzung der unterschiedlichen Ergebnisvariablen eingefügt wird. Diesen Selektionsterm bezeichnen wir mit  $\lambda_1$ . Auf der Arztseite besteht die Möglichkeit, dass die Ärzte, die sich an Managed-Care-Modellen beteiligen, sich durch bestimmte Charakteristika von den anderen Ärzten unterscheiden. So zeigte die Schätzung der Effizienz (Abschnitt 9.2), dass Ärzte in Managed-Care-Modellen einen höheren Effizienzwert aufweisen als andere Ärzte. Allerdings konnte diese Schätzung noch nicht ermitteln, aus welchem Grund die Ärzte einen höheren Effizienzwert hatten als andere. Durch die Berücksichtigung dieses Effizienzwertes ( $Eff$ ) im Gesamtmodell kann es gelingen, den Selektionseffekt vom Managed-Care-Effekt zu trennen. In diesem Sinne lässt sich die Berücksichtigung des Effizienzeffektes ebenfalls als eine Art Kontrollfunktionsansatz auffassen.

Obwohl die Selektionsproblematik damit im Modell explizit berücksichtigt wird, sei darauf hingewiesen, dass die Endogenität der beiden Effekte ohnehin durch die gewählte Modellierung zum Teil bereits „geheilt“ wird. In der Analyse betrachten wir den direkten Arzt-Patienten-Kontakt einzelner Ärzte mit relativ vielen Patienten, berücksichtigen aber gleichzeitig durch die *random effects* die nicht zu beobachtende Verschiedenartigkeit der Ärzte. Nichtbeobachtbare Verschiedenartigkeit der Ärzte kann natürlich auch die Selektion von bestimmten Ärzten in MC-Verträge beinhalten. Für die Patienten ist festzuhalten, dass wir zwar wissen, dass Versicherte mit einem relativ guten Gesundheitszustand eher an MC-Verträgen teilnehmen (vgl. wiederum Abschnitt 9.3), dieser Umstand für den direkten Arzt-Patienten-Kontakt aber weniger von Belang ist, weil wir hier ausschließlich Patienten betrachten. D.h. mit anderen Worten, dass der Teil der Selektion, der sich darin äußert, dass „gute“ Risiken überhaupt nicht erst zum Arzt gehen, bei der Untersuchung des Arzt-Patienten-Kontaktes wegfällt, weil nur kranke Versicherte betrachtet werden.

Insgesamt erhalten wir schließlich unter Einbeziehung der beschriebenen Selektionsterme und unter Berücksichtigung der besonderen Datenstruktur das allgemeine Modell der nachfolgenden Schätzungen:

$$\begin{aligned}
 y_{ijk} = & \alpha_0 + \alpha_1' A_i + \alpha_2' P_j + \alpha_3' AP_{ij} + \alpha_4' Gem_k \\
 & + \alpha_5 MC\_A_i + \alpha_6 MC\_P_j + \alpha_7 Eff_i + \alpha_8 \lambda_{1j} + u_i + \varepsilon_{ijk}.
 \end{aligned}
 \tag{9.29}$$

---

<sup>226</sup> Vgl. zur Beschreibung der einzelnen Variablen die Ausführungen zu den Daten im Abschnitt 9.4.3 wie auch die Schätzungen der einzelnen Modelle, in denen diese Variablen verwendet werden (Abschnitte 9.4.5, 9.4.6 und 9.4.7. Darüber hinaus sind alle Variablen des Gesamtmodells mit einer Kurzbeschreibung im Anhang III zusammengefasst.

Gleichung (9.29) beschreibt somit den Einfluss von Managed Care auf die jeweilige Ergebnisvariable unter Berücksichtigung der besonderen Datenstruktur und der möglichen Endogenität der interessierenden Effekte. Darüber hinaus wird für bestimmte wichtige Charakteristika von Ärzten, Patienten und Gemeinden kontrolliert

Die Schätzung für unterschiedliche Ergebnisvariablen erfolgt in einer Stichprobe, in der viele konventionell tätige Ärzte und einige MC-Ärzte enthalten sind. Allerdings ist eine eindeutige Abgrenzung nicht möglich. So weiß man bspw. nicht, ob ein Arzt, der in unserer Stichprobe ausschließlich konventionelle Patienten behandelt hat, nicht auch in einem hier nicht betrachteten Hausarztmodell beteiligt ist. Außerdem werden nur CSS-Versicherte untersucht, sodass ebenfalls die Möglichkeit besteht – obwohl der Arzt zwar keine CSS-Versicherten eines Hausarztnetzes behandelt hat –, dass er dennoch Versicherte anderer Krankenkassen des gleichen Netzes versorgt.

Die Untersuchung wird unter Einbeziehung der einzigen in der Stichprobe befindlichen HMO durchgeführt. Da die Daten der HMO für die interne Leistungserstellung nur begrenzt aussagefähig sind, muss eine mögliche Kontrolle auf Verzerrungen der Schätzergebnisse erfolgen. Aus diesem Grund werden alle hier vorgestellten Schätzungen sowohl mit als auch ohne die Einbeziehung der HMO geschätzt.

#### **9.4.3 Daten**

Es stehen wiederum Daten über 3 Jahre zur Verfügung (1997-1999). Die Daten für die Jahre 1997 und 1998 gehen nicht direkt in die Schätzungen ein, da sie zum einen für die Konstruktion der Korrekturterme für die Selektion bereits verbraucht worden sind (insbesondere für die Abschätzung des Gesundheitszustands des Patienten – vgl. Abschnitt 9.3)<sup>227</sup> und zum anderen zur Konstruktion weiterer Hilfsvariablen dienen.

Im Vergleich zur Untersuchung des Nachfrageverhaltens aller Versicherten im Untersuchungsgebiet (vgl. Abschnitt 9.3) wird die Untersuchungsstichprobe dadurch weiter eingeschränkt, dass jetzt nur diejenigen Patienten betrachtet werden, die von Grundversorgern entsprechend der Definition aus Abschnitt 7.3 behandelt worden sind. Darüber hinaus ergibt sich eine weitere Einschränkung aus der Untersuchung der Effizienz der Arztpraxen (vgl. Abschnitt 9.2). Dort wurden nur Ärzte betrachtet, von denen man zumindest annähernd davon ausgehen kann, dass sie hauptsächlich in einer Arztpraxis arbeiten. Diese Einteilung schloss Ärzte aus der Untersu

chung aus, die weniger als 20 Patienten im Jahr behandelten und die einen geringeren Umsatz als 20,000 SFr. im Jahr erzielten. So ergibt sich für die folgenden Untersuchungen eine Untersuchungsstichprobe von zunächst 100,914 Fällen. Außerdem wurden noch alle Fälle ausgeschlossen, die zu einer Einweisung in ein Krankenhaus führten. Da – wie bereits erwähnt – die Versicherten im konventionellen Bereich prinzipiell freie Arztwahl besitzen, ist zu erwarten, dass die in der Stichprobe befindlichen Ärzte bei den konventionell versicherten Patienten weniger Leistungen veranlassen als bei MC-Patienten. Das Problem verschärft sich noch bei der Betrachtung von Krankenhauseinweisungen, weil Patienten oft nicht von Allgemeinmediziner\*innen ins Krankenhaus überwiesen werden, sondern von den hier nicht betrachteten Spezialisten. Alles in allem umfasst das Untersuchungskollektiv daher 100,523 Fälle.

Die erklärenden Variablen, die zur Schätzung der unterschiedlichen Ergebnisvariablen verwendet werden, lassen sich entsprechend der Modellherleitung in vier Gruppen einteilen: 1.) Variablen, die den Gesundheitszustand und die Heterogenität der *Patienten* abbilden, 2.) Variablen, die mit spezifischen *Ärzte*-charakteristika und Eigenschaften der Arztpraxis verbunden sind, 3.) Variablen, die – zumindest zum Teil – den unterschiedlichen *Casemix* der Arztpraxen und die Leistungszusammensetzung einer konkreten Behandlung erklären und schließlich 4.) Variablen, die besondere *Markteigenschaften* anzeigen.

Da im Folgenden die konkrete Arzt-Patienten-Beziehung betrachtet wird, können jetzt Variablen einbezogen werden, die aufgrund der höheren Aggregationsebene bei den anderen Schätzungen nicht berücksichtigt wurden. Betrachten wir zunächst Variablen auf der Patientenebene. Eine besonders wichtige Variable stellt hier die Distanz zwischen dem Patienten und der Praxis des Arztes dar. Diese Variable bietet nicht nur direkt eine Abschätzung für die Zeitkosten eines Patienten, sondern auch (in technischer Hinsicht) die Verbindung unterschiedlicher Aggregationsebenen der Daten. Die Entfernung wird gemessen durch die (logarithmierte) Distanz zwischen der Wohngemeinde des Patienten und der Gemeinde des Arztes (*LD*). Interessant ist in diesem Zusammenhang außerdem die Beobachtung, ob ein Patient einen Arzt außerhalb seiner Wohngemeinde aufsucht (*KA*). Die Leistungsanspruchnahme der Vorjahre wurde noch genauer erfasst, indem zusätzlich zur Inanspruchnahme in den beiden Vorjahren auch die Leistungsanspruchnahme in Abhängigkeit vom Alter berücksichtigt wurde (durch das Produkt aus den Leistungen im Vorjahr und dem Alter des Patienten *ALT97*, *ALT98*). Außerdem wird berücksichtigt, dass die Leistungsanspruchnahme stark von der gewählten Kostenbeteiligungs-

---

<sup>227</sup> Die Beschreibung aller übrigen aufgeführten Variablen findet sich im Abschnitt 7.4.



regelung abhängen kann. *FR98* wird durch das Produkt der Franchisevariablen *FR* und *LLST98* gebildet. Aufgrund der Individualdatenstruktur kann außerdem die Variable *UNT\_VRS* einbezogen werden. Sie misst, ob der Patient im Betrachtungszeitraum bei mehreren Allgemeinmedizinern gewesen ist. Wegen der Gatekeeperfunktion des Allgemeinarztes in MC-Modellen dürfte der Erwartungswert dieser Variablen für konventionell versicherte Patienten höher liegen als für MC-Patienten. Wie bei der Nachfrageschätzung wurde auch bei den Schätzungen des Gesamtmodells das Alter des Versicherten genauer spezifiziert: Eine Dummy-Variable identifiziert Kinder von 0-18 Jahre (*KIND*) und eine weitere junge Erwachsene von 19-26 Jahren (*JUGEND*).

Auf Arzzebene werden zusätzlich zu den schon bekannten Variablen weitere entsprechend der theoretischen Anforderungen berücksichtigt. So wurden aus den Umsatzzahlen der beiden vergangenen Jahre zwei Variablen gebildet, die das kurzfristige wie das langfristige Einkommen des Arztes approximieren und daher auch als ein Indikator für die Vermögenssituation des Arztes dienen können. *L\_ARZT98* ist der logarithmierte Umsatz des Vorjahrs und *VERM* bezeichnet das durchschnittliche Arzteinkommen der letzten beiden Jahre. Die Verbindung mit der Berufserfahrung des Arztes (*BERUFSER*) soll zusätzlich die unterschiedliche Vermögenssituation der Ärzte in unterschiedlichen Lebensphasen einfangen. Dabei wurde *BERUFSER* aus der schon bekannten Variablen *EINTRITT* durch die logarithmierte Differenz mit dem Untersuchungsjahr gebildet ( $BERUFSER = \text{Log}(1999 - EINTRITT)$ ). Außerdem wird die Veränderung des Patientenstamms der Arztpraxis berücksichtigt. *LPAT98* bezeichnet die logarithmierte Patientenzahl des Vorjahres. *P\_DIFF* gibt hingegen die Veränderung des Patientenstammes über die Zeit an (Anzahl Patienten 1998 – Anzahl Patienten 1997 in 1000 Patienten). Die Differenz ist positiv, wenn der Arzt seinen Patientenstamm vergrößern konnte und negativ, wenn er im letzten Jahr Patienten verloren hat. Schließlich berücksichtigen wir auch die Zusammensetzung des Patientenstamms hinsichtlich des Anteils von MC-Versicherten in der Praxis. *PAT\_ANT* ist der Quotient aus der Anzahl MC-Patienten und der Anzahl behandelte konventioneller Versicherter. Die so konstruierte Variable bezieht sich nur auf die betrachtete Stichprobe, da keine Angaben über die Gesamtzahl von MC-Versicherten der Ärzte verfügbar sind.

Zwei weitere Variablen dienen der Charakterisierung des Patientenstamms auf Praxisebene: zum einen das Durchschnittsalter der Patienten einer Arztpraxis (*ALT\_PR*) und zum anderen der Anteil der Männer am gesamten Patientenstamm einer Praxis (*MANN\_PR*).

Wie an früherer Stelle bereits erwähnt, verfügt die Untersuchung über keine Diagnoseinformation. Immerhin kann aber versucht werden, den Arzt-Patienten-Kontakt so genau wie möglich mit den vorhandenen Daten zu beschreiben. Die Anzahl an Leistungsnummern, die ein Arzt für einen Patienten verrechnet, wird daher genauso berücksichtigt (*LST\_NR*) wie die Konsultationen der Patienten bei einem Arzt (logarithmierte Konsultationen *LKONS*). Wobei eine Leistungsnummer alle Leistungen innerhalb einer abgeschlossenen Behandlungszeit umfasst, die von einem Leistungserbringer erbracht bzw. veranlasst wurden.<sup>228</sup> Diese Variable stellt somit einen Zusammenhang zwischen Patientenstatus und Behandlungsintensität her. Außerdem ist auch bekannt, wie regelmäßig der Patient bei einem Arzt war. Ist der Patient jeden Monat mindestens einmal beim Arzt gewesen, bekommt die Variable (*REGEL*) den Wert 12. Ihren kleinsten Wert mit 1 nimmt die Variable demnach dann an, wenn der Patient nur in einem einzigen Monat bei dem betreffenden Arzt war. Aus der Betrachtung der Häufigkeit, der Intensität und der Regelmäßigkeit des Arzt-Patienten-Kontaktes können Rückschlüsse über die Krankheitsschwere gezogen werden.

Zwei weitere Variablen runden den casemix ab. Die Variable *M\_V* gibt Aufschluss über die Möglichkeit des Arztes, die Patienten direkt mit Medikamenten zu versorgen. Die Selbstdispensation kann zwar selber nicht direkt gemessen werden, dafür gibt aber *M\_V* das Verhältnis zwischen selbstabgegebenen Medikamenten und Medikamenten von der Apotheke an. Ein höherer Wert dieser Variable korrespondiert dabei mit einem höheren Anteil der direkt verkauften Medikamenten durch den Arzt. Die Variable *L98* schließlich verknüpft den Vorjahresumsatz des Arztes mit den gesamten Leistungen des Patienten im Vorjahr. Sie soll zusätzlich für eine mögliche Korrelation zwischen Gesundheitszustand des Patienten und dem Einkommen des Arztes kontrollieren.

Die Einbeziehung der Durchschnittskosten der medizinischen Versorgung (*DK*) auf Marktebene – getrennt nach Arzt- und Patientengemeinde (*DK\_A*, *DK\_P*) – dient zusätzlich der allgemeinen Charakterisierung der Morbidität der Bevölkerung. Auf Marktebene werden ebenfalls Arztdichte und Bevölkerungsdichte sowie der Herfindahl-Index zur Charakterisierung der Wettbewerbssituation in den Gemeinden herangezogen.

---

<sup>228</sup> Genauer gesagt, fasst eine Leistungsnummer alle Leistungen eines veranlassenden Leistungserbringers bei einem Patienten zusammen, deren Beginn in einem Monat lag und deren Ende in einem Monat lag. Überweisungen zu unterschiedlichen Ärzten in oben genannter Periode werden zusammengezählt.

### III Empirische Untersuchung

Tabelle 20 Gesamtmodell: Deskriptive Statistik erklärender Variablen

N=100523	Mittelwert	Std. Fehler	Min	Max		Mittelwert	Std. Fehler	Min	Max
<b>Patienten</b>					<b>Ärzte</b>				
LND	3.101	4.092	0	12.435	ROENT	0.736	0.441	0	1
KA	0.020	0.139	0	1	ZH	0.676	0.468	0	1
FR	1.415	0.829	0	5	PA100	0.604	0.489	0	1
FR98	2.604	5.389	0	44.866	PA105	0.283	0.450	0	1
RISK	0.746	0.436	0	1	GEM	0.234	0.423	0	1
SEX	0.399	0.490	0	1	GEM_ANZ	0.568	1.090	0	6
ALTER	41.574	23.988	2	103	TITEL_D	0.918	0.274	0	1
ALTER2	2303.830	2116.750	4	10609	BERUFSER	2.452	0.582	0.693	3.091
KIND	0.217	0.412	0	1	MANN	0.791	0.406	0	1
JUGEND	0.069	0.254	0	1	FRAU	0.166	0.372	0	1
SPT	0.117	0.321	0	1	LALTPR	3.627	0.513	1.730	4.490
LLST97	6.400	2.516	0	12.097	SEX_PR	0.394	0.101	0	1
LLST98	6.578	2.422	0	12.096	L_ARZT98	6.157	0.412	4.317	8.599
ALT97	281.318	217.100	0	1014.060	VERM	12.939	0.727	10.017	14.820
ALT98	289.398	218.568	0	1030.740	VERMB	31.852	8.023	7.477	43.888
LPR	2.534	3.302	0	9.320	LPAT98	6.819	0.677	3.258	8.822
UNT_VRS	1.667	1.013	1	20	P_DIFF	0.720	0.527	-0.741	5.714
UNT_V2	3.803	9.750	1	400	PAT_ANT	0.246	1.649	0	17.204
<b>Case</b>					<b>Gemeinde</b>				
M_V	209.144	1857.340	0	31394.2	HERF	0.128	0.202	0.005	1
L98	40.563	15.408	0	97.231	LDK_P	5.092	0.194	4.585	5.400
LNK	1.183	0.791	0	4.898	LDK_A	5.108	0.193	4.652	5.400
LNNR	0.847	0.780	0	3.829	ARZTD	14.417	5.642	2.460	29.676
K_NR	1.401	1.798	0	15.139	BEVD	0.037	0.036	6.6E-04	0.108
REGEL	2.433	1.915	1	12	ST_P_K	1154	765	364	6980
<b>MC-Variablen</b>									
MC_A	0.091	0.288	0	1					
MC_P	0.040	0.196	0	1					
EFF	0.884	0.067	0.250	0.980					
LAMBDA1	-0.009	0.379	-1.290	4.030					

Viele der in Tabelle 20 dargestellten erklärenden Variablen sind abgeleitete Variablen, die inhaltlich interessante Aspekte des Arztverhaltens abbilden oder besondere Kontrollfunktionen in den Schätzungen wahrnehmen. Das ohnehin umfangreiche Variablen-set in Verbindung mit diesen abgeleiteten Variablen sowie der damit verbundenen hohen Korrelation einzelner Variablen untereinander könnte die Güte der Schätzungen beeinflussen. Daher ist es notwendig, bei allen Schätzungen die Stärke der Multikollinearität zu messen und gegebenenfalls, die eine oder andere Variable auszuschließen.

Das Hauptinteresse bei den erklärenden Variablen liegt auf den beiden Variablen *MC\_A* und *MC\_P*. Die Variable *MC\_A* zeigt an, ob der Patient von einem Arzt behandelt wurde, der zu

einem der untersuchten Hausarztmodelle bzw. zur HMO in Zürich gehört, und *MC\_P* identifiziert Mitglieder dieser Hausarztmodelle oder der untersuchten HMO.

Entsprechend der Herleitung des empirischen Modells wird für eine mögliche Arztselektion durch die Variable *EFF* kontrolliert. Der geschätzte Effizienzwert aus Abschnitt 9.2 liegt nahe 1, wenn der Arzt im Vergleich zu anderen Ärzten Leistungen mit einem geringeren Ressourcenverbrauch anbietet. Versteht man unter Selektion der Ärzte, dass es gerade die Ärzte mit MC-Beteiligung sind, die aufgrund ihres Praxismanagements Arztdienstleistungen kostengünstiger produzieren können, so kann die Effizienz-Variable ein guter Indikator für die Selektion der Ärzte sein. M.a.W. unterstellen wir eine positive Korrelation zwischen MC-Beteiligung der Ärzte und ihrer Effizienz.

Um für eine mögliche Selektion der Patienten zu kontrollieren, greifen wir auf die Ergebnisse der Nachfrageschätzung im Abschnitt 9.3 zurück. Dort wurde im ersten Schritt die Versicherungswahl der Versicherten erklärt. Diese Wahl kann über die geschätzte Mills-Ratio als zusätzlicher Regressor in den Schätzungen berücksichtigt werden (*LAMBDA1*). Die Mills-Ratio ist dabei das Verhältnis aus Dichte- und Verteilungsfunktion der geschätzten Wahl. Durch die Einbeziehung dieser Variablen kann der Endogenität der MC-Wahl durch die Patienten begegnet werden (vgl. hierzu die Ausführungen in Abschnitt 8.2.3 und wiederum die Ausführungen zur Nachfrageschätzung in Abschnitt 9.3). Alle Variablenbeschreibungen sind in einer Tabelle im Anhang nochmals zusammengefasst (vgl. Anhang III).

Während in Tabelle 20 die Statistiken der erklärenden Variablen dargestellt sind, werden in Tabelle 21 die wichtigsten Outputgrößen charakterisiert. Entsprechend der theoretischen Ableitung des empirischen Modells werden insgesamt vier Modelle geschätzt – zwei, die sich mit Nutzenkategorien beschäftigen und zwei, die direkt mit dem Einkommen bzw. dem Gewinn der Arztpraxis verbunden sind. Zu der letztgenannten Kategorie gehören der direkte Umsatz des Arztes, welcher das Bruttoeinkommen aus ärztlicher Tätigkeit approximiert (*UMSATZ*) sowie die veranlassten Kosten, die kein direktes Einkommen für die Ärzte bedeuten (*VERANL*). Dabei lassen sich die direkten wie auch die veranlassten Kosten weiter in ihre Bestandteile zerlegen. So beinhaltet der Umsatz der Ärzte mit Selbstdispensation auch den Verkauf von Medikamenten, während der Umsatz der Ärzte ohne Medikamentenverkauf nur die Leistungen der ärztlichen Behandlungen umfassen. Bei den veranlassten Leistungen lassen sich die Kosten für Medikamente (*MED\_APO*), für Physiotherapie (*PHYS*) und Laborleistungen (*LAB*)

gesondert betrachten. Diese Aufteilung erfolgt hier der Vollständigkeit halber. In der empirischen Untersuchung wird nur das direkte Einkommen (*UMSATZ*) und der gesamte Umsatz der Arztpraxis (*UMSATZ\_GES*) betrachtet. Der mögliche Einfluss der Medikamentenabgabe durch den Arzt wird durch die oben beschriebene Variable *M\_V* erfasst. Als Nutzenkategorien werden die Konsultationen und die gesamten Überweisungen betrachtet, die sich aus den Überweisungen zu anderen Ärzten und den externen Laborleistungen zusammensetzen.

Tabelle 21 Gesamtmodell: Deskriptive Statistik der Leistungen

<b>Konventionelle Patienten</b>				
<b>N=96495</b>	<b>Mittelwert</b>	<b>Std.Fehler</b>	<b>Min</b>	<b>Max</b>
UMSATZ_GES	836.38	1,411.64	0.5	76805.1
UMSATZ	497.27	793.08	0	72796.6
KONSU_IN	3.64	4.85	0	133
ÜBERWEIS	0.28	0.45	0	1
LAB_EXT	39.54	135.21	0	5721
MED_ARZT	117.02	426.38	0	28307.9
M_APO	224.85	778.62	0	31572.6
PHYS	43.66	238.85	0	11516.3
<b>HAM- und HMO-Patienten</b>				
<b>N=4028</b>	<b>Mittelwert</b>	<b>Std.Fehler</b>	<b>Min</b>	<b>Max</b>
UMSATZ_GES	1,182.52	1,602.56	0.3	26592.6
UMSATZ	285.40	474.02	0	10490
KONSU_IN	2.57	4.04	0	73
ÜBERWEIS	0.76	0.43	0	1
LAB_EXT	79.17	193.08	0	2722
MED_ARZT	59.69	203.13	0	3003.6
MED_APO	297.75	718.45	0	17963.4
PHYS	44.84	204.39	0	5383.2

Alle Angaben in Tabelle 21 sind Werte pro Patient. Man erkennt deutliche Unterschiede in den einzelnen Kategorien. So beträgt der durchschnittliche Umsatz der hier betrachteten Ärzte für konventionelle Patienten knapp 500 SFr. pro Patient im Jahr und für MC-Patienten 285 SFr. Demgegenüber ist der durchschnittliche Gesamtumsatz der Ärzte für MC-Patienten wesentlich höher als für konventionell versicherte Patienten. Der Unterschied hängt direkt mit der Gatekeeperfunktion des Allgemeinarztes im MC-Bereich zusammen. Interessant ist in dem Zusammenhang auch das Überweisungsverhalten: Während es in 76% der Fälle bei MC-Patienten zu einer Überweisung zu einem anderen Arzt oder zu einem externen Labor kommt, beträgt der Wert im konventionellen Bereich nur 28%. Die Konsultationen liegen mit durchschnittlich 2.57 pro Patient um mehr als einer Konsultation unter dem Durchschnittswert für konventionelle Patienten (3.64).

Aus den Daten geht hervor, dass für einzelne Patienten die eine oder andere Position nicht verrechnet wurde. Bspw. gab es Patienten, für die keine direkten Arztleistungen abgerechnet worden sind, sondern nur Medikamentenkosten anfielen. Auch kann es Arzt-Patienten-Kontakte gegeben haben, bei denen nur ein Ratschlag des Arztes ohne weitere Leistungsanspruchnahme erfolgte. Außerdem kann es während eines Arzt-Patienten-Kontaktes zwar zu einer Leistungsanspruchnahme gekommen sein, allerdings ohne dass eine Konsultation abgerechnet wurde.

#### ***9.4.4 Überweisungen und weitere veranlasste Leistungen***

Ob ein Arzt einen Patienten zu einem anderen Arzt überweist, hängt von der Art und der Schwere der Erkrankung ab. Die Überweisung eines Patienten deutet auf eine Erkrankung des Patienten hin, die von ihrer Art entweder über das medizinische Fachwissen oder die technischen Möglichkeiten des behandelnden Arztes hinausgehen. Geht man davon aus, dass HAM- bzw. HMO-Patienten im Durchschnitt einen besseren Gesundheitszustand aufweisen, als die Versicherten im konventionellen Sektor, so kann erwartet werden, dass Patienten in MCOs weniger häufig zu anderen Ärzten überwiesen werden. Andererseits muss aber berücksichtigt werden, dass Patienten im konventionellen Sektor die Möglichkeit besitzen, auch ohne die Konsultation eines Allgemeinarztes einen Spezialisten aufzusuchen. Auf der Arztseite kommt hinzu, dass die Ärzte in HAM oder HMOs den gesamten Ressourcenverbrauch der Versorgung im Auge haben. Daher kann es besser sein, einen Patienten rechtzeitig zu einem Spezialisten zu überweisen, um so die Gesamtkosten der Behandlung zu begrenzen. Natürlich geht eine solche Strategie auch immer zu Lasten des Arzteinkommens, da sich jede Überweisung als potenzielle Einkommenseinbuße auffassen lässt. In diesem Sinne bedeutet also eine höhere Überweisungsquote eine stärkere Sachwalterrolle des Arztes für die Patienten.

Außer den Überweisungen zu anderen ambulant tätigen Ärzten werden auch die veranlassten Labortests in diesem Zusammenhang betrachtet. Denn diese Leistungen führen ebenfalls nicht zu einem direkten Einkommen für die Ärzte. Allerdings kann sich in der Häufigkeit von externen Labortests einfach die Tatsache widerspiegeln, dass der betreffende Arzt über kein eigenes Labor verfügt und deshalb externe Labors beauftragen muss.

Wir betrachten für die Untersuchung des Überweisungsverhaltens der Ärzte die Nutzenfunktion (9.27). Ferner unterstellen wir, dass der Arzt genau dann einen Patienten überweisen wird, wenn der Nutzen aus einer Überweisung für den Arzt höher ist, als wenn er den Patienten nicht

überwiesen hätte:  $U_A(\text{Überweisung}) > U_A(\text{keine Überweisung})$ . Damit ist die Differenz der beiden Nutzen bei einer Überweisung des Patienten positiv. Bezeichnen wir die Differenz mit  $y^*$ , so lässt sich die Entscheidung des Arztes  $i$  bei der Behandlung von Patient  $j$  in einer bestimmten Gemeinde  $k$  in Abhängigkeit von erklärenden Variablen, die sowohl wichtige ärztliche Determinanten als auch Patienteneigenschaften umfassen,<sup>229</sup> folgendermaßen darstellen:

$$\begin{aligned}
 y_{ijk}^* &= \alpha_0 + \alpha_1' A_i + \alpha_2' P_j + \alpha_3' Gem_k \\
 &\quad + \alpha_4 MC_{-} A_i + \alpha_5 MC_{-} P_j + \alpha_6 Eff_i + \alpha_7 \lambda_{1j} + v_i + \varepsilon_{ijk} \\
 y_{ijk} &= 1 \quad \text{falls} \quad y_{ijk}^* > 0 \\
 y_{ijk} &= 0 \quad \text{falls} \quad y_{ijk}^* \leq 0.
 \end{aligned}
 \tag{9.30}$$

Der Nutzen einer Überweisung ist nicht beobachtbar, vielmehr kann nur die Entscheidung des Arztes beobachtet werden, ob der Arzt den Patienten überwiesen hat oder nicht ( $y$ ). Somit haben wir es mit einem Schätzproblem zu tun, bei dem die abhängige Variable binär ist und die geeignete Methode der Schätzung daher ein Probit- oder Logitmodell darstellt.

Die Auswahl der erklärenden Variablen erfolgte sowohl unter Berücksichtigung inhaltlicher Aspekte als auch unter Verwendung statistischer Tests. Die Tests beziehen sich dabei nicht nur auf den Erklärungsgehalt möglicher zusätzlicher Variablen, sondern auch auf eine mögliche perfekte Korrelation von erklärenden Variablen mit einem der beiden Modellausgänge. Für die Erklärung des Überweisungsverhaltens sind aus den theoretischen Überlegungen vor allem diejenigen Variablen wichtig, die den Gesundheitszustand der Patienten möglichst gut abbilden, die Arztbesonderheiten beschreiben und für bestimmte Markteffekte kontrollieren können.

Wie bei der Vorstellung des allgemeinen Modells bereits angemerkt, müssen wir bei der Schätzung die hierarchische Datenstruktur beachten. Im Allgemeinen kann das durch ein *random effects model* geschehen (vgl. Abschnitt 8.2.4). Allerdings übersteigt der hier verwendete Datenumfang die Möglichkeiten des bei den Untersuchungen verwendeten Programmpakets für diese Modellklasse (Probit-Modell).<sup>230</sup> Aus diesem Grund wurde ein Verfahren angewendet, welches nicht direkt die random effects berücksichtigt, sondern versucht, die gruppenspezifische Varianzen durch eine geeignete Gewichtung einzufangen. Dazu wird eine Gewichtungsvariable konstruiert, die sowohl die Besonderheiten des Arztes, als auch die Korrelation zwischen den Patienten einer Arztpraxis berücksichtigen soll. Es wird unterstellt, dass das Produkt aus dem

---

<sup>229</sup> Vgl. zu den erklärenden Variablen sowohl die Diskussion in Abschnitt 9.4.1 als auch die Beschreibung in 9.4.3.

Umsatz des Arztes in der Vorperiode und der gesamten Leistungsanspruchnahme des Patienten in der Vorperiode diesen Kriterien entspricht.<sup>231</sup>

#### **Ergebnisse**

Patienten in HAM bzw. in der HMO weisen eine viel größere Überweisungswahrscheinlichkeit auf als Patienten im konventionellen Bereich (vgl. Tabelle 22).<sup>232</sup> Offenbar macht es dabei einen Unterschied, ob der behandelnde Arzt sich an einer MCO beteiligt oder nicht: MC-Ärzte besitzen im Durchschnitt eine geringere Überweisungswahrscheinlichkeit als ihre Kollegen im traditionellen Bereich. Dieser Effekt bleibt auch dann erhalten, wenn wir ausschließlich konventionell versicherte Patienten betrachten. Denn eine erneute Schätzung mit einer Stichprobe, in der nur konventionelle Patienten einbezogen werden, zeigt ebenfalls einen signifikant negativen Einfluss der MC-Beteiligung des Arztes.

Für mögliche Selektionseffekte wird durch die Einbeziehung der beiden Variablen *EFF* und *LAMBDA1* kontrolliert. Der signifikant positive Einfluss von *EFF* bedeutet, dass kostengünstig produzierende Ärzte eine höhere Wahrscheinlichkeit einer Überweisung aufweisen als Ärzte mit einem niedrigeren Effizienzwert. Da diese Variable positiv mit der MC-Beteiligung der Ärzte verbunden ist, kann sie die Selektion der Ärzte in MC-Verträgen effektiv kontrollieren.

Bei den Patienten scheint die Selektion für das Überweisungsverhalten der Ärzte dagegen keine Rolle zu spielen: Der Korrekturfaktor *LAMBDA1* ist nicht signifikant von Null verschieden. Dieses Ergebnis kann dahin gehend interpretiert werden, dass die Wahl des Versicherungsvertrags durch die Versicherten keinen Einfluss auf das Überweisungsverhalten der Ärzte hat. Vielmehr bedeutet der signifikant positive Koeffizient von *MC\_P*, dass MC-Patienten aufgrund der Gatekeeper-Funktion des Arztes bzw. der eingeschränkten Arztwahl im MC-Bereich eine höhere Wahrscheinlichkeit einer Überweisung besitzen als konventionelle Patienten.

---

<sup>230</sup> Es wurde für alle Schätzungen das Programmpaket Limdep 7.0 verwendet.

<sup>231</sup> Der dargestellte Ansatz wird als Weighted Least Square (WLS) bezeichnet und wird häufig bei Problemen der Heteroskedastizität eingesetzt (vgl. bspw. Gujarati 2000).

<sup>232</sup> Dieses Ergebnis wird bspw. auch durch eine Studie von Iversen und Luras (2000) gestützt, die für norwegische Allgemeinmediziner herausfanden, dass bei einem Vergütungswechsel von einer eher einzelleistungsbasierten Vergütung zu einem System mit Kopfpauschalen die Überweisungen signifikant angestiegen sind.



Tabelle 22 Gesamtmodell: Wahrscheinlichkeit einer Überweisung

N = 100,522	Koeffizient	Std.Fehler		Koeffizient	Std.Fehler
<b>Patienten</b>			<b>Ärzte</b>		
ONE	-1.236 **	0.473	ZH	-0.269 **	0.025
LND	1.37E-03	1.22E-03	NEUTRAL	-0.363 **	0.026
ALTER	9.73E-03 **	4.30E-04	BERUFSE	-0.758 **	0.149
KIND	-0.312 **	0.018	GEM_ANZ	-0.018 **	0.005
JUGEND	-0.063 **	0.018	L_ARZT98	-0.281 **	0.061
FR	0.022 **	0.005	VERM	0.377 **	0.063
FR98	-2.19E-03	1.30E-03	VERMB	0.055 **	0.012
LLST98	0.024 **	0.003	LPAT98	-0.559 **	0.060
ALT97	7.23E-05	4.00E-05	P_DIF	0.059 **	0.012
ALT98	-5.02E-04 **	7.17E-05			
<b>MC- und Kontroll-Variablen</b>			<b>Gemeinde</b>		
MC_ARZT	-0.136 **	0.021	HERF	0.020	0.027
MC_PAT	1.213 **	0.049	LGEM_P	0.126 *	0.060
EFF	0.582 **	0.079	ARZTD	-2.95E-03 *	1.26E-03
LAMBDA1	7.91E-03	2.20E-02	ST_P_K	-2.56E-06	6.79E-06
			BEVD	0.002 **	2.71E-04
<b>Pseudo R<sup>2</sup></b>	0.426				

\*)  $p < 0.05$ , \*\*)  $p < 0.01$

Die Fähigkeit des Modells einer richtigen Vorhersage von Überweisungen ist relativ schwach. Insgesamt kommt es in 30,313 Fällen zu einer Überweisung bzw. zu einer externen Laboruntersuchung. Davon erkennt das Modell 4603 Fälle, also ungefähr 15% der Überweisungen. Obwohl das  $R^2$  relativ hoch ist (42.6%)<sup>233</sup>, besitzt das Modell gerade für die uns interessierende Frage der richtigen Vorhersage von Überweisungen keine sehr hohe Erklärungskraft. Dies ist allerdings nicht weiter verwunderlich, da weder die Diagnosen noch andere Informationen über die konkrete Erkrankung des Patienten bekannt sind. Vielmehr müssen wir auf relativ allgemeine Information über den Gesundheitszustand des Patienten zurückgreifen, um Überweisungen zu erklären, die in hohem Maße von der konkreten Diagnose abhängen.

Trotzdem sind die meisten der berücksichtigten Faktoren signifikant. Bei den Eigenschaften der Patienten fällt vor allem der signifikante Einfluss der Vorleistungen (*LLST98*) im Zusammenhang mit dem Alter ins Auge. So steigt mit einer höheren Vorleistung des Patienten die Wahrscheinlichkeit, dass der Patient überwiesen wird. Dieser Zusammenhang ist aber nicht für alle Altersklassen gleich. Vielmehr sinkt mit steigendem Alter und gleichzeitig hohen Vorleistungen der Patienten die Wahrscheinlichkeit einer Überweisung im Verhältnis zu jüngeren Patienten mit

<sup>233</sup> Wir verwenden wieder die Formulierung des Pseudo- $R^2$  nach McKelvey und Zavoina (1975). Vgl. LIMDEP (1998).

ähnlich hohen Vorleistungen (negativer Koeffizient von *ALT98*). Bei der Interpretation von Koeffizienten derartiger Interaktionsvariablen ist Vorsicht geboten, da das Vorzeichen des Effekts nicht allein durch das Vorzeichen des Koeffizienten bestimmt wird. Aufgrund der nichtlinearen Spezifikation des Modells kann es durchaus sein, dass der Effekt für bestimmte Kombinationen der erklärenden Variablen gerade das entgegengesetzte Vorzeichen aufweist. Auf den vorliegenden Fall trifft dies allerdings nicht zu.<sup>234</sup> Nicht signifikant ist hingegen *LND*. Demnach spielt die Entfernung des Wohnorts der Patienten von der Arztpraxis beim Überweisungsverhalten der Ärzte keine Rolle.

Alle zur Erklärung des Überweisungsverhaltens herangezogenen Arztvariablen sind signifikant. Interessant ist im Lichte der Theorie vor allem der Umstand, dass mit einem höheren langfristigen Einkommen (*VERM*) die Wahrscheinlichkeit einer Überweisung steigt. Eine Erklärung hierfür kann die Einkommenswirkung von Überweisungen liefern. Es wurde vermutet, dass eine hohe Überweisungsrate zu einem geringeren Einkommen der Ärzte führt. Deshalb lässt sich der beobachtete positive Effekt dahingehend interpretieren, dass Ärzte mit einem hohen langfristigen Einkommen sich immer weniger in ihrem Verhalten nach dem Einkommensmotiv ausrichten. Diese Interpretation wird dadurch gestützt, dass die Wirkung eines hohen Vermögens im Zusammenhang mit einer großen Berufserfahrung (*VERMB*) ebenfalls einen positiven Einfluss auf die Wahrscheinlichkeit einer Überweisung hat. Dabei ist zu beachten, dass die Berufserfahrung (*BERUFSE*) an sich die Überweisungswahrscheinlichkeit senkt, sodass bei einem gegebenen durchschnittlichen Einkommen mit zunehmender Berufserfahrung der Effekt negativ ist.<sup>235</sup> Erst bei sehr hohen langfristigen Umsätzen kehrt sich dieser Zusammenhang um und die Überweisungswahrscheinlichkeit steigt. Hingegen führt ein hoher Umsatz pro Patient im Vorjahr (*L\_ARZT98*) zu einer Verringerung der Überweisungswahrscheinlichkeit, was den Zusammenhang zwischen Einkommensmotiv und Überweisungsverhalten nochmals verdeutlicht: Hatte der Arzt im Vorjahr einen hohen pro-Kopf-Umsatz, so hatte er entsprechend dieser Interpretation weniger Überweisungen als andere Ärzte mit einem niedrigen pro-Kopf-Umsatz.

---

<sup>234</sup> Ai und Norton (2003) haben jüngst gezeigt, dass der marginale Effekt einer Interaktionsvariablen nicht nur von der Dichtefunktion – wie bei normalen Effekten in Probitmodellen –, sondern auch noch von der ersten Ableitung der Dichtefunktion abhängt.

<sup>235</sup> Auch bei diesem Effekt ist die oben erwähnte Schwierigkeit der korrekten Interpretation des marginalen Effektes zu berücksichtigen. Die Abhängigkeit des Vorzeichens von der Dichtefunktion und deren erster Ableitung kann hier zu einem Vorzeichenwechsel führen. D.h. das Vorzeichen des marginalen Effektes einer Änderung des Vermögens auf die Überweisungswahrscheinlichkeit bei gegebener Berufserfahrung hängt von den konkreten Werten aller erklärenden Variablen ab.

Geht man von relativ geringen Schwankungen im Arztverhalten aus, so setzt sich sein Überweisungsverhalten auch in der Untersuchungsperiode fort.

Schließlich ist bezüglich des Einflusses der Arztvariablen festzuhalten, dass eine hohe Patientenzahl im Vorjahr (*LPAT98*) zu einer Verringerung der Überweisungswahrscheinlichkeit in der Untersuchungsperiode führt, ein positives Wachstum der Patientenzahlen (*P\_DIFF*) diese Wahrscheinlichkeit dagegen erhöht. Damit wird die Vermutung bestätigt, dass der Patientenstamm bzw. die Veränderung des Patientenstamms einen Einfluss auf das Arztverhalten ausüben.

Hinsichtlich des Markteinflusses auf die Überweisungswahrscheinlichkeit fällt zunächst auf, dass von fünf einbezogenen Variablen drei signifikant von Null verschieden sind. Die Wettbewerbssituation des einzelnen Arztes in der Gemeinde (gemessen über den Herfindahl-Index *HERF*) und der Steuerertrag pro Kopf (*ST\_P\_K*) haben keinen signifikanten Einfluss auf die Überweisungswahrscheinlichkeit.<sup>236</sup> Dagegen ist in Gemeinden mit einer hohen Arztdichte (*ARZTD*) die Überweisungswahrscheinlichkeit geringer als in Gemeinden mit einer niedrigen Arztdichte, während eine hohe Bevölkerungsdichte (*BEVD*) einen signifikant positiven Einfluss auf die Überweisungswahrscheinlichkeit ausübt. Dass die Arztdichte einen signifikant negativen Einfluss auf die Überweisungswahrscheinlichkeit hat, kann Ausdruck eines angebotsinduzierenden Verhaltens der Ärzte sein. Denn offensichtlich bewerten die Ärzte den Nutzen einer Überweisung in Abhängigkeit der Arztdichte: Ist die Konkurrenz groß, so sinkt die Überweisungswahrscheinlichkeit, weil die Ärzte das durch steigende Konkurrenz gesunkene Einkommen durch die Erhöhung eigener Behandlungen – und damit durch weniger Überweisungen – kompensieren wollen. Außerdem ist die Überweisungswahrscheinlichkeit für diejenigen Patienten höher, die aus Gemeinden mit hohen Durchschnittskosten kommen (*LDK\_P*).

#### **Zusammenfassung der wichtigsten Ergebnisse**

1. Ärzte, die sich an MC-Modellen beteiligen, weisen eine geringere Überweisungswahrscheinlichkeit auf, als Ärzte, die ausschließlich Patienten aus dem konventionellen Bereich versorgen. Dies trifft auch dann zu, wenn man ausschließlich konventionelle Patienten betrachtet.

---

<sup>236</sup> Allerdings kann die Hypothese, dass alle Markteinflüsse gemeinsam nicht von Null verschieden sind auf dem 1%-Signifikanzniveau verworfen werden.

2. Patienten einer MC-Organisation besitzen eine höhere Überweisungswahrscheinlichkeit als konventionelle Patienten. Darin äußert sich einerseits die Gatekeeperfunktion der MC-Ärzte als auch der Umstand, dass Patienten des konventionellen Bereichs direkt einen anderen Arzt aufsuchen können.

#### **9.4.5 Konsultationen**

In einem zweiten Schritt werden die angebotenen Konsultationen (einschließlich der Hausbesuche) untersucht. Entsprechend den theoretischen Resultaten erwartet man, dass Versicherte in HAM oder HMOs weniger Konsultationen erhalten als in der konventionellen Versorgung. Für eine geringere Inanspruchnahme kommt das ärztliche Bemühen um Ressourceneinsparung in Frage, es kann aber auch durch die Patienteneigenschaften oder das Verhalten der Patienten in den HAM bzw. HMOs hervorgerufen werden.<sup>237</sup> So können sehr viele Konsultationen pro Patient Ausdruck von sehr vielen chronischen Patienten sein, da chronisch Kranke vergleichsweise viele Konsultationen im Jahr benötigen.<sup>238</sup> Da nicht von einer vollständigen Homogenität der Patientenstämme der Ärzte ausgegangen werden kann, wird für die mögliche Heterogenität auf Praxisebene kontrolliert.

Um die Analyse zu vereinfachen, untersuchen wir die logarithmierten Arztbesuche, für die wir Normalverteilung unterstellen.<sup>239</sup> Allerdings gilt das nur für diejenigen Arzt-Patienten-Kontakte, bei denen es zu mindestens einer Konsultation des Arztes durch den Patienten gekommen ist. In ca. 15,000 Fällen – d.h. 15,000 Arzt-Patienten-Kontakten in einem Jahr – kam es zu keiner Konsultation. Bei diesen Fällen ist davon auszugehen, dass der Arzt entweder nur Medikamente verschrieben oder Überweisungen ausgefüllt hat. Darüber hinaus gibt es auch Leistungspositionen, für die keine Extra-Konsultation abgerechnet werden kann (vgl. hierzu auch Abschnitt 7.1). In jedem Fall kann davon ausgegangen werden, dass es sich bei einer Großzahl dieser Fälle um Bagatelleerkrankungen oder Routineuntersuchungen gehandelt haben muss. Diese Sicht der Dinge wird auch durch Tabelle 23 bestätigt: Man erkennt deutlich, dass die gesamten Leistungen, die ein Arzt pro Patient veranlasst, wie auch der direkte Umsatz des Arztes für Arzt-Patienten-Kontakte ohne eine Konsultation deutlich geringer sind als für diejenigen Arzt-Patienten-Kontakte mit Konsultation.

---

<sup>237</sup> Vgl. auch Abschnitt 9.4.1.

<sup>238</sup> Vgl. Marty (1997).

### III Empirische Untersuchung

Tabelle 23 Gesamtmodell: Leistungen pro Patient und Arzt – ohne und mit mindestens 1 Konsultation

<b>Keine Konsultation</b>				
N=15199	Mittelwert	Std.Fehler	Minimum	Maximum
UMSATZ_GES	332.419	666.365	0	18111.1
UMSATZ	99.776	247.493	0	7492.3
<b>Mindestens eine Konsultation</b>				
N=85323	Mittelwert	Std.Fehler	Minimum	Maximum
UMSATZ_GES	942.412	1498.13	3.15	76805.1
UMSATZ	558.039	825.352	3.15	72796.6

Es ist daher anzunehmen, dass sich Leistungsfälle mit Konsultationen von denen ohne Konsultationen fundamental unterscheiden. Diesen Unterschied gilt es in der Schätzung zu berücksichtigen. Wir verwenden dafür wieder einen Selektionsmechanismus: Auf der ersten Stufe wird die Wahrscheinlichkeit betrachtet, dass es während einer Arzt-Patienten-Beziehung zu mindestens einer Konsultation kommt, auf der zweiten Stufe wird dann schließlich nach Erklärungen für die Anzahl der Konsultationen selbst gesucht – gegeben, dass mindestens eine Konsultation stattgefunden hat.

Auch bei den Konsultationen können wir das allgemeine Modell zur Anwendung bringen. Zunächst geht es für den Arzt darum, eine Nutzenabwägung zu treffen, ob er eine Konsultation durchführen sollte oder nicht. Dabei muss er vor allem den Gesundheitszustand seines Patienten beachten, sowie – ganz allgemein – die Präferenzen seines Patienten. Er wird er bei seiner Entscheidung auch berücksichtigen, dass eine Konsultation in der Regel mit einem bestimmten Einkommen für ihn verbunden ist. Bezeichnen wir mit  $u^*$  das Nutzenniveau des Arztes, welches aus der Abwägung der Nutzenniveaus resultiert, wenn er mindestens eine oder keine Konsultation durchführt, so wird er genau dann eine Konsultation durchführen, wenn  $u^*$  positiv ist. Allerdings sind wir nicht in der Lage, diesen Nutzen zu beobachten, vielmehr zeigen uns die Daten, ob der Arzt mindestens eine Konsultation durchgeführt hat oder nicht ( $z = 1$  oder  $z = 0$ ).

<sup>239</sup> Die Konsultationen mit den Ausprägungen 0, 1, 2, ... folgen einer diskreten Verteilung. Eine angemessene Verteilungsannahme wäre daher die Poisson-Verteilung. Aufgrund der sehr großen Fallzahl scheint die Approximation der diskreten durch eine stetige Verteilung gerechtfertigt.

Der Prozess lässt sich wie folgt formal beschreiben:

$$\begin{aligned}
 u_{ijk}^* &= \alpha_0 + \alpha_1' A_i + \alpha_2' P_j + \alpha_3' Gem_k \\
 &\quad + \alpha_4 MC\_A_i + \alpha_5 MC\_P_j + \alpha_6 Eff_i + \alpha_7 \lambda_{1j} + v_i + \varepsilon_{ijk} \\
 z_{ijk} &= 1 \quad \text{falls} \quad u_{ijk}^* > 0 \\
 z_{ijk} &= 0 \quad \text{falls} \quad u_{ijk}^* \leq 0.
 \end{aligned} \tag{9.31}$$

Falls er diesen Abwägungsprozess positiv abgeschlossen hat ( $z = 1$ ) geht es dann um die Erklärung der Anzahl der Konsultationen ( $y$ ) für diejenigen Fälle bei denen es mindestens zu einer Konsultation gekommen ist.:

$$\begin{aligned}
 y_{ijk} &= \beta_0 + \beta_1' A_i + \beta_2' P_j + \beta_3' Gem_k \\
 &\quad + \beta_4 MC\_A_i + \beta_5 MC\_P_j + \beta_6 Eff_i \\
 &\quad + \beta_7 \lambda_{1j} + \beta_8 \lambda_{2j} + v_i + \varepsilon_{ijk} \quad \text{für } z = 1,
 \end{aligned} \tag{9.32}$$

wobei  $\lambda_2$  für die entsprechende Selektionskorrektur bezüglich der Fälle ohne eine Konsultation steht.

Das Modell in (9.31) wird wie das Modell zur Schätzung der Überweisungswahrscheinlichkeiten mit Hilfe eines Probit-Modells geschätzt. Auch bei diesem Probit-Modell wurde die Heterogenität der Ärzte nicht nur durch direkte Einbeziehung von arztrelevanten Variablen, sondern auch wieder durch den beschriebenen Gewichtungsfaktor berücksichtigt. Hingegen wird (9.32) mit einem verallgemeinerten OLS-Ansatz unter Berücksichtigung der *random effects* geschätzt.<sup>240</sup>

## Ergebnisse

### **Wahrscheinlichkeit einer Arztkonsultation**

Zunächst betrachten wir die Wahrscheinlichkeit, dass es für einen Patienten während eines Jahres überhaupt zu einer Konsultation gekommen ist. Der Einfachheit und der Vergleichbarkeit halber wurden für die Erklärung dieser Wahrscheinlichkeit die gleichen Variablen ausgewählt wie für die Erklärung der Überweisungswahrscheinlichkeit.

Das zentrale Ergebnis der Schätzung ist in der linken unteren Ecke von Tabelle 24 dargestellt. Demnach ist die Wahrscheinlichkeit, mindestens eine Konsultation zu erhalten, für Patienten aus dem MC-Bereich geringer als für Patienten im konventionellen Sektor. Zudem ist die

---

<sup>240</sup> D.h. die besondere Zählstruktur wird hier nicht berücksichtigt. Verschiedene Zählstrukturmodelle zur Schätzung der Konsultationen werden von vielen Autoren untersucht, so z.B. auch Winkelmann (2001), der unter anderem auch ein Zählstrukturmodell unter Berücksichtigung der Selektion verwendet.

Wahrscheinlichkeit ebenfalls geringer, wenn der behandelnde Arzt sich an einer der in der Stichprobe betrachteten MC-Organisationen beteiligt. Letzteres Ergebnis ist damit ein erster Beleg für Spill-over-Effekte im Sinne der theoretischen Betrachtung im Abschnitt 6.4: Ärzte, die sowohl MC-Patienten als auch konventionell versicherte Patienten versorgen, reduzieren ihre Anstrengungen auch für konventionell versicherte Patienten, zumindest was die Wahrscheinlichkeit mindestens einer Konsultation betrifft.

Tabelle 24 Gesamtmodell: Wahrscheinlichkeit mindestens einer Konsultation

N = 100522	Koeffizient	Std.Fehler		Koeffizient	Std.Fehler
<b>Patienten</b>			<b>Ärzte</b>		
<b>KONST</b>	6.23 **	0.51	<b>ZH</b>	-1.05E-01 **	2.63E-02
<b>LND</b>	-5.86E-03 **	1.33E-03	<b>NEUTRAL</b>	-0.22 **	0.04
<b>ALTER</b>	8.48E-04	4.79E-04	<b>BERUFSER</b>	-2.01 **	0.16
<b>KIND</b>	0.08 **	0.02	<b>GEM_ANZ</b>	-0.08 **	0.01
<b>JUGEND</b>	-0.12 **	0.02	<b>L_ARZT98</b>	-0.03	0.06
<b>FR</b>	-0.08 **	0.01	<b>VERM</b>	-0.11	0.07
<b>FR98</b>	1.20E-02 **	1.40E-03	<b>VERMB</b>	0.15 **	0.01
<b>LLST98</b>	1.14E-02 **	3.27E-03	<b>LPAT98</b>	-0.22 **	0.06
<b>ALT97</b>	2.11E-04 **	4.59E-05	<b>P_DIF</b>	0.01	0.01
<b>ALT98</b>	-4.66E-04 **	7.59E-05			
<b>MC- und Kontroll-Variablen</b>			<b>Gemeinde</b>		
<b>MC_A</b>	-0.14 **	0.02	<b>HERF</b>	0.02	0.03
<b>MC_P</b>	-0.25 **	0.05	<b>LDK_P</b>	-0.59 **	0.06
<b>EFF</b>	1.24 **	0.08	<b>ARZTD</b>	-2.50E-03	1.34E-03
<b>LAMBDA1</b>	-0.11 **	0.02	<b>ST_P_K</b>	7.81E-05 **	7.86E-06
			<b>BEVD</b>	-1.37E-03 **	2.89E-04
<b>Pseudo R<sup>2</sup></b>	0.316				

\*) p < 0.05, \*\*) p < 0.01

Natürlich kann mit den zur Verfügung stehenden Daten die Wahrscheinlichkeit mindestens einer Konsultation während einer Arzt-Patienten-Beziehung innerhalb eines Jahres nicht vollständig beschrieben werden. Denn auch hier sind die konkreten Diagnosen – und damit die Art der Erkrankung des Patienten – nicht bekannt. Um so bemerkenswerter ist es, dass mit den zur Verfügung stehenden Daten ein R<sup>2</sup> von knapp 0.32 erreicht wird. Allerdings ist die Erklärungskraft des Modells hinsichtlich der richtigen Prognose, dass Patienten keine einzige Konsultation in ihrer Arzt-Patienten-Beziehung haben, wiederum gering. Nur 16 Fälle von 15,199 Arzt-Patienten-Beziehungen, bei denen es zu keiner Konsultation gekommen ist, können richtig erkannt werden. Andererseits werden fast alle Fälle mit mindestens einer Konsultation richtig erkannt.

Im Folgenden werden wieder einige wichtige Ergebnisse der Schätzung kurz beleuchtet: Zunächst fällt bei den Patientenvariablen auf, dass die Entfernung zwischen Patient und Arzt (*LND*) einen signifikant negativen Einfluss auf die Wahrscheinlichkeit mindestens einer Konsultation hat. Dies kann damit erklärt werden, dass Patienten, die eine weite Strecke zum nächsten Arzt zurücklegen müssen, höchstwahrscheinlich aus Regionen kommen, wo es relativ wenig Ärzte gibt. Sind also die Suchkosten – d.h. hier die Entfernungskosten – hoch, so ist die Wahrscheinlichkeit gering, dass es zu einer Konsultation kommt. Dieses Ergebnis wird allerdings durch den signifikant negativen Koeffizienten von *LDK\_P* – den durchschnittlichen Kosten eines Versicherten innerhalb einer Gemeinde – relativiert. Denn die Durchschnittskosten sind gerade in ländlichen Gegenden eher gering. Daher steigt mit abnehmenden Durchschnittskosten die Wahrscheinlichkeit, dass es zu mindestens einer Konsultation kommt. In Verbindung mit den Entfernungskosten bedeutet das Ergebnis, dass die betrachtete Wahrscheinlichkeit auf jeden Fall sinkt, wenn einerseits die Entfernung groß ist und die Durchschnittskosten der Wohngemeinde des Patienten hoch sind. Man beachte in diesem Zusammenhang auch die Möglichkeit einer inversen Beziehung zwischen Entfernung und Suchkosten (vgl. Abschnitt 6.2). Demnach sind die Suchkosten dort am höchsten, wo die Auswahl der Anbieter am größten ist – also die Entfernung zum nächsten Arzt am geringsten.

Die Leistungs- und Altersvariablen sind überwiegend signifikant und weisen in der Regel auch das erwartete Vorzeichen auf. Die Ausnahme bildet das Produkt aus Alter und Leistungen im Vorjahr (*ALT98*). Der Koeffizient dieser Variablen hat ein signifikant negatives Vorzeichen. Damit sinkt offenbar die Wahrscheinlichkeit mindestens einer Konsultation im Jahr, wenn zu einem hohen Alter auch ein schlechter Gesundheitszustand tritt. Wichtig ist, dass damit nicht notwendigerweise auch eine geringere Leistungsanspruchnahme einhergehen muss, vielmehr kann sich im negativen Vorzeichen einfach die Tatsache ausdrücken, dass schwer kranke ältere Menschen oft an ihr Bett gebunden sind und dann durch Pflegedienste betreut werden.<sup>241</sup>

Auch bei den Arztvariablen sind die meisten Koeffizienten signifikant von Null verschieden. Von besonderem Interesse ist hier wieder der Einfluss von Berufserfahrung und Vermögen der Ärzte. Eine größere Berufserfahrung (*BERUFSERF*) der Ärzte übt demnach einen deutlich negativen Effekt auf die Wahrscheinlichkeit positiver Konsultationen aus. Im Zusammenhang

---

<sup>241</sup> Bei der Interpretation des Vorzeichens als marginalen Effekt muss man wieder vorsichtig sein, da es auch von den konkreten Werten der erklärenden Variablen abhängt. Vgl. die Ausführungen zu den Überweisungswahrscheinlichkeiten.



mit dem Vermögen lassen sich zwei Effekte unterscheiden. Der Effekt des Vermögens (*VERM*) für sich betrachtet, ist zwar nicht signifikant von Null verschieden ( $p = 0.11$ ), die verbundene Hypothese, dass dieser Effekt und das Produkt mit der Berufserfahrung (*VERMB*) gemeinsam gleich Null sind, kann aber auf dem 5%-Signifikanzniveau abgelehnt werden. Daher sinkt mit steigendem Vermögen die Wahrscheinlichkeit positiver Konsultationen. Dieses Ergebnis hängt aber im entscheidenden Maße von der Berufserfahrung der Ärzte ab. So hat ein hoher Umsatz in den zurückliegenden Jahren in Verbindung mit zunehmender Berufserfahrung eher einen positiven Effekt auf die Wahrscheinlichkeit positiver Konsultationen, während bei einem durchschnittlichen Vermögen mit steigender Berufserfahrung die Wahrscheinlichkeit abnimmt.

Hinsichtlich des Markteinflusses auf die Konsultationswahrscheinlichkeit kann festgehalten werden, dass sowohl die Wettbewerbsintensität auf dem Markt (*HERF*) als auch die Arztdichte (*ARZTD*) keinen Einfluss ausüben. Interessanterweise ist aber die Konsultationswahrscheinlichkeit in Gemeinden mit einem hohen Steueraufkommen pro Kopf besonders hoch (signifikant positiver Koeffizient von *ST\_P\_K*). Auch bei dieser Schätzung kann wieder die Nullhypothese abgelehnt werden, dass die Marktvariablen zusammen keinen Einfluss auf die Konsultationswahrscheinlichkeit haben.

#### ***Anzahl Arztkonsultationen***

Unter Berücksichtigung der Wahrscheinlichkeit, dass es in der Arzt-Patienten-Beziehung während eines Jahres zu mindestens einer Konsultation gekommen ist, wird in der folgenden Schätzung die Anzahl der Konsultationen selbst erklärt. Die gewählte Modellierung wird durch die Schätzung bestätigt. Die Spezifizierung des Störterms über random effects ist hoch signifikant (die Null-Hypothese, dass es keine derartigen Effekte gibt, kann mit einem LM-Test<sup>242</sup> überprüft werden). Daher sind die Störterme innerhalb einer Praxis miteinander korreliert (*RHO*).

Beide MC-Variablen weisen ein signifikant positives Vorzeichen auf. Damit bieten Ärzte mit MC-Beteiligung ihren Patienten mehr Konsultationen an, als Ärzte, die (in der Stichprobe) ausschließlich konventionell versicherte Patienten versorgen, unabhängig davon, ob sie konventionelle Patienten oder MC-Patienten behandeln. Gleichzeitig werden MC-Versicherten deutlich

---

<sup>242</sup> LM-Test steht für Lagrange Multiplier test und basiert auf der Steigung der log likelihood function (Vgl. Greene 1997, S. 160 ff.

mehr Konsultationen angeboten als konventionellen Patienten. Man beachte allerdings, dass dieses Ergebnis unter Berücksichtigung der Wahrscheinlichkeit, dass ein Patient überhaupt eine Konsultation erhält, erzielt wurde. Damit kann festgehalten werden, dass die ohne Berücksichtigung von Kontrollvariablen dokumentierte geringere Anzahl von Konsultationen für den MC-Bereich (vgl. Tabelle 21) ausschließlich auf die geringere Wahrscheinlichkeit mindestens einer Konsultation zurückzuführen ist. Hierin kann ein wichtiger Selektionsmechanismus gesehen werden. Denn offenbar suchen MC-Patienten erst viel später ihren Arzt auf als Patienten im konventionellen Bereich (geringere Konsultationswahrscheinlichkeit), obwohl sie ein gesundheitliches Problem haben, was sich in einer höheren Behandlungsintensität ausdrückt (höhere Anzahl an Konsultationen unter Berücksichtigung der Konsultationswahrscheinlichkeit).<sup>243</sup>

---

<sup>243</sup> Vgl. zu diesem Zusammenhang Altman et al. (2003).

### III Empirische Untersuchung

Tabelle 25 Gesamtmodell: Regression pos. Konsultationen (random effects)

N = 85,323	Koeffizient	Std.Fehler		Koeffizient	Std.Fehler
<b>Patienten</b>			<b>Ärzte</b>		
<b>KONSTANT</b>	-0.899 **	0.242	<b>ROENTGEN</b>	-0.060 **	0.014
<b>LND</b>	-3.45E-04	6.66E-04	<b>ZH</b>	0.149 **	0.022
<b>KA</b>	-2.15E-04	1.56E-02	<b>PA100</b>	0.020	0.042
<b>FR</b>	0.104 **	0.007	<b>PA105</b>	-0.011	0.045
<b>FR98</b>	-2.01E-02 **	1.13E-03	<b>GEM_ANZ</b>	-7.57E-03	5.31E-03
<b>UNFALL</b>	2.04E-03	9.39E-03	<b>NEUTRAL</b>	0.295 **	0.058
<b>SEX</b>	0.011 **	0.004	<b>TITEL_D</b>	-0.033	0.026
<b>ALTER</b>	2.30E-03 **	7.26E-04	<b>BERUFSER</b>	0.094 **	0.011
<b>ALTER2</b>	1.37E-05 *	6.28E-06	<b>LALTPR</b>	-0.027	0.027
<b>KIND</b>	-0.072 **	0.014	<b>SEX_PR</b>	0.117 **	0.054
<b>JUGEND</b>	0.069 **	0.012	<b>VERM</b>	-0.061 **	0.010
<b>SPT</b>	0.026 **	0.007	<b>P_DIFF</b>	-0.075 **	0.034
<b>LLST97</b>	0.015 **	0.002	<b>PAT_ANT</b>	-0.037 **	0.010
<b>LLST98</b>	0.022 **	0.002			
<b>ALT97</b>	-1.54E-05	4.68E-05			
<b>ALT98</b>	5.01E-04 **	5.38E-05			
<b>LPR</b>	8.55E-03 **	1.40E-03			
<b>UNT_VRS</b>	-0.067 **	0.003			
<b>UNT_V2</b>	2.53E-03 **	3.41E-04			
<b>MC- und Kontrollvariablen</b>			<b>Gemeinde</b>		
<b>MC_A</b>	0.106 **	0.023	<b>HER</b>	-0.094 **	0.036
<b>MC_P</b>	0.499 **	0.037	<b>LDK_P</b>	0.568 **	0.048
<b>EFF</b>	0.258 **	0.106	<b>ST_P_K</b>	-6.50E-05 **	8.77E-06
<b>LAMBDA1</b>	0.059 **	0.016	<b>ARZTD</b>	6.27E-03 **	1.60E-03
<b>LAMBDA2</b>	-2.569 **	0.149	<b>BEVD</b>	2.17E-03 **	4.06E-04
<b>R<sup>2</sup></b>	0.166				
<b>Rho</b>	0.094 **		<b>LM-Test</b>	4085 **	

\*)  $p < 0.05$ , \*\*)  $p < 0.01$

Bei den Patienten-Variablen sind einige, aber nicht alle signifikant von Null verschieden. Wichtig für die Anzahl der Konsultationen ist wieder der Gesundheitszustand des Patienten. Die in Anspruch genommenen Vorleistungen des Patienten (*LLST97* und *LLST98*) sind genauso signifikant positiv wie die Leistungen im Vorjahr in Verbindung mit dem Alter (*ALT98*). Darüber hinaus erhalten Patienten, die im Vorjahr im Spital waren (*SPT*), signifikant mehr Konsultationen als Patienten ohne einen Spitalaufenthalt im Vorjahr. Schließlich fällt auf, dass

im Unterschied zur Bestimmung der Konsultationswahrscheinlichkeit die Entfernung des Patienten zum Arzt (*LND*) für die Konsultationsanzahl offenbar keine Rolle spielt.<sup>244</sup>

Bei den Arztvariablen zeigt sich zunächst, dass Ärzte in Zürich signifikant mehr Konsultationen anbieten als ihre Kollegen in Genf (*ZH*). Dieses Ergebnis steht direkt in Verbindung mit den unterschiedlichen Vergütungsverträgen in den beiden Kantonen: In dem Kanton, in dem die einzelne Konsultation höher bewertet wird, werden auch mehr Konsultationen durch die Ärzte angeboten (vgl. auch die Ausführungen in Abschnitt 7.1). Des Weiteren ist bemerkenswert, dass eine größere Berufserfahrung (*BERUFSERF*) positiv mit der Anzahl Konsultationen korreliert ist. Der durchschnittliche Umsatz der letzten beiden Jahre (*VERM*) wirkt dagegen negativ auf die in der laufenden Periode angebotenen Konsultationen. Außerdem erkennt man, dass ein höherer MC-Patientenanteil in einer Praxis (*PAT-ANT*) die Anzahl der Konsultationen für beide Patientengruppen senkt. Letzteres Ergebnis lässt sich als Bestätigung der Vorhersagen des theoretischen Modells aus Abschnitt 6.4.1 interpretieren. Dort wurde im ersten Fall untersucht, wie sich das Anstrengungsniveau verändert, wenn der MC-Anteil innerhalb der Arztpraxis steigt. Demnach sinkt das Anstrengungsniveau für konventionell Versicherte, während es für MC-Versicherte steigt. Die empirischen Ergebnisse bestätigen in gewisser Hinsicht die Vorhersagen des Modells: Ein höherer MC-Anteil in der Praxis (*PAT-ANT*) senkt die Konsultationen für alle Versicherte, die Anzahl der Konsultationen für MC-Versicherte steigen gleichzeitig (*MC\_P*).

Alle Variablen auf Marktebene sind bei dieser Schätzung signifikant. Insbesondere fallen der negative Koeffizient der Wettbewerbsvariablen (*HERF*) und der positive Koeffizient der Arztdichte (*ARZTD*) auf: Mit steigendem Wettbewerb (kleiner Wert von *HERF*) und hoher Arztdichte (großer Wert von *ARZTD*) steigt die Zahl der Konsultationen. Das Ergebnis kann Ausdruck größerer Qualitätsanstrengungen auf einem Wettbewerbsmarkt sein, weil die Ärzte ohne diese Anstrengungen fürchten müssten, einen Teil ihres Patientenstammes zu verlieren. Es kann aber auch auf angebotsinduzierte Nachfrage hindeuten, wonach die Ärzte versuchen, den ihnen verbleibenden Patienten (aufgrund des schärferen Wettbewerbs) immer mehr Konsultationen anzubieten.

---

<sup>244</sup> Im Allgemeinen zeigt die empirische Literatur, dass zwischen Entfernung zur Arztpraxis und Leistungsanspruchnahme ein negativer Zusammenhang besteht (vgl. für einen Überblick Scott 2000). Das abweichende Ergebnis für die Anzahl an Konsultationen in dieser Untersuchung lässt sich wiederum mit der Unterscheidung zwischen Konsultationswahrscheinlichkeit (negativer Zusammenhang) und Konsultationsanzahl (kein signifikanter Zusammenhang) erklären (vgl. auch Pohlmeier und Ullrich 1994).

Um der hohen Multikollinearität in dieser Schätzung Herr zu werden, mussten einige wichtige Variablen von der Analyse ausgeblendet werden; so z.B. die Verbindung des Vermögens eines Arztes mit dessen Berufserfahrung (*VERMB*), der Umsatz pro Patient des letzten Jahres (*L\_ARZT98*) oder die Patientenzahl der Arztpraxis im letzten Jahr (*LPAT98*). Gerade die Einbeziehung letzterer Variable wäre wichtig gewesen, um die Frage der Angebotsinduzierung beantworten zu können. Denn in Verbindung mit dem Wachstum des Patientenstammes im letzten Jahr (*P\_DIF*) würde *LPAT98* Rückschlüsse auf die Kapazitätsgrenze der Arztpraxis zulassen und – in Verbindung mit der Arztdichte – auch auf die Angebotsinduzierung.

#### **Zusammenfassung der einzelnen Ergebnisse**

1. Ärzte, die sich an MCOs beteiligen, weisen zwar eine geringere Wahrscheinlichkeit positiver Konsultationen auf als Ärzte, die in der Stichprobe ausschließlich konventionell versicherte Patienten behandeln, bieten aber denjenigen Patienten, bei denen es zu mindestens einer Konsultation gekommen ist, rund 10% mehr Konsultationen an.
2. Managed-Care-Patienten haben im Vergleich zu konventionellen Patienten eine geringere Wahrscheinlichkeit im Jahr mindestens eine Konsultation zu bekommen. Wenn sie allerdings mindestens eine Konsultation erhalten haben, bekommen sie von ihrem Arzt im Durchschnitt 64% mehr Konsultationen als vergleichbare Patienten im konventionellen Bereich.
3. Während die Suchkosten der Patienten eine entscheidende Rolle bei der Bestimmung der Wahrscheinlichkeit mindestens einer Konsultationen spielen, ist die Entfernung des Patienten vom Arzt offenbar nicht für die Bestimmung der Anzahl an Konsultationen relevant. Darin äußert sich die unterschiedliche Beeinflussungsmöglichkeit des Patienten durch den Arzt: Die Wahrscheinlichkeit einer Arztkonsultation kann er nicht direkt beeinflussen, daher sind die Suchkosten für die Erklärung der Konsultationswahrscheinlichkeit wichtig. Der Arzt kann hingegen die Anzahl an Konsultationen im hohen Grade selbst festlegen, sodass die Suchkosten, also die Entfernung des Patienten vom Arzt, immer weniger für die Anzahl Konsultationen von Belang sind.

#### **9.4.6 Gesamtleistungen**

In diesem Abschnitt betrachten wir das gesamte Leistungsgeschehen in der Arztpraxis unabhängig davon, ob die erbrachte bzw. veranlasste Leistung ein direktes Einkommen für den Arzt bedeutet oder nicht. In jeder Arzt-Patienten-Beziehung kam es mindestens in einer Kategorie zu

einer Leistungsanspruchnahme durch den Patienten, sodass bei der Schätzung kein Selektionsmechanismus für mögliche 0-Leistungen betrachtet werden muss. Die Schiefe der Leistungsverteilung wird durch die Logarithmierung der abhängigen Variablen berücksichtigt.<sup>245</sup> Zur Erklärung der Gesamtleistungen kommen fast alle zur Verfügung stehenden Variablen zum Einsatz. Neben den Patienten- und Arzt-Variablen sind das Markt-faktoren und auch Variablen, die den *casemix* und die Besonderheiten jeder einzelnen Arzt-Patienten-Beziehung berücksichtigen. Insbesondere werden dabei Variablen als erklärende Variablen betrachtet, die für die Schätzung der Gesamtleistungen nicht exogen sind. Zu diesen Variablen gehören die logarithmierten Konsultationen (*LNK*), die logarithmierten Leistungsnummern (*LNNR*), das Produkt aus beiden Faktoren sowie die Regelmäßigkeit der Arzt-Patienten-Beziehung (*REGEL*). Obwohl diese Variablen bestimmte Besonderheiten des einzelnen Arzt-Patienten-Kontakts einfangen, kann durch die Einbeziehung von mehr oder weniger endogenen Faktoren eine Verzerrung bei der Schätzung aller Koeffizienten entstehen. Daher wurde das Modell zur Überprüfung auch ohne die kritischen Variablen geschätzt und das Ergebnis mit dem allgemeinen Modell verglichen.<sup>246</sup>

Darüber hinaus werden die Ergebnisse der vorangegangenen Wahrscheinlichkeitsschätzungen als Kontrollvariablen berücksichtigt. Denn es zeigte sich, dass Fälle mit einer hohen Überweisungswahrscheinlichkeit sich fundamental von den Fällen mit einer geringen Überweisungswahrscheinlichkeit unterscheiden. Das Gleiche gilt für die Wahrscheinlichkeit, mindestens einer Konsultation im Jahr. Unter Verwendung der geschätzten Probitkoeffizienten wurde daher die jeweilige Mills-Ratio als erklärende Variable im Modell berücksichtigt. Wobei wir mit *LAMBDA2* den Korrekturfaktor für die Konsultationswahrscheinlichkeit und mit *LAMBDA3* den für die Überweisungswahrscheinlichkeit bezeichnen. Beide Faktoren sind ebenfalls endogen, sodass auch hier eine mögliche Verzerrung der Schätzergebnisse beachtet werden muss.

Alles in allem hat das Modell zur Erklärung der Gesamtleistungen ( $y$ ) eines Patienten in einer Arztpraxis damit folgendes Aussehen:

$$\begin{aligned}
 y_{ijk} = & \alpha_0 + \alpha_1' A_i + \alpha_2' P_j + \alpha_3' AP_{ij} + \alpha_4' Gem_k \\
 & + \alpha_5 MC_{-} A_i + \alpha_6 MC_{-} P_j + \alpha_7 Eff_i + \alpha_8 \lambda_{1j} \\
 & + a_9 \lambda_2 + a_{10} \lambda_3 + v_i + \varepsilon_{ijk}.
 \end{aligned} \tag{9.33}$$

---

<sup>245</sup> Zu Problemen der logarithmischen Transformation von abhängigen Variablen vgl. Manning (1998), Jones (2000),

<sup>246</sup> Diese Vorgehensweise wurde bspw. auch von Scott und Shiell (1997a) verwendet.

Das Modell stellt wiederum ein *random effects model* dar (vgl. Abschnitt 9.4.2). Dabei werden verschiedene beobachtbare Faktoren zur Erklärung des Leistungsgeschehens in einer Arztpraxis herangezogen. Neben Arztfaktoren ( $\mathcal{A}$ ) und Variablen, welche die Patienten beschreiben ( $\mathcal{P}$ ), kommen als weitere Kontrollvariablen Faktoren hinzu, die bestimmte Eigenheiten des konkreten Arzt-Patienten-Kontaktes als auch der Arztpraxis beschreiben sollen ( $\mathcal{AP}$ ). Außerdem wird für bestimmte Faktoren auf Gemeindeebene kontrolliert, die einen Einfluss auf das Leistungsgeschehen innerhalb der einzelnen Arztpraxis haben können. Gleichzeitig wird aber im stochastischen Teil des Modells durch die arzt spezifische Fehlerkomponente ( $\nu$ ) der unbeobachtbaren Heterogenität der Ärzte Rechnung getragen.

#### **Ergebnisse**

In Tabelle 26 sind die Ergebnisse für die Schätzung der Gesamtleistungen dargestellt. Die Modellierung über arzt spezifische Störterme wird bestätigt. Ein LM-Test verwirft die Hypothese keiner derartigen Effekte: Der Wert der LM-Statistik ist hochsignifikant. Somit sind die Störterme innerhalb der Arztpraxis – wenn auch schwach – miteinander korreliert ( $Rho$ ).

Das Hauptergebnis ist eindeutig: Während für MC-Patienten insgesamt wesentlich mehr Leistungen durch den einzelnen Arzt abgerechnet werden (42%), besitzt der Status des Arztes hinsichtlich einer MC-Beteiligung keinen Einfluss auf die Gesamtleistungen. Ersteres Ergebnis ist wiederum direkte Folge der eingeschränkten Arztwahl im MC-Bereich. Für die Patienten in Hausarzt netzen und in der HMO sind in der untersuchten Stichprobe so gut wie alle Leistungen erfasst, da der Arzt hier als Gatekeeper alle weiteren Leistungen veranlasst. Dagegen kann der Patient im konventionellen Bereich jeder Zeit auch weitere Allgemeinmediziner oder Spezialisten konsultieren. Ob ein Patient im Untersuchungszeitraum auch noch weitere Grundversorger konsultiert hat, wurde dabei in der Schätzung berücksichtigt ( $UNT\_VRS$ ). Die Gesamtleistung bei einem einzelnen Arzt geht zurück, wenn der Patient mehrere Grundversorger konsultierte. Für die Inanspruchnahme der Patienten des konventionellen Bereichs von Facharztleistungen kann hingegen bei dieser Schätzung nicht ausreichend kontrolliert werden, da im konventionellen Sektor prinzipiell freie Arztwahl herrscht und Patienten nicht notwendigerweise von einem Hausarzt zum Facharzt überwiesen werden müssen.

Die fehlende Signifikanz der MC-Variable der Ärzte ( $MC\_ARZT$ ) kann mit Hinweis auf die gewählte Kontrollvariable für diesen MC-Effekt – nämlich die Effizienz der Ärzte ( $EFF$ ) – erklärt werden: Da die MC-Beteiligung mit der Effizienz-Variable positiv korreliert ist und die

Effizienz-Variable hoch signifikant negativ ist, wird der interessierende Effekt insignifikant. Die Effizienz-Variable ist deshalb hoch signifikant negativ, weil die Schätzung der Gesamtleistung für einzelne Patienten wiederum positiv korreliert ist mit der Leistungserstellung auf Praxisebene, auf welcher die Effizienz der Ärzte geschätzt worden ist (vgl. Abschnitt 9.2). Mit anderen Worten verursachen MC-Ärzte in der Tat weniger Leistungen als Ärzte, die (in der Stichprobe) ausschließlich konventionell versicherte Patienten behandelt haben. Die beobachtete geringere Leistungsverursachung hat seinen Grund aber vollständig in der Tatsache, dass diese Ärzte ohnehin kostengünstiger Arztdienstleistungen produzieren als andere Ärzte.



### III Empirische Untersuchung

Tabelle 26 Gesamtmodell: Gesamtleistungen Arzt (random-effects)

N=100522	Koeffizient	Std.Fehler		Koeffizient	Std.Fehler
<b>Patienten</b>			<b>Ärzte</b>		
KONST	4.397 **	0.220	ROENTG	-2.44E-03	1.26E-02
LND	5.46E-03 **	5.78E-04	ZH	-0.096 **	0.020
KA	-5.02E-03	1.53E-02	PA100	0.018	0.037
FR	0.024 **	0.006	PA105	0.054	0.040
FR98	-6.88E-03 **	8.44E-04	GEM_ANZ	-0.012 *	0.005
RISK	0.018 *	0.009	NEUTRAL	-0.014	0.050
SEX	0.055 **	0.004	TITEL_D	-0.036	0.024
ALTER	0.013 **	0.001	BERUFSER	-0.042 **	0.009
ALTER2	-7.54E-05 **	6.15E-06	LALTPR	-0.045	0.024
KIND	8.25E-03	0.013	SEX_PR	0.052	0.048
JUGEND	0.019	0.010	VERM	0.028 **	0.009
SPT	7.89E-03	6.69E-03	P_DIFF	-0.027	0.023
LLST97	9.70E-04	1.97E-03	PAT_ANT	-0.010	0.009
LLST98	-0.148 **	0.009			
ALT97	2.39E-04 **	4.40E-05			
ALT98	1.66E-04 **	4.65E-05			
LPR	6.86E-04	1.38E-03			
UNT_VRS	-6.71E-03	3.46E-03			
UNT_V2	3.57E-04	3.40E-04			
<b>Gemeinde</b>			<b>Casemix und Besondere Eigensch.</b>		
HERF	0.010	0.032	M_V	1.95E-07	2.65E-06
LDK_P	0.114 **	0.033	L98	0.025 **	0.001
ST_P_K	-2.16E-06	6.89E-06	LNK	1.059 **	0.007
ARZTD	5.02E-04	1.43E-03	LNNR	1.011 **	0.008
BEVD	0.421	0.275	K_NR	-0.394 **	0.005
			REGEL	0.106 **	0.003
<b>MC-Variablen</b>			<b>Kontrollvariablen</b>		
MC_A	-0.012	0.020	EFF	-1.612 **	0.067
MC_P	0.412 **	0.029	LAMBDA1	-4.44E-04	1.26E-02
			LAMBDA2	-0.182 **	0.005
			LAMBDA3	0.273 **	0.003
R^2	0.764				
Rho	0.064 **				
LM-Stat	4160 **				

\*) p < 0.05, \*\*) p < 0.01

Schaut man sich nun die Wirkung der anderen erklärenden Variablen an, so fällt auf, dass die Arztcharakteristika bzw. die Eigenschaften der Arztpraxis nur in vier von dreizehn Fällen einen signifikanten Beitrag zur Erklärung der Gesamtleistungen liefern können. Aufgrund des relativ hohen  $R^2$  deutet dies auf eine hohe Multikollinearität hin. In der Tat sind viele Arztvariablen

stark miteinander korreliert. So mussten aufgrund einer extrem hohen Korrelation mit anderen Variablen, die Variablen *VERMB* und *L\_ARZT98* aus der Menge der erklärenden Variablen entfernt werden. Die verbliebenen Arztvariablen besitzen zwar nach wie vor eine hohe Korrelation, Multikollinearitätstests zeigen aber, dass diese Kollinearität die Interpretation der Koeffizienten nicht beeinflussen. Das bedeutet, dass die nicht von Null verschiedenen Koeffizienten tatsächlich nicht signifikant sind und nicht wegen der hohen Multikollinearität. Im Einzelnen zeigt sich, dass Ärzte im Kanton Zürich (*ZH*) weniger direkte Leistungen ihren Patienten anbieten als ihre Kollegen im Kanton Genf. Arbeitet der Arzt in einer Gruppenpraxis (*GEM\_ANZ*), erhalten die Patienten ebenfalls weniger Leistungen als bei Ärzten, die in einer Einzelpraxis tätig sind. Dieser Effekt nimmt mit der Ärztezahl in der Gruppenpraxis zu. Ärzte mit einer großen Berufserfahrung (*BERUFSERF*) verursachen mehr Leistungen als Ärzte mit einer vergleichsweise geringen Berufserfahrung. Schließlich steht der durchschnittliche Umsatz in den beiden letzten Jahren (*VERM*) ebenfalls in einem positiven Zusammenhang mit den verursachten Leistungen in der Untersuchungsperiode.

Von den einbezogenen Patienten-Variablen sind ungefähr die Hälfte signifikant. Von besonderem Interesse ist hier vor allem die Entfernung des Patienten von der Arztpraxis (*LND*). Je größer die Entfernung des Patienten zu seinem Arzt ist, desto höher fallen auch die Gesamtleistungen aus. Dieser positive Effekt kann dahingehend interpretiert werden, dass Patienten, die aus eher ländlichen Regionen kommen und daher durchschnittlich einen weiteren Weg zum nächsten Arzt zurücklegen müssen, eine höhere Leistungsanspruchnahme aufweisen als Patienten, die einen Arzt in ihrer Nähe konsultieren können. Dabei ist allerdings zu berücksichtigen, dass Versicherte in ländlichen Gebieten im Durchschnitt wesentlich geringere Leistungen verursachen als Versicherte in der Stadt. Dieser Umstand spiegelt sich in der einzigen signifikanten Markt-Variable wider, wonach die Gesamtleistungen pro Patient in einem positiven Zusammenhang mit den Durchschnittskosten in der Wohngemeinde der Patienten stehen. Wohnt ein Patient in einer eher ländlichen Gemeinde (= verhältnismäßig geringe Durchschnittskosten) sind seine Gesamtleistungen geringer als bei städtischen Patienten. Auf der anderen Seite steigt aber mit der Entfernung die Gesamtleistungsanspruchnahme an.

Interessant ist ferner der altersabhängige langfristige Gesundheitszustand des Patienten (*Alter*, *Alter2*, *LLST98*, *ALT97* und *ALT98*). Die Gesamtleistungen steigen zunächst mit abnehmenden Raten bis zu einem Alter von ca. 83 an, um dann wieder zu sinken (positiver linearer Effekt, negativer quadratischer Effekt). Berücksichtigt man den langfristigen Gesundheitszustand des

Patienten (hier gemessen durch die Leistungsausgaben der beiden Vorjahre), so wird deutlich, dass mit zunehmendem Alter und schlechtem Gesundheitszustand die gesamten Leistungsausgaben steigen. Von besonderem Interesse ist in diesem Zusammenhang, dass offenbar nicht das einfache Niveau der Gesundheitsausgaben der Vorperioden für die Inanspruchnahme in der laufenden Periode verantwortlich ist (nicht signifikanter Effekt für 1997 und sogar negativer Effekt für 1998), sondern vielmehr die Betrachtung der Gesundheitsausgaben in Zusammenhang mit dem Alter ausschlaggebend ist.

Abschließend werfen wir einen Blick auf die Variablen zur Erklärung der speziellen Arzt-Patienten-Beziehung. Naturgemäß besitzen diese Variablen eine sehr hohe Erklärungskraft für die Gesamtleistungen, was hier aber nicht weiter interessieren soll. Interessant sind vielmehr Vorzeichen und Stärke der Effekte. So erkennt man, dass die Gesamtleistungsausgaben mit den Konsultationen und Leistungsnummern fast 1:1 steigen ( $LNK$  und  $LNNR$ ). Allerdings besitzt der gemischte Effekt ( $K_{NR}$ ) ein signifikant negatives Vorzeichen, was als eine Art substitutive Beziehung der beiden Faktoren gedeutet werden kann. Damit ist gemeint, dass eine gleichzeitige proportionale Steigerung von Konsultationen und Leistungsnummern zu einer Verringerung der Gesamtleistungen führen kann. Zur Verdeutlichung des Sachverhaltes kann ein kleines Beispiel dienen. Nimmt man der Einfachheit halber an, dass während einer Konsultation 2 Leistungsnummern abgerechnet werden, so würde zunächst mit zunehmenden Konsultationen und dementsprechend steigender Leistungsnummernzahl auch die Gesamtleistung pro Patient steigen. Ab ungefähr 11 Konsultationen würden dann aber die Gesamtleistungen pro Patient sinken. Dieser wichtige Punkt wird bei der Betrachtung der direkten Arztleistungen, also dem Umsatz des Arztes, noch eingehender beleuchtet. In dem Zusammenhang sind außerdem die beiden Korrekturterme von Interesse, die aus den Wahrscheinlichkeitsmodellen für Überweisungen und Konsultationen hervorgegangen sind. Während eine höhere Überweisungswahrscheinlichkeit erwartungsgemäß mit höheren Gesamtleistungen verbunden ist, führt eine hohe Wahrscheinlichkeit, dass der Patient mindestens eine Konsultation hatte, zu einer Verringerung der Gesamtleistungen. Dieser etwas überraschende Effekt kann nur im Zusammenhang mit den anderen, oben betrachteten Konsultations-Variablen richtig interpretiert werden: Patienten mit mindestens einer Konsultation im Jahr haben höhere Gesamtleistungen im Jahr, als wenn keine Konsultation abgerechnet worden wäre. Dieser Effekt muss mit dem negativen Wahrscheinlichkeitseffekt verrechnet werden.

Wie erwähnt, sind die zuletzt genannten Variablen einbezogen worden, um die speziellen Eigenheiten der einzelnen Arzt-Patienten-Beziehung einzufangen. Da wir keine Daten über Diagnosen besitzen, hilft ihre Einbeziehung zumindest zum Teil, Korrelationen zwischen der Leistungsvariablen und bestimmten Krankheitsbildern zu berücksichtigen.<sup>247</sup> Allerdings wird durch die Einbeziehung solcher Variablen unterstellt, dass sie für die Erklärung der Gesamtleistungen exogene erklärende Faktoren darstellen. In aller Regel wird dies nicht der Fall sein, sodass durch die Einbeziehung möglicherweise endogener Faktoren eine Verzerrung der Schätzergebnisse hervorgerufen wird. Aus diesem Grund wurde das Modell nochmals ohne diese Variablen geschätzt und die Ergebnisse mit den vorgestellten Resultaten verglichen. Dabei zeigte sich, dass die absolute Höhe einzelner Koeffizienten zwar variiert, der prinzipielle Zusammenhang zwischen erklärenden Variablen und den Gesamtleistungen aber nicht beeinflusst wird.<sup>248</sup>

#### **Zusammenfassung der wichtigsten Ergebnisse**

1. Ob Ärzte sich an Managed-Care-Modellen beteiligen oder nicht, hat keinen signifikanten Einfluss auf die gesamten Leistungen, die von einem Arzt bei einem Patienten in der Stichprobe verursacht werden.
2. Patienten, die in Hausarztmodellen eingeschrieben bzw. Mitglied einer HMO sind, weisen signifikant höhere Gesamtleistungen auf als Patienten des konventionellen Bereichs (bezogen auf einen Arzt). Dieses Ergebnis ist einerseits Ausdruck der Regelungen in MC-Modellen (Gatekeeperfunktion) und andererseits auf die freie Arztwahl der Versicherten im konventionellen Sektor zurückzuführen.

#### **9.4.7 Direkte Arztleistungen**

Die oben betrachteten Gesamtleistungen setzen sich aus Leistungen, die der Arzt selbst erbringt (direkte Arztleistungen), und veranlassten Leistungen zusammen. In der abschließenden Untersuchung werden nur die Leistungen betrachtet, die direkt vom Arzt erbracht worden sind. Damit wird in einem viel stärkeren Maße das Einkommensmotiv bei der Bereitstellung von Arztdienstleistungen als bei den vorangegangenen Untersuchungen analysiert. Dabei kann die Betrachtung aber nicht losgelöst von der gesamten Leistungserstellung in einer Arztpraxis

---

<sup>247</sup> So dürften in der Regel chronisch Kranke relativ häufig und vor allem regelmäßig ihren Hausarzt aufsuchen, eventuell aber nur relativ wenige Leistungsnummern verrechnet bekommen.

<sup>248</sup> Damit ist gemeint, dass sich weder Vorzeichen der Koeffizienten ändern noch die Signifikanzniveaus.

erfolgen. Denn es bestehen wichtige Wechselwirkungen zwischen den hier betrachteten direkten Arztleistungen auf der einen Seite und dem Überweisungsverhalten, den Konsultationen sowie weiteren veranlassten Leistungen auf der anderen Seite. Daher werden wie bei der Untersuchung der Gesamtleistungen wieder bestimmte Faktoren, welche die Arzt-Patienten-Beziehung genauer beschreiben, und die Korrekturterme aus den vorangegangenen Schätzungen als erklärende Variablen verwendet.

Das betrachtete Modell unterscheidet sich damit in den erklärenden Faktoren nicht vom Modell im letzten Abschnitt. Es ist nun aber zu berücksichtigen, dass es nicht bei allen Arzt-Patienten-Beziehungen zu direkten Leistungen des Arztes kommt. So können Patienten ausschließlich eine Überweisung erhalten oder ein Rezept ausgestellt bekommen. Diese Fälle dürften daher vor allem relativ gesunde Patienten umfassen. Allerdings kann es sich bei solchen Fällen auch um Patienten mit chronischen Erkrankungen handeln, die regelmäßig bestimmte Medikamente beziehen, aber nicht notwendigerweise Arztkonsultationen in Anspruch nehmen müssen. Der erstere Erklärungsversuch wird durch die Beobachtung bestätigt, dass Fälle ohne direkte Arztleistungen auch geringere Gesamtleistungen aufweisen als Fälle mit abgerechneten direkten Leistungen. Die zweite Erklärung wird dagegen durch die Beobachtung gestützt, dass die Patienten ohne direkte Arztleistungen in der Untersuchungsperiode in den vergangenen Jahren eine vergleichsweise hohe Leistungsanspruchnahme aufwiesen. Darüber hinaus hatten sie auch noch relativ regelmäßige Arzt-Kontakte. Obwohl also nicht mit Sicherheit gesagt werden kann, was dazu führt, dass einige Patienten keine direkten Arztleistungen in Anspruch genommen haben, ist es doch von Wichtigkeit diesen Selektionseffekt zu berücksichtigen. Daher wird für die Erklärung der direkten Arztleistungen wieder das herkömmliche Selektionsmodell verwendet, um den beschriebenen Effekt zumindest teilweise zu berücksichtigen (vgl. auch die Ausführungen zu den Konsultationen in Abschnitt 9.4.5).

Der mögliche Selektionseffekt wird wieder über die Einbeziehung der Mills-Ratio in der Schätzung der (positiven) direkten Arztleistungen berücksichtigt. Damit lässt sich das zu schätzende Modell in folgender Form schreiben:

$$\begin{aligned}
 y_{ijk} = & \alpha_0 + \alpha_1' A_i + \alpha_2' P_j + \alpha_3' AP_{ij} + \alpha_4' Gem_k \\
 & + \alpha_5 MC_{-A_i} + \alpha_6 MC_{-P_j} + \alpha_7 Eff_i + \alpha_8 \lambda_{1j} \\
 & + \alpha_9 \lambda_2 + \alpha_{10} \lambda_3 + \alpha_{11} \lambda_4 + v_i + \varepsilon_{ijk}.
 \end{aligned}
 \tag{9.34}$$

Zur Erklärung der direkten Arztleistungen, d.h. dem Einkommen des Arztes  $i$  pro Patient  $j$  in einer bestimmten Gemeinde  $k$ , werden wieder bestimmte Arztcharakteristika  $A$ , Eigenschaften der Patienten  $P$  sowie Eigenheiten der konkreten Arzt-Patienten-Beziehung  $AP$  und Markteigenschaften  $GEM$  herangezogen. Ob Managed Care einen signifikanten Einfluss auf die direkten Arztleistungen hat, wird durch zwei Effekte bestimmt: die MC-Beteiligung der Ärzte und die MC-Beteiligung der Patienten. Eine mögliche Endogenität dieser Variablen wird durch die Effizienz der Ärzte ( $Eff$ ) und die Wahlwahrscheinlichkeit von MC-Verträgen der Patienten ( $\lambda_1$ ) kontrolliert. Zusätzlich wird ebenfalls die Wahrscheinlichkeit mindestens einer Konsultation während der Arzt-Patienten-Beziehung ( $\lambda_2$ ) und für Besonderheiten im Überweisungsverhalten ( $\lambda_3$ ) kontrolliert. Der Korrekturterm  $\lambda_4$  trägt der möglichen Selektion von bestimmten Fällen Rechnung, die während einer Arzt-Patienten-Beziehung keine direkten Arztleistungen aufgewiesen haben.

Das Modell wird wieder als ein *random effects model* geschätzt. D.h. es wird explizit die besondere Datenstruktur modelliert, die dadurch gekennzeichnet ist, dass die gleichen Ärzte viele unterschiedliche Patienten behandeln. Die mögliche Korrelation von Beobachtungen innerhalb einer Arztpraxis wird daher durch einen arzt-spezifischen Störterm berücksichtigt ( $\mu_i$ ).

#### **Ergebnisse**

Zunächst wurde die Wahrscheinlichkeit bestimmt, mit der es überhaupt zu direkten Arztleistungen kommt. Die Schätzung erfolgte durch ein Probit-Modell unter Verwendung des gleichen Variablensets wie bei der Schätzung der Überweisungswahrscheinlichkeit und der Wahrscheinlichkeit positiver Arztkonsultationen. Die Ergebnisse der Schätzung bestätigen die Vermutung, dass eine Identifizierung derjenigen Fälle, bei denen es zu keinen direkten Arztleistungen gekommen ist, nicht einfach sein würde: Das Modell erkennt nur 17 von 8,668 Fällen richtig, die keine direkten Arztleistungen aufweisen. Andererseits besitzen die gewählten Variablen einen relativ hohen Erklärungsgehalt für die Wahrscheinlichkeit positiver direkter Arztleistungen, da das Pseudo  $R^2$  mit 0.31 relativ hoch ist.

Da die Schätzung keinen besonderen Wert für sich alleine besitzt, wird auf die explizite Darstellung der Ergebnisse hier verzichtet. Wichtig für die Schätzung der positiven Arztleistungen ist allein die Berücksichtigung einer möglichen Selektion aufgrund der Fälle mit keinen direkten

Arztleistungen. Dies geschieht – wie gesagt – durch die Einbeziehung der geschätzten Wahrscheinlichkeiten über die Mills-Ratio.

Die Schätzung der direkten Arztleistungen ergibt, dass sowohl die MC-Beteiligung der Ärzte als auch die Teilnahme der Versicherten an MC-Modellen einen signifikant negativen Effekt auf die direkten Arztleistungen besitzen (vgl. Tabelle 27). Die Ärzte mit MC-Beteiligung verursachen wesentlich weniger direkte Leistungen als Ärzte ohne MC-Beteiligung. Von besonderem Interesse ist hierbei der Spill-over-Effekt: Managed-Care-Ärzte behandeln konventionelle Patienten kostengünstiger als Ärzte des konventionellen Bereichs. Damit wird wiederum die Vermutung des theoretischen Modells bestätigt, wonach Managed Care auch auf den konventionellen Sektor ausstrahlt. Wichtig für die Ergebnisse ist die Berücksichtigung einer möglichen Selektion von Ärzten und Patienten in die MC-Modelle. Ob der Patient an einem MC-Modell teilnimmt oder nicht, spielt dabei für das Ergebnis keine Rolle. Geht man davon aus, dass für eine mögliche Selektion der Patienten hinreichend kontrolliert wurde, entspricht der Koeffizient der Patienten dem reinen Managed-Care-Effekt. Dieses Ergebnis bekommt durch die Betrachtung der Vergleichsgruppe besonderes Gewicht. Denn während wir für die MC-Patienten in der Regel bei einem Arzt sämtliche direkten Leistungen beobachten, ist das aufgrund der freien Arztwahl für konventionelle Patienten nicht der Fall.

### III Empirische Untersuchung

Tabelle 27 Gesamtmodell: Direkte Leistungen des Arztes (random effects)

N = 91,854	Koeffizient	Std.Fehler		Koeffizient	Std.Fehler
<b>Patienten</b>			<b>Ärzte</b>		
KONST	7.08 **	0.489	ROENTGE	0.017	0.014
LND	4.68E-04	6.85E-04	ZH	-0.014	0.021
KA	-0.014	0.015	PA100	0.092 *	0.040
FR	-6.40E-03	6.31E-03	PA105	0.105 *	0.043
FR98	-1.97E-03 *	9.20E-04	GEM_ANZ	-0.015 **	0.005
UNFALL	0.017	0.009	NEUTRAL	0.066	0.055
SEX	0.046 **	0.004	TITEL_D	-0.076 **	0.025
ALTER	0.013 **	0.001	BERUFSER	-0.084 **	0.010
ALTER2	-4.34E-05 **	5.90E-06	LALTPR	-0.221 **	0.026
KIND	0.063 **	0.014	SEX_PR	-0.083	0.052
JUGEND	0.028 **	0.010	VERM	0.076 **	0.009
SPT	0.021 **	0.006	P_DIFF	-2.06E-03	2.55E-02
LLST97	5.81E-03 **	1.89E-03	PAT_ANT	-6.30E-03	9.36E-03
LLST98	-0.185 **	0.009			
ALT97	1.83E-04 **	4.40E-05			
ALT98	-2.44E-04 **	4.64E-05			
LPR	-3.71E-03 **	1.33E-03			
UNT_VRS	-6.77E-04	3.29E-03			
UNT_V2	5.37E-04	3.17E-04			
<b>Gemeinde</b>			<b>Casemix und Besondere Eigensch.</b>		
HERF	0.099 **	0.034	M_V	3.16E-06	2.85E-06
LGEM_P	-0.319 **	0.047	L98	0.032 **	0.001
ST_P_K	2.32E-05 **	7.57E-06	LNK	1.168 **	0.007
ARZTD	-4.42E-03 **	1.54E-03	LNNR	0.574 **	0.008
BEVD	-0.109	0.303	K_NR	-0.212 **	0.005
			REGEL	0.018 **	0.003
<b>MC-Variablen</b>			<b>Kontrollvariablen</b>		
MC_A	-0.106 **	0.022	EFF	-2.316 **	0.073
MC_P	-0.257 **	0.037	LAMBDA1	0.015	0.013
			LAMBDA2	0.071 **	0.003
			LAMBDA3	1.078 **	0.109
			LAMBDA4	-0.408 **	0.005
R^2	0.704		Rho	0.092 **	
			LM-Test	12260 **	

\*)  $p < 0.05$ , \*\*)  $p < 0.01$

Auch hier betrachten wir einige ausgewählte interessante Ergebnisse etwas genauer. Zunächst wieder ein Blick auf die Patienten-Variablen: Die Suchkosten der Patienten (bzw. die Entfernung zwischen Patient und Arzt *LND*) spielen bei den direkten Arztleistungen keine Rolle. Hierin wird wiederum die Beeinflussungsmöglichkeit des Leistungsgeschehens durch den Arzt deutlich. Denn betrachtet man den nichtsignifikanten Effekt für die direkten Leistungen im



Vergleich mit dem signifikant positiven Effekt bei den Gesamtleistungen, kann man Rückschlüsse über das Arztverhalten ziehen. So lassen sich die beiden Effekte nun dahingehend interpretieren, dass Ärzte Patienten mit einer größeren Entfernung weit häufiger zu anderen Ärzten überweisen als Patienten, die nicht so weit weg wohnen. Andererseits bieten die Ärzte ihren Patienten aber unabhängig von der Entfernung der Patienten zur Arztpraxis Leistungen in gleicher Höhe an. Dies lässt sich als ein starker Nachweis für das Einkommensmotiv der Ärzte interpretieren.

Die Leistungsanspruchnahme der Patienten nimmt nicht notwendigerweise mit dem Alter zu (*ALT98*). Ebenso sind die Vorleistungen nicht mehr nur als einfache Prädiktoren des Gesundheitszustands aufzufassen, da der Koeffizient der Leistungen im Vorjahr jetzt signifikant negativ ist (*LLST98*).

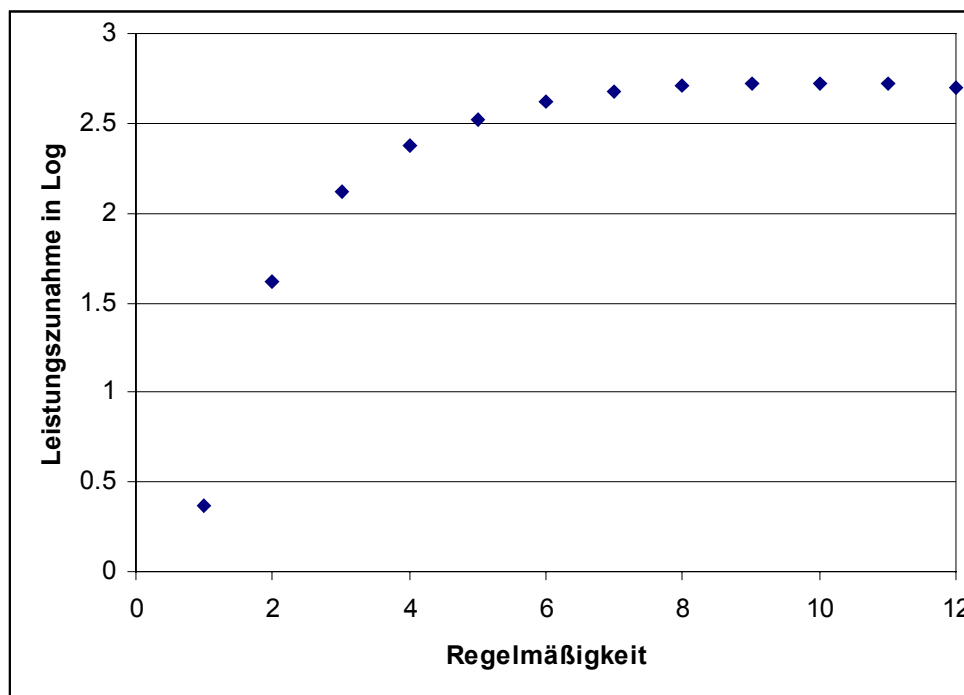
Bei der Betrachtung der Arztvariablen muss wieder beachtet werden, dass einige Variablen wegen ihrer sehr hohen Korrelation mit anderen Variablen aus der Analyse ausgeschlossen werden mussten. Es fällt vor allem auf, dass – wie bei der Untersuchung der Gesamtleistungen – auch wieder der Koeffizient der Berufserfahrung (*BERUFSERF*) signifikant negativ ist, während der Koeffizient des Vermögens (*VERM*) signifikant positiv ist. Außerdem bedeutet der signifikant negative Koeffizient von (*GEM\_ANZ*), dass Ärzte in Gemeinschaftspraxen weniger Leistungen direkt verursachen als Ärzte in Einzelpraxen.

Die Koeffizienten der Marktvariablen sind fast alle signifikant (außer die Bevölkerungsdichte *BEVD*). Von besonderem Interesse sind wieder die Arztdichte (*ARZTD*) und die Wettbewerbsvariable (*HERF*). Im Unterschied zur Untersuchung der Konsultationsanzahl wechseln jetzt die Vorzeichen der Koeffizienten – beide sind signifikant negativ. Damit führt ein hoher Monopolisierungsgrad zu höheren Leistungen, wenn gleichzeitig die Arztdichte gering ist. Dieses Ergebnis bestätigt somit die Vorhersagen des theoretischen Modells aus Abschnitt 6.3. Dort wurde gezeigt, dass ein schärferer Wettbewerb zwischen den Ärzten zu einer Mengensenkung führt, wenn der Arzt pro Einzelleistung vergütet wird.

Die Variablen zur Erklärung der speziellen Arzt-Patienten-Beziehung ergeben folgendes Bild. Im Unterschied zu den Gesamtleistungen erkennt man, dass die direkten Leistungen mit den Konsultationen zwar auch wieder ungefähr im Verhältnis 1:1 steigen, das Verhältnis bei den Leistungsnummern aber wesentlich geringer ausfällt (*LNK* und *LNNR*). Allerdings besitzt der gemischte Effekt (*K\_NR*) auch hier ein signifikant negatives Vorzeichen, was als eine Art

substitutive Beziehung der beiden Faktoren gedeutet wurde (vgl. die Ausführungen zu den Gesamtleistungen im Abschnitt 9.4.6). Der beschriebene Wirkungszusammenhang bedeutet für die direkten Arztleistungen konkret, dass es für einen Arzt wenig Sinn macht, Konsultationen und Leistungsnummern proportional zu steigern, da es andere Kombinationen gibt, die ihm ein höheres Einkommen verschaffen würden. Von Interesse ist in diesem Zusammenhang auch die Regelmäßigkeit einer Arzt-Patienten-Beziehung (*REGEL*). Ein Patient, der sehr regelmäßig (bspw. jeden Monat mindestens ein Mal) seinen Arzt aufsucht, erhält auch wesentlich mehr direkte Leistungen als ein Patient, der nur ein paar Mal im Jahr einen Arzt aufsucht. Betrachtet man das Ergebnis allerdings unter Berücksichtigung der Konsultationen und der Leistungsnummern, ergibt sich folgendes Bild: Mit einer höheren Regelmäßigkeit der Arztbesuche steigen die direkten Arztleistungen für die Patienten an, wobei der Anstieg mit zunehmender Regelmäßigkeit abnimmt. Abbildung 9 verdeutlicht den Zusammenhang. Dabei wurden zur Evaluation des Effektes die Mittelwerte der Konsultationen und der Leistungsnummern auf den 12 Regelmäßigkeitsstufen verwendet. Der Effekt wird in logarithmierten Leistungsausgaben gemessen.

Abbildung 9 Leistungseffekt in Abhängigkeit der Regelmäßigkeit unter Berücksichtigung von Konsultationen und Leistungsnummern



Hinsichtlich der zuletzt genannten Variablen gilt das Gleiche wie bei der Schätzung der Gesamtleistungen: Die Variablen wurden einbezogen, um die speziellen Eigenheiten der einzelnen Arzt-Patienten-Beziehung einzufangen. Da keine Daten über Diagnosen vorhanden sind, hilft ihre Einbeziehung, die Korrelationen zwischen der Leistungsvariablen und bestimmten Krankheitsbildern zu berücksichtigen. Allerdings sind diese Variablen möglicherweise endogen, sodass durch ihre Berücksichtigung eine Verzerrung der Schätzergebnisse hervorgerufen wird. Daher wurde das Modell auch hier nochmals ohne diese Variablen geschätzt und die Ergebnisse mit den vorgestellten Resultaten verglichen. Diese Ergebnisse unterscheiden sich nicht signifikant von den hier dargestellten.

#### **Zusammenfassung der wichtigsten Ergebnisse**

1. Eine MC-Beteiligung von Ärzten führt zu einer signifikanten Leistungsreduktion für alle Patienten. Managed-Care-Ärzte behandeln daher auch konventionelle Patienten kostengünstiger als Ärzte des konventionellen Sektors
2. Patienten, die in Hausarztmodellen eingeschrieben bzw. Mitglied der Zürcher HMO sind, weisen signifikant niedrigere direkte Arztleistungen auf als Patienten des konventionellen Bereichs.

#### **9.4.8 Diskussion der Resultate**

In diesem Abschnitt wurde der direkte Arzt-Patienten-Kontakt untersucht. Das Hauptinteresse lag auf der Wirkung von Managed Care auf die einzelnen Ergebnisvariablen (Überweisungen, Konsultationen, Leistungen des Arztes). Die Managed-Care-Wirkung dokumentiert sich auf der Arztseite durch Kostenkontrolle und Qualitätsüberwachung. Auf der Patientenseite lässt sich die Wirkung vor allem bei der eingeschränkten Arztwahl der Versicherten festmachen. Auch empirisch lassen sich zwei MC-Effekte unterscheiden. So wird zwischen Ärzten und Patienten differenziert, die an MCOs teilnehmen. Eine solche Unterscheidung ist wichtig, da Ärzte, die an Hausarztmodellen beteiligt sind, neben MC-Patienten auch andere Patienten behandeln können.

In Tabelle 28 wird die Wirkung von Managed Care in den einzelnen Modellen nochmals zusammengefasst. Man erkennt deutlich, dass Managed Care Wirkung zeigt. Denn die vorgestellten Ergebnisse wurden sowohl für Selektionserscheinungen auf der Arztseite als auch auf der Patientenseite kontrolliert. Die MC-Beteiligung der Ärzte hat demnach eine negative

Wirkung auf die Wahrscheinlichkeit von Überweisungen und ebenso auf die Wahrscheinlichkeit, dass es während einer Arzt-Patienten-Beziehung zu mindestens einer Konsultation kommt. Dagegen bieten MC-Ärzte mehr Konsultationen ihren Patienten an als Ärzte, die nur (in der Stichprobe) konventionelle Patienten behandeln. Von besonderem Interesse sind jedoch die Ergebnisse hinsichtlich der gesamten bzw. der direkten Arztleistungen. Während der Status des Arztes keinen signifikanten Einfluss auf die gesamten Leistungen ausübt, werden die direkten Leistungen des Arztes – also die Leistungen, die für den Arzt ein direktes Einkommen bedeuten – durch den Beteiligungsstatus des Arztes an MC-Organisationen eindeutig negativ beeinflusst.

Tabelle 28 Managed-Care-Effekte im Gesamtmodell

	Überweisung	Konsultation		Gesamt	Direkt
<b>MC_ARZT</b>	- <sup>a</sup>	- <sup>a</sup>	+	0	-
<b>MC_PAT</b>	+ <sup>a</sup>	- <sup>a</sup>	+	+	-

<sup>a)</sup> Wahrscheinlichkeiten

Die Wirkung des Patientenstatus hinsichtlich einer MC-Beteiligung ist bei allen vier Untersuchungen eindeutig. Am interessantesten ist hier sicherlich der Umstand, dass MC-Patienten weniger direkte Leistungen vom Arzt erhalten als Patienten im konventionellen Bereich. Darin kommt zum Ausdruck, dass aufgrund der besonderen institutionalen Rahmenbedingungen im MC-Bereich (Gatekeeperfunktion des Arztes, Kostenkontrolle) tatsächlich Kosteneinsparungen realisiert werden können. Demgegenüber ist ebenso eindeutig, dass für MC-Patienten insgesamt mehr Leistungen veranlasst werden als für Patienten des konventionellen Sektors. Da die direkte Arzt-Patienten-Beziehung von Allgemeinmedizinern untersucht wurde, entspricht dieses Ergebnis den Erwartungen. Denn während für MC-Patienten so gut wie alle Leistungen erfasst worden sind, trifft das für den konventionellen Bereich nur zum Teil zu. Dort fehlen in der Regel viele Leistungen von Spezialisten, da konventionelle Patienten prinzipiell freie Arztwahl genießen und daher auch ohne Rücksprache mit ihrem Hausarzt einen Spezialisten aufsuchen können.

Die Ergebnisse lassen weitere interessante Rückschlüsse zu. So ist bemerkenswert, dass die beobachteten geringeren Konsultationszahlen für MC-Patienten ausschließlich auf die geringere Konsultationswahrscheinlichkeit von diesen Patienten zurückgehen. Hat ein Patient dagegen mindestens eine Konsultation erhalten, werden den MC-Patienten mehr Konsultationen angeboten als Patienten im konventionellen Bereich. Dieses Ergebnis kann im Zusammenhang

mit den geringeren direkten Arztleistungen auch als positiver Qualitätsaspekt der Managed-Care-Versorgung gedeutet werden. Des Weiteren bestätigen die Untersuchungen der Überweisungswahrscheinlichkeiten und der direkten Arztleistungen die Existenz von so genannten Spill-over-Effekten. Einerseits weisen Ärzte mit einer Managed-Care-Beteiligung für alle behandelten Patienten eine geringere Überweisungswahrscheinlichkeit auf als Ärzte des konventionellen Sektors. Andererseits werden auch direkte Arztleistungen von ihnen unabhängig von der Managed-Care-Beteiligung der Patienten kostengünstiger angeboten als durch ihre Kollegen im konventionellen Bereich.

Um die Managed-Care-Effekte abschätzen zu können, musste verschiedenen Schwierigkeiten bei den Schätzungen Rechnung getragen werden. Neben der möglichen Selektion von Ärzten und Patienten galt hierbei das Hauptaugenmerk der besonderen Datenstruktur. Durch die Berücksichtigung individueller Arzteffekte konnte für die Heterogenität der Ärzte kontrolliert werden. Allerdings zeigte die Datenstruktur auch, dass es im Grunde genommen noch eine zweite Datenhierarchie gibt. So sind Ärzte wie auch Patienten in bestimmten Regionen ansässig und dürften daher bestimmte Gemeinsamkeiten aufweisen. Ein solcher regionaler Faktor kann ebenfalls einen Einfluss auf das Verhalten der Akteure ausüben. Der möglichen Korrelation auf Gemeindeebene mit Patienten- bzw. Arztvariablen wurde durch die Einbeziehung von Variablen begegnet, die eine Verbindung zwischen den einzelnen Ebenen herstellen sollen. In diesem Sinne besitzt die Entfernung zwischen Arzt und Patient eine herausragende Stellung, da sie nicht nur die Suchkosten der Patienten approximieren soll, sondern auch eine Verbindung zwischen der einzelnen Arzt-Patienten-Beziehung zur übergeordneten Gemeindeebene herstellt.

Die betrachteten Modelle wurden konsequent aus den theoretischen Untersuchungen hergeleitet. Dies führte dazu, dass viele abgeleitete Variablen in den empirischen Analysen berücksichtigt werden mussten – wie etwa das Vermögen des Arztes (durchschnittlicher Umsatz der letzten beiden Jahre), die Verbindung des Vermögens mit der Berufserfahrung oder aber das Wachstum des Patientenstammes im letzten Jahr. Durch die Einbeziehung dieser – aus theoretischer Sicht – wünschenswerten Variablen wurde die ohnehin hohe Gefahr der Multikollinearität noch verschärft, sodass von Fall zu Fall der Ausschluss der einen oder anderen Variablen notwendig wurde.

Ein weiteres Problem ergab sich aus der simultanen Betrachtung von Ärzten, die sich an Hausarztmodellen beteiligen und der einzigen im Untersuchungssample befindlichen HMO. Da

die Ärzte innerhalb der HMO mit einem Gehalt vergütet werden, sind die intern verrechneten Leistungen (in der hier gewählten Sprachwahl die direkten Arztleistungen) für die HMO nur bedingt verlässlich. Die vorgestellten Schätzungen berücksichtigen aber die HMO. Damit besteht die Gefahr, dass der MC-Effekt verzerrt geschätzt wird. Um dies auszuschließen, wurden alle Schätzungen auch ohne die Einbeziehung der HMO durchgeführt. An den prinzipiellen Ergebnissen ändert sich dadurch nichts. Insbesondere gilt dies für die interessierenden MC-Effekte. Daneben ändert sich aber auch bei allen anderen Koeffizienten nichts an der Wirkungsrichtung oder der Signifikanz.

Die Vernachlässigung der HMOs in den Schätzungen bietet zusätzlich die Möglichkeit, die Hausarztnetze untereinander zu vergleichen. Zu diesem Zweck wurde abschließend das Modell für die Gesamtleistungen und die direkten Leistungen unter Einbeziehung einer zusätzlichen Dummy-Variablen für das Hausarztnetz Genf geschätzt. Dabei zeigt sich, dass die Genfer Hausärzte sowohl die gesamten Leistungen als auch die direkten Leistungen signifikant kostengünstiger anbieten als ihre Kollegen in den Zürcher Hausarztmodellen. Dieses Ergebnis bestätigt somit die Schlussfolgerungen aus Abschnitt 7, wo gezeigt wurde, dass sich die Netze nicht nur durch vertragliche Regelungen, sondern auch im Selbstverständnis der Ärzte voneinander unterscheiden.

## **10 Zusammenfassung und Ausblick**

Allgemeinmediziner, sind in der Regel die Ersten, die vom Patienten bei gesundheitlichen Problemen aufgesucht werden. Sie besitzen daher eine wichtige Funktion hinsichtlich der gesamten vom Patienten in Anspruch genommenen Leistung. In Hausarztmodellen (HAM) und Gesundheitszentren (HMOs) fungiert der Hausarzt als Lotse des Gesundheitswesens. Im Unterschied zum konventionellen Sektor müssen die Patienten bei gesundheitlichen Beschwerden zuerst ihn aufsuchen. Damit besitzt der Hausarzt die Kontrolle über die gesamte Behandlungskette eines Patienten. Ziel der Untersuchung ist es, mögliche Unterschiede im Verhalten von Ärzten, die sich an Managed Care beteiligen, im Vergleich zu ihren Kollegen im konventionellen Bereich zu erkennen und zu erklären

Die Rolle des Arztes wird hauptsächlich durch doppelseitige Informationsasymmetrien bestimmt. Aufgrund seiner Expertenstellung besitzt der Arzt gegenüber dem Patienten einen Informationsvorsprung. Denn nur er ist in der Lage, die richtige Diagnose zu stellen und

anschließend die notwendige Behandlung durchzuführen oder zu veranlassen. Da die Patienten darüber hinaus Versicherungsschutz genießen, besitzt der Arzt andererseits auch gegenüber den Krankenkassen einen Informationsvorteil. Die Untersuchung derartiger Informationsasymmetrien führt zur Herleitung optimaler Vergütungsregeln der Leistungserbringer. Das optimale Design der Vergütung hängt zudem von der Reputation der Ärzte und den Suchkosten der Patienten ab.

Traditionelle mikroökonomische Modelle des Arztverhaltens verzichten auf die explizite Modellierung der beschriebenen Informationsasymmetrien und nehmen dafür das Verhalten der Ärzte unter unterschiedlichen Marktgegebenheiten (vollständige Konkurrenz, Monopol, monopolistische Konkurrenz) ins Blickfeld. Dabei sind die besonderen Eigenschaften des Wirtschaftsgutes Gesundheit, Unteilbarkeit und die Nicht-Wiederverkäuflichkeit, ausschlaggebend dafür, dass der Markt für Arztdienstleistungen als ein Markt monopolistischer Konkurrenz aufgefasst werden kann: Der Arzt verfügt – auch bei Preisen, die durch den einzelnen Arzt selber nicht beeinflusst werden können – über Marktmacht und ist daher in der Lage, mehr als die von den Patienten gewünschte Gesundheitsleistung bereitzustellen. Unter solchen Marktbedingungen lassen sich wiederum optimale Vergütungsregeln von Ärzten herleiten.

In beiden Modellen wird die optimale Vergütung durch einen zweigeteilten Tarif charakterisiert, der sich aus einer Pauschale und einem leistungsbezogenen Element zusammensetzt. Darüber hinaus bieten die theoretischen Modelle unterschiedliche Einblicke in die Struktur der Nutzenfunktion der Ärzte, die für die empirische Analyse von Wichtigkeit sind.

Die vorhandenen Daten erlauben die Untersuchung des Arztverhaltens auf unterschiedlichen Ebenen. In einer Effizienzanalyse werden die Leistungserbringer über ihre gesamten abgerechneten Leistungen miteinander verglichen. Das Gesamtmodell betrachtet hingegen konkret den einzelnen Arzt-Patienten-Kontakt. Darüber hinaus wird auch eine separate Untersuchung der Nachfrage durchgeführt, bei der sämtliche vom Patienten nachgefragten Leistungen analysiert werden

Bei der Beurteilung der Effizienz steht der Ressourcenaufwand pro Konsultation im Blickpunkt. Es zeigt sich, dass Ärzte, die sich an Managed-Care-Modellen beteiligen, Konsultationen kostengünstiger anbieten, als Ärzte, die nur im konventionellen Bereich tätig sind. Dies muss allerdings nicht notwendigerweise auf die Anreize zurückgeführt werden, die Managed Care den Leistungserbringern bietet. Es könnte sein, dass die Ärzte gerade wegen ihres kostengünstigen

Praxismanagements an derartigen Modellen teilnehmen. Auch eine Selektion der Nachfrager in die Managed-Care-Programme ist möglich. Das Verhalten der Versicherten wird in einer separaten Nachfrageschätzung untersucht. Neben dem Nachweis einer Selektion guter Risiken in Managed-Care-Verträge erbringt die Schätzung der Nachfrage auch die Bestätigung einer kostengünstigeren Versorgung der Versicherten im Managed-Care-Bereich (unter Berücksichtigung des Selektionseffektes).

Das Gesamtmodell des Arztverhaltens untersucht das Verhalten der Ärzte durch die Betrachtung der Überweisungen zu Fachärzten, der Konsultationen, der veranlassten Gesamtleistungen und der direkten Arztleistungen auf der Ebene des Arzt-Patienten-Kontakts. Bei allen Untersuchungen wird für eine mögliche Selektion seitens der Ärzte und der Patienten kontrolliert.

In den beiden Hauptuntersuchungen werden alle veranlassten Leistungen sowie die Leistungen, die direkt zu einem Einkommen für den Arzt führen, analysiert. Beide Schätzungen beinhalten neben den Korrekturtermen für die mögliche Selektion von Patienten und Ärzten weitere Variablen, welche die spezielle Arzt-Patienten-Beziehung abbilden. Dazu gehört neben den Ergebnissen hinsichtlich der Kontakthäufigkeit und der Überweisungswahrscheinlichkeit aus den vorangegangenen Schätzungen auch die Einbeziehung einer Variable für die Regelmäßigkeit der Arzt-Patientenbeziehung.

Das Gesamtmodell zeigt, dass die Anreize in Managed Care Wirkung entfalten: Unter Berücksichtigung der Selektion von Patienten und Ärzten, der Besonderheiten der konkreten Arzt-Patienten-Beziehung und den Unterschieden zwischen veranlassten und direkten Arztleistungen ist die Leistungserstellung im Managed-Care-Bereich hinsichtlich der direkten Arztleistungen deutlich kostengünstiger als im konventionellen Sektor.

Die empirische Analyse belegt damit insgesamt, dass es durch die Anwendung von Managed-Care-Instrumenten zu einer besseren Versorgung der Patienten kommt. Sie bietet zugleich Wege an, die Möglichkeiten von Managed Care noch besser zu nutzen. So ist durch die Einbeziehung echter Kostenbeteiligungsmodelle auf der Angebotsseite – wie sie in der Schweiz teilweise bereits vollzogen wurde – ein weiterer Effizienzgewinn zu erwarten. Eine Aufhebung des Kontrahierungszwangs kann ebenfalls zu einer kostengünstigeren Versorgung der Patienten beitragen: einerseits durch die gezielte Auswahl der wirtschaftlichsten Anbieter durch die Versicherungen und andererseits durch die entstehenden Anreize für alle Ärzte, ihren Praxisstil verstärkt auf eine kostengünstigere Versorgung und auf die Bedürfnisse der Patienten auszu-



### III Empirische Untersuchung

---

richten. Obwohl die Voraussetzungen in Deutschland vollkommen anders sind, könnten die angesprochenen Punkte wesentliche Bestandteile einer Reform der ambulanten Versorgung auch in Deutschland sein.

## Literaturverzeichnis

- AGZ (1999), Statuten der Ärztesgesellschaft Zürich, [www.aerzte-zh.ch/content/deu/documents/d200212301017076541.pdf](http://www.aerzte-zh.ch/content/deu/documents/d200212301017076541.pdf), Stand. 30.4.03.
- Ai, C. und E.C. Norton (2003), Interaction terms in logit and probit models, *Economics Letters* 80, 123-129.
- Altman, D., D. Cutler und R. Zeckhauser (2003), Enrollee mix, treatment intensity, and cost in competing indemnity and HMO plans, *Journal of Health Economics* 22, 23-45.
- Amelung, V. und H. Schumacher (2000), *Managed Care*, Gabler, München.
- Angrist, J. (2000), Estimation of Limited-Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice, NBER Technical Working Paper 248, NBER, Cambridge MA.
- Arrow, K. (1963), Uncertainty and the welfare economics of medical care, *American Economic Review* 53, 941-973.
- Arrow, K. (1971), *Essays in the Theory of Risk Bearing*, Markham, Chicago.
- Baker, L. (1999), Association of Managed Care Market Share and Health Expenditures for Fee-for-Service Medicare Patients, *JAMA* 281, 5, 432-437.
- Battese, G. E. und T. J. Coelli (1992), Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India, *Journal of Productivity* 3, 153-169.
- Battese, G. E. und T. J. Coelli (1995), A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics* 20, 325-332.
- Beck, K. (2000), Wenn der Leistungsanstieg statistisch gar nicht nachweisbar ist – Prognosemodelle der Krankenversicherer im Test, *Managed Care* 6, 43-46.
- Beck, K. (2001), Capitationberechnung in der Schweiz: The State of the Art, *Managed Care I*, 12-16.
- Beck, K. (2003), Risiko Krankenversicherung – Risikomanagement in einem regulierten Krankenversicherungsmarkt, *mimeo*, Luzern.
- Benstetter, F. (2002), *Health Care Economics: The Market for Physician Services*, Lang, Frankfurt am Main u.a.
- Bhattacharyya, S. und F. Lafontaine (1995), Double-sided moral hazard and the nature of share contracts, *Rand Journal of Economics* 26, 4, 461-781.
- Blomquist, A. (1991), The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics* 10, 411-432.
- Böhlert, I., I. Adam. und B.-P. Robra (1997), Das Schweizer Gatekeepersystem – ein Modell zur Verbesserung der Leistungsentwicklung und Wirtschaftlichkeit, *Das Gesundheitswesen* 59, 488-493.
- Brasure, M., S. Stearns, E. C. Norton, und Th. III. Ricketts (1999), Competitive Behavior in Local Physician Markets, *MCR&R* 56, 4, 395-414.

- Bresnahan, T. und P. Reiss (1991), Entry and competition in concentrated markets, *Journal of Political Economy* 99, 977-1009.
- Breyer, F. und P. Zweifel (1999), *Gesundheitsökonomie*, Springer, Berlin u.a.
- Breyer, F., P. Zweifel und M. Kifmann (2003), *Gesundheitsökonomie*, Springer, Berlin u.a.
- Buchs, L. (2001), Managed Care, in: Kocher, G. (Hrsg.), *Gesundheitswesen Schweiz 2001/2002 – ein aktueller Überblick*, Konkordat der Schweizerischen Krankenversicherer (KSK), Solothurn.
- Bundesamt für Sozialversicherung (1998), Evaluation neuer Formen der Krankenversicherung, Synthesebericht, Forschungsbericht Nr1/98, BSV, Bern.
- Bundesamt für Sozialversicherung (2000a), Statistik der wählbaren Franchisen in der Krankenversicherung 1999, BSV, Bern.
- Bundesamt für Sozialversicherung (2000b), Bestandsaufnahme besonderer Versicherungsformen in der obligatorischen Krankenversicherung, Forschungsbericht Nr. 15/00, BSV, Bern.
- Bundesamt für Sozialversicherung (2001a), Wirkungsanalyse Krankenversicherungsgesetz - Synthesebericht, BSV, Bern.
- Bundesamt für Sozialversicherung (2001b), Schweizerische Sozialversicherungsstatistik 2001, BSV, Bern.
- Bundesamt für Statistik (2000), Analyse der Auswirkungen des KVG auf die Finanzierung des Gesundheitswesens und anderer Systeme der sozialen Sicherheit, Bundesamt für Statistik (BfS), Neuchâtel.
- Coelli, T. J. (1996), A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation, CEPA Working Paper 96/07, University of New England, Australia.
- Coelli, T. J., D. S. Prasada Rao und G.E. Battese (1998), *An Introduction to efficiency and Productivity Analysis*, Kluwer, Boston MA u.a.
- CSS-Versicherung (1998), Geschäftsbericht für 1997, CSS-Versicherung, Luzern.
- Cutler, D., M. McClellan, und J. P. Newhouse (2000), How does Managed Care do it?, *RAND Journal of economics* 27, 1, 99-123.
- Danthine, J.-P. und J.B. Donaldson (2002), *Intermediate Financial Theory*, Prentice-Hall, New York.
- Deaton, A. und J. Muelbauer (1980), *Economics and consumer behavior*, Cambridge University Press, New York.
- Dranove, D. und A. Satterthwaite (2000), The Industrial Organization of Health Care Markets, in: A. Culyer und J. P. Newhouse (Hrsg.), *Handbook of Health Economics*, Elsevier, Amsterdam.
- Dranove, D. und P. Wehner (1994), Physician-induced demand for childbirds, *Journal of Health Economics*, 13, 61-73.
- Dudley, R. A. und H. S. Luft (2001), Managed Care in Transition, *The New England Journal of Medicine* 344, 14, 1087-1092.
- Egglestone, K. (2001), Multitasking, Competition and Provider Paying, Working Paper, Tufts University, Medford.

- Eidgenössisches Departement des Innern (2002), Soziale Krankenversicherung Analyse, <http://www.bsv.admin.ch/aktuell/presse/2002/d/02052301.pdf>, Stand: 02.09.02.
- Ellis, R. P. (1998), Creaming, Skimping and Dumping: provider competition on the intensive and extensive margins, *Journal of Health Economics* 17, 537-555.
- Emons, W. (2001), Credence goods monopolists, *International Journal of Industrial Organization* 19, 375-389.
- Encinosa III, W.E., M. Gaynor und J.B Rebitzer (1997), The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems, Working Paper 5953, NBER, Cambridge MA.
- Evans, R. G. (1974), Supplier-induced demand: some empirical evidence and implications, in Perlman, M. (Hrsg.), *The Economics of Health and Medical Care*, Macmillan, Edinburgh.
- Farrell (1957), The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A* 120, 253-281.
- Felder, S. (2001), Zeitgespräch – Reform des Risikostrukturausgleichs in der GKV – Wehret den Anfängen! Gegen Morbiditätsfaktoren im Risikostrukturausgleich, *Wirtschaftsdienst: Zeitschrift für Wirtschaftspolitik* 81, 4, 98-102.
- Feldstein, P. J. (1999), *Health Care Economics*, Albany, New York u.a.
- Folland, Sh. (1997), *The economics of Health and Health Care*, Prentice Hall, New York.
- Forum Managed Care (2000a), Managed-Care-Modelle in der Schweiz, HMOs (2000) [www.fmc.ch/mc00-3hmos.htm](http://www.fmc.ch/mc00-3hmos.htm), Stand: 22.05.2000.
- Forum Managed Care (2000b), Managed-Care-Modelle in der Schweiz, Hausarztmodelle (2000) [www.fmc.ch/mc99-1ham.htm](http://www.fmc.ch/mc99-1ham.htm), Stand: 22.05.2000.
- Frech III, H. E. (1996), *Competition and Monopoly in Medical Care*, The AEI Press, Washington D.C.
- Frech III, H. E. (Hrsg.) (1991), *Regulating Doctors' Fees: Competition, Benefits, and Controls under Medicare*, The AEI Press, Washington, D.C.
- Frech III, H. E., Langenfeld, J. und M. Corbett (2000), Managed Health Care Effects: Medical Care Costs and Access to Health Insurance, Working Paper, American Association of Health Plans.
- Frei, W. (2002), Krankenversicherer, in: Kocher, G. (Hrsg.) *Gesundheitswesen Schweiz 2001/2002 – ein aktueller Überblick*, Solothurn: Konkordat der Schweizerischen Krankenversicherer (KSK).
- Fuchs, V. (1974), *Who shall live?* Basic Books, New York.
- Gaynor, M. (1994), Issues in the Industrial Organization of the Market for Physician Services, *Journal of Economics and Management Strategy* 3, 211-255.
- Gaynor, M. und P. Gertler (1995), Moral hazard and risk spreading in partnerships, *RAND Journal of Economics* 26, 4, 591-613.
- Getzen, T. (1997), *Health Economics*, Wiley, New York u.a.
- Giuffrida, A. und H. Gravelle (2001), Measuring performance in primary care: econometric analysis and DEA, *Applied economics* 33, 2, 163-176.

- Glazer, J. und T. G. McGuire (2000), Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care, *American Economic Review* 90, 4, 1055-1071.
- Glied, S. (2000), Managed Care, in: A. Culyer und J. P. Newhouse (Hrsg.), *Handbook of Health Economics*, Elsevier, Amsterdam.
- Glied, S. und J. Zivin (2000), *How do doctors behave when some (but not all) of their patients are in managed care?* NBER Working Paper 7907, NBER, Cambridge MA.
- Glied, S. und J. Zivin (2002), How do doctors behave when some (but not all) of their patients are in managed care? *Journal of Health Economics* 21, 337-353.
- Goethe, W. von, Faust, in: Bibliothek Deutscher Klassiker, Goethes Werke in zwölf Bänden (1988), Aufbau-Verlag, Berlin und Weimar.
- Gonzales, X.M. und D. Miles (2002), Statistical precision of DEA: a bootstrap application to Spanish public services, *Applied economics letters* 9, 2, 127-132.
- Greene, W. (1997), *Econometric Analysis*, Prentice-Hall, London u.a.
- Grossman, S.J. und O. D. Hart (1983), An Analysis of the Principal-Agent Problem, *Econometrica* 51, 1, 7-45.
- Gujarati, D. N. (2000), *Basic Econometrics*, McGraw-Hill, New York et al.
- Hadley, J. (1991), Theoretical and Empirical Foundations of the Resource-Based Relative Value Scale, in H. E. Frech (Hrsg.), *Regulating doctor's fees: Competition, Benefits, and Controls under Medicare*, The AEI Press, Washington, D.C.
- Halvorsen R. und R. Palmquist (1980), The Interpretation of Dummy-Variables in Semilogarithmic Equations, *American Economic Review* 70, 3, 474-475.
- Hänggeli, Chr., J. Jau, E. Eicher und S. Bradke (2001), Freipraktizierende Ärztinnen und Ärzte, in: G. Kocher und W. Oggier, *Gesundheitswesen Schweiz 2001/2002: Ein aktueller Überblick*, Konkordat der Schweizer Krankenkassen (KSK), Solothurn.
- Heckman, J. (1979), Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-161.
- Heckman, J., H. Ichimura und P. Todd (1998), Matching As An Econometric Evaluation Estimator, *Review of Economic Studies* 65, 261-294.
- HEDIS (2000) [www.ncqa.org](http://www.ncqa.org) (Stand: 23.08.02).
- Hellinger, F.J. (1996), The Impact of Financial Incentives on Physician Behavior in Managed Care Plans: A Review of the Evidence, *Medical Care Research and Review* 53, 3, 294-314.
- Hillman, A. L. (1995), The Impact of Physician Financial Incentives on High-Risk Populations in Managed Care, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 8, 1, 23-30.
- Hillman, A., P. Welch, und M. Pauly (1992), Contractual Arrangements Between HMOs and Primary Care Physicians: Three-Tiered HMOs and Risk Pools, *Medical Care* 30, 2, 136-148.
- Hirshleifer, J. und J. G. Riley (1979), The Analytics of Uncertainty and Information – An Expository Survey, *Journal of Economic Literature* 17, 1375-1421.
- Hirshleifer, J. und J. Riley (1992), *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge.

- Hollingsworth, B., P. J. Dawson und N. Maniadakis (1999), Efficiency measurement of health care: a review of non-parametric methods and applications, *Health Care Management Science* 2, 161-172.
- Huber, F., C. Marti, S. Götschi und A. Weber (2002), Managed Care in der Schweiz, *Schweizerische Ärztezeitung* 83, 2629-2632.
- IGES, D. Cassel und J. Wasem (2001), Zur Wirkung des Risikostrukturausgleichs in der gesetzlichen Krankenversicherung, Untersuchung im Auftrag des Bundesministeriums für Gesundheit, Endbericht.
- Iversen, T. und H. Luras (2000), The Effect Of Capitation On GPs' Referral Decision, *Health Economics* 9, 199-210.
- Jacobs, R. (2001), Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis, *Health Care Management Science* 4, 103-115.
- Jelovac, I. (2001), Physician's payment contracts, Treatment decisions and diagnosis accuracy, *Health Economics* 10, 9-25.
- Jones, A. (2000), Health Econometrics, in: A. Culyer und J. P. Newhouse (Hrsg.), *Handbook of Health Economics*, Elsevier, Amsterdam.
- Karmann, A. (1994), Multiple-tasks and multiple-agent models: Incentive contracts and an application to point pollution control, *Annals of Operations Research* 54, 57-78.
- Kim, K. S. und S. Wang (1998), Linear Contracts and the Double Moral-Hazard, *Journal of Economic Theory* 82, 342-378.
- Kimball, M. S. (1990), Precautionary Savings in the Small and in the Large, *Econometrica* 58, 53-73.
- Krauth C, F. W. Schwartz, M. Perleth, K. Buser, R. Busse und J.-M. Graf von der Schulenburg (1997), Zur Weiterentwicklung des Vergütungssystems in der ambulanten ärztlichen Versorgung, Medizinische Hochschule Hannover.
- Kristiansen, I. (1995), What is in the doctor's utility function? A theoretical and empirical investigation into what influences doctor's decision making, PhD-Thesis, University of Tromsø, Norway.
- Kühn, H. (1997), Managed Care: Medizin zwischen kommerzieller Bürokratie und integrierter Versorgung. Am Beispiel USA, Arbeitspapier P 97-202, Wissenschaftszentrum Berlin.
- Kumbhakar, S. C. und C. A. K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.
- Kupsch, S., A. Kern, J. F. Hallauer und F. Beske (1998), Selbstbeteiligung im Gesundheitswesen im westeuropäischen Vergleich, *Gesundheitsökonomie und Qualitätsmanagement* 3, 20-29.
- KVG, Bundesgesetz über die Krankenversicherung (Stand 1999), Bern.
- Labelle, R., G. Stoddart, und T. Rice (1994), A re-examination of the meaning and importance of supplier-induced demand, *Journal of Health Economics* 13, 347-368.
- Lechner, M. (2002), Mikroökonomische Evaluation arbeitsmarktpolitischer Maßnahmen, Working Paper, Universität St. Gallen.

- Leonard, K.L. und J. Zivin (2000), How To Compensate Physicians When Both Patient and Physician Effort are Unobservable, Working Paper, Columbia University, New York.
- Li, T. und R. Rosenman (2001), Cost Inefficiency in Washington Hospitals: A Stochastic Frontier Approach Using Panel Data, *Health Care Management Science* 4, 73-81.
- LIMDEP (1998), LIMDEP, Version 7.0, Economic Software Inc., New York.
- Luft, H. S. (1995), HMOs, Market Competition and Premium Cost, *Journal of Health Economics* 14, 115-119.
- Luft, H. S. (1999), Why Are Physicians So Upset about Managed Care?, *Journal of Health Policy and Law* 24, 5, 957-965.
- Maddala, G.S. (1985), A survey of the literature on selectivity bias as it pertains to health care markets, *Advances in Health Economics and Health Services Research* 6, 3-18.
- Manning, W. (1998), The logged dependent variable, heteroscedasticity, and the retransformation problem, *Journal of Health Economics* 17, 283-295.
- Manning, W. G., Newhouse, J. P., Duan, N., Keeler E. B., Leibowitz, A. and S. Marquis (1987), Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment, *American Economic Review* 77, 3, 251-277.
- Marty, F. (1997), Zur Verrechnungspraxis der Ärzte, Working Paper, Universität Bern.
- McGuire, T.G. (2000), Physician Agency, in: A. Culyer, und J. P. Newhouse (Hrsg.), *Handbook of Health Economics*, Elsevier, Amsterdam.
- McKelvey, R. D. und W. Zavoina (1975), A Statistical Model for the Analysis of Ordinal Level Dependent Variables, *Journal of Mathematical Sociology* 4, 103-120.
- Mechanic, D., D.D. McAlpine und M. Rosenthal (2001), Are Patients' Office Visits with Physicians Getting Shorter?, *The New England Journal of Medicine* 344, 3, 198-204.
- Milgrom, P. (1981), Good news and bad news: representation theorems and applications, *The Bell Journal of Economics* 12, 380-391.
- Milgrom, P. (1987), Adverse Selection without hidden Information, Working Paper 8742, University of California, Berkeley.
- Milgrom, P. und J. Roberts (1992), *Economics, organizations, and management*, Printice-Hall, London u.a.
- Miller, R. H. und H. S. Luft (2002), HMO plan performance update: An analysis of the literature, 1997-2001, *Health Affairs* 21(4), 63-86.
- Mojon-Azzi, S. (2001a), Gruppenpraxen – Was versteht man darunter?, *Schweizerische Ärztezeitung* 82, 22, 1146-1150.
- Mojon-Azzi, S. (2001b), Gruppenpraxen – Ihre Vor- und Nachteile, *Schweizerische Ärztezeitung* 82, 22, 1151-1155.
- Mooney, G. und M. Ryan (1993), Agency in health care: getting beyond first principles, *Journal of Health Economics* 12, 125-135.
- Mühlenkamp, H. (2001), Die Rolle von Managed Care im US-amerikanischen Gesundheitswesen - aktuelle Entwicklungen und gegenwärtige Erkenntnisse, Arbeitsbericht, 3/2000, Institut für Haushalts- und Konsumökonomik, Universität Hohenheim.

- Newhouse, J. P. (1994), Frontier estimation: How useful a tool for health economics, *Journal of Health Economics* 13, 317-322.
- Newhouse, J. P. (1996), Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection, *Journal of Economic Literature* 34, 1236-1263.
- Nicolet, Ch., Geoffard, P.-Y. und L. Gardiol (2001), Testing for Adverse Selection and Moral Hazard: an Econometric Study of Deductibles and Health Expenditures on Swiss Insurance Claims Data, Working Paper, University Lausanne.
- Pai C.-W., Y. A. Ozcan und H. J. Jiang (2000), Regional Variation in Physician Practice Pattern: An Examination of Technical and Cost Efficiency for Treating Sinusitis, *Journal of Medical Systems* 24, 2, 103-117.
- Pauly, M. (1980), *Doctors and Their Workshops*, University Press, Chicago.
- Pauly, M. und S. Nicholson (1999), Adverse Consequences of Adverse Selection, *Journal of Health Politics, Policy and Law* 24, 5, 921-930.
- Perneger, Th., J-F. Etter, und B. Schaad (1995), Evaluation der HMO der Universität Genf, ISPN, Universität Genf, Studien und Forschungen 6/1995.
- Pharma Information (1999), Das Gesundheitswesen in der Schweiz, Pharma Information – Verband der forschenden pharmazeutischen Firmen der Schweiz, Basel.
- Phelps, C. E. (1997), *Health Economics*, Addison-Wesley, New York u.a.
- Plato (1991), *The republic*, Vintage Books, New York.
- Pohlmeier, W. und V. Ullrich (1994), An Econometric Model of the Two-Part Decision Making Process in the Demand for Health Care, *The Journal Of Human Resources* XXX, 2, 339-361.
- Polsky, D. und S. Nicholson (2000), HMOs and Risk Selection, Working Paper, University Pennsylvania.
- Pratt, J.W. (1964), Riskaversion in the Small and in the Large, *Econometrica* 32, 122-136.
- Preisüberwacher (1999), Studie zu den Tarifen GRAT/Infra, Eidgenössisches Volkswirtschaftsdepartement, Bern.
- Räbiger, J., U. Hasenbein, S. Klatt, M. Sinha, H. M. Brenner und K.-D. Henke (2002), Konzept für eine standardisierte Evaluation Managed-Care-orientierter Versorgungsmodelle in Deutschland: ein Werkstattbericht, in Preuß, K.-J., Räbiger, J. und J. H. Sommer (Hrsg.) *Managed Care, Evaluation und Performance-Measurement integrierter Versorgungsmodelle*, Schattauer, Stuttgart.
- Reinhardt, U. (1975), *Physician Productivity and Demand for Health Manpower*, Ballinger Publishing Company, Cambridge MA.
- Rice, N. und A. Jones (1997), Multilevel Models And Health Economics, *Health Economics* 6, 561-575.
- Ritzmann, I. (1999), Der Verhaltenskodex des „Savoir faire“ als Deckmantel ärztlicher Hilflosigkeit?, *Gesenerus* 56, 197-219.
- Robinson, J. (2001), The End of Managed Care, *JAMA* 285, 2622-2628.



- Robinson, R. (2000), Managed Care in the United States: A Dilemma for Evidence-Based Policy?, *Health Economics* 9, 1-7.
- Robinson, R. und A. Steiner (1998), *Managed Health Care*, Open University Press, Buckingham (Phil).
- Rosko, M. D. (2001) Cost Efficiency of US Hospitals: A Stochastic Frontier Approach, *Health Economics* 10, 539-551.
- Salanié, B. (1997), *The Economics of Contracts, A Primer*, MIT Press, Cambridge.
- Schaller, Ph. und M.-A. Raetzo (2002), Réseau Delta: Betrachtung nach zehn Jahren Erfahrung, *Managed Care* 2, 19-22.
- Schellhorn (2001), The Effect Of Variable Health Insurance Deductibles on the Demand For Physician Visits, *Health Economics* 10, 441-456.
- Schmidt, C.M. (2000), Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme, *Vierteljahreshefte zur Wirtschaftsforschung* 69, 3, 425-437.
- Schneider, U. (2002), *Theorie und Empirie der Arzt-Patient-Beziehung*, Peter Lang, Frankfurt a.M u.a.
- Schulenburg, H.-M. Graf von der (1981), *Systeme der Honorierung frei praktizierender Ärzte und ihre Allokationswirkungen*, Mohr, Tübingen.
- Schweizer, U. (1999), *Vertragstheorie*, Mohr Siebeck, Tübingen.
- Schweizerische Ärztezeitung (1999), Das Hausarztssystem WintiMed, Bilanz und Auswertung nach vier Jahren, *Schweizerische Ärztezeitung* 80, 27, 1663-1668.
- Scott, A. (2000), Economics of General Practice, in: A. Culyer und J. P. Newhouse (Hrsg.), *Handbook of Health Economics*, Elsevier, Amsterdam.
- Scott, A. und A. Shiell (1997a), Analysing the Effect of Competition on General Practitioners' Behaviour using a Multilevel Modelling Framework, *Health Economics* 6, 577-588.
- Scott, A. und A. Shiell (1997b), Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model, *Journal of Health Economics* 16 (1997) 323-342.
- Scott, A. und S. Vick (1999), Patients, Doctors and Contracts: An Application of Principal-Agent Theory to the Doctor-Patient Relationship, *Scottish Journal of Political Economy* 46, 2, 111-134.
- Selden, Th. (1990), A Model of Capitation, *Journal of Health Economics* 9, 397-409.
- Shaw, B. (1991) *Doktors Dilemma*, Suhrkamp-Taschenbuch, Frankfurt a.M.
- Sinn, H.-W. (1989), *Economic Decisions under Uncertainty*, Second Edition, Physica-Verlag, Heidelberg.
- Snow und Thurston (2000), *The „Micro-Macro Problem“: Implication for Research on Physician Behavior*, Working Paper, Brigham Young University, Provo, Utah.
- Staat, M. (2000), Die relative Effizienz ärztlicher Leistungen: eine empirische Analyse praktischer Ärzte in Österreich, in: *Ökonomie und Politikberatung im Spannungsfeld von Theorie und Praxis*, INFER-Jahrstagung 2000, S. 199-220, Berlin.

- Stigler, G. und R. Sherwin (1985), The extent of the Market, *Journal of Law and Economics* 28, 555-585.
- Sue, M. (2000), Measuring Practice Performance: Are You Providing Quality Care?, *Clinicians Review* 10, 5, 132-134, 137, 138.
- Thiele, H. und A. Wambach (1999), Wealth Effects in the Principal Agent Model, *Journal of Economic Theory* 89, 247-260.
- Tirole, J. (1995), *Industrieökonomik*, Oldenbourg-Verlag, München.
- Varian, H. (1992), *Microeconomic Analysis*, Norton & Company, Inc., New York.
- Vella, F. und M. Verbeek (1999), Estimating and interpreting models with endogenous treatment effects, *Journal of business & economic statistics* 17, 473-478.
- Verbeek, M. (2001), *A Guide To Modern Econometrics*, Wiley, Chichester u.a.
- Wedig, G. (1993), Ramsey pricing and supply-side incentives in physician markets, *Journal of Health Economics* 12, 365-384.
- Werbblow, A. und S. Felder (2003), Der Einfluss von freiwilligen Selbstbehalten in der gesetzlichen Krankenversicherung: Evidenz aus der Schweiz, *Schmollers Jahrbuch* 123, 2, 235-264.
- Whynes, D., Ch. T. Ennew und T. Feighan (1999), Entrepreneurial attitudes of primary health care physicians in the United Kingdom, *Journal of Economic Behavior & Organization*, 38, 331-347.
- Winkelmann, R. (2001), Health Care Reform and the Number of Doctor Visits – An econometric analysis, Working Paper No. 317, IZA, Bonn.
- Wintimed (2000), Statuten des Hausarztnetzes Wintimed, [www.wintimed.ch/statuten.htm](http://www.wintimed.ch/statuten.htm), Stand: 30.4.03.
- Wolinsky, A. (1993), Competition in a market for informed experts' services, *RAND Journal of Economics* 24, 3, 380-398.
- Wong, H. S. (1996), Market structure and the role of consumer information in the physician services industry: An empirical test, *Journal of Health Economics* 15, 139-160.
- Zweifel, P. (1994), Eine Charakterisierung von Gesundheitssystemen: Welche sind von Vorteil bei welchen Herausforderungen? In: P. Oberender (Hrsg.), *Probleme der Transformation im Gesundheitswesen*, Nomos, Baden-Baden.
- Zweifel, P. und R. Eisen (2001), *Versicherungsökonomie*, Springer Verlag, Heidelberg u.a.
- Zweifel, P., H. Lehman und L. Steinmann (2002), Patching up the physician- patient relationship: insurers vs. Governments as complementary agents, in: B. Lindgren (Hrsg.), *Individual Decisions for health*, Taylor & Francis, Routledge.
- Zweifel, P., S. Felder und M. Meier (1999), Aging of Population and Health Care Expenditure: A Red Herring?, *Health Economics* 8, 485-496.

## Anhang I

### Soziale Krankenversicherung in der Schweiz

Erste Krankenkassen entstanden in der Schweiz bereits im 19. Jahrhundert. Das erste Krankenversicherungsgesetz wurde 1911 verabschiedet.<sup>249</sup> Seit 1996 ist ein neues Krankenversicherungsgesetz (KVG) in Kraft.<sup>250</sup> Während das alte Krankenversicherungsrecht nur begrenzt soziale Aspekte berücksichtigte,<sup>251</sup> führte das neue Gesetz den sozialen Ausgleich zwischen Frauen und Männern sowie zwischen Alten und Jungen durch die Einführung gleicher Pro-Kopf-Beiträge ein. Die Versicherten können ihre Krankenkasse frei wählen. Die Versicherer hingegen unterstehen einem Kontrahierungszwang, der sie zwingt, alle Versicherten aufzunehmen. Die Versicherer stehen damit im Wettbewerb um die Gunst der Versicherten. Dieser Wettbewerb zwischen den Kassen wird durch einen Risikoausgleich begleitet, der – trotz des Kontrahierungszwanges – einen möglichen Wettbewerb um die sogenannten guten Risiken verhindern soll.

Die Prämie entspricht in der Schweiz – wie erwähnt – einer Kopf-Abgabe, d.h. jeder Versicherte – unabhängig von Alter, Geschlecht und Einkommen – muss innerhalb der Kasse und einer Region den gleichen Beitrag leisten. Nur Kinder (0-18 Jahre) und junge Erwachsene (19-25 Jahre) zahlen (unterschiedlich) ermäßigte Prämien.<sup>252</sup> Im Unterschied zu Deutschland können die Prämien innerhalb einer Kasse über die Regionen entsprechend der Kosten variieren.<sup>253</sup> Über alle Kassen betrachtet, ergibt sich dabei folgendes Bild: Zum einen unterscheiden sich die

---

<sup>249</sup> Vgl. Frei (2002).

<sup>250</sup> Vgl. KVG Art. 107. Des Weiteren sei darauf hingewiesen, dass es sich bei der vorliegenden Darstellung der Schweizer Krankenversicherung nur um einen groben Überblick handelt. Mittlerweile existieren eine Reihe von Publikationen, die das Schweizer System nach der Einführung des neuen Krankenversicherungsgesetzes in einzelnen Teilen oder im vollen Umfang darstellen, so z.B. Böhlert et al. (1997) oder Bundesamt für Sozialversicherung (1998).

<sup>251</sup> Es konnten zwar differenzierte Prämien bezüglich Alter und Geschlecht erhoben werden, die Prämien zwischen Männern und Frauen durften aber nicht mehr als 10% differieren. Auch für eine Altersdifferenzierung gab es Grenzen, da sich die Prämie nach dem Eintrittsalter der Versicherten richten musste (vgl. Bundesamt für Sozialversicherung 2001a).

<sup>252</sup> Bei den folgenden Vergleichen wird nur auf die Erwachsenenprämie Bezug genommen.

<sup>253</sup> Der Gesetzgeber gestattet es den Kassen, in jedem Kanton bis zu drei Risikozonen einzurichten, wobei die höchste Prämie in einem Kanton nicht höher als ein bestimmter Prozentsatz der niedrigsten Prämie sein darf (immer bezogen auf eine Kasse). Die noch gültige Regelung wird ab dem 1.1.2004 durch eine einheitliche Prämienregionengestaltung abgelöst werden. Vgl. auch die Mitteilungen des BSV hierzu ([www.bsv.admin.ch](http://www.bsv.admin.ch)).

durchschnittlichen Prämienhöhen zwischen den Kantonen erheblich. So betrug im Jahre 2001 die durchschnittliche Monatsprämie in den günstigsten Kantonen weniger als 170 SFr., während sie in den teuersten Kantonen 230 SFr. überstieg (bei einem gesamtschweizerischen Mittel von 223 SFr.). Der absolut teuerste Kanton ist Genf mit durchschnittlich 330 SFr.<sup>254</sup> Insgesamt lässt sich bei der Verteilung der Prämien in der Schweiz ein deutliches Gefälle zwischen der deutschsprachigen Schweiz (eher günstig) und der französisch- bzw. italienischsprachigen Schweiz (eher teuer) ausmachen. Darüber hinaus besteht auch eine starke Differenzierung zwischen eher ländlichen Kantonen und eher städtischen Kantonen. Neben kantonalen Unterschieden existieren auch innerhalb der Kantone erhebliche Unterschiede. Während bspw. im Kanton Zürich die Differenz zwischen höchster und niedrigster Prämie ungefähr 170 SFr. beträgt, macht diese Differenz in Appenzell gerade mal 30 SFr. aus.<sup>255</sup> Abbildung 10 macht die Zusammenhänge in einem so genannten Box-Plot-Diagramm deutlich. Innerhalb der Rechtecke befinden sich für jeden Kanton 50% der von den Versicherern verlangten Prämien. Die horizontale Markierung innerhalb der Rechtecke markiert die Median-Prämie für den Kanton. Die Enden der vertikalen Striche kennzeichnen das 5%- bzw. 95%-Percentil.<sup>256</sup>

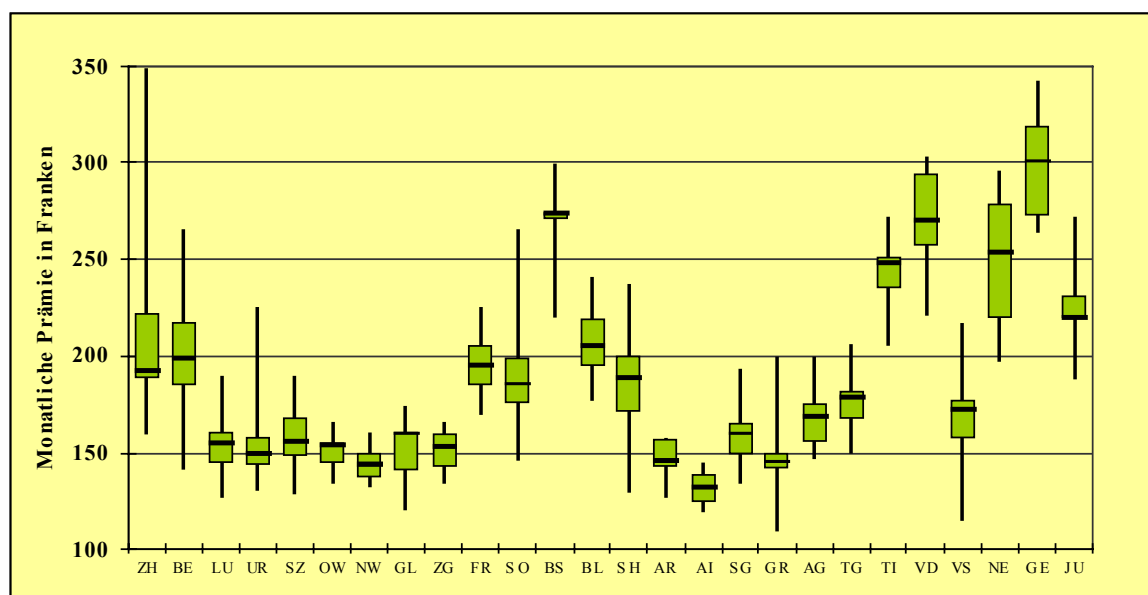
---

<sup>254</sup> Vgl. Bundesamt für Sozialversicherung (2001b).

<sup>255</sup> Vgl. auch hier Bundesamt für Sozialversicherung (2001b).

<sup>256</sup> Es ist zu beachten, dass es nicht in allen Kantonen 3 unterschiedliche Prämienregionen gibt. So besitzen die Kantone Uri, Schwyz, Obwalden, Nidwalden, Glarus, Appenzell Innerrhoden und Appenzell Ausserrhoden nur eine Prämienregion.

Abbildung 10 Kantonale und interkantonale Unterschiede in den Erwachsenenprämien 1999



Quelle: Bundesamt für Sozialversicherung (2001b).

Betrachtet man die Prämienentwicklung seit Einführung des neuen KVGs, so ergibt sich eine Gesamtteuerung von ca. 34% (166. SFr. im Jahre 1996, 223 SFr. im Jahre 2001), wobei die durchschnittliche jährliche Teuerung bei ca. 5.6 % lag.

Des Weiteren gibt es im Unterschied zu Deutschland eine allgemeine Selbstbeteiligung der Versicherten in der sozialen Krankenversicherung der Schweiz. Die Selbstbeteiligungsregelung besteht aus zwei Teilen: der sogenannten *Franchise*, einem Festbetrag, bis zu dem alle Leistungen aus der Tasche des Patienten bezahlt werden, und einer prozentualen Selbstbeteiligung (*Selbstbehold*), die für jede Inanspruchnahme oberhalb der Franchise gezahlt werden muss.<sup>257</sup> Dabei hat der Gesetzgeber vorgesehen, dass die Wahl einer höheren Franchisestufe mit einem Prämienrabatt verbunden werden kann. Es zeigt sich, dass die meisten Versicherer die maximal möglichen Rabatte gewähren.<sup>258</sup> Die prozentualen Rabattsätze führen in Kantonen mit vergleichsweise hohen Prämien zu höheren Prämieinsparungen (in absoluten Zahlen) als in Kantonen mit relativ niedrigen Prämien. Es ist deshalb auch kein Wunder, dass es in den „teuren“ Kantonen

<sup>257</sup> Die Franchisestufen in der Schweiz betragen 230, 400, 600, 1200 und 1500 Franken. Der Selbstbehold beträgt 10% (Stand: 2001).

<sup>258</sup> Bei der Franchisestufe 400 Franken betrug 1999 dieser Maximalrabatt 8%, bei 600 Franken 15%, bei 1200 Franken 30% und bei der höchsten Stufe 40%, vgl. Bundesamt für Sozialversicherung (2001b).

einen höheren Anteil Versicherter mit Wahlfranchisen gibt als in „billigen“ Kantonen.<sup>259</sup> Die Kostenbeteiligung der Patienten ist nach oben beschränkt.<sup>260</sup> Darüber hinaus können Personen mit geringem Einkommen Prämienverbilligungen in Anspruch nehmen, die zum Teil vom Bund und zum anderen Teil vom jeweiligen Kanton getragen werden.

Das neue Krankenversicherungsrecht wird laufend evaluiert. So gibt es bspw. Evaluierungen des Gesamtsystems (1998 und 2000), Untersuchungen über die Wirkung der Wahlfranchisen (2000) und auch die schon erwähnte Untersuchung der besonderen Versicherungsformen (2000). Diese Publikationen werden alle vom Bundesamt für Sozialversicherung (BSV) herausgegeben.<sup>261</sup> Schließlich kommt es regelmäßig zu Anhörungen im Bundesrat, um das KVG weiter zu entwickeln (vgl. Eidgenössisches Departement des Inneren 2002).

---

<sup>259</sup> Aus diesem Grund wurden 2001 so genannte Plafonds eingezogen, deren Höhe durch den Rabatt nicht überstiegen werden darf. So darf der 8%-ige Rabatt der ersten Wahlfranchisestufe 170 SFr. nicht übersteigen. Für die Zukunft (ab 1.1.2004) wurden schließlich weitere Änderungen der Franchiseregulierung beschlossen (höhere obligatorische Franchise, veränderte Rabattsätze und veränderte Plafonds). Vgl. die Mitteilungen des Bundesamts für Sozialversicherung hierzu ([www.bsv.admin.ch](http://www.bsv.admin.ch)).

<sup>260</sup> Für die einzelnen Franchisestufen ergeben sich max. Selbstbeteiligungen von 830, 1000, 1200, 1800 oder 2100 Franken im Jahr. Kinder brauchen keine Franchise zahlen, können aber auch zwischen 4 Wahlfranchisestufen wählen. Eine Übersicht über Selbstbeteiligungsregelungen in Europa bietet Kupsch, S. et al. (1998).

<sup>261</sup> Die genannte Literatur findet sich unter Bundesamt für Sozialversicherung (1998, 2000a, 2000b).

Anhang II

**Inputfaktoren für die Effizienzschätzung (Abschnitt 9.2)**

Variable	Beschreibung	Variable	Beschreibung
<b>LARZT</b>	Log(Direkte Arztleistungen)	<b>LALAB</b>	LARZT*LLAB
<b>LMARZT</b>	Log(Medizin vom Arzt)	<b>LALMA</b>	LARZT*LMARZT
<b>LMED</b>	Log(Medikamente aus Apotheke)	<b>LALM</b>	LARZT*LMED
<b>LLAB</b>	Log(Laborleistungen)	<b>LALOUT</b>	LARZT*LOUT
<b>LOUT</b>	Log(Physio)*Log(Roentgen)	<b>LAT</b>	LARZT*T
<b>T</b>	Zeitindex	<b>LABLMA</b>	LLAB*LMARZT
<b>LARZT2</b>	LARZT <sup>2</sup>	<b>LABLM</b>	LLAB*LMED
<b>LMARZT2</b>	LMARZT <sup>2</sup>	<b>LABLOUT</b>	LLAB*LOUT
<b>LMED2</b>	LMED <sup>2</sup>	<b>LABT</b>	LLAB*T
<b>LLAB2</b>	LLAB <sup>2</sup>	<b>LMALM</b>	LMARZT*LMED
<b>LOUT2</b>	LOUT <sup>2</sup>	<b>LMALOUT</b>	LMARZT*LOUT
<b>T2</b>	T <sup>2</sup>	<b>LMAT</b>	LMARZT*T
		<b>LMLOUT</b>	LMED*LOUT
		<b>LMT</b>	LMED*T
		<b>LOUTT</b>	LOUT*T

## Anhang III

## Gesamtmodell: Deskriptive Statistik, Variablenbeschreibung

N=100523	Mittelwert	Std. Fehler	Min	Max	Beschreibung
<b>Patienten</b>					
LND	3.101	4.092	0	12.435	Log (Entfernung Patient – Arzt)
KA	0.020	0.139	0	1	1 = Arzt in Patientengemeinde
FR	1.415	0.829	0	5	Franchisestufe
FR98	2.604	5.389	0	44.866	Franchise * Log(Leistungen des Vorjahres)
RISK	0.746	0.436	0	1	1 = Unfalleinschluss
SEX	0.399	0.490	0	1	1 = Mann
ALTER	41.574	23.988	2	103	Alter in Jahren
ALTER2	2303.830	2116.750	4	10609	Alter zum Quadrat
KIND	0.217	0.412	0	1	1 = Patient unter 18 Jahre
JUGEND	0.069	0.254	0	1	1 = Patient zwischen 18 und 26 Jahren
SPT	0.117	0.321	0	1	1 = Patient im Vorjahr mind. 1 Tag im Spital war
LLST97	6.400	2.516	0	12.097	Log (Leistungen 1997)
LLST98	6.578	2.422	0	12.096	Log (Leistungen 1998)
ALT97	281.318	217.100	0	1014.060	Alter * LLST97
ALT98	289.398	218.568	0	1030.740	Alter * LLST98
LPR	2.534	3.302	0	9.320	Log (Prämienverbilligung)
UNT_VRS	1.667	1.013	1	20	Anzahl in Anspruch genommener Grundversorger im Jahr
UNT_V2	3.803	9.750	1	400	UNT_V2 zum Quadrat
<b>Ärzte</b>					
ROENT	0.736	0.441	0	1	1 = Arztpraxis mit Röntgengerät
ZH	0.676	0.468	0	1	1 = Arzt im Kanton Zürich
PA100	0.604	0.489	0	1	1 = Allgemeinmediziner (Referenz: Kinderarzt)
PA105	0.283	0.450	0	1	1 = Allgemeine Innere Medizin (Referenz: Kinderarzt)
GEM	0.234	0.423	0	1	1 = Gemeinschaftspraxis (Referenz: Kinderarzt)
GEM_ANZ	0.568	1.090	0	6	Anzahl der Ärzte in Gemeinschaftspraxen
TITEL_D	0.918	0.274	0	1	1 = mind. Arzt mit Dokortitel
BERUFSE	2.452	0.582	0.693	3.091	Log(Eintrittsjahr (KSK) – Untersuchungsjahr)
MANN	0.791	0.406	0	1	1 = Mann (Referenz: Neutral)
FRAU	0.166	0.372	0	1	1 = Frau (Referenz: Neutral)
LALTPR	3.627	0.513	1.730	4.490	Log(Durchschnittsalter Patienten einer Praxis)
SEX_PR	0.394	0.101	0	1	Männeranteil unter Patienten einer Praxis
L_ARZT98	6.157	0.412	4.317	8.599	Log(Umsatz des Vorjahrs)
VERM	12.939	0.727	10.017	14.820	Log((Umsatz97+Umsatz98)/2)
VERMB	31.852	8.023	7.477	43.888	VERM*BERUFSE
LPAT98	6.819	0.677	3.258	8.822	Log(Patientenzahl des Vorjahres)
P_DIFF	0.720	5.274	-0.75	5.71	(Patienten 1998 – Patienten 1997)/1000
PAT_ANT	0.246	1.649	0	17.204	Anteil MC-Patienten in einer Praxis
<b>MC-Variablen</b>					
MC_ARZT	0.091	0.288	0	1	1 = Arzt behandelt MC-Patienten
MC_PAT	0.040	0.196	0	1	1 = Patient ist Mitglied eines HAMs oder der HMO
EFF	0.884	0.067	0.250	0.980	Effizienzwert der Arztpraxis
LAMBDA1	-0.009	0.379	-1.290	4.030	Mills-Ratio aus Nachfrageschätzung



Anhang III

Case					
<b>M_V</b>	209.144	1857.340	0	31394.2	Verhältnis von selbstabgegebenen Medikamenten zum Apothekenverkauf
<b>L98</b>	40.563	15.408	0	97.231	Log(Umsatz des Jahres im Vorjahr)*LLST98
<b>LNK</b>	1.183	0.791	0	4.898	Log(Konsultationen)
<b>LNNR</b>	0.847	0.780	0	3.829	Log(Leistungsnummern)
<b>K_NR</b>	1.401	1.798	0	15.139	LNK * LNNR
<b>REGEL</b>	2.433	1.915	1	12	12 = Patient ist mindestens einmal in jedem Monat beim Arzt gewesen
Gemeinde					
<b>HERF</b>	0.128	0.202	0.005	1	1 = Monopol
<b>LDK_P</b>	5.092	0.194	4.585	5.400	Durchschnittskosten in der Patientengemeinde
<b>LDK_A</b>	5.108	0.193	4.652	5.400	Durchschnittskosten in der Arztgemeinde
<b>ARZTD</b>	14.417	5.642	2.460	29.676	Arztdichte (pro 10,000 Einwohner)
<b>LBEV</b>	10.480	1.714	5.930	12.720	Log(Bevölkerungsdichte (Bevölkerung pro Fläche) Fläche in ha)
<b>LHA</b>	7.371	1.139	4.850	9.080	Log(Fläche der Gemeinde in ha)