

Effizienzvergleiche und Wirtschaftlichkeit: Evaluation einer Pilotstudie

In jüngster Zeit wurde die Data Envelopment Analysis (DEA), eine nicht-parametrische Methode zur Bestimmung technischer Effizienz, von Lovell (2006) sowie von Agrell und Bogetoft (2005) als Verfahren, das einer anreizkompatiblen Vergütung von Leistungserbringern im Gesundheitswesen zugrunde liegen kann, propagiert. Agrell und Bogetoft (2005) wiesen darauf hin, dass vor dem Hintergrund der schwerlich mit parametrischen Funktionen nachzubildenden (Produktions-) Prozesse im Gesundheitswesen und gleichzeitig wenig konkret formulierten gesellschaftlichen Prioritäten, nicht-parametrische Methoden klare Vorteile gegenüber parametrischen Alternativen aufweisen. Lovell (2006) geht auf verschiedene von Newhouse (1994) vorgetragene Argumente, die eine skeptische Haltung gegenüber Effizienzschätzungen als Grundlage von Vergütungsmechanismen im Gesundheitswesen reflektieren, ein. Er führt in diesem Zusammenhang aus, dass die Argumente von Newhouse (1994), insbesondere die Sorge, eine starke Fokussierung auf die Effizienz der Anbieter könne sich negativ auf die Qualität des Angebots auswirken, ebenso auf andere Bereiche, bspw. die Energieversorgung, zuträfen, in denen eine Anreizregulierung auf Basis von Effizienzschätzungen gang und gäbe ist (vgl. Pollitt und Jamasb, 2001, für einen Überblick über den Bereich der Energieversorger).¹ Gleichzeitig ist festzuhalten, dass jede Vergütungsform, die nicht explizit für die Qualität der Leistungserbringung vergütet, die Gefahr birgt, dass Qualitätsaspekte der Leistungserbringung in den Hintergrund treten. Die Qualitätssicherung ist somit immer ein gesondert zu behandelnder Aspekt der Leistungserbringung.

Die Mehrheit der derzeit vorliegenden Studien zur Evaluierung der Effizienz von Leistungserbringern im Gesundheitswesen mittels DEA bezieht sich auf den stationären Bereich

¹ Ein prominentes Beispiel für die Anwendung solcher Verfahren ist der Modus für die Regulierung der Netznutzungsentgelte in Deutschland, den die Bundesnetzagentur (2006) vorgestellt hat.

(vgl. Chilingerian und Sherman, 2004, sowie Hollingsworth, 2003, für neuere Übersichtsartikel). Hollingsworth (2003) vermutet, dass die geringe Zahl der Studien im ambulanten Bereich auf einen Mangel an Daten zurückzuführen ist. Allerdings stellt bereits Eisenberg (1986) fest, dass für ambulante Leistungen zwar lediglich etwa 20 % der Gesamtkosten des Gesundheitswesens (bezogen auf die USA) zu veranschlagen seien, die im ambulanten Bereich getroffenen Entscheidungen in Bezug auf die weiterführende Therapie aber bis zu 80 % der Gesamtkosten zumindest mittelbar beeinflussen. Die offensichtliche Bedeutung der Entscheidungen, die im ambulanten Bereich gefällt werden unterstreicht die Notwendigkeit von Studien zur Effizienz niedergelassener Ärzte. Die *health funds* in den USA scheinen sich in puncto Kostenmonitoring im Wesentlichen auf einfache Kennzahlenvergleiche zu beschränken „... *despite the fact that DEA is a superior methodology.*“ (Defelice und Bradford, 1997).

Die vorliegende Studie hat die Analyse eines Pilotprojekts einer österreichischen Gebietskrankenkasse zur Effizienzregulierung niedergelassener Ärzte zum Inhalt. Beginn der Pilotphase war das Jahr 2004. Die erste Projektphase beinhaltete die Analyse der im Folgenden zu beschreibenden Daten für niedergelassene Allgemeinmediziner und Fachärzte unter Verwendung der DEA. Die zweite Phase beinhaltete eine Ansprache aller Ärzte, deren Effizienz im unteren Viertel der Verteilung lag, weil hier die größten Optimierungspotenziale vermutet werden. Langfristig war an das Einführen einer Bonus-Malus-Regelung in Abhängigkeit von der Effizienz der betreffenden Ärzte gedacht.

Folgende Festlegungen wurden im Rahmen des Pilotprojekts getroffen: Die Allgemeinmediziner wurden nach Praxisstandort in vier Regionen unterteilt, die sich hinsichtlich der Mortalität der Bewohner unterscheiden. Die Ärzte in den einzelnen Regionen werden in der Folge getrennt voneinander analysiert. Ebenso werden für alle Facharztgruppen getrennte Analysen durchgeführt und jeweils die Ärzte im unteren Viertel der Verteilung der Effizienz in die Ansprache einbezogen. Der diesen Festlegungen zugrunde liegende Gedanke war, die

Ärzte mit der geringsten Effizienz aus der *jeweiligen Gruppe* zu identifizieren und diese durch eine gezielte Ansprache einer Steigerung ihrer Effizienz zu motivieren.

Das Ziel der vorliegenden Studie ist die Bewertung der beschriebenen Vorgehensweise im Hinblick darauf, ob auf diese Art tatsächlich die maximalen Effizienzreserven im Bereich niedergelassener Ärzte identifiziert werden. Fraglich ist, ob die insgesamt über alle Stichproben vorhandenen Effizienz- und Wirtschaftlichkeitspotenziale durch die beschriebene Vorgehensweise identifiziert werden. Zum einen kann das Unterteilen von Stichproben in Unterstichproben dazu führen, dass sich die Mehrheit der Ärzte mit geringer Effizienz auf eine oder wenige Unterstichproben konzentriert – dann würde ein Teil dieser nicht effizienten Ärzte, solange sie nicht im unteren Viertel der Effizienzverteilung „ihrer“ Unterstichprobe erscheinen, von einer Ansprache ausgenommen. Zum anderen ist die Behandlung durch unterschiedliche Fachärzte mit im Schnitt deutlich unterschiedlichen Kosten verbunden. Dies bedeutet, dass ein relativ effizienterer Facharzt auch einer Disziplin, die hohe durchschnittliche Kosten aufweist, höhere Wirtschaftlichkeitspotenziale ($\hat{=}$ Kosten \times Grad der Ineffizienz) aufweist, als ein Arzt mit geringerer Effizienz aber gleichzeitig geringeren Gesamtkosten.

Daher werden zwei Aspekte des skizzierten Vorgehens, die im Zusammenhang mit der Effizienzbewertung zum Zweck der Regulierung von praktischer Relevanz sind, aufgegriffen und einer Überprüfung unterzogen: zum einen die Unterteilung in verschiedene Regionalgruppen, zum anderen die Identifikation von Optimierungspotenzialen *getrennt nach Fachgruppe*.

Die vorliegende Arbeit ist wie folgt aufgebaut: Zunächst erfolgt ein kurzer Überblick über die Literatur zur Bewertung der Effizienz niedergelassener Ärzte. Eine Beschreibung des Datenmaterials schließt sich an. Die verwendeten Methoden werden in einem gesonderten Abschnitt vorgestellt. Nach der Dokumentation der Resultate wird in einem abschließenden Abschnitt eine Bewertung der Ergebnisse vorgenommen.

1. Empirische Studien zur Effizienz niedergelassener Ärzte

Die Autoren von Studien zur Effizienz niedergelassener Ärzte² bezeichnen den Ressourcenmix, mit dem diese Ärzte ihre Leistungen erbringen, als „Behandlungsstrategie“ (*treatment strategy*). Ein Arzt stellt Behandlungsleistungen entweder selbst bereit oder ordnet Leistungen von Fachärzten an. Die Verfasser der genannten Studien gehen davon aus, dass der Inputmix eines Arztes aus (nicht perfekten) Substituten besteht und die Transformation von Inputs (Ressourcen) in Outputs (Leistungen) eine „Behandlungstechnologie“ konstituiert.

Chilingerian und Sherman (1997) publizieren die erste Studie dieser Form. Sie verwenden ein Input-orientiertes DEA-Modell mit konstanten Skalenerträgen (CRS, vgl. Abschnitt 3), das den nachfolgenden Studien als Modell gedient zu haben scheint. Die Information über den Fallmix beschränkt sich auf das Alter der Patienten. Im Rahmen ihrer Untersuchungen zeigen Chilingerian und Sherman (1997), dass einige Ärzte, insbesondere Internisten und andere Fachärzte, zu häufigen. Sie interpretieren dies als Hinweis darauf, dass diese Ärzte durch Laboranforderungen den eigenen zeitlichen Einsatz substituieren, den eine andernfalls zwingend notwendige Wiedervorstellung des Patienten in der eigenen Praxis mit sich brächte. Wirtschaftlicher aus Sicht des *sickness funds*, für den die Leistungen erbracht werden, wäre die im Vergleich zur Laborleistung günstigere Wiedervorstellung. Chilingerian und Sherman (1997) schließen, dass niedergelassene Ärzte sich nicht ausschließlich am Wirtschaftlichkeitsprinzip orientieren und bei der Wahl des Inputmix die Präferenzen der Ärzte eine bedeutsame Rolle spielen. Die durchschnittliche Effizienz der Ärzte in der Stichprobe wird mit 91 % angegeben, wobei sich 43 der insgesamt 326 untersuchten Ärzte als vollständig effizient erweisen.

Die Untersuchung von Andes et al. (2002) unterscheidet sich von den anderen publizierten Studien, indem sie den Ertrag einer Praxis als Output spezifiziert. Im vorhandenen Datenkranz ist weder das Alter noch ein anderer Fallmixindikator vorhanden. Die Ärzte in

² Rosenman und Friesner (2004), Wagner et al. (2003), Andes et al. (2002), Pai et al. (2000), Ozcan (1998) sowie Chilingerian and Sherman (1997); vgl. Tabelle A1 im Anhang.

dieser Stichprobe weisen eine durchschnittliche Effizienz von lediglich 46 % auf, wobei 7 der 115 untersuchten Praxen als vollständig effizient eingestuft werden. Die Resultate deuten auf die Existenz von Skaleneffekten hin, wobei sich Praxen mittlerer Größe gegenüber kleineren bzw. größeren Praxen als vergleichsweise effizienter erweisen.

Eine Reihe von Studien beinhaltet detailliertere Informationen zum Fallmix. Während Pai et al. (2000) sowie Ozcan (1998) die Effizienz von Ärzten in Bezug auf eine spezifische Diagnose (Nebenhöhlen bzw. Mittelohr) unter Berücksichtigung eines Index für die Schwere des Falls bestimmen, verwenden Wagner et al. (2003) Informationen zum Allgemeinzustand der Patienten. Sowohl bei Ozcan (1998) als auch bei Pai et al. (2000) umfasst die Stichprobe – wie bereits bei Chilingirian und Sherman (1997) – Ärzte unterschiedlicher Fachrichtungen. Beide Studien ermitteln eine substantielle regionale Variation in Bezug auf die Effizienz der Ärzte in den untersuchten Stichproben. Bei Ozcan ergibt sich auf Basis einer CRS-Spezifikation eine durchschnittliche Effizienz von 79.6 %, wobei sich 46 von 160 Ärzten als effizient erweisen. Pai et al. (2000) erhalten mit einer Effizienz von durchschnittlich 78.1 % bei 43 (von 156) effizienten Ärzten ein nahezu identisches Ergebnis, ebenfalls auf Basis von CRS.

Wagner et al. (2003) verfügen neben einem Fallmixindex noch – als einzige der hier aufgeführten Studien – über einen subjektiven, durch Befragung erhobenen Qualitätsindikator. Ihre Stichprobe umfasst allerdings lediglich 27 Beobachtungen, von denen 11 als effizient eingestuft werden. Die durchschnittliche Effizienz liegt bei 90.3 %.

Die Studie von Rosenman und Friesner (2004) ist die einzige, für die ein Modell mit variablen Skalenerträgen (VRS; vgl. wiederum Abschnitt 3) geschätzt wird. Die durchschnittliche Effizienz der untersuchten Allgemeinmediziner liegt bei 79.8 %; hier sind 56 der 156 Beobachtungen effizient. Die Autoren versuchen im Rahmen eines zweistufigen Verfahrens Charakteristika der Praxen zu identifizieren, die mit deren Effizienz signifikant korrelieren. Allerdings verwenden sie auf der zweiten Stufe einen Tobit-Ansatz, der für das betreffende Schätzproblem nicht geeignet ist (vgl. Simar und Wilson, 2007).

Es lässt sich festhalten, dass die vorliegenden Studien die Effizienz niedergelassener Ärzten anhand von Querschnitten beurteilen und mit einer Ausnahme auf der CRS-Annahme beruhen. Die verwendeten Stichproben beinhalten i.d.R. Ärzte verschiedener Fachrichtungen. Gelegentlich wird für den Fallmix kontrolliert und z. T. wird die Analyse auf bestimmte Diagnosen eingeschränkt. Lediglich eine Studie verfügt über (subjektive) Informationen zur Qualität der Behandlung. Die Mehrzahl der untersuchten Ärzte ist nicht vollständig effizient, wobei die durchschnittliche Effizienz, berechnet mit Stichproben, deren Umfang zwischen 26 und 346 Beobachtungen variiert, mit einer Ausnahme zwischen 78 % und 91 % liegt.

Für die folgende Analyse wird auf eine Reihe in Abschnitt 2 näher zu beschreibender Datensätze zurückgegriffen, die sich für den Zweck der Effizienzbewertung in besonderer Weise eignen. Bei den Beobachtungen handelt es sich um Ärzte einer österreichischen Gebietskrankenkasse, die alle demselben Vertrag unterliegen; dieser reglementiert die Leistungen, die erbracht werden können, sehr strikt. Daher ist davon auszugehen, dass das tatsächliche Leistungsspektrum der Ärzte weitgehend homogen ist. Eine Stichprobe von Allgemeinmediziner umfasst über 800 Beobachtungen und ist damit wesentlich größer als die in der Literatur bisher verwendeten Datensätze. Diese eignet sich auch für einen Test im Hinblick auf eine etwaige regionale Variation der Effizienz. Gleichzeitig sind für Allgemeinmediziner (Sub-)Spezialisierungen untersagt, was zusätzlich auf eine große Homogenität innerhalb dieser Gruppe schließen lässt.

Neben den Daten für Allgemeinmediziner liegen mehrere Stichproben unterschiedlichen Umfangs für Fachärzte vor. Auch diese unterliegen qua Vertrag einer strikten Reglementierung ihrer Leistungen. Anhand dieser Stichproben kann überprüft werden, inwieweit bei einer alleinigen Fokussierung auf die Effizienz der Leistungserbringer die Wirtschaftlichkeitspotenziale insgesamt identifiziert werden können. Die Effekte der Stichprobengröße im Hinblick auf die Verzerrung des Schätzers (vgl. wiederum Abschnitt 3) und der Konsequenzen hieraus werden ebenfalls untersucht. Beiden Datensätzen gemeinsam ist, dass Informationen über

Zuweisungen zur weiterführenden Therapie an Fachärzte vorhanden sind. Auf diese Weise kann ein wesentlicher Teil der Folgekosten, die durch einen Arzt ausgelöst werden und diesem direkt zurechenbar sind, berücksichtigt werden. Insgesamt beinhaltet das hier verwendete Datenmaterial im Vergleich zu den Daten, die im Rahmen der zitierten Studien verwendet wurden, deutlich homogenere Beobachtungen und ist damit für einen Effizienzvergleich besser geeignet.

2. Daten

Für die Analyse stehen zwei Datensätze aus dem Jahr 2005 zur Verfügung: eine Stichprobe mit insgesamt 801 Allgemeinmediziner in vier verschiedenen Regionen sowie 11 weitere Stichproben niedergelassener Ärzte verschiedener Fachrichtungen, deren Umfang zwischen 26 und 115 Beobachtungen liegt. Sämtliche Ärzte sind Vertragsärzte einer österreichischen Gebietskrankenkasse. Für alle Ärzte liegen Informationen über die Zugehörigkeit ihrer Patienten zu verschiedenen Alterskategorien vor; allerdings sind keine darüber hinaus gehenden Informationen zu Fallmix, Diagnosen oder Komplexität einzelner Fälle verfügbar, wie das in den o. g. Studien z. T. der Fall war.

Des Weiteren sind die Einnahmen der Ärzte aus einer pauschalierten Grundvergütung pro Patient sowie einer zusätzlichen Einzelfallvergütung bekannt. Eine Einzelfallvergütung kann zusätzlich zur Pauschale geltend gemacht werden, sobald eine Leistung erbracht wird, die von dieser nicht abgedeckt ist. Eine Einteilung kann – je nach Leistung – in Form von Festbeträgen bzw. über ein Punktesystem, abgerechnet werden.

Die Grundvergütung ist nach der Zahl der Patienten gestaffelt, wobei die Staffeln für die Pauschale bei 500, 800 bzw. 1.100 Patienten liegen. Für die ersten 500 Patienten liegt die Pauschale bei ca. 17 € pro Patient und Quartal; dieser Betrag sinkt ab dem 1.101. Patienten auf unter 5 €. Die Abrechnung erfolgt direkt zwischen Arzt und Kasse. Die gesondert abzurechnenden Einzelleistungen sind vertraglich fixiert und strikt reglementiert. Es ist den Ärzten unter dem gegebenen Vertrag bis auf wenige Ausnahmen im Facharztbereich nicht

gestattet, bspw. größere Diagnostikgeräte zu betreiben – letztere würden es erlauben, Leistungen zu erbringen, die deutlich höher als es dem zeitlichen Aufwand entspricht, vergütet werden – und entsprechend abzurechnen. Die Vergütung kann daher als grobe Approximation für den zeitlichen Einsatz des Arztes angesehen werden. Gleichzeitig liegen Informationen über die Kosten der verordneten Medikamente sowie über die Kosten für Facharztzuweisungen vor.

Tabelle 1: Kosten pro Patient für Allgemeinmediziner nach Regionen (in €)

Gesamtkosten (in €)	Region			
	1	2	3	4
Minimum	74,89	87,09	70,16	87,58
5%	108,17	113,84	109,43	115,84
Mittelwert	155,83	154,7	157,35	161,06
95%	216,61	203,06	206,02	203,52
Maximum	538,59	224,06	352,85	377,44
Gebühr	37,93	36,59	36,77	36,7
Heilmittel	82,24	79,95	80,8	86,29
Zuweisungen	35,66	38,15	39,77	38,07
Patienten	3.484,68	3.361,37	3.339,01	3.090,55
Anteil Patienten bis 50 J.	55,24 %	53,92 %	54,96 %	50,45 %
Beobachtungen	246	190	246	119

Zunächst werden die Daten für die Allgemeinmediziner beschrieben; für sie liegt zusätzlich zu den genannten Variablen ein Indikator für die Region, in der ein Arzt praktiziert, vor. Das gesamte Versorgungsgebiet wird hier in vier Klassen, die sich hinsichtlich der Mortalität der Bevölkerung unterscheiden, unterteilt (Region 1: sehr hohe Mortalität bis Region 4: weit unterdurchschnittliche Mortalität). Durch diese Einteilung soll ein ggf. über die Praxen variierender Fallmix berücksichtigt werden. Die Einteilung erfolgte so, dass die ersten beiden Klassen i.d.R. einen signifikant gegenüber dem Bundesdurchschnitt erhöhten Mortalitätsindex aufwiesen, während dies für die beiden anderen Klassen nicht der Fall ist. Demzufolge liegt der Index für die beiden ersten Klassen aggregiert um 15 % über demjenigen der beiden restlichen Klassen, während sich die beiden Klassen mit jeweils hoher bzw. niedriger Mortalität untereinander hinsichtlich des Index nicht substantiell unterscheiden. Die Zahl der

Beobachtungen für die einzelnen Klassen bewegt sich zwischen 119 (Region 4) und 246 (Region 1 bzw. 3; vgl. die letzte Zeile von Tabelle 1).

Bereits ein erster Blick auf den Mittelwert der Gesamtkosten (vgl. die obere Hälfte von Tabelle 1) zeigt, dass dieser über die verschiedenen Gruppen kaum streut; der höchste pro-Kopf Wert wird zudem in Region 4 (geringste Mortalität) erzielt. Hier fallen auch die höchsten durchschnittlichen Heilmittelkosten an (vgl. die untere Hälfte der Tabelle), bei den Zuweisungen lässt sich in der Gruppe 1 (höchste Mortalität) der geringste Wert beobachten. Die Patientenzahlen³ pro Praxis sind in Regionen mit höherer Mortalität höher als in solchen mit niedrigerer Mortalität, wobei der Anteil der Patienten bis 50 Jahre in Region 1 mit über 55 % am höchsten liegt; am niedrigsten ist dieser Anteil in Region 4.

Tabelle 2: Deskriptive Statistiken für die Fachärzte

FG	Zahl Ärzte	Honorar	Anteil		Mittelwerte (in €)		
			Heilmittel	Zuwei- sungen	Kosten, gesamt	Zahl Patienten	Kosten pro Patient
Augen	89	,85	,14	,012	238.458	4.385	54,38
Chirurgie	37	,75	,10	,16	177.459	1.555	114,12
Dermatologie	76	,62	,34	,041	248.556	4.393	56,58
Gynäkologie	115	,64	,1	,26	211.674	3.198	66,19
Innere	99	,53	,29	,19	357.193	2.336	152,91
Kinder	86	,71	,16	,13	212.114	3.257	65,13
HNO	61	,81	,09	,092	222.851	3.327	66,98
Lungen	39	,52	,42	,054	371.331	2.631	141,14
Orthopädie	26	,6	,08	,32	432.859	3.168	136,63
Urologie	46	,59	,3	,11	335.563	3.062	109,59
Neurologie	29	,51	,36	,13	374.010	2.187	171,02
Psychiatrie	26	,5	,44	,061	366.283	1.763	207,76

Tabelle 2 gibt einen Überblick über die für die verschiedenen Fachdisziplinen (vgl. Spalte 1 der Tabelle) zur Verfügung stehenden Informationen. In der zweiten Spalte ist die Zahl der Ärzte in der jeweiligen Fachgruppe aufgeführt. In den letzten drei Spalten finden sich die Kosten, die pro Jahr in den Praxen insgesamt erfasst werden, die Zahl der Patienten sowie die Kosten pro Patient. In den Spalten unter „Anteil“ sind die Gesamtkosten anteilig nach

³ Die angegebenen Zahlen beziehen sich auf Kontakte pro Quartal und wurden auf Jahresebene aggregiert.

Honorar, Medikamenten und Zuweisungen aufgeschlüsselt. Die Kosten pro Patient und auch deren Aufteilung streuen über die Fachgebiete erheblich, wobei bspw. der Wert für Augenärzte lediglich ein Viertel des entsprechenden Wertes in der Psychiatrie beträgt.

Es wird deutlich, dass die verschiedenen Facharztgruppen Kosten in unterschiedlicher Höhe (vgl. hierzu auch die Tabellen A1 bis A3 im Anhang) verursachen. Dies bezieht sich sowohl auf die Gesamtkosten, als auch auf die Kosten für Honorare ohne Berücksichtigung von Zuweisungen oder Heilmitteln. So entfallen auf die Fachgebiete der Augenheilkunde, der Dermatologie und der Gynäkologie weniger als ein Drittel der Patienten pro Quartal;⁴ gemeinsam erreichen diese drei Fachgebiete aber einen Kostenanteil von nahezu 50 %. Diese Unterschiede sind dann bedeutsam, wenn der Versicherer – dem Wirtschaftlichkeitsprinzip folgend – seine Analyse auf die Gruppen konzentrieren möchte, bei denen die größten Optimierungspotenziale zu erwarten sind.

Die beschreibenden Statistiken für die Fachärzte lassen vermuten, dass die Optimierungspotenziale in den einzelnen Gruppen unterschiedliche Größenordnungen aufwiesen. Daher ist zu hinterfragen, ob die Ansprache eines gleichen Anteils von Ärzten mit geringer Effizienz die Optimierungspotenziale insgesamt sinnvoll identifiziert. Werden nun pro Region die jeweils 25 % Ärzte mit der geringsten Effizienz innerhalb ihrer Vergleichsgruppe angesprochen, ist dies nicht gleichbedeutend mit der Ansprache von 25 % der Ärzte, die die geringste Effizienz insgesamt aufweisen.

3. Methode

Die Ausführungen zur Methode sind in drei Abschnitte untergliedert. Zunächst werden anhand einer Abbildung die grundlegenden Eigenschaften der DEA verdeutlicht. Darauf wird die Methode formal beschrieben; der dritte Unterabschnitt geht auf spezielle methodische Aspekte, die für die vorliegende Untersuchung relevant sind, ein: (i) den Vorschlag von Zhang und Bartels (1998) zum Vergleich von Effizienzwerten, die mit Stichproben

⁴ Die Zahl der Patienten registriert die Erstkontakte eines Arztes mit seinen Patienten pro Quartal.

unterschiedlicher Größe ermittelt wurden; dieser ist für die Analyse der Facharztgruppen einschlägig (ii) auf den Test auf Effizienzunterschiede zwischen einzelnen Unterstichproben mit einem Bootstrap-Verfahren nach Simar und Wilson (2000); der Test ist für das Überprüfen der regionalen Einteilung der Allgemeinarztpraxen von Bedeutung.

3.1 Illustration der DEA

Eine Reihe von in der Literatur regelmäßig zitierten Faustregeln im Hinblick auf die Spezifikation von DEA-Modellen (vgl. Cooper et al., 2000 sowie Dyson et al., 2001) beziehen sich explizit auf das Verhältnis von Stichprobenumfang und Zahl der Variablen; diese Faustregeln geben (unterschiedliche) Schranken für dieses Verhältnis vor, die einzuhalten seien, wenn die Resultate hinsichtlich der Effizienz hinreichend differenzieren sollen. Die Regeln spiegeln den Umstand wider, dass Modelle mit einer großen Zahl von Variablen in Kombination mit kleinen Stichproben üblicherweise viele Beobachtungen als effizient einstufen und eine insgesamt hohe durchschnittliche Effizienz für die Beobachtungen ermitteln; in einem solchen Fall ist fraglich, ob die Resultate hinreichend zwischen den Beobachtungen in Bezug auf deren Effizienz differenzieren. Es ist zu beachten, dass die Ergebnisse der in Abschnitt 1 zitierten Studien daher einem direkten Vergleich nicht zugänglich sind, da sich Stichproben- und Modelleigenschaften z. T. unterscheiden.⁵

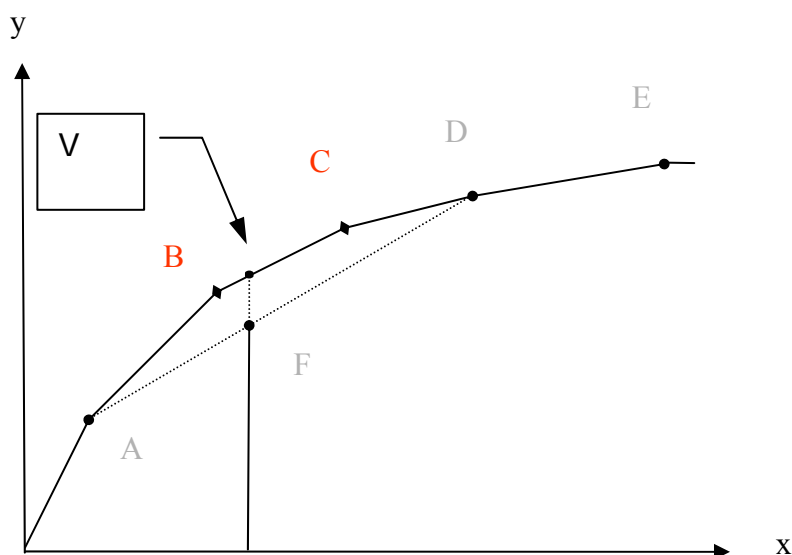
Aufgrund der im Folgenden erläuterten Verzerrung des DEA-Schätzers würden zwei Studien bei unterschiedlicher Stichprobengröße auch dann eine unterschiedliche Effizienz ermitteln, wenn die tatsächliche Effizienz der Ärzte in den Stichproben exakt dieselbe wäre (vgl. hierzu Kneip et al., 1998, Pedraja-Chaparro, Salinas-Jiménez und Smith, 1999, Gstach, 1995; Zhang und Bartels, 1998). Kneip et al. (1998) zeigen für den Fall einer Technologie mit einem Input und einem Output, dass die Verzerrung von drei Faktoren abhängt: vom Stichprobenumfang, von der Dichte der Beobachtungen am für die zu bewertende Einheit

⁵ Insofern sind Vergleiche einzelner Gruppen von Kliniken anhand eines Literaturüberblicks, wie sie bspw. Hollingsworth (2003) vornimmt, nicht aussagekräftig und führen ggf. in die Irre.

relevanten Segment der Technologie sowie von der Krümmung des effizienten Randes. In kleineren Stichproben ist für Beobachtungen mit einer geringen Dichte von Vergleichspunkten an einem stark gekrümmten Segment der Technologie eine höhere Verzerrung zu erwarten als in großen Stichproben mit einer entsprechend höheren Dichte bei einer geringen Krümmung. Kneip et al. (1998) zeigen, wie die Verzerrung des Schätzers korrigiert werden kann, allerdings ausschließlich für den Fall mit einem Input und einem Output. Für den höherdimensionalen Fall ist eine Korrektur lediglich durch das Bootstrap-Verfahren möglich.

Abbildung 1 illustriert, wie DEA-Effizienzwerte berechnet und wie die Ergebnisse vom Stichprobenumfang und der Zahl im Modell verwendeten Variablen beeinflusst werden. Sie zeigt eine Technologie mit einem Input, x und einem Output, y (x kann man sich als durch einen Arzt erbrachte Leistungen vorstellen, y bspw. als die Zahl der Patienten unter 50 Jahren). Gleichzeitig unterscheiden sich die Beobachtungen hinsichtlich eines exogenen Merkmals und werden dementsprechend entweder mit einem Punkt (A, D und E) oder einer Raute (B und C) gekennzeichnet. Die Beobachtungen, die mit einer Raute markiert sind, seien durch das exogene Merkmal hinsichtlich ihrer Leistungserbringung bevorteilt (bspw. niedrige Mortalität).

Abbildung 1: Die Verzerrung des DEA- Schätzers



Der eingezeichnete Rand der Technologie beruht nicht auf der CRS-Annahme, d. h. es wird nicht unterstellt, dass ein bestimmtes Verhältnis von Output zu Input, wie das der Beobachtung A, über die gesamte Skala der Produktion realisierbar sei. Im CRS-Fall wäre die Technologie durch einen Fahrstrahl durch den Punkt A beschrieben. Die hier eingezeichnete Grenze der Technologie stellt hingegen auf VRS ab; dies führt dazu, dass alle Referenzpunkte im Mittel dieselben Zahlen von Patienten in den einzelnen Alterskategorien behandeln müssen, wie der zu bewertende Arzt. Die VRS-Annahme führt im Input-orientierten Fall immer zu Effizienzwerten, die mindestens ebenso hoch sind, wie Effizienzwerte, die mit der Annahme konstanter Skalenerträge ermittelt würden (vgl. den gegenüber den anderen Studien *c.p.* höheren Anteil effizienter Beobachtungen in den Analysen von Rosenman und Friesner, 2004).

Die Beobachtungen A bis E sind Teil des effizienten Randes und werden daher als effizient eingestuft. Sie bekommen einen Effizienzwert von 1 zugewiesen. Es ist aber zu beachten, dass die in Abbildung 1 eingezeichnete Linie, die als effizienter Rand bezeichnet wurde, eine *Schätzung* des effizienten Randes darstellt. Bei einem mittels DEA geschätzten Rand einer Technologie handelt es sich notwendigerweise immer um eine *innere* Approximation der betreffenden Technologie, da eine Realisation einer Input-Output-Kombination nicht besser sein kann als das theoretische Optimum. Demzufolge *überschätzen* DEA-Effizienzwerte die tatsächliche Effizienz, d. h. der Schätzer ist systematisch verzerrt. Die Verzerrung fällt *c. p.* umso geringer aus, je mehr Daten zur Verfügung stehen.

Ein Input-orientierter Effizienzwert wird für einen ineffizienten Arzt (F) als Verhältnis zweier Strecken berechnet: Der horizontale Abstand des (effizienten) Referenzpunktes, V, zur y-Achse im Verhältnis zum horizontalen Abstand von F selbst zur Achse. Der Referenzpunkt liegt immer auf dem effizienten Rand; er weist mindestens denselben Output und maximal denselben Input wie die zu bewertende Beobachtung auf. Das Streckenverhältnis sei mit $OV/OF < 1$ bezeichnet. Der Faktor $(OV/OF - 1) \times 100\%$ ist derjenige, um den x_1 reduziert

werden müsste, damit F im x_1 - y_1 -Raum gerade effizient wäre; er beschreibt das Optimierungspotenzial in Bezug auf den betreffenden Inputfaktor.

Allerdings ist F als Punkt markiert, d. h. der gedachte Arzt praktiziert in einer Region mit hoher Mortalität. Daher sollte er u. U. nicht mit anderen Ärzten, die in Regionen mit vergleichsweise geringer Mortalität praktizieren (in der Abbildung mit Rauten gekennzeichnet, bspw. B und C) verglichen werden. Die Beobachtungen B und C, die dem Referenzpunkt V zugrunde liegen, sind durch Rauten gekennzeichnet und somit hinsichtlich der Mortalität bevorteilt. Deshalb verläuft die für F relevante Referenzlinie nun durch die Punkte (nicht: Rauten) A und D. Anhand dieser Referenzlinie wird F als effizient eingestuft. Das Einführen einer exogenen Größe, anhand derer die Vergleichbarkeit unter den Beobachtungen eingeschränkt wird, führt somit potenziell zu höheren Effizienzwerten für einzelne Beobachtungen; im vorliegenden Beispiel erlangt F den Status eines effizienten Arztes.

Zusammenfassend lässt sich festhalten: Offensichtlich beeinflusst das Einführen neuer Variablen in einem Modell das durchschnittliche geschätzte Effizienzniveau und zwar unabhängig davon, ob diese exogener Natur sind (Mortalität) oder Bestandteil der Technologie (Zahl der Patienten). Im Input-orientierten Fall des obigen Beispiels werden die Effizienzwerte positiv beeinflusst. Dies ist beim Vergleich von Ergebnissen über verschiedene Stichproben hinweg zu berücksichtigen. Außerdem ist es von entscheidender Bedeutung, die Relevanz von Variablen, die in einem Modell Verwendung finden, zu überprüfen. Die methodische Vorgehensweise wird in den folgenden Abschnitten beschrieben.

3.1.1 Formale Darstellung

Um diese Zusammenhänge zur Bewertung der technischen Effizienz mittels DEA näher zu betrachten, beginnt dieser Abschnitt mit der Definition einiger grundlegender Begriffe. Eine Produktionstechnologie

$$(1) \quad T = \{(x, y) \in \mathbb{R}_+^{i+r} \mid x \text{ kann } y \text{ produzieren}\},$$

beschreibt welche Menge von i Inputs x welche Mengen von r Outputs y produzieren kann.

Die Menge der Inputerfordernisse $X(y)$ wird definiert als:

$$(2) \quad X(y) = \{x \in \mathbb{R}_+^i \mid (x, y) \in T\}.$$

Die bezüglich dieser Mengen getroffenen Annahmen besagen, dass a) T abgeschlossen und konvex ist, die Outputmöglichkeiten $Y(x)$ für alle x abgeschlossen, konvex und beschränkt sind und die Menge der erforderlichen Inputs $X(y)$ für alle y abgeschlossen und konvex ist; dass b) ein positiver Output positive Inputs verlangt, und dass c) Inputs und Outputs frei verschwendbar sind. Die Isoquante, $\partial X(y)$, ist definiert als:

$$(3) \quad \partial X(y) = \{x \mid x \in X(y), \theta x \notin X(y) \forall 0 < \theta < 1\},$$

wobei $\theta_j = \min \{\theta \mid \theta x_j \in X(y_j)\}$ der Input-orientierte Effizienzwert (das eben beschriebene Streckenverhältnis) für eine gegebene Kombination (x_j, y_j) ist.

Weder die Menge der Produktionsmöglichkeiten T , noch die Inputerfordernisse $X(y)$ noch die Isoquante $\partial X(y)$ sind bekannt, aber für jede Stichprobe $S = \{(x_j, y_j) \mid j = 1, \dots, J\}$ können die Stichproben-Äquivalente von (2), $\hat{X}(y)$, und (3), $\partial \hat{X}(y)$, sowie $\hat{\theta}$ bestimmt werden. Das lineare Programm, mit dessen Hilfe θ_j bestimmt wird, enthält (mindestens) ebenso viele Nebenbedingungen, wie Variablen im Modell vorhanden sind (vgl. Formel (4)). Werden zu einem bestehenden Input-orientierten Modell Variablen hinzugefügt, kann die Zielgröße des Programms nicht sinken, d. h. die Effizienzwerte werden i. d. R. steigen.

$$(4) \quad \hat{\theta}_k = \min \left\{ \theta \mid y_k \leq \sum_{j=1}^n \lambda_j y_j; \theta x_k \geq \sum_{j=1}^n \lambda_j x_j; \theta > 0; \sum_{j=1}^n \lambda_j = 1; \lambda_j \geq 0, i = 1, \dots, n \right\}.$$

Gleichzeitig wurde ausgeführt, dass weniger Beobachtungen ebenfalls zu *c.p.* höheren Effizienzwerten führen. Auch dieser Effekt ist leicht anhand von Formel (4) erkennbar: je weniger Beobachtungen in die möglichen Referenzpunkte (die Größen, die ein λ beinhalten)

eingehen, desto weniger wahrscheinlich ist es, dass alle Bedingungen simultan für ein $\hat{\theta}_k < 1$ erfüllt sind.

Zhang und Bartels (1998) schlagen ein pragmatisches Verfahren vor, diesen Effekt auszugleichen, wenn die Ergebnisse von Stichproben mit unterschiedlicher Größe verglichen werden sollen. Ihr Vorschlag besteht darin, aus größeren Stichproben durch Ziehen ohne Zurücklegen eine große Zahl von Unterstichproben zu erzeugen, die den Umfang der kleinsten der zu vergleichenden Stichproben aufweisen. Der Mittelwert aller Effizienzwerte, die für die Beobachtungen in den einzelnen Unterstichproben ermittelt wurde, kann dann als Grundlage für einen Vergleich über die Stichproben hinweg dienen.

3.1.2 Das Bootstrap-Verfahren

Verschiedene Autoren haben Verfahren zum DEA-Bootstrap vorgeschlagen. Es sei angemerkt, dass ein einfaches Ziehen mit Zurücklegen aus den Rohdaten kein konsistentes Verfahren darstellt (vgl. hierzu die Ausführungen in Simar und Wilson 1999a, b). Ein naiver Bootstrap, d. h. ein einfaches Bilden von Pseudo-Stichproben durch Ziehen mit Zurücklegen aus den Originaldaten und das erneute Berechnen von Effizienzwerten auf Basis der Pseudo-Stichproben, wäre mit dem Vorteil verbunden, robust im Hinblick auf die Annahmen bezüglich des Daten-generierenden Prozesses zu sein. Diese Eigenschaft weist auch die methodische Alternative des so genannten subsampling-Bootstrap auf; sie lässt sich für eine Korrektur der Verzerrung in kleinen Stichproben aber nicht anwenden, weil die Wahl der Größe der Unterstichproben kritisch ist und keine befriedigende Lösung für die Bestimmung der optimalen Größe der Unterstichproben existiert (Kneip und Simar, 2003).⁶

Die Schätzung eines Effizienzwerts $\hat{\theta}_j$ und die entsprechende Bootstrap-Schätzung $\hat{\theta}_j^*$ stehen in folgender Beziehung zueinander, die den grundlegenden Bootstrap-Zusammenhang widerspiegelt:

⁶ Eine weitere gangbare Alternative wäre das in Simar und Wilson (2000) dargestellte Verfahren.

$$(5) \quad (\hat{\theta}_j - \theta_j) \Big| S \stackrel{\text{approx.}}{\sim} (\hat{\theta}_j^* - \hat{\theta}_j) \Big| S^*.$$

Die Beziehung des DEA-Schätzers zum wahren Effizienzwert (vgl. die linke Seite von (5)) wird, wie im Rahmen von Bootstrap-Ansätzen üblich, approximiert durch die Beziehung des Bootstrap-Schätzers zum DEA-Schätzer (vgl. die rechte Seite des obigen Ausdrucks). Deshalb kann die Verzerrung (*bias*), $bias_{S,j} = E_S(\hat{\theta}_j) - \theta_j$, der DEA-Effizienzwerte im Allgemeinen Fall durch die Bootstrap-Entsprechung, $\widehat{bias}_{S^*,j} = E_{S^*}(\hat{\theta}_j) - \hat{\theta}_j$, approximiert werden. So kann eine Verzerrungskorrektur mit $\tilde{\theta}_j = \hat{\theta}_j - \widehat{bias}_{S^*,j} = 2\hat{\theta}_j - \bar{\theta}_j^*$, mit $\bar{\theta}_j^* = R^{-1} \sum_R \bar{\theta}_j^*$ durchgeführt werden, wobei S^* die jeweilige Stichprobe bezeichnet. Natürlich sollte, damit eine Verzerrungskorrektur in einer Verbesserung der Ergebnisse des gewöhnlichen DEA-Ansatzes resultiert, die Korrektur nicht zu einem höheren *mean square error* (MSE) als bei der Ausgangsschätzung mit der Standardmethode führen. Damit dies der Fall ist, muss die Bedingung $\widehat{bias}/\hat{\sigma} > 3^{-1/2}$ erfüllt sein. Für den Fall, dass das MSE-Kriterium durch die Stichprobe nicht erfüllt wird, muss man schlussfolgern, dass die Information in den Daten ungenügend ist.

Der Algorithmus zur Ableitung der verzerrungskorrigierten DEA-Schätzfunktion ist im Detail in Simar und Wilson (1998, Abschnitte 4.1 und 4.2) beschrieben. Hier werden lediglich die Eckpunkte der Methode dargestellt. Zunächst muss die Dichte $f(\cdot)$ der Effizienzwerte über einen *Kernel smoother* geschätzt werden; hierfür ist das Bestimmen einer optimalen Bandbreite (*bandwidth*) notwendig. Die Kenntnis dieser Bandbreite genügt für die Simulation von Effizienzwerten aus der entsprechenden Verteilung. Auf Basis der optimalen Bandbreite und zufällig gezogener, entsprechend verteilter Störgrößen können Pseudo-Effizienzwerte erzeugt werden, die der Verteilung der Effizienzwerte aus der Ausgangsschätzung folgen. Hierbei werden die Störgrößen zu Effizienzwerten addiert, die durch Ziehen mit Zurücklegen aus Effizienzwerten der Ausgangsschätzung erhalten werden. Durch Anwenden der *reflection technique* nach Silverman (1986) wird sichergestellt, dass die Pseudo-

effizienzwerte innerhalb des Wertebereichs der Verteilung der Originalwerte liegen. Auf Basis dieser neuen Werte werden – im Input-orientierten Fall – Inputs berechnet, die wiederum den Pseudoeffizienzwerten entsprechen. Unter Verwendung dieser neuen Daten werden nun durch ein erneutes Durchführen einer DEA die Bootstrap-DEA Effizienzwerte berechnet. Dieses Verfahren wird B mal wiederholt, wobei B ausreichend groß gewählt wird. Schließlich kann die Verzerrung des Schätzers auf Basis der o. a. Formeln korrigiert werden.

Dieses eröffnet die Möglichkeit, bestimmte Hypothesentests in Bezug auf DEA-Modelle und ihre Spezifikationen durchzuführen. Insbesondere ist es möglich, Tests auf Effizienzunterschiede hinsichtlich bestimmter Teilstichproben durchzuführen. Werden Ärzte analysiert, deren Praxen in unterschiedlichen Regionen angesiedelt sind, deren Bewohner sich hinsichtlich effizienzrelevanter Merkmale unterscheiden, so kann wie folgt vorgegangen werden: Es wird ein Modell für die gesamte Stichprobe gemeinsam geschätzt. Anschließend wird die durchschnittliche Effizienz $\bar{\theta}_i$ für jede Teilstichprobe, die mit A und B bezeichnet seien, bestimmt. Die Prüfgröße $\tau = \bar{\theta}_A / \bar{\theta}_B$ ist, falls die Nullhypothese zutrifft, dass beide Unterstichproben sich im Hinblick auf die durchschnittliche Effizienz nicht unterscheiden, gleich 1. Die Prüfgröße τ_B kann nun im Rahmen des Bootstrap-Verfahrens in jeder einzelnen der B Replikationen berechnet werden und zwar sowohl mit als auch ohne Korrektur der Verzerrung. Aus der Verteilung der τ_B kann nun ein kritischer Wert entnommen werden, anhand dessen sich beurteilen lässt, ob τ signifikant von 1 abweicht.

Für die folgende Analyse wird auf ein Input-orientiertes VRS-Modell abgestellt. Auch wenn es langfristig wünschenswert ist, zu einer Konfiguration mit optimalen Praxisgrößen im ambulanten Bereich zu kommen, kann nicht unterstellt werden, dass ein einzelner Arzt die Zahl seiner Patienten kurzfristig wesentlich beeinflussen kann. Somit wären die auf Basis eines CRS-Modells abgeleiteten Zielgrößen unrealistisch.

4. Resultate

Alle der im Folgenden vorgestellten Resultate beruhen auf derselben Spezifikation: als Inputs werden die drei Kostengrößen für Honorare, Medikation und Zuweisungen verwendet, als Outputs die Zahl der Patienten bis 50 Jahre und die Zahl der Patienten über 50 Jahre. Für Allgemeinmediziner wird ein Modell für die gesamte Stichprobe berechnet sowie vier Modelle für die vier Regionen. Für das Modell auf Basis der gesamten Stichprobe wird zum Berechnen der kritischen τ -Werte das Bootstrap-Verfahren mit 500 Replikationen verwendet. Für jede Facharztgruppe wird ein Modell mit der vollständigen Stichprobe evaluiert. Anschließend werden für alle Datensätze, die mehr als das Minimum von 26 Beobachtungen aufweisen, Ergebnisse auf Basis von kleineren, durch Ziehen ohne Zurücklegen erzeugten Datensätzen nach dem Verfahren von Zhang und Bartels (1998) abgeleitet. Mit den Ergebnissen auf Basis von 26 Beobachtungen ist es möglich, einen Vergleich über alle Fachgruppen auf Basis einer einheitlichen Grundlage durchzuführen.

4.1 Regionale Einteilung der Allgemeinmediziner

Die Resultate, die auf der Grundlage der in Abschnitt 3 vorgestellten Methoden berechnet wurden, zeigen für die Allgemeinmediziner, dass kein nachweisbarer Zusammenhang zwischen Region und Effizienz besteht. Obwohl die Regionen durch eine unterschiedliche Mortalität charakterisiert sind, lassen sich keine signifikanten Unterschiede hinsichtlich der Effizienz, mit der die Leistungen erbracht werden, erkennen.

In Tabelle 3 sind die Effizienzwerte, die für die 801 Allgemeinmediziner in der vorliegenden Stichprobe ermittelt wurden, getrennt nach Regionen aufgeführt. Hierbei ist zu beachten, dass die Information über die Region weder durch Teilung des Datensatzes noch durch Einfügen eines gesonderten Indikators im Rahmen der Berechnungen berücksichtigt wurde. Die Tabelle ist in zwei Teile untergliedert, wobei die obere Hälfte die nicht korrigierten Ergebnisse enthält während in der unteren Hälfte die verzerrungskorrigierten Ergebnisse (vgl.

Abschnitt 3) aufgeführt sind. Im Mittel beträgt die Verzerrung etwa 4.5 % Prozentpunkte (vgl. die Zeilen „gesamt“ in der jeweiligen Hälfte der Tabelle).

Tabelle 3: Ergebnisse für die Allgemeinmediziner

Tests für vier Regionen					Tests Region 1 und 2 vs. Region 3 und 4		
Ø Effizienz		τ			Ø Effizienz		τ
Test-region	Testregion	andere Regionen	Prüf-größe	kritischer Wert	Test-region	Prüf-größe	kritischer Wert
Region 1	,81837	,80958	1,0109	1,0274	,8148	1,0070	1,1014
Region 2	,82028	,8129	,99678	,9842			
Region 3	,80998	,8133	,99592	,9898			
Region 4	,80762	,81309	,99328	,9864			
gesamt	,8123				,8092		
mit Verzerrungskorrektur							
Region 1	,76763	,76733	1,0004	1,0083	,7692	1,0051	1,0110
Region 2	,7713	,76623	1,0066	1,0108			
Region 3	,7659	,7681	,99713	,9915			
Region 4	,76401	,76802	,99478	,9886			
gesamt	,7674				,7653		

Sowohl für die nicht korrigierten (vgl. Spalte 2, obere Hälfte der Tabelle) als auch für die korrigierten Ergebnisse (vgl. untere Hälfte) liegen die Regionen 1 und 2, in denen die Mortalität deutlich über derjenigen der beiden anderen Regionen liegt, über den Werten der Regionen 3 und 4. Die Vermutung, eine höhere Mortalität im Einzugsbereich der Praxis würde zu einem höheren Aufwand bei der Behandlung führen, wird hier in ihr Gegenteil verkehrt.

Die Tests, ob die Effizienz der Leistungserbringer in einer Region signifikant von den übrigen abweicht, fallen allesamt negativ aus, unabhängig davon, ob diesen die nicht korrigierten oder die korrigierten Effizienzwerte zugrunde liegen. Wie in Abschnitt 3 beschrieben, wird für den Test der Quotient aus dem Durchschnitt der Effizienzwerte pro Region und dem Durchschnitt für die übrigen Regionen (beim Test für 4 Regionen „Ø Effizient Rest“) gebildet; dieser wird mit τ bezeichnet (vgl. die entsprechenden Spalten der Tabelle). Diese Größe wird dann im Rahmen eines zweiseitigen Tests mit dem kritischen Wert, der im Rahmen des

Bootstrap-Verfahrens ermittelt wurde, verglichen (vgl. die Spalten „kritisches τ “ in der Tabelle). Der Test wurde so angelegt, dass eine signifikante Abweichung vom Wert 1 für den Quotienten – der Wert 1 reflektiert die im Rahmen der Nullhypothese getroffenen Annahme, beide Durchschnitte seien identisch – überprüft wurde, wobei ein moderates Signifikanzniveau von 10% gefordert war. Sollte der Wert des Quotienten größer 1 sein, wie dies bspw. für die Regionen 1 (nicht korrigiert) bzw. die Regionen 1 und 2 (korrigiert) der Fall ist, so wird getestet, ob der Quotient den Wert 1 signifikant übersteigt (im Fall von Werten größer 1) bzw. unterschreitet (im Falle von Werten kleiner 1). Dies ist bei keinem der Tests der Fall.

Es bleibt zu überprüfen, ob die Ärzte, die anhand der Auswertungen über die gesamte Stichprobe für eine Ansprache ausgewählt worden wären, über die Gruppen unterschiedlich verteilt sind, d. h. dass bei der Auswertung nach den einzelnen Regionen andere Ärzte ausgewählt worden wären als nach der Auswertung über die gesamte Stichprobe.

Tabelle 3 führt die Ergebnisse für die Allgemeinmediziner getrennt nach Regionen auf und beinhaltet einen Vergleich der beiden Vorgehensweisen, dem Schätzen auf Basis einer Gesamtstichprobe und getrennt nach Regionen. Die oben vorgestellten Ergebnisse des Tests weisen nach, dass es keinen signifikanten Unterschied zwischen den Regionen gibt, wenn die Schätzung auf Basis einer Stichprobe erfolgt. Daher wird diese Vorgehensweise bevorzugt. Die 200 bei jeder Vorgehensweise im unteren Viertel der Verteilung der Effizienzwerte identifizierten Ärzte verteilen sich unabhängig von der Art der Schätzung in etwa gleich über die Gruppen (vgl. die Zeile „ges. Stichprobe“ mit dem Referenzwert in der Zeile „25 %“). Dies war auch nicht anders zu erwarten, da die Effizienz pro Region nur geringfügige Abweichungen aufwies. Dies gilt ebenso für den 25 %-Wert der Schätzung mit der gesamten Stichprobe, der nur unwesentlich über die Regionen streut (vgl. die letzte Zeile der Tabelle).

Bedeutsam sind allerdings die Unterschiede im Hinblick darauf, welcher Arzt als im unteren Viertel befindlich identifiziert wurde. So finden sich in Region 1 lediglich 25 (vgl. die Zeile „Schnittmenge“) der 56 Ärzte, die auf Basis der gesamten Stichprobe als im unteren

Quartil befänglich eingestuft worden waren, auch bei den Ergebnissen, die getrennt nach Regionen abgeleitet wurden, wiederum im unteren Vierte; wie auch in allen anderen Regionen liegt die Übereinstimmung hier bei unter 50 %. Offensichtlich wird eine Reihe von Ärzten auf Basis der kleineren Stichproben nicht mehr als stark ineffizient eingestuft, obwohl sie dies ausweislich der Ergebnisse auf Basis der gesamten Stichprobe tatsächlich sind. Die Tatsache, dass hier nur eine geringe Überschneidung zu verzeichnen ist, wirft die Frage auf, ob tatsächlich alle Ärzte aus dem unteren Viertel der Verteilung in die Ansprache einbezogen werden sollten, da die Trennschärfe der Effizienzmessung im Bereich des unteren Fünftel der Verteilung möglicherweise nicht mehr hinreichend ist.

Tabelle 4: Vergleich der Resultate für die Aufteilung nach Regionen vs. gesamte Stichprobe

	Region 1	Region 2	Region 3	Region 4	Gesamt
Beobachtungen	246	190	246	119	801
Gesamte Stichprobe	56	48	64	32	200
25 %	61	48	61	30	200
Schnittmenge	25	21	30	11	87
% Übereinstimmung	42,62 %	43,75 %	49,18 %	36,67 %	43,50 %
25%-Wert	,7441	,7328	,7312	,7351	,7360

Die Ergebnisse für die unterschiedlichen Gruppen von Fachärzten sind in Tabelle 5 dokumentiert. Die Tabelle führt in ihrer oberen und unteren Hälfte jeweils dieselben Ergebnisse auf, um eine übersichtliche Markierung hervorstechender Effekte zu ermöglichen. Die jeweiligen Hälften sind wie folgt strukturiert: die Fachgruppen sind in aufsteigender Reihenfolge nach der Zahl der Beobachtungen (diese sind in der ersten Spalte aufgeführt) geordnet, die Ergebnisse sind dabei in Form einer oberen Dreiecksmatrix abgetragen. Auf der Hauptdiagonalen sind die Ergebnisse für die jeweiligen tatsächlichen Stichprobengrößen aufgeführt, in den Zellen oberhalb der Hauptdiagonalen sind die Ergebnisse für Pseudo-Stichproben, deren Umfang geringer ist als derjenige der eigentlichen Stichprobe, dokumentiert.

4.2 Analyse der Fachärzte

Der Sinn dieser Vorgehensweise wurde in Abschnitt 3 erläutert: Um die Vergleichbarkeit der Ergebnisse zu gewährleisten, müssen diese auf Basis identischer Stichprobengrößen bei einer gleichen Anzahl von Variablen abgeleitet werden. Jede Zeile in der Matrix enthält Ergebnisse, denen jeweils Stichproben identischen Umfangs, wie in Spalte 2 angegeben, zugrunde liegen. Die schattierten Felder der Hauptdiagonalen des oberen Tabellenteils markieren die Facharztgruppen mit der höchsten Effizienz. Im Lichte der Ausführungen in Abschnitt 3 verwundert es nicht, dass die Facharztgruppe mit der geringsten Stichprobengröße (Psychiatrie) die höchste Effizienz erzielt. Auch der zweite Rang in der Reihenfolge der durchschnittlichen Effizienz wird von der Facharztgruppe mit der zweitkleinsten Stichprobe belegt (Neurologie). Hingegen weisen zwei der Stichproben, deren durchschnittliche Effizienz auf den letzten drei Rängen liegt, überdurchschnittliche Stichprobengrößen auf. Es ist demnach nicht auszuschließen, dass allein die Unterschiede der Stichprobengrößen ursächlich für einen Teil der Unterschiede bezüglich der *gemessenen* Effizienz sind.

Ein Vergleich mit den Rangfolgen, die in der unteren Tabellenhälfte abgetragen sind, lässt vermuten, dass dies z. T. der Fall ist. Legt man eine Stichprobengröße von 26 Beobachtungen – dies entspricht dem Umfang der kleinsten Stichprobe in den Originaldaten (Psychiatrie) – zugrunde, dann weisen die Fachgebiete der Inneren Medizin sowie der Gynäkologie die höchsten durchschnittlichen Effizienzwerte auf, gefolgt von der Urologie (vgl. die erste Zeile der unteren Tabellenhälfte). Gleichzeitig findet sich die Neurologie, die auf Basis der Originaldaten den zweiten Rang einnahm, auf einem der hinteren Plätze wieder. Für eine simulierte Stichprobengröße von 29 Beobachtungen (dies entspricht dem Umfang der Stichprobe für die Neurologie) wird die Psychiatrie mit 26 Beobachtungen nicht mehr berücksichtigt. Es ergibt sich ansonsten dasselbe Bild wie in der Zeile „26“. Auch wenn die Stichprobengröße im Rahmen der Simulationen auf 37 erhöht wird, ändert sich an den ersten beiden Plätzen der Rangfolge nichts, statt der Urologie folgt nun die Dermatologie auf Rang 3.

Anhand dieser Ergebnisse ist nicht auszuschließen, dass eine Auswahl von Fachgebieten, die einer eingehenderen Analyse zu unterziehen wären, fehlgeleitet worden würde, solange diese anhand der Originalstichprobengrößen erfolgte.

Es wurde bereits einleitend erwähnt, dass die Fokussierung auf die relative Effizienz innerhalb einer Gruppe nicht notwendigerweise zur Aufdeckung der Wirtschaftlichkeitsreserven über die Gruppen insgesamt führt. In Tabelle 6 ist für die einzelnen Fachgebiete der Übersichtlichkeit halber nochmals die Zahl der Beobachtungen aufgeführt. Des Weiteren sind die gesamten Kosten, die für die jeweilige Facharztgruppe im unteren Viertel der Effizienzverteilung anfallen, in der Spalte „Kosten 25 %“ tabelliert, ebenso die Kosten pro Praxis. In den jeweils nachfolgenden Spalten findet sich die Rangfolge des Fachgebiets für den entsprechenden Indikator in absteigender Reihenfolge der entsprechenden Kostengröße. In der Spalte „DEA 25 %“ ist der 25%-Wert der Effizienz eingetragen, dahinter die entsprechende Rangziffer (in aufsteigender Reihenfolge der Effizienz). Die folgende Spalte „DEA 25 % korr.“ gibt den durchschnittlichen Wert der Effizienz des unteren Viertels aus den Simulationen für 26 Beobachtungen (vgl. Tabelle 5) an, die folgende Spalte beinhaltet einen Korrekturfaktor, der die höchste Effizienz der vorangegangenen Spalte mit 100 % bewertet und alle anderen Werte anteilig berechnet. Grundgedanke dieses Korrekturfaktors ist, dass die auf Basis der vollständigen Stichproben berechneten Größen um die auf einer einheitlichen Stichprobengröße verbleibenden Effizienzunterschiede korrigiert werden sollten. Schließlich weist die Spalte „Potenzial“ das Optimierungspotenzial, das anhand der in „DEA 25 %“ eingetragenen Werte berechnet wurde, aus, die vorletzte Spalte enthält das entsprechende Potenzial, das mit dem Korrekturfaktor gewichtet wurde. Da die beiden Rangfolgen identisch sind (nach dieser Reihenfolge sind die Fachgebiete in der Tabelle sortiert), sind die Ergebnisse qualitativ dieselben. Es versteht sich von selbst, dass dieses anhand eines einzigen Vergleichs einer Reihe von Fachgruppen ermittelte Ergebnis nicht verallgemeinert werden kann.

Tabelle 5: Ergebnisse für die Facharztgruppen

Beobach- tungen	Psychiatrie	Neurologie	Chirurgie	Lungen	Urologie	HNO	Dermato- logie	Orthopädie	Augen- heilkunde	Innere Medizin	Gynä- kologie
26	92.741%	92.713%	93.426%	92.814%	94.735%	88.964%	94.379%	92.173%	92.940%	95.234%	96.410%
29		91.784%	93.028%	92.161%	93.997%	88.197%	93.742%	91.254%	91.832%	94.681%	96.031%
37			91.077%	90.079%	92.613%	85.778%	93.060%	89.278%	90.441%	93.327%	94.928%
39				89.751%	92.329%	85.616%	92.967%	89.309%	90.594%	93.216%	95.093%
46					91.342%	84.222%	92.221%	86.868%	89.171%	91.833%	94.264%
61						82.297%	91.002%	84.763%	87.634%	90.476%	93.170%
76							90.059%	82.034%	86.956%	89.281%	91.946%
85								81.059%	86.280%	88.412%	91.539%
89									86.052%	88.235%	91.185%
99										87.510%	90.627%
115											89.764%
	1	2	4	7	3	10	5	11	9	8	6
	8	9	5	7	3	11	4	10	6	2	1
Beobach- tungen	Psychiatrie	Neurologie	Chirurgie	Lungen	Urologie	HNO	Dermato- logie	Orthopädie	Augen- heilkunde	Innere Medizin	Gynä- kologie
26	92.741%	92.713%	93.426%	92.814%	94.735%	88.964%	94.379%	92.173%	92.940%	95.234%	96.410%
29		91.784%	93.028%	92.161%	93.997%	88.197%	93.742%	91.254%	91.832%	94.681%	96.031%
37			91.077%	90.079%	92.613%	85.778%	93.060%	89.278%	90.441%	93.327%	94.928%
39				89.751%	92.329%	85.616%	92.967%	89.309%	90.594%	93.216%	95.093%
46					91.342%	84.222%	92.221%	86.868%	89.171%	91.833%	94.264%
61						82.297%	91.002%	84.763%	87.634%	90.476%	93.170%
76							90.059%	82.034%	86.956%	89.281%	91.946%
85								81.059%	86.280%	88.412%	91.539%
89									86.052%	88.235%	91.185%
99										87.510%	90.627%
115											89.764%

Tabelle 6: Resultate für die Facharztgruppen II

FG	Beobach- tungen	Kosten				Effizienzwerte						Potenzial			
		25% in Mio. €	Rang	pro Praxis in €	Rang pro Praxis	25 % Rang	25 % Rang	25 % korr.	Korrektur- faktor	korr. Rang	Rang korr.	in €	Rang	korr. in €	Rang korr.
HNO	85	9.26	2	440952	2	,6735	2	,8648	,9116	,7576	5	136430	1	149666	1
Orthopädie	46	3.94	7	328333	5	,6906	3	,9132	,9626	,7174	2	101586	2	105535	2
Urologie	39	4.51	6	451000	1	,8060	7	,8680	,9149	,8809	1	87494	3	95628	3
Augen	99	10.21	1	408400	3	,7872	5	,9487	1,0000	,7872	4	86907	4	86907	4
Innere	29	2.95	9	368750	4	,8098	8	,8427	,8883	,9117	3	70136	5	78958	5
Dermatologie	61	3.19	8	199375	10	,6727	1	,7846	,8270	,8134	7	65255	6	78903	6
Lungen	88	5.34	3	242727	8	,7486	4	,8776	,9251	,8092	9	61021	7	65965	7
Neurologie	76	5.16	5	271578	7	,7961	6	,8912	,9394	,8475	10	55374	8	58947	8
Chirurgie	26	2.22	10	317142	6	,8653	11	,8653	,9121	,9487	8	42719	9	46836	9
Gynäkologie	115	6.06	4	202000	9	,8207	10	,9454	,9965	,8236	6	36218	10	36345	10
Psychiatrie	37	1.57	11	174444	11	,8180	9	,9040	,9529	,8584	11	31748	11	33318	11

Ein Blick auf die Rangfolgen der Kostengrößen sowie der ursprünglichen, auf Basis der jeweils vollständigen Datensätze ermittelten DEA-Werte zeigt, dass weder eine Fokussierung auf die gruppenbezogene Effizienzposition, wie sie sich in der Vorgehensweise der Kasse widerspiegelt, noch eine Konzentration auf die Facharztgruppen mit den höchsten Kosten pro Patient in eine sinnvolle Optimierung münden. Die Fachärzte der HNO, die ausweislich der anhand der Potenziale ermittelten Rangfolge die höchsten Effizienzreserven aufweisen, stehen in keiner der anderen Rangfolgen an erster Stelle. Die ebenfalls hohe Potenziale aufweisenden Fachgebiete der Orthopädie und der Urologie nehmen – je nach Betrachtungsweise – hintere Plätze in den entsprechenden Rangfolgen ein.

Das gezielte Ansprechen von Ärzten mit hohem Optimierungspotenzial bietet offensichtlich die Möglichkeit, entweder bei gleichem Mitteleinsatz höhere erwartete Optimierungserfolge zu realisieren – wird statt eines Psychiaters ein weiterer HNO-Arzt angesprochen, steigt das erwartete Potenzial um über 100.000 € – oder aber bei geringerem Mitteleinsatz ein gleich bleibendes erwartetes Optimierungspotenzial zu realisieren: Das Potenzial bleibt mindestens ebenso hoch, wenn statt vier Psychiatern / Gynäkologen (drei Chirurgen) ein HNO-Arzt angesprochen wird.

Abschließend sei angemerkt, dass es wahrscheinlich nicht praktikabel ist, einen zu hohen Anteil von Ärzten einer Facharztgruppe bei der Ansprache zu berücksichtigen, auch wenn diese hohes wirtschaftliches Potenzial aufweisen, so dass einer Optimierung unter rein wirtschaftlichen Gesichtspunkten hier Grenzen gesetzt sind. Genauso wenig ist es als sinnvoll zu erachten, eine Facharztgruppe aufgrund geringer Potenziale vollständig ohne Ansprache zu lassen. Auch offensichtlich problematische Effizienzwerte, hinter denen sich nur wenig Potenzial in einem wirtschaftlichen Sinne verbirgt, können nicht einfach ignoriert werden.

5. Fazit

Die Auswertung von Daten für Vertragsärzte einer österreichischen Gebietskrankenkasse mittels DEA zeigte auf, dass die Vorgehensweise im Rahmen einer Pilotstudie zur Effizienz-

regulierung vorhandene Optimierungsreserven nicht vollständig aufdeckt. Für Allgemeinmediziner wurde eine Unterteilung der Stichprobe nach Praxisstandort vorgenommen. Es lässt sich jedoch kein empirischer Beleg dafür finden, dass die Effizienz über die Standorte signifikant variiert. Die Unterteilung führt dazu, dass nur eine Minderheit der 200 Ärzte, die auf Basis der Analysen mit der Gesamtstichprobe dem Viertel der Ärzte mit der geringsten Effizienz zugerechnet werden müssen, auch tatsächlich identifiziert wird. Hier muss überdacht werden, ob eine Beschränkung auf das unter Fünftel bzw. auf einen noch kleineren Anteil im Rahmen einer Ansprache sinnvoll wäre, da die mangelnde Überdeckung der beiden Analysen ein Hinweis darauf sein könnte, dass die Effizienzergebnisse bereits im Bereich des unteren Viertels / Fünftels nicht mehr hinreichend trennen.

Die Analyse der Facharztgruppen wurde einer Neubewertung unter zwei Aspekten unterzogen: zum einen wurden Simulationen durchgeführt, die einen Vergleich der Facharztgruppen auf Basis von (Pseudo-) Stichproben einheitlicher Größe ermöglichten. Diese ergaben eine völlig andere Reihung der Facharztgruppen nach Effizienz als die Analysen auf Basis der Originalstichproben. Zudem wurden bei der Bewertung des Optimierungspotenzials die durchschnittlichen Kosten je Praxis berücksichtigt. Hier konnte aufgezeigt werden, dass mit den eingesetzten Mitteln durch Umschichtungen zwischen den Facharztgruppen vermutlich höhere Optimierungspotenziale realisiert werden könnten.

Literatur

- Agrell PJ, Bogetoft P (2005) Should Health Regulators use DEA? In: González Fidalgo E, González López-Valcárcel B, Meneu R, Ventura J (eds.) *Coordinación e incentivos en sanidad*. Asociación de Economía de la Salud, Barcelona, pp. 133-154
- Andes S, Metzger LM, Kralewski J, Gans D (2002) Measuring Efficiency of Physician Practices Using Data Envelopment Analysis. *Managed Care* 11: 48-56
- Bogetoft P, Hougaard JL (1999) Efficiency Evaluations Based on Potential (Non-Proportional) Improvements. *Journal of Productivity Analysis* 12: 233-247
- Banker RD, Morey RC (1986) Efficiency Analysis for exogenously fixed inputs and outputs. *Operations Research* 34: 513-521
- Bundesnetzagentur (2006) Entwurf des Berichtes der Bundesnetzagentur nach § 112a EnWG zur Einführung der Anreizregulierung nach § 21a ENWG (http://www.zfk.de/zfk/infothek/pdf/Anreizregulierung_02_05_06.pdf)
- Burgess JF, Wilson PW (1995) Decomposing Hospital Productivity Changes, 1985–1988: A Nonparametric Malmquist Approach. *Journal of Productivity Analysis* 6: 343- 363
- Chilingerian JA, Sherman HD (1997) DEA and primary care physician report cards: deriving preferred practice cones from managed care service concepts and operation strategies. *Annals of Operations Research* 73:35–66
- Chilingerian JA, Sherman HD (2004) Health Care Applications: From Hospitals to Physicians, From Productive Efficiency to Quality Frontiers. In: Cooper WW, Seiford LM, Zhu J (eds.) *Handbook on Data Envelopment Analysis*. Kluwer, Boston, pp. 349–400
- Cooper WW, Seiford L, Tone K (2000) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer, Boston
- Defelice LC, Bradford WD (1997) Relative inefficiency in production between solo and group practice physicians. *Health Economics* 6: 455–465

- Dyson RG, R Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. *European Journal of Operational Research* 132: 3-17.
- Eisenberg JM (1986) Doctors' decisions and the cost of medical care: the reasons for doctors' practice patterns and ways to change them. Health Administration Press, Ann Arbor
- Gstach D (1995) Comparing Structural Efficiency of Unbalanced Subsamples: A Resampling Adaptation of Data Envelopment Analysis. *Empirical Economics* 20:531-542
- Hollingsworth B (2003) Non-Parametric and Parametric Applications Measuring Efficiency in Health Care. *Health Care Management Science* 6:203-218
- Kneip A, Park B, Simar L (1998) A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores in *Econometric Theory* 14: 783 - 793
- Kneip A, Simar L. (2003) Asymptotics for DEA Estimators in Nonparametric Frontier Models, Discussion paper #317, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve.
- Lovell K (2006) Frontier Analysis in Health Care. *International Journal of Healthcare Technology and Management* 7: 5 - 14
- Newhouse JP (1994) Frontier Estimation: How Useful a Tool for Health Economics. *Journal of Health Economics* 13:317–322
- Ozcan YA (1998) Physician benchmarking: measuring variation in practice behaviour in treatment of otitis media. *Health Care Manage Science* 1: 5–17
- Pai CW, Ozcan YA, Jiang HJ (2000) Regional Variation in Physician Practice Patterns: An Examination of Technical and Cost Efficiency for Treating Sinusitis. *Journal of Medical Systems* 24:103-117
- Pedraja-Chaparro F, Salinas-Jiménez J, Smith P (1999) On the Quality of the Data Envelopment Analysis Model, *Journal of the Operational Research Society* 50: 636-644.
- Pollitt MG, Jamasb T (2001) Benchmarking and Regulation of Electric Utilities: Lessons from International Experience. *Utilities Policy* 9:107-130.

- Rosenman R, Friesner D (2004) Scale and Scope Inefficiencies in Physician Practices. *Health Economics* 13:1091-1116
- Simar L, Wilson PW (1998) Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Model. *Management Science* 44:49-61.
- Simar L, Wilson PW (1999a) Some Problems with the Ferrier/Hirschberg Bootstrap Idea. *Journal of Productivity Analysis* 11:67-80
- Simar L, Wilson, PW (1999b) Of Course We Can Bootstrap DEA Scores! But Does It Mean Anything? Logic Trumps Wishful Thinking. *Journal of Productivity Analysis* 11:93-97
- Simar L, Wilson PW (2000) A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics* 27:779-802
- Simar L, Wilson PW (2006) Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives, unveröffentlichtes Manuskript
- Simar L, Wilson PW (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136: 31-64
- Wagner J, Shimshak D, Novak MA (2003) Advances in physician profiling: the use of DEA. *Socio-economic Planning Sciences* 37:141–163
- Zhang Y, Bartels R (1998) The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand. *Journal of Productivity Analysis* 9:187-204

Anhang

Anhand von Tabelle A2 lässt sich erkennen, dass allein die Honorarleistung sowohl innerhalb der verschiedenen Facharztgruppen als auch über die Facharztgruppen hinweg stark differiert. Deutliche Unterschiede zwischen dem 5 %- und dem 95 %-Wert der Verteilung der Honorare ergeben sich insbesondere für den Bereich Chirurgie (nahezu Faktor 4, vgl. die schattierten Zellen in der Zeile „Chirurgie“) sowie für die Innere Medizin (annähernd Faktor 3; vgl. die schattierten Zellen in der Zeile „Innere“). Inwieweit diese Unterschiede auf Verschiebungen im Fallmix über die einzelnen Praxen zurückzuführen sind, lässt sich erst durch das Berücksichtigen der Altersverteilung der Patienten beurteilen.

Tabelle A1: Verteilung der Honorare über die Facharztgruppen

FG	Minimum	5%	Mittelwert	95%	Maximum
Augen	27.73	32.61	45.92	67.56	81.32
Chirurgie	39.17	39.50	82.09	156.81	208.63
Dermatologie	18.11	23.03	34.87	47.36	61.38
Gynäkologie	29.92	34.59	41.69	52.20	58.72
Innere	36.32	39.52	75.14	118.99	137.25
Kinder	31.83	33.05	46.30	58.17	165.01
HNO	25.13	31.68	52.63	76.61	88.73
Lungen	39.14	49.40	71.91	88.03	95.34
Orthopädie	66.13	66.26	83.39	143.85	145.80
Urologie	35.57	44.21	63.88	87.01	92.60
Neurologie	52.39	58.28	86.87	125.69	136.19
Psychiatrie	48.97	56.67	106.86	234.89	335.58

Auch im Bereich der Heilmittel fallen deutliche Unterschiede ins Auge, wobei hier wiederum der Fachbereich der Inneren Medizin sowie die Fachgruppen der Kinderheilkunde und der HNO betroffen sind.

Tabelle A2: Heilmittel pro Kopf

FG	Minimum	5%	Mittelwert	95%	Maximum
Augen	2.21	3.26	7.59	13.59	21.59
Chirurgie	1.31	1.87	10.59	30.33	35.70
Dermatologie	9.47	10.34	18.82	27.97	42.17
Gynäkologie	2.70	3.38	6.79	12.33	19.65
Innere	3.48	13.89	42.38	86.40	122.30
Kinder	3.87	4.75	10.83	20.61	66.37
HNO	0.71	2.18	6.09	13.43	16.21
Lungen	29.10	33.12	60.23	115.98	117.08
Orthopädie	4.02	5.09	11.35	22.16	30.64
Urologie	0.18	21.17	32.32	45.08	62.20
Neurologie	13.38	27.89	61.94	98.72	106.78
Psychiatrie	34.17	49.78	100.92	142.50	551.30

Die Zuweisungen weisen sowohl innerhalb als auch im Vergleich über die Facharztgruppen hinweg die stärksten Schwankungen auf.

Tabelle A3: Zuweisungen pro Kopf

FG	Minimum	5%	Mittelwert	95%	Maximum
Augen	0.04	0.12	0.65	1.44	3.41
Chirurgie	2.23	4.12	15.91	46.71	47.55
Dermatologie	0.55	0.79	2.28	4.96	6.77
Gynäkologie	7.07	8.45	18.32	33.16	72.87
Innere	3.32	11.83	28.27	75.25	97.87
Kinder	3.65	4.67	8.33	11.43	21.36
HNO	0.78	1.61	6.16	12.06	20.91
Lungen	1.85	2.28	7.46	12.96	13.50
Orthopädie	21.86	25.64	44.57	77.18	82.82
Urologie	3.59	4.80	12.52	29.64	33.80
Neurologie	6.44	6.62	22.11	41.05	42.71
Psychiatrie	1.99	2.32	11.51	25.92	28.13