

Empirische Untersuchungen zu morbiditätsorientierten Regelleistungsvolumina in regionaler Perspektive

Mit der Einführung morbiditätsorientierter Regelleistungsvolumina beabsichtigt der Gesetzgeber, die Vergütung der gesamten vertragsärztlichen Versorgung umfassend zu reformieren. In dem folgenden Beitrag geht es um die Frage, wie gut auf der Ebene von Bundesländern die Leistungsausgaben von Versicherten geschätzt werden können. Die empirischen Ergebnisse zeigen erhebliche regionale Differenzen auf, die mit sehr unterschiedlichen Einflussfaktoren zusammenhängen können. Hier besteht empirischer Untersuchungsbedarf und es ist zu diskutieren, ob und wie diese Einflussfaktoren ggf. im neuen Vergütungssystem berücksichtigt werden können und sollen.

■ Hendrik Dräther, Bettina Gerste und Antje Schwinger

1. Vertragsärztliche Vergütung: veränderte Rahmenbedingungen und Untersuchungsfrage

Gegenwärtig werden zwischen Krankenkassenverbänden und Kassenärztlichen Vereinigungen (KVen) Vereinbarungen zur Fortschreibung der vertragsärztlichen Gesamtvergütungen u.a. unter Berücksichtigung der Mitgliederentwicklung der Krankenkassen getroffen (§ 85 SGB V). Diese Regelung stößt auf Kritik, weil dabei weder Veränderungen der Gesamtzahl der zu versorgenden Versicherten (einschließlich der mitversicherten Familienangehörigen) noch deren behandlungsrelevante Zusammensetzung berücksichtigt werden. In den 90er Jahren hat es zur Sicherung von Beitragssatzstabilität verschiedentlich sogar explizite gesetzliche Vorschriften zur strikten Budgetierung der Gesamtvergütungen gegeben. Ein regelhafter Zusammenhang zwischen der Höhe der kassenspezifischen Gesamtvergütungen und dem tatsächlichen vertragsärztlichen Versorgungsgeschehen besteht daher heute kaum noch.¹

Nach den Vorstellungen des Gesetzgebers soll die Höhe der Gesamtvergütung in Zukunft auf einer anderen Grundlage bestimmt und ausgehandelt werden.² Über „arztgruppenbezogene Regelleistungsvolumina“ (§ 85 a SGB V) soll zunächst für die Versicherten einer Kasse ein vergütungsrelevantes Punktzahlvolumen auf der Grundlage des einheitlichen Bewertungsmaßstabs (EBM) berechnet werden. Dazu sollen die Versicherten auf Basis von Behandlungsdiagnosen in Risikoklassen eingruppiert werden, denen jeweils ein spezifischer Behandlungsbedarf in

EBM-Punkten zugeschrieben wird. Die Risikoklassen sollen mithilfe eines „international anerkannten Patientenklassifikationsverfahrens“ gebildet werden. Der für die Gesamtvergütung einer Kasse relevante Behandlungsbedarf ergibt sich aus dem Behandlungsbedarf je Risikoklasse und der Anzahl ihrer entsprechend eingruppierten Versicherten.

Neben der Ebene der Gesamtvergütung sollen auch auf der Einzelarztebene Regelleistungsvolumina eingeführt werden (§ 85 b SGB V). Diese fungieren als Grenze, bis zu der ein Arzt die von ihm erbrachten Leistungen auf der Grundlage des EBM zu einem festen Punktwert („Regelpunktwert“) abrechnen kann; darüber hinausgehende Leistungen sollen nur noch mit zehn Prozent des Regelpunktwerts vergütet werden. Somit sind auf der Ebene der Gesamtvergütung versichertenbezogene und auf der Einzelarztebene arztbezogene Regelleistungsvolumina zu berechnen, die generell als morbiditätsorientierte Regelleistungsvolumina (M-RLV) bezeichnet werden. Für die weiterhin vorgesehene Aufteilung der Gesamtvergütung auf die Arztgruppen fehlen nähere Vorgaben.

Die folgende Untersuchung soll sich zunächst auf einen bestimmten Aspekt von M-RLV konzentrieren. Es wird empirisch geprüft, inwieweit ausgewählte international anerkannte Patientenklassifikationsverfahren (PKS) geeignet sind, den Versicherten Risikoklassen zuzuordnen, sodass ihre vertragsärztlichen Leistungsausgaben gut geschätzt werden können. Weil es laut Gesetz explizit vorgesehen ist, dass M-RLV nach Versorgungsregionen differenziert werden können, soll die Schätzbarkeit der Leistungsausgaben insbesondere unter regionalen Gesichtspunkten betrachtet werden.

Hendrik Dräther, Bettina Gerste, Antje Schwinger, Mitarbeiter des Wissenschaftlichen Instituts der AOK (WIdO)

2. Methodische Anmerkungen zu den Modellrechnungen

1.1 Datenbasis

Datengrundlage der folgenden Berechnungen sind Abrechnungsdaten einer bundesweit repräsentativen Stichprobe von AOK-Versicherten aus den Jahren 2001 und 2002. Berücksichtigt werden sowohl die Ausgaben für die in EBM-Punkten abgerechneten Leistungen als auch die Sachausgaben, die mit Euro-Beträgen bewertet sind.³ Für alle Berechnungen wurden die EBM-Punkte auf der Basis eines einheitlichen Punktwertes in Euro-Beträge umgerechnet, welcher sich rechnerisch aus den Abrechnungsdaten ableitet. Sachleistungen und in EBM-Punkten bewertete Leistungen wurden für die im Folgenden dargestellten Schätzmodelle in Euro aufsummiert, sofern hierzu keine abweichenden Hinweise gegeben werden.

1.2 Modellansätze und Schätzmaße

Es werden drei PKS getestet. In den beiden Zellenansätzen von ACG (Adjusted Clinical Groups) und 3M werden Versicherte in sich gegenseitig ausschließende Zellen gruppiert.⁴ Das HCC-Modell (Hierarchical Condition Categories) von DxCG ist dagegen ein Zuschlagsmodell, das Versicherten bei entsprechender Morbidität mehrere Risikozuschläge zuweist. Alle PKS klassifizieren die Versicherten auf Basis von Behandlungsdiagnosen, Alter und Geschlecht. Beim Zellenansatz wird jeder Zelle ein bestimmter Geldbetrag zugewiesen, der als Schätzwert für die Leistungsausgaben einer Person in der betreffenden Risikoklasse herangezogen wird. Hierbei handelt es sich um die Durchschnittsausgaben der in eine Zelle eingruppierten Versicherten. Die Geldbeträge je Zuschlag werden mittels einer linearen Regression errechnet, bei der das R^2 maximiert wird.⁵

Bei der empirischen Bewertung der Schätzergebnisse ist die Schätzgüte von Bedeutung, mit der Leistungsausgaben prognostiziert werden können. Dazu werden zwei Gütemaße verwendet.

- Zum einen ist von Interesse, wie gut die *individuellen* Leistungsausgaben geschätzt werden können. Auf einer Skala von 0 bis 100 % wird durch das R^2 angegeben, wie hoch der Anteil der in der Regressionsanalyse durch die unabhängigen Variablen erklärten Varianz an der Gesamtvarianz der individuellen Ausgaben jedes Versicherten ist.⁶ Ein R^2 von 100 % würde eine hundertprozentige Vorhersage der individuellen Leistungsausgaben bedeuten.
- Auf der Ebene der Gesamtvergütung sollen Leistungen für ein Versichertenkollektiv vergütet werden. Hier ist folglich die Frage relevant, wie gut die Leistungsausgaben des jeweiligen Versichertenkollektivs insgesamt getroffen werden. Um dies auszudrücken, wird das Predictive Ratio (PR) herangezogen.⁷ Das PR zeigt die „Vorhersage-Relation“ zwischen geschätzten Ausgaben und Ist-Ausgaben von Personenkollektiven (z. B. Altersgruppen) an.

1.3 Regionalisierung

Die regionalen Untersuchungen beschränken sich auf AOK-Daten aus sieben Bundesländern, auf zeitgleiche Schätzungen und werden exemplarisch auf Basis der HCC-Klassifikation vorgenommen, die in den nachfolgend dargestellten Analysen die besten Schätzergebnisse vorweist. Es wurden mit den hier betrachteten Bundesländern Regionen ausgewählt, bei denen die statistischen Kennzahlen aus der Stichprobe mit denen aus anderen Datenquellen am besten übereinstimmen.⁸

Für die Beurteilung der regionenspezifischen Schätzgüte ist von zentraler Bedeutung, ob die Zuschlagshöhe für die 184 HCC-Zuschläge regional variiert wird oder nicht. Am Beispiel Diabetes bedeutet dies, dass der entsprechende Zuschlag entweder bundeseinheitlich mit demselben Eurobetrag oder für jede Region mit unterschiedlichen Eurobeträgen versehen wird. Letzteres würde z. B. erfolgen, um das jeweils regionenspezifische Vergütungs-, Versorgungs- und Dokumentationsgeschehen gesondert zu berücksichtigen.

3. Allgemeine Schätzungen von Leistungsausgaben

Wie Staffeldt (2004) für BKK-Versicherte, Maydell und Schulte (2005) für Versicherte der Barmer Ersatzkasse und Stillfried und Ryll (2004) für GKV-Versicherte im Zuständigkeitsbereich der Kassenärztlichen Vereinigung Niedersachsen zeigen, können mit den in Nordamerika entwickelten PKS die individuellen Leistungsausgaben auf Basis von Behandlungsdiagnosen gut prognostiziert werden. Auch wenn Vergleiche zwischen diesen Ergebnissen auf einige methodische Probleme stoßen, die u. a. aufgrund unterschiedlicher Abrechnungsdaten, Versichertenkollektive und Modelleinstellungen entstehen, beziehen sich die am R^2 dargestellten Schätzergebnisse immerhin auf das gleiche Versorgungsgeschehen aus einem vergleichbaren Zeitraum (Jahre 2001 bis 2003) in Deutschland und sind somit wesentlich aussagekräftiger als etwa Vergleiche zu Schätzergebnissen aus den USA, die in anderen Anwendungskontexten gewonnen wurden.⁹

Auf Basis der jeweiligen bundesweiten Stichprobe errechnen Maydell und Schulte (2005) R^2 -Werte in prospektiven Modellrechnungen von über 32 % und (Staffeldt 2004) von bis zu 36 %. Die R^2 -Werte bei den Berechnungen von Stillfried und Ryll (2004) bewegen sich grundsätzlich im gleichen Bereich. Im WIdO wurden ebenfalls vergleichbare Schätzergebnisse erzielt (siehe Tabelle 1). Die Schätzgüte erreicht beim prospektiven Modell und unter Anwendung des HCC-Zuschlagsmodells ein R^2 von 35,8 %. Mithilfe des CRG-Zellenmodells wurde ein R^2 von 30,1 % errechnet.

Bei den Schätzergebnissen fällt auf, dass die drei PKS die versichertenbezogenen Leistungsausgaben unterschiedlich gut schätzen. Im zeitgleichen Modell bspw. können die versichertenbezogenen Leistungsausgaben durch das PKS von ACG von der Johns Hopkins Universität Boston am wenigsten erklärt werden (R^2 von 8,4 %), während die

Tabelle 1: R2 auf Basis eines bundesweiten Versichertenkollektivs in %

R-Quadrat zeitgleich	HCC	ACG	CRG
Modell 1: Ausgaben gesamt	43,7	8,4	39,8
Modell 2: Ausgaben gesamt (bis 14.250 je Versicherten)	49,5	21,6	41,1
Modell 3: Ausgaben ohne Sachausgaben	44,9	36,7	33,2
Modell 4: Ausgaben gesamt ohne HKV* (>1.500 €)	57,9	---	---
R-Quadrat prospektiv			
Modell 1: Ausgaben gesamt	35,8	6,0	30,1
Modell 2: Ausgaben gesamt (bis 14.250 je Versicherten)	35,1	13,7	28,7
Modell 3: Ausgaben ohne Sachausgaben	30,7	23,7	23,5

*: HKV = Hochkostenversicherte

Quelle: WIdO 2005

PKS von 3M und DxCG eine relativ hohe Erklärungskraft zeigen (R^2 von 43,7% bzw. 39,8%).

Hier haben die verschiedenen methodischen Ansätze und der Grad der Morbiditätsdifferenzierung der PKS eine erhebliche Bedeutung. Aus der Tatsache, dass das eine oder andere getestete PKS bessere R^2 -Werte erzielt, kann nicht geschlossen werden, dass ein bestimmtes PKS zur Schätzung von individuellen Leistungsausgaben in der vertragsärztlichen Versorgung am besten geeignet ist. So ordnen alle getesteten Klassifikationsverfahren Versicherte entweder in Zellen ein oder weisen ihnen bestimmte Zuschläge zu, wobei allen PKS gemeinsam ist, die Bildung homogener Gruppen entlang des medizinischen Aufwandes zu ermöglichen. Sie greifen dafür auf Regeln und Annahmen zurück, die auf Nordamerikanische Versorgungssysteme und nicht auf die sektorspezifische vertragsärztliche Versorgung in Deutschland geeicht sind. Ein anderer zu beachtender Aspekt ist, dass das Ziel der Bildung ausgabenhomogener Gruppen – eine notwendige Voraussetzung für hohe R^2 -Werte – tendenziell umso weniger gut erreicht wird, je kleiner die Anzahl an relevanten Zellen bzw. Zuschlägen ist. Das betrachtete PKS von ACG weist diesbezüglich die höchste Aggregationsdichte auf, weil jeder Versicherte genau in eine der insgesamt 93 Zellen eingeordnet wird. Bei den hier durchgeführten Berechnungen waren zudem nur die Hälfte dieser Zellen tatsächlich besetzt. Dagegen erhalten bei der HCC-Klassifikation die Versicherten bei entsprechender Morbidität mehrere von insgesamt 184 Zuschlägen, obgleich auch hier mehr als 50 Risikozuschläge kaum oder gar nicht vergeben wer-

den. Da aber die Zuschläge nahezu beliebig miteinander kombiniert werden können, wird das Morbiditätsspektrum von Kollektiven bei diesem Klassifikationsverfahren wesentlich differenzierter abgebildet, was ungleich günstigere Voraussetzungen für das Erreichen von guten Schätzergebnissen schafft. Der vorgenommene Vergleich der betrachteten PKS sagt also lediglich etwas über die unterschiedliche Vorhersagegüte individueller Ausgaben auf der Basis des vorliegenden Zustands der PKS aus.

Die am R^2 gemessene Erklärungskraft schwankt aber nicht nur in Abhängigkeit von dem ausgewählten Klassifikationsverfahren. Es macht sich auch bemerkbar, ob Leistungsausgaben für die gleiche Periode, aus der die Behandlungsdiagnosen stammen (zeitgleiche Berechnungen), oder für die Folgeperiode geschätzt werden sollen (prospektive Berechnungen) (siehe Tab. 1).¹⁰

Wie auch andere Untersuchungen zeigen, werden besonders hohe Leistungsausgaben von Versicherten teilweise deutlich unterschätzt.¹¹ Diese im Folgenden als Hochkostenversicherte bezeichneten Personen (HKV) können einen erheblichen Einfluss auf die am R^2 gemessene Schätzgüte ausüben.¹²

Im Folgenden betrachten wir zwei Typen von Hochkostenversicherten. Einmal solche, bei denen die Leistungsausgaben ca. das 50fache der Durchschnittsausgaben des gesamten Versichertenkollektivs übersteigen. Dies ist bei der bundesweiten Stichprobe bei einer Grenze von etwa 14.250 € je Versicherten der Fall. Davon sind ca. 0,07% der Versicherten betroffen, bei denen ca. 7% der gesamt-

ten Leistungsausgaben anfallen. Ein zweiter Typ eines Hochkostenversicherten wird über eine Leistungsausgabengrenze definiert, bei der die individuellen Leistungsausgaben im Durchschnitt um mehr als 50% unterschätzt werden. Bei zeitgleichen Modellrechnungen auf Basis der HCC-Klassifikation tritt eine solche Fehlschätzung ab einer Leistungsausgabengrenze von 1.500 € auf.¹³ Davon sind ca. 1,2% der Versicherten betroffen, denen insgesamt ca. 16,2% der Leistungsausgaben zuzuordnen sind.

Wenn Ausgaben oberhalb von 14.250 € nicht berücksichtigt bzw. abgeschnitten werden, steigt die am R² gemessene Erklärungskraft beim HCC-Zuschlagsmodell bei zeitgleichen Berechnungen von 43,7% auf 49,5% signifikant an (siehe Tab. 1). Bei dem Zellenmodell von ACG steigt das R² von 8,4% auf 21,6% noch deutlicher an. Dagegen steigt die Erklärungskraft des CRG-Zellenmodells von 39% auf 41,1% nur leicht an. Bei prospektiven Berechnungen mit gekappten Leistungsausgaben oberhalb von 14.250 €. treten nur bei der Klassifikation von ACG deutliche bessere Schätzergebnisse auf (das R² steigt von 6% auf 14% an). Sowohl das HCC-Zuschlagsmodell als auch das CRG-Zellenmodell reagieren hier nur wenig. Sortiert man Versicherte mit Leistungsausgaben oberhalb von 1.500 € aus, so erhöht sich das R² beim HCC-Zuschlagsmodell in zeitgleichen Berechnungen von 43,7% sogar auf 57,9%.

Insgesamt können wir uns der Aussage von Stillfried und Ryll (2004), S. 48 anschließen, dass sich die im Ausland etablierten PKS zur Schätzung der Leistungsausgaben in der vertragsärztliche Versorgung anwenden lassen. Mit

der vorliegenden vergleichenden Analyse unadaptierter Klassifikationssysteme ist jedoch noch keine Aussage darüber getroffen, welches System für die Schaffung morbiditätsbasierter Regelleistungsvolumina in der vertragsärztlichen Versorgung besser geeignet ist.

4. Schätzung von Leistungsausgaben auf regionaler Ebene

1.4 Regionale Schätzungen mit bundeseinheitlichen Zuschlägen¹⁴

So gut sich die Ergebnisse des HCC-Verfahrens von DxCG im Vergleich der Klassifikationsverfahren auf der Bundesebene darstellen, so unterschiedlich sind die Schätzergebnisse bei einer regionalen Betrachtung. Wie aus Tabelle 2 zu entnehmen ist, werden die Leistungsausgaben von Versichertenkollektiven, die in verschiedenen Regionen versichert sind, teilweise erheblich über- und unterschätzt. Die Leistungsausgaben der Region 6 werden insgesamt um ca. 11% überschätzt (PR von 1,115), während die Leistungsausgaben der Region 4 insgesamt um ca. 27% unterschätzt werden (PR von 0,728).¹⁵ Andererseits werden die Leistungsausgaben in Region 1 oder 2 mit einem PR von 1,022 bzw. 0,986 insgesamt nahezu exakt geschätzt. Da ein einheitlicher Punktwert zur Umrechnung der EBM-Punkte in Geldeinheiten verwendet wurde, liegen keine Preisunterschiede für EBM-Punktmengen vor. Diese Über- und Unterschätzungen von Leistungsausgaben müssen daher auf andere Ursachen zurückzuführen sein.

Tabelle 2: Predictive Ratios auf Basis der HCC-Klassifikation, bundeseinheitlicher Zuschläge und Punktwerte

	PR – alle Ausgaben	Anteil Versicherte mit Ausgaben > 14.500 €	PR -Ausgaben bis 14.250 €	Anteil Versicherte mit Ausgaben > 1.500 €	PR – ohne Versicherte mit Ausgaben > 1500 €
R1	1,022	0,09 %	1,047	1,5 %	1,045
R2	0,986	0,06 %	0,989	1,5 %	0,985
R3	1,054	0,03 %	1,031	2,4 %	1,032
R4	0,728	0,06 %	0,725	2,3 %	0,749
R5	0,935	0,09 %	0,951	0,7 %	0,942
R6	1,115	0,13 %	1,134	0,5 %	1,126
R7	0,997	0,09 %	1,006	0,4 %	1,008

Zeitgleiche Berechnungen für das Jahr 2001, Versicherte mit mindestens 270 Versichertentagen und ohne Diagnosen mit den Zusatzkennungen „A“, „V“ und „Z“.

Quelle: WIdO 2005

Generell können die am R^2 gemessenen regional unterschiedlichen Schätzergebnisse mit der regional unterschiedlichen Bedeutung von Hochkostenversicherten zusammenhängen. Ein systematischer Zusammenhang ist aber nicht ersichtlich, da die Leistungsausgaben von Versichertenkollektiven mit einem hohen Anteil an Hochkostenversicherten aus den Regionen R1 bis R4 entweder relativ gut getroffen (R1, R2 und R3) oder auffällig unterschätzt (R4) werden. Die Leistungsausgaben von Versichertenkollektiven mit niedrigem Anteil an Hochkostenversicherten (R5 bis R7) werden entweder deutlich überschätzt (R6), relativ gut getroffen (R7) oder unterschätzt (R5). Werden Leistungsausgaben oberhalb von 14.250 € abgeschnitten, dann bleiben die Über- und Unterschätzungen bei den Versichertenkollektiven praktisch bestehen. Auch wenn Hochkostenversicherte mit Leistungsausgaben oberhalb von 1.500 Euro ausgeschlossen werden, treten weiterhin erhebliche regionale Über- und Unterschätzungen auf.¹⁶ Auch hier fallen die Regionen 3 und 4 mit einem hohen Anteil an Hochkostenfällen auf. Die Herausnahme von Hochkostenfällen ändert an den Unterschätzungen der Leistungsausgaben der Versicherten aus der Region 4 wenig, während die Leistungsausgaben aus Region 3 insgesamt mit einem PR von 1,054 weiterhin gut getroffen werden.¹⁷

Als zweites Zwischenfazit halten wir daher fest: Bei Verwendung eines bundeseinheitlichen Punktwerts und bundeseinheitlichen Zuschlägen werden die Leistungsausgaben der Regionen deutlich unterschiedlich gut geschätzt. Dies ändert sich auch nicht nennenswert, wenn der Anteil der Hochkostenversicherten, der einen möglichen Erklärungsfaktor für die unterschiedliche Güte der Schätzungen darstellt, explizit berücksichtigt wird.

1.5 Regionale Schätzungen mit regional differenzierten Zuschlägen

In die Schätzung der Leistungsausgaben eines Versicherten können im Zuschlagsmodell mehrere individuelle Zuschläge einfließen. Dies Zuschläge ergeben sich aus der Multiplikation des Basiswertes mit einem Relativgewicht. Das Relativgewicht drückt das Verhältnis aus, in dem eine bestimmte Morbidität eines Versicherten zu den gesamten Leistungsausgaben eines Durchschnittsversicherten steht. Der Basiswert stellt die durchschnittlichen Leistungsausgaben eines Versicherten dar. Das kann zum Beispiel bedeuten, dass ein insulinpflichtiger Diabetiker ein Relativgewicht von 2 zugewiesen bekommt, ein Versicherter mit einer kurzen akuten Atemwegsinfektion eines von 0,5. Diese Relativgewichte bestimmen, in welchem Verhältnis die verschiedenen Erkrankungsbilder vergütungsrelevant werden. Für einen Diabetiker ist in diesem Beispiel doppelt soviel zu zahlen wie für den Durchschnittsversicherten, der das Relativgewicht 1 erhält, für den akut Erkrankten, der das Relativgewicht 0,5 hat, lediglich die Hälfte. Versicherte können mehrere Zuschläge zugewiesen bekommen, so dass sich die gesamten errechneten individuellen Leistungsausgaben aus der Summe seiner individuellen Zuschläge ergeben.

Nach § 85 a Abs. 5 S. 4 SGB V wird im Gesetz zwar die Möglichkeit eröffnet, M-RLV an Versorgungsregionen anzupassen, allerdings werden Methode und Kriterien der Anpassung offen gelassen. Es stehen dabei verschiedene Optionen der Anpassung von Risikozuschlägen zur Verfügung, unter denen für die folgende Betrachtung zwei ausgewählt wurden.

• Anpassungsmethode 1 (Regionalisierung des Basiswertes)

Hier werden zunächst Zuschläge auf Basis des bundesweiten Versichertenkollektivs berechnet. Ohne die relative Position der Zuschläge zueinander zu verändern werden die Geldbeträge korrigiert. Zum Beispiel bleibt das Verhältnis der Risikozuschläge für insulinpflichtige Diabetiker zu denen für akut Erkrankte konstant. Die Geldbeträge werden aber ins Verhältnis zu den regionalen Durchschnittsausgaben je Versicherten gesetzt. Das bedeutet: Relativgewichte gelten weiterhin bundeseinheitlich, während die Basiswerte, die hier als Durchschnittsausgaben je Versicherten gelten, regional variieren.

• Anpassungsmethode 2 (Regionalisierung des Basiswertes und der Relativgewichte)

Hier werden die Zuschläge für jedes regionale Versichertenkollektiv unabhängig voneinander mittels linearer Regression (vgl. Abschnitt 2.2) neu kalkuliert. Die Zuschlagsbeträge fallen dabei regional unterschiedlich hoch aus und auch deren relative Positionen zueinander variieren zwischen den Regionen. Mit anderen Worten: Relativgewichte und Basiswerte variieren regional. Da bei diesem Vorgehen die geschätzten Ausgaben insgesamt den tatsächlichen Ausgaben in der Region entsprechen, entstehen bei diesem Vorgehen keine regionalen Fehlschätzungen auf der Ebene von Versichertenkollektiven. Aufgrund der regionalen Anpassung des Basiswertes und der Relativgewichte werden für die einzelnen Regionen bessere Schätzergebnisse erzielt.

1.5.1 Anpassungsmethode 1

Mit Anpassung der bundesweiten Risikozuschläge an die Durchschnittsausgaben je Versicherter je Region verändern sich die Schätzergebnisse (siehe Tabelle 3). Die Leistungsausgaben der Region 1 werden weiterhin insgesamt relativ gut geschätzt (PR von 1,025), während die Leistungsausgaben der Region 3 um fast 20% überschätzt werden (PR von 1,198), nachdem sie vorher (auf der Basis eines bundeseinheitlichen Basiswertes) relativ gut getroffen waren. Die Leistungsausgaben der Versicherten aus Region 4 werden nicht mehr deutlich (PR von 0,728), sondern nur noch leicht unterschätzt (PR von 0,947). Die Leistungsausgaben des Versichertenkollektivs aus der Region 5 werden nicht mehr unter- (PR von 0,935), sondern deutlich überschätzt (PR von 1,114). Und in Region 7 werden die Gesamtausgaben weiterhin relativ gut geschätzt.

Auch wenn Hochkostenversicherte mit Leistungsausgaben oberhalb von 1.500 € aus den Berechnungen herausgelassen werden, treten erhebliche Fehlschätzungen in den Regionen 2, 3, 4 und 7 auf.

Tabelle 3: Anpassungsmethode 1 - Predictive Ratios auf Basis regional an die Durchschnittsausgaben je Versicherten und je Region angepasster Zuschläge

	Alle Ausgaben		Alle Ausgaben, ohne Versicherte mit Ausgaben > 1500 €
	bundeseinheitlicher Basiswert	Regionaler Basiswert	Regionaler Basiswert
R1	1,022	1,025	0,983
R2	0,986	0,947	0,915
R3	1,054	1,198	1,134
R4	0,728	0,947	0,878
R5	0,935	1,114	1,037
R6	1,115	1,164	1,052
R7	0,997	0,986	0,895

Zeitgleiche Berechnungen für das Jahr 2001, Versicherte mit mindestens 270 Versichertentagen und ohne Diagnosen mit den Zusatzkennungen „A“, „V“ und „Z“.

Quelle: WIdO 2005

Es lässt sich also keine am PR gemessene systematische Verbesserung erreichen. Zuschläge auf Basis eines bundesweiten Versichertenkollektivs zu berechnen und anschließend an die jeweiligen regionalen Durchschnittsausgaben anzupassen, führt nicht dazu, dass die regionalen Über- und Unterschätzungen der Ist-Leistungsausgaben wegfallen.

1.5.2 Anpassungsmethode 2

Zuschläge können auch für jede Region unabhängig voneinander berechnet werden. Die Geldbeträge je Zuschlag und die relativen Positionen der Zuschläge zueinander unterscheiden sich dann zwischen den Regionen, nur das PKS bleibt das gleiche. Für das bundesweite Versichertenkollektiv wurde in zeitgleichen Berechnungen ein R^2 von 43,7% errechnet (siehe Kapitel 3). Die regionalen R^2 -Werte streuen bei zeitgleichen Modellrechnungen in einem Intervall von 37,9% bis 67,3% (siehe Tabelle 4). Es fallen hier sowohl die erheblichen Unterschiede beim R^2 als auch die außergewöhnlich hohen R^2 -Werte von 64,9% und 67,3% in den Regionen 6 und 7 auf.

Die individuellen Leistungsausgaben können somit regional unterschiedlich gut mit den 184 Zuschlägen des HCC-Verfahrens geschätzt werden. In der Region 3 besteht mit einem R^2 von 37,9% ein wesentlich niedrigerer signifikanter Zusammenhang zwischen geschätzten und tatsächlichen individuellen Leistungsausgaben als beispielsweise in der Region 6 oder 7. Das könnte z. B. be-

deuten, dass in der Region 3 der Zuschlag für insulinpflichtige Diabetiker in weit geringerem Ausmaß die diesbezüglichen individuellen Leistungsausgaben abbildet als in den Regionen 6 oder 7.

Auch hier soll der Einfluss der Hochkostenversicherten kurz betrachtet werden. Hochkostenversicherte mit Leistungsausgaben oberhalb von 1.500 € üben zunächst keinen eindeutig negativen Einfluss auf die Schätzgüte aus (vgl. Tabelle 4). Zum einen zeigt sich, dass die R^2 -Werte der Regionen 3 und 4 weit auseinander liegen, obwohl deren Anteil an Hochkostenversicherten vergleichbar hoch ist. Das R^2 in der Region 4 ist auch deutlich höher als das in der Region 5, obwohl dort die relative Bedeutung an Hochkostenversicherten viel geringer ist.

Das Herausnehmen von Versicherten mit Leistungsausgaben oberhalb von 1.500 € zeigt dagegen deutliche Wirkungen bei der Schätzgüte. Die R^2 -Werte liegen zwischen 56,9% in Region 1 und 62,5% in Region 5 und damit insgesamt höher und zudem dichter beieinander. Es werden aber auch partiell schlechtere Schätzergebnisse erzielt, da die R^2 -Werte in den Regionen 6 und 7 sinken. Der Rückgang der Schätzgenauigkeit in diesen Regionen zeigt, dass auch die Gruppe der hier betrachteten Hochkostenfälle nicht bundeseinheitlich fehlgeschätzt wird. Dies verweist auf die Notwendigkeit regionaler Differenzierung bei der Abgrenzung von Hochkostenfällen, will

Tabelle 4: Anpassungsmethode 2 - R2 mit regional differenzierten Relativgewichten und Basiswerten

	R ² , alle Ausgaben	Anteil Hochkostenfälle ab 1.500 €	R ² ohne Hochkostenversicherte ab 1.500 €
R1	54,6 %	1,5 %	56,9 %
R2	44,1 %	1,5 %	58,8 %
R3	37,9 %	2,4 %	59,1 %
R4	60,4 %	2,3 %	61,4 %
R5	51,6 %	0,7 %	62,5 %
R6	64,9 %	0,5 %	60,4 %
R7	67,3 %	0,4 %	58,7 %

Zeitgleiche Berechnungen für das Jahr 2001, Versicherte mit mindestens 270 Versichertentagen und ohne Diagnosen mit den Zusatzkennungen „A“, „V“ und „Z“.

Quelle: WIdO 2005

man durch deren Berücksichtigung die Schätzgüte in allen Regionen verbessern.

Als drittes Zwischenfazit kann daher für den betrachteten Ansatz (zeitgleiches HCC-Zuschlagsmodell, Schätzung der individuellen Ausgaben) festgehalten werden: Die Ermittlung von regionsspezifischen Zuschlägen mittels regionaler Regressionen führt im Regelfall (Ausnahme Region 3) zu besseren Schätzergebnissen als der bundesweite Ansatz. Die regionalen R²-Werte weichen jedoch stark voneinander ab. Das Herauslassen von Hochkostenversicherten führt zu einer Homogenisierung der R²-Werte auf einem insgesamt höheren Niveau. Insgesamt werfen die Ergebnisse weitere Fragen zu den regional unterschiedlichen Vergütungs- und Versorgungsstrukturen sowie Kodierpraxen auf.

5. Zusammenfassung und Diskussion

Die empirische Untersuchung hat zu drei zentralen Ergebnissen geführt:

1. Die tatsächlichen Leistungsausgaben der vertragsärztlichen Versorgung können mit den drei betrachteten PKS auf Grundlage von Behandlungsdiagnosen insgesamt recht gut geschätzt werden. Die Schätzergebnisse decken sich weitgehend mit anderen publizierten Berechnungen.

2. Bei der für das HCC-Verfahren exemplarisch erfolgten regionalen Differenzierung kommt es zu erheblichen Fehlschätzungen, wenn mit bundeseinheitlichen Zuschlägen und bundeseinheitlichen Punktwerten geschätzt wird.¹⁸

3. Deutliche regionale Unterschiede in den Schätzergebnissen bleiben auch bei regionalisierten Zuschlägen bestehen.

Am PR gemessene regionale Über- und Unterschätzungen und regional stark abweichende R²-Werte deuten daher einmal mehr an, dass die vertragsärztliche Versorgung regional heterogene Vergütungs- und Versorgungsstrukturen sowie Kodierpraxen aufweist. Worauf die teilweise beträchtlichen Unterschiede in den regionalen Schätzergebnissen im Einzelfall zurückzuführen sind, konnte im Rahmen der hier erfolgten Untersuchung nicht geklärt werden. Wenn die vertragsärztliche Vergütung auf das neue System der M-RLV umgestellt werden soll und dies weithin friktionslos erfolgen soll, sind noch umfassende Untersuchungen zu den Zusammenhängen zwischen Diagnosen, Leistungsanspruchnahme, Leistungsausgaben und PKS erforderlich. Im Folgenden sollen – ohne Anspruch auf Vollständigkeit – einige mögliche Einflussfaktoren genannt werden.

So muss bei der Interpretation der Schätzergebnisse zunächst berücksichtigt werden, dass zur Schätzung der Leistungsausgaben individuelle Behandlungsdiagnosen

verwendet wurden, die erst seit kurzem in dieser Form erhoben und verwaltungstechnisch aufbereitet wurden. Hier können diverse Probleme der Vollständigkeit und bei der Qualität der Daten vorliegen.

Generell sind die niedergelassenen Ärzte erst seit dem 1.1.2000 dazu verpflichtet, die von ihnen erbrachten Leistungen in den Abrechnungsunterlagen mit den nach ICD-10 verschlüsselten Diagnosen zu dokumentieren. Einige Arztgruppen sind von der Verschlüsselungspflicht befreit; zu nennen sind hier u. a. Laborärzte, Pathologen, Zytologen und Radiologen, da von diesen in der Regel keine Diagnosen gestellt werden. Daneben gibt es Arztgruppen, die die Anwendung der ICD-10 sehr kritisch sehen, weil diese aus ihrer Sicht keine geeignete Grundlage zur Abbildung ihres Versorgungsgeschehens darstellen. So haben es insbesondere Hausärzte häufig mit unspezifischen Befindlichkeitsstörungen oder auch funktionellen Störungen bei zunächst unklarem Hintergrund zu tun. Da eine Stellung von Verdachtsdiagnosen ggf. eine zeitlich aufwendige und teure Differentialdiagnostik ohne therapeutische Konsequenzen nach sich ziehen würde, wird hierauf vielfach verzichtet. So liegt es weitgehend im Ermessen des Hausarztes, welche Diagnose(n) er bei nicht näher präzisierten Störungen und Beeinträchtigungen stellt. Ob es hierdurch ggf. zu regionalen Unterschieden im Kodierverhalten der Ärzte kommt, ist unbekannt.

Deutliche Hinweise auf regionale Unterschiede hinsichtlich der Diagnosequalität ergeben sich daraus, dass die Diagnosezusätze „A“ (Ausschlussdiagnose), „V“ (Verdachtsdiagnose) und „Z“ (symptomloser Zustand nach der betreffenden Diagnose) in einigen Regionen gar nicht und in anderen Regionen in unterschiedlicher Häufigkeit verwendet wurden. Die Zuverlässigkeit, mit der Diagnosen eine Erkrankung oder nur einen entsprechenden Verdacht anzeigen, schwankt somit zwischen den Regionen.¹⁹ Da Diagnosen mit den genannten Zusätzen nicht in die Berechnungen einfließen, kann es durchaus sein, dass z. B. ein bloßer Verdacht auf Diabetes in einer Region als solcher kodiert wurde und deswegen bei den Berechnungen unberücksichtigt blieb, in einer anderen Region dagegen in die Schätzung einfließt, weil keine explizite Kennzeichnung als Verdachtsdiagnose erfolgte.

Weithin bekannt ist dagegen, dass das Versorgungsgeschehen selbst regionale Unterschiede aufweist, etwa als Folge unterschiedlicher Angebotsstrukturen. In Regionen mit einer vergleichsweise hohen Arztdichte (etwa im städtischen Raum) werden andere Kapazitäten an medizinischen Leistungen bereitgestellt, angeboten und nachgefragt als in Regionen mit einer relativ niedrigen Arztdichte. Vom Angebot und von der Erreichbarkeit niedergelassener Ärzte hängt ab, ob eine Erkrankung überhaupt erkannt und diagnostiziert wird sowie ob und wie eine diagnostizierte Erkrankung behandelt wird. Auch regionale Unterschiede im Angebot an Haus- und Fachärzten sind hier von Bedeutung, weil bestimmte Leistungen nur nachgefragt und erbracht werden können, wenn entsprechende Fachärzte in der Nähe sind.

Schließlich dürften auch sektorübergreifende Interdependenzen (insbesondere mit der stationären Versorgung) eine wesentliche Rolle in Bezug auf das regionenspezifische Behandlungsgeschehen im ambulanten ärztlichen Bereich spielen. Welche Leistungen innerhalb der vertragsärztlichen Versorgung (je Risikoklasse) erbracht werden, hängt auch von den komplementären und substitutiven Angeboten des stationären Sektors ab, einschließlich der Existenz bestimmter Spezialkliniken. Beispielsweise erfolgt die Versorgung von Dialysepatienten regional sehr unterschiedlich in Bezug auf das jeweilige Mix der beteiligten Leistungssektoren (ambulant/stationär). Umfang und Ausrichtung der stationären Versorgung einer Region könnten somit bereits Auswirkungen darauf haben, welche Diagnosen in der vertragsärztlichen Versorgung überhaupt gestellt werden, und dürften sich in jedem Fall darin niederschlagen, in welchem Sektor Behandlungsleistungen bei gegebenem Morbiditätsspektrum erbracht werden.

Neben primär angebotsseitigen Einflussfaktoren können aber auch vorrangig nachfrageseitige Faktoren von Bedeutung sein, die zum Beispiel in der sozioökonomischen Struktur der Bevölkerung begründet sind („schichtspezifische Inanspruchnahme“) und – bei gleicher Morbidität – zu erkennbaren Regionalunterschieden im Leistungs- und Ausgabengeschehen führen.

Regionale Unterschiede im Behandlungsgeschehen der Ärzte können auch aus regional unterschiedlichen Vereinbarungen zwischen Krankenkassen und KVen in Bezug auf extrabudgetäre Leistungen und sonstige Sondervereinbarungen resultieren. Extrabudgetäre Leistungen sind prinzipiell nicht durch die Gesamtvergütung begrenzt, sodass bestimmte medizinische Leistungen in einer Region mengenmäßig gedeckelt und in einer anderen Region ungedeckelt sein können. Daneben gibt es Sondervereinbarungen über Leistungen, die nicht über den EBM bundeseinheitlich abgerechnet und überwiegend als Sachleistungen vereinbart werden; der regional deutlich variierende Sachkostenanteil (zwischen 11 und 24 Prozent der Gesamtausgaben) gibt einen Hinweis auf die Dimension der entsprechenden Leistungsanteile. Schließlich können unterschiedliche Verhaltensanreize auch durch regionale Unterschiede in der Höhe des rechnerischen Punktwerts in Verbindung mit den regionalen Sondervereinbarungen ausgelöst werden.

Die Liste möglicher Einflussfaktoren, die sich in regionalen Besonderheiten im jeweiligen Dokumentations-, Versorgungs- und Vergütungsgeschehen niederschlagen und – wie gezeigt – zu den teilweise beträchtlichen Unterschieden bei den regionalen Schätzergebnissen führen können, ließe sich noch leicht erweitern. Da sich zudem auch die Kassen und Kassenverbände hinsichtlich ihrer Vertrags- und Vergütungsvereinbarungen mit den KVen, im Morbiditätsspektrum und nicht zuletzt auch – u. a. schichtenspezifisch bedingt – im Inanspruchnahmeverhalten ihrer Versicherten unterscheiden, ergibt sich somit eine Vielzahl an offenen Untersuchungsfragen für weitergehende empirische Untersuchungen.

Eine andere, nicht empirisch, sondern letztlich normativ zu beantwortende Frage ist es dagegen, welche der beobachteten regionalen Besonderheiten im Rahmen der M-RLV bei der vertragsärztlichen Vergütung in Zukunft berücksichtigt werden sollen (und welche nicht) und wie dies ggf. konkret geschehen soll. Dabei ist stets zu beachten, dass entsprechende Entscheidungen nicht nur Konsequenzen für das Verhältnis zwischen den miteinander konkurrierenden Krankenkassen haben, sondern immer auch unmittelbar auf die relativen Vergütungspositionen der Vertragsärzte – nach Arztgruppen und Regionszugehörigkeit – einwirken.

Literaturverzeichnis

- Buchner, Florian; Ryll, Andreas und Wasem, Jürgen (2002): Periodenbezogene Vergütungssysteme: Die risikoadjustierte Kopfpauschalvergütung. In: Eberhard Wille (Hrsg.): Gesundheitsökonomische Beiträge (38). Baden-Baden, Nomos Verlags-Gesellschaft: 63-95.
- Cumming, Robert B.; Knutson, David; Cameron, Brian A. und Derrick, Brian (2002): A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations. Society of Actuaries, www.soa.org.
- Felder, Stefan (2004): „Reformbedarf der ärztlichen Gesamtvergütung in Deutschland.“ Gesundheits- und Sozialpolitik, 11-12, 32-34.
- IGES; Lauterbach, Karl W und Wasem, Jürgen (2004): Klassifikationsmodelle für Versicherte im Risikostrukturausgleich – Endbericht – im Auftrag des Bundesministeriums für Gesundheit.
- Jacobs, Klaus; Cassel, Dieter und Wasem, Jürgen (2002): Zur Wirkung des Risikostrukturausgleichs in der gesetzlichen Krankenversicherung – Eine Untersuchung im Auftrag des Bundesministeriums für Gesundheit. Schriftenreihe des Bundesministeriums für Gesundheit (140). Baden-Baden, Nomos Verlagsgesellschaft.
- Maydell, Boris von und Schulte, Claudia (2005): „Möglichkeiten zur Justierung von Patientenklassifikationssystemen für die ambulante Vergütung – Eine Analyse mit Hilfe der Clinical Risk Risk Groups.“ Gesundheit und Sozialpolitik 2005(1-2), 28-35.
- Schönbach, Karl-Heinz (2000): „Arzthonorare: Anpassungsbedarf des Systems der Kopfpauschalen.“ Die BKK 2000(9), 375-378.
- Schröder, Wilhelm und Ryll, Andreas (2002): Pauschalierende Vergütungssysteme in der Integrierten Versorgung. In: Tophoven, C. und Lieschke, L. (Hrsg.): Integrierte Versorgung. Köln, Deutscher Ärzte-Verlag: 131-169.
- Staffeldt, Thomas (2004): „Morbiditätsbezogene Regeleistungsvolumina in der ärztlichen Vergütung – Eine Machbarkeitsanalyse.“ Gesundheits- und Sozialpolitik 11-12, 51-59.
- Stillfried, Dominik Graf von und Ryll, Andreas (2004): „Umsetzbarkeit morbiditätsbezogener Regeleistungsvolumina in der vertragsärztlichen Versorgung.“ Gesundheits- und Sozialpolitik 2004(11-12), 36-50.
- SVRG, Sachverständigenrat, für die Konzertierte Aktion im Gesundheitswesen (2001): Gutachten 2001, Band III: Über-, Unter- und Fehlversorgung. Baden-Baden, Nomos.
- Van den Ven, Wynand und Randall, Ellis (2000): Risk adjustment in competitive health markets. In: A.J. Culyer und J.P. Newhouse (Hrsg.): Handbook of Health Economics: 91-95.

Fußnoten

- 1 Vgl. hierzu Schönbach (2000), Jacobs et al. (2002: 55ff.) und Felder (2004).
- 2 Einführung von M-RLV nach §§ 85 a SGB V durch das Gesetz vom 14.11.2003 (BGBl I, S. 2190 ff.).
- 3 Ausgaben für vertragsärztliche Leistungen trennen sich in vertragsärztliche (Dienst-) Leistungen und Sachkosten. Vertragsärztliche Dienstleistungen werden grundsätzlich durch Punkte, Sachkosten grundsätzlich durch Eurobeträge bewertet. Eine Ausnahme bildet das Kapitel U des EBM. Die dort aufgeführten Eurobeträge sind Aufwandsentschädigungen und zählen in den alten EBM Regelungen nicht zu Sachkosten.
- 4 Siehe hierzu ausführlicher IGES, Lauterbach und Wasem (2004).
- 5 Es wurde darauf verzichtet, für die Berechnung der Risikozuschläge eine Lernmenge zu bilden und für die anschließenden Schätzungen von individuellen Leistungsausgaben auf eine Validierungsmenge zurückzugreifen.
- 6 Vgl. hierzu Buchner, Ryll und Wasem (2002: 72).
- 7 Siehe hierzu Staffeldt (2005).
- 8 Die ausgewählten Regionen kommen sowohl aus den Neuen als auch aus den Alten Bundesländern und umfassen auch einen Stadtstaat. Bei der Abbildung der Regionen liegt eine gewisse Unschärfe darin, dass hier die Daten der AOK einer Region herangezogen werden, die auch die Daten der Versicherten enthalten, die außerhalb des betreffenden Bundeslandes Leistungen in Anspruch genommen haben. Insofern werden hier genau genommen Schätzunterschiede zwischen AOKs abgebildet, die aber natürlich eine weitgehende Annäherung an regionale Unterschiede darstellen. Für die Zwecke der vorliegenden Analyse ist dies völlig ausreichend.
- 9 Siehe z. B. Cumming et al. (2002).
- 10 Zeitgleiche Berechnungen sind erwartungsgemäß schätzgenauer als prospektive, u.a. weil (individuelle) Akuterkrankungen in der Folgeperiode durch Behandlungsdiagnosen aus der Vorperiode nicht erklärt werden können.
- 11 Siehe hierzu Van den Ven und Randall (2000: 32 ff.).
- 12 Siehe hierzu auch Maydell und Schulte (2005).
- 13 Die Grenze, ab der noch höchstens 50 % der Leistungsausgaben geschätzt werden – hier 1.500 € –, ist vom verwendeten Klassifikationsverfahren, den zu schätzenden Leistungsausgaben (z. B. mit oder ohne Sachausgaben) oder der Zeitperspektive (zeitgleiche oder prospektive Berechnungen) abhängig.
- 14 Auf Basis von bundesweit einheitlichen Zuschlägen wird die regionale Schätzgüte mithilfe des PR verglichen.
- 15 Es handelt sich um einen empirischen Befund, der auch bei Staffeldt (2004: 56) zu sehen ist.
- 16 Hier handelt es sich um ein andere Definition von Hochkostenfällen. Die Grenze korrespondiert ungefähr mit derjenigen Ausgabenklassen, bei der im zeitgleichen Modell die versichertenbezogenen Ausgaben um mehr als 50 % unterschätzt wurden.
- 17 Hier können weitere methodische Aspekte eine erhebliche Rolle spielen. So ergeben eigene Untersuchungen deutliche Unterschiede bei den Schätzergebnissen, wenn die Risikozuschläge nach Herausnahme der Hochkostenversicherten neu angepasst werden oder die auf Grundlage der gesamten Versichertenstichprobe berechneten Risikozuschläge erhalten bleiben. Die am PR gemessenen regionalen Fehlschätzungen bleiben aber grundsätzlich bestehen.
- 18 Es ist nicht zu erwarten, dass bei einem der beiden anderen PKS grundsätzlich andere Schätzergebnisse erzielt werden.
- 19 Der Gesetzgeber ist hier inzwischen aktiv geworden und hat die Verwendung der Diagnosenzusätze „A“, „V“, „Z“ und „G“ (gesicherte Diagnose) seit dem 1.1.2005 zwingend vorgeschrieben.