

Instrumente zur Beurteilung des «Case-Mix»: eine systematische Literaturanalyse

Karin Huwiler-Müntener^a,
Georg von Below^b,
Martina Hersperger^b,
Ludwig T. Heuss^b,
Matthias Egger^a

a Institut für Sozial- und Präventivmedizin, Universität Bern

b Verbindung der Schweizer Ärztinnen und Ärzte FMH, Abteilung Daten, Demographie und Qualität DDQ

Interessenbindungen und mögliche Interessenkonflikte

Die Autoren des Instituts für Sozial- und Präventivmedizin (ISPM) der Universität Bern wie auch das ISPM sind politisch vollständig unabhängig. Die Arbeit wurde von der FMH finanziert.

* Die Begriffe Instrument, Modell und Indikator werden synonym verwendet. Es handelt sich um Erfassungssysteme, die auf verschiedenen Variablen wie Alter, Geschlecht, Diagnosen usw. beruhen und dazu verwendet werden, die Kosten einer Patientenpopulation vorherzusagen.

** Im Bericht ist oft vom Risikoausgleich die Rede, weil ein grosser Teil der untersuchten Literatur in diesem Kontext anzusiedeln ist.

Korrespondenz:
Prof. Dr. med. Matthias Egger
Institut für Sozial- und Präventivmedizin
Universität Bern
Finkenhübelweg 11
CH-3012 Bern
Tel. 031 631 35 11
Fax 031 631 35 20

egger@ispm.unibe.ch

www.ispm.ch

Zusammenfassung

Alter und Geschlecht bilden die Morbidität einer Patientenpopulation nur ungenügend ab. In der vorliegenden systematischen Übersichtsarbeit analysierten wir publizierte Instrumente, welche in der Grundversorgung für die Vorhersage der Kosten verwendet werden. Als Mass der Vorhersagekraft wurde der durch die Instrumente erklärte Anteil der Kostenvarianz (R^2) gewählt. Insgesamt wurden 68 Publikationen mit rund 500 R^2 analysiert. Die am häufigsten vorkommenden Variablen waren soziodemographische Angaben, Diagnosen und Medikamente. Die R^2 schwankten zwischen 0,3 und 64%, der Median lag bei 11%. Höhere R^2 resultieren bei Instrumenten, welche Angaben über Diagnosen oder Medikamente enthielten. Ambulante Kosten können tendenziell besser vorhergesagt werden als Gesamtkosten. Bei der Auswahl eines geeigneten Instrumentes zur Abbildung der Patientenstruktur sollte nicht nur die Vorhersagekraft bezüglich Kosten, sondern auch der Aufwand bei der Datenerhebung und gegebenenfalls die Setzung von (falschen) Anreizen berücksichtigt werden.

Hintergrund

Die unterschiedliche Morbidität von Patientenpopulationen wird heute in der Schweiz vor allem im Zusammenhang mit dem Risikoausgleich zwischen Krankenversicherern und mit den unterschiedlichen Kosten in der Betreuung von scheinbar vergleichbaren Patienten diskutiert. Um unterschiedliche Morbiditäten abzubilden, kommen aktuell hauptsächlich Alter und Geschlecht zur Anwendung, die Morbidität einer Population kann jedoch nur durch Berücksichtigung weiterer Aspekte genauer abgebildet werden. Die vorliegende Übersichtsarbeit ist Teil des Projektes «Morbiditätsindikatoren» der FMH, welches geplant wurde, um «ein schweizweit einheitliches System zur Dokumentation der Morbidität der versorgten Patientenpopulation zu entwickeln bzw. zu adaptieren und zum Einsatz zu bringen» [1]. Ziel des Projektes ist die «Schaffung von Transparenz bezüglich Strukturen und Prozessen in der medizinischen Versorgung» [1]. Die vorliegende Literaturübersicht soll einen Überblick über bereits existierende Instrumente*

zur Messung der Patientenstruktur in der Grundversorgung und ihre Eignung zur Kostenprognostizierung geben**.

Methoden

Kostenvorhersage

Kosten werden hauptsächlich mit Hilfe von Regressionsmodellen vorhergesagt: man versucht, Gesundheitskosten aufgrund der Patientenstruktur vorherzusagen. Die Patientenstruktur wird dabei anhand von Alter, Diagnosen, Medikamenten oder früheren Kosten abgebildet. In der Regressionsanalyse wird untersucht, welche Variablen für die Kosten relevant sind und welcher Anteil der Kosten mit diesem Modell erklärt werden kann. Dies wird auch als *Modellentwicklung* oder *Modellschätzung* bezeichnet. Um zu beurteilen, wie gut ein Modell die Kosten in verschiedenen Patientengruppen erklärt, muss die Vorhersagekraft in einer unabhängigen Population überprüft werden. Dies wird als *Modellvalidierung* bezeichnet.

Beurteilung der Vorhersagekraft von Modellen

Zur Beurteilung der Vorhersagekraft von Modellen wird häufig das R^2 (das Quadrat des multiplen Korrelationskoeffizienten) verwendet. R^2 gibt an, welcher Anteil der Kostenvarianz mit dem untersuchten Modell erklärt werden kann; es kann Werte von 0% bis 100% annehmen. Je grösser R^2 , umso besser kann ein Modell Kosten vorhergesagen. R^2 hängt nicht nur von den Variablen ab, die in das Modell eingeflossen sind. Beispielsweise sind R^2 aus Modellentwicklungen tendenziell höher als R^2 aus Validierungen [2, 3]. Weiter ist der zeitliche Zusammenhang zwischen der Erhebungsperiode und der Periode, für welche die Kosten vorhergesagt werden, wichtig: Werden die Kosten für die gleiche Periode modelliert, in der die Variablen erhoben werden (zeitgleiche Vorgehensweise), resultieren höhere R^2 , als wenn die Kosten für eine spätere Periode vorhergesagt werden (prospektive Vorgehensweise). Weiter resultieren höhere R^2 , wenn die Kosten auf einen Maximalwert limitiert werden (Kosten, die diesen Wert übersteigen, werden auf den Maximalwert zurückgesetzt).

Berücksichtigte Studien

Die vorliegende Arbeit basiert auf publizierten Studien, die in elektronischen Datenbanken (Pubmed, Embase, Business Source Premier) und aufgrund der Referenzlisten relevanter Artikel identifiziert wurden. Originalarbeiten, welche die Vorhersagekraft von Instrumenten bezüglich Kosten in der Primärversorgung untersuchten, wurden berücksichtigt. Wir nahmen nur Studien auf, die repräsentative Patientengruppen untersuchten, die die Modelle mittels R^2 beurteilten und die Gesamtkosten (also ambulante und stationäre Versorgung) oder die Kosten der ambulanten Versorgung untersuchten. Publikationen aus den Jahren 1990 bis 2005 in englischer, deutscher oder französischer Sprache wurden analysiert. Wir erfassten Angaben zu den Instrumenten (z.B. enthaltene Variablen), zur untersuchten Population (z.B. Alter) und zu den vorhergesagten Kosten.

Statistische Analyse

Die R^2 der verschiedenen Studien werden als Median und Interquartilrange (Bereich, der die mittleren 50% der R^2 abdeckt: 25% der Werte sind kleiner als die untere Grenze, 25% grösser als die obere Grenze des IQR) dargestellt. Median und IQR wurden auch für Untergruppen berechnet: prospektive bzw. zeitgleiche Studienanlage, Gesamt- bzw. ambulante Kosten und rein diagnosebasierte bzw. rein medikamentenbasierte Instrumente. Als «rein diagnosebasiert» wurden Instrumente definiert, die Diagnosen enthielten, aber frühere Kosten, Medikamente oder selbstberichtete Erkrankungen nicht berücksichtigten. An weiteren Variablen durften einzig Alter und Geschlecht enthalten sein. Als «rein medikamentenbasiert» wurden Instrumente definiert, die den Gesundheitszustand über Medikamente einschätzten und weder Diagnosen, frühere Kosten oder selbstberichtete Erkrankungen enthielten. Alter und Geschlecht konnten wiederum enthalten sein. Schliesslich beschreiben wir die Eigenschaften von Studien mit einem relativ hohen R^2 (>40%).

Resultate

Literatursuche

Die Literatursuche ergab über 1000 Publikationen, wobei nur 115 potentiell relevant waren. Von den 115 Publikationen wurden 47 nach näherer Analyse ausgeschlossen, und 68 Publikationen in die Übersichtsarbeit aufgenommen. Die Studien wurden zwischen 1990 und 2005 publiziert, rund die Hälfte nach 2000. Insgesamt 51 Studien stammen aus den USA, 9 aus den Nie-

derlanden, je zwei aus Deutschland und der Schweiz, und je eine aus Australien, Belgien, Kanada und Spanien. Eine Liste der 68 Studien ist unter www.ispm.ch/index.php?id=1191 erhältlich.

Übersicht über die identifizierten Instrumente

Die *soziodemographischen* Instrumente enthalten nur soziodemographische Variablen. Am häufigsten werden Alter und Geschlecht verwendet; diese Variablen sind auch in den meisten Instrumenten der anderen Gruppen enthalten. Weitere Beispiele sind eine gesundheitlich bedingte Rentenberechtigung sowie der Invaliditätsstatus.

Zur Gruppe der *diagnosebasierten* Instrumente haben wir Instrumente gezählt, die ICD-codierte («International Classification of Diseases») Diagnosen enthalten. Die Instrumente dieser Gruppe unterscheiden sich darin, dass bei einigen Instrumenten nur die Diagnose mit den höchsten Kosten von Bedeutung ist, bei anderen werden mehrere Diagnosen berücksichtigt. Weitere Unterschiede betreffen die Gruppierung oder die Herkunft der Diagnosen. Während einige Modelle ausschliesslich auf im ambulanten Bereich gestellten Diagnosen beruhen, schliessen andere nur anlässlich einer Hospitalisierung gestellte Diagnosen ein. Neuere Modelle berücksichtigen oft Diagnosen aus beiden Bereichen.

Bei den *medikamentenbasierten* Instrumenten wird über verordnete Medikamente auf zugrundeliegende Erkrankungen geschlossen. Wie bei den diagnosebasierten Modellen können nicht alle Medikamente berücksichtigt werden; die meisten Instrumente fokussieren auf Medikamente, die bei längerdauernden Erkrankungen eingesetzt werden und möglichst spezifisch sind, das heisst, bei nur einer oder wenigen Erkrankungen zum Einsatz kommen (z.B. Insulin). Oft ist auch eine minimale Dauer der Verordnung erforderlich.

Andere Modelle basieren auf Angaben der Patienten zu ihrem *selbsteingeschätzten Gesundheitszustand*. Dies kann die Antwort auf die Frage sein «Wie geht es Ihnen gesundheitlich?», berichtete Symptome oder Einschränkungen in den alltäglichen Verrichtungen. Eine weitere Gruppe von Modellen verwendet die früheren Kosten einer Person zur Vorhersage der zukünftigen Kosten (*kostenbasierte Instrumente*).

Instrumente bzw. einzelne Variablen können auch *kombiniert* werden. Relativ häufig untersucht wurde die Kombination aus diagnose- und medikamentenbasierten Instrumenten. Dabei werden beispielsweise Spitaldiagnosen mit ambulant verordneten Medikamenten kombiniert.

Andere Instrumente (z.B. der Thurgauer Morbiditätsindex, TMI) [4] konnten wir keiner der obigen Gruppen zuordnen. Der TMI bildet den Gesundheitszustand von Patienten ab, basiert aber nicht auf einer ICD-Codierung. Andere Modelle dieser Gruppe beinhalten nebst Diagnosen auch einzelne chirurgische Interventionen.

Beschreibung der Studien

Aus den 68 Publikationen haben wir 492 R² extrahiert. Die Publikationen enthielten zwischen einem und 28 R². Rund 40% der R² stammten aus einer Validierung. 70% der R² wurden prospektiv berechnet und in rund 80% wurden die Gesamtkosten als Endpunkt verwendet. Die zur Schätzung von R² untersuchten Populationen hatten im Mittel eine Grösse von 397 504 Personen. Der Median war 37 000 Personen (IQR 9479–374 572). Diagnosebasierte Instrumente wurden am häufigsten untersucht. Die R² schwankten zwischen 0,3% und 64,4% (Median 11%, IQR 6,9–21%).

Tabelle 1 zeigt den erklärten Anteil der Kostenvarianz (Median und IQR der R² in %) für Gesamt- bzw. ambulante Kosten bei prospektiver bzw. zeitgleicher Erhebung. Insgesamt lag der Median der Studien mit prospektiver Erhebung bei 9% (IQR 6–13%), bei Studien mit zeitgleicher Erhebung bei 36% (IQR 24–43%). Auch in-

nerhalb einer Kostengruppe resultieren höhere R² bei zeitgleicher Erhebung, und ambulante Kosten können tendenziell besser vorhergesagt werden. Bei diesen Vergleichen muss allerdings beachtet werden, dass weitere, hier nicht berücksichtigte Faktoren ebenfalls einen Einfluss auf die R² haben.

Tabelle 2 unterscheidet zwischen diagnose- und medikamentenbasierten Instrumenten. Bei letzteren konnten nur die Gesamtkosten untersucht werden; bezüglich ambulanter Kosten waren nicht genügend Daten vorhanden. Wiederum sind die R² bei zeitgleicher Erhebung höher als bei prospektiver Erhebung. Ambulante Kosten können mit den rein diagnosebasierten Instrumenten tendenziell besser vorhergesagt werden als die Gesamtkosten. Hier muss jedoch berücksichtigt werden, dass die Anzahl Beobachtungen (n) in einzelnen Untergruppen klein ist. Auch sind andere Variablen, welche die R² beeinflussen, nicht berücksichtigt.

Insgesamt 49 R² waren grösser als 40%; diese R² stammten aus 16 Publikationen mit jeweils einem bis sechs R². Vierzig (82%) waren zeitgleich berechnet worden, 9 (18%) prospektiv. Von den 9 prospektiv berechneten R² enthielten 8 (89%) Modelle frühere Kosten. Bei 14 (29%) R² wurden die ambulanten Kosten untersucht; 13 (27%) R² stammen aus einer Validierung, 36 (73%) aus einer Modellschätzung.

Tabelle 1
Erklärter Anteil der Kostenvarianz (R²) für prospektive bzw. zeitgleiche Erhebung; nach Kostenart. Basierend auf 438 R².

	Ambulante Kosten	Gesamtkosten
Prospektive Erhebung	(n = 41)	(n = 299)
– Median	14%	9%
– IQR	11–21%	6–12%
Zeitgleiche Erhebung	(n = 27)	(n = 71)
– Median	38%	31%
– IQR	36–42%	21–43%

Tabelle 2
Erklärter Anteil der Kostenvarianz (R²) für prospektive bzw. zeitgleiche Erhebung; nach Kostenart und Inhalt der Modelle. Basierend auf 147 R².

	Ambulante Kosten Rein diagnosebasiert	Gesamtkosten Rein diagnosebasiert	Gesamtkosten Rein medikamentenbasiert
Prospektive Erhebung	(n = 8)	(n = 63)	(n = 21)
– Median	18%	9%	11%
– IQR	10–19%	8–14%	8–12%
Zeitgleiche Erhebung	(n = 17)	(n = 32)	(n = 6)
– Median	39%	36%	21%
– IQR	38–42%	29–43%	18–25%

Anwendung mehrerer Instrumente in der gleichen Population: ein Beispiel

Da sich viele Studien in der Patientenpopulation, dem zeitlichen Verhältnis zwischen Erhebungsperiode der Patientenvariablen und der Zeitperiode, für die die Kosten prognostiziert werden, sowie weiteren Aspekten unterscheiden, sind Vergleiche von R² über verschiedene Studien mit Vorsicht zu interpretieren. Wir stellen deshalb hier eine Studie vor, die mehrere Instrumente im gleichen Datenset untersucht hat.

Diese Studie wurde von Reschke et al. im Zusammenhang mit dem Risikostrukturausgleich (RSA) zwischen Krankenversicherern in Deutschland durchgeführt [5]. Aktuell basiert der RSA auf den Variablen Alter, Geschlecht, einer Rentenberechtigung sowie einer allfälligen Einschreibung in strukturierte Behandlungsprogramme. Der RSA soll künftig morbiditätsorientiert durchgeführt werden. Deshalb wurden Instrumente berücksichtigt, welche sowohl den ambulanten als auch den stationären Bereich abdecken. Die Modelle wurden in einer Population von etwa 850 000 Personen mit Daten aus den Jahren 2001/02 evaluiert. Tabelle 3 zeigt die in den jeweiligen Modellen enthaltenen Variablen.

Alle Modelle beinhalten die Variablen Alter, Geschlecht und Rentenberechtigung. Die diagnosebasierten Modelle ACG, ACG-PM, CDPS und HCC enthalten zusätzlich ambulante und stationäre Diagnosen, die Kombinationsmodelle PCG+DCG und RxGroups+IP-HCC Diagnosen aus dem stationären Bereich sowie ambulant verordnete Medikamente. Das Modell ACG-PM ist eine Weiterentwicklung der ACGs und beinhaltet u.a. den sog. HOSDOM-Parameter, welcher angibt, ob ein Patient eine Diagnose aufweist, welche erfahrungsgemäss bei $\geq 50\%$ der Patienten zu einer Spitaleinweisung führt.

In der Studie wurde die Vorhersagbarkeit der Gesamtkosten wie auch der ambulanten Kosten (ärztliche Leistungen) zeitgleich und prospektiv untersucht, wobei es sich um Modellschätzungen handelt. Tabelle 4 zeigt die resultierenden R^2 .

Erwartungsgemäss schneidet der RSA am schlechtesten ab. Zeitgleich berechnete R^2 sind

innerhalb einer Kostenart jeweils höher als die prospektiv berechneten. Von den diagnosebasierten Modellen schneidet ACG am schlechtesten ab, ACG-PM und CDPS sind vergleichbar, HCC zeigt die höchsten R^2 . Sowohl bezüglich Gesamtkosten als auch ambulanten Kosten schneidet die Kombination aus RxGroups und IP-HCC am besten ab. Die Autoren empfehlen deshalb diese Kombination für den morbiditätsorientierten Risikostrukturausgleich in Deutschland.

Diskussion

Im Rahmen der vorliegenden Übersichtsarbeit analysierten wir verschiedene Instrumente, die zur Vorhersage von Kosten verwendet werden.

Limitationen

Unsere Arbeit beruht nicht auf Originaldaten, sondern auf der publizierten Literatur. Die meisten Publikationen stammen aus den USA und

Tabelle 3

Studie von Reschke et al. [5] zu einem morbiditätsorientierten Risikostrukturausgleich in der gesetzlichen Krankenversicherung in Deutschland: Übersicht über die in den Modellen enthaltenen Variablen.

	Alter, Geschlecht	Medikamente	Ambulante Diagnosen	Stationäre Diagnosen	Rentenberechtigung
RSA	*				*
ACG	*		*	*	*
ACG-PM	*		*	*	*
CDPS	*		*	*	*
HCC	*		*	*	*
PCG + DCG	*	*		*	*
RxGroups + IP-HCC	*	*		*	*

RSA: Risikostrukturausgleich; ACG: «Adjusted clinical groups»; ACG-PM: «ACG-Predictive model»; CDPS: «Chronic illness and disability payment system»; HCC: «Hierarchical chronic conditions»; PCG+DCG: «Pharmacy cost groups + Diagnostic cost groups»; RxGroups+IP-HCC: «RxGroups + Inpatient-HCC».

Tabelle 4

Studie von Reschke et al. [5]: Erklärter Anteil der Kostenvarianz (R^2) für verschiedene Modelle, unter Berücksichtigung von Gesamtkosten, ärztlichen Leistungen und zeitgleicher bzw. prospektiver Erhebung.

	R^2 für Gesamtkosten, zeitgleich	R^2 für Gesamtkosten, prospektiv	R^2 für ärztliche Leistungen, zeitgleich	R^2 für ärztliche Leistungen, prospektiv
RSA	7%	6%	3%	3%
ACG	14%	9%	8%	6%
ACG-PM	21%	12%	13%	11%
CDPS	19%	12%	15%	14%
HCC	26%	15%	41%	38%
PCG + DCG	30%	15%	36%	33%
RxGroups + IP-HCC	48%	24%	44%	43%

RSA: Risikostrukturausgleich; ACG: «Adjusted clinical groups»; ACG-PM: «ACG-Predictive model»; CDPS: «Chronic illness and disability payment system»; HCC: «Hierarchical chronic conditions»; PCG+DCG: «Pharmacy cost groups + Diagnostic cost groups»; RxGroups+IP-HCC: «RxGroups + Inpatient-HCC».

anderen Ländern, deren Gesundheitswesen nur bedingt mit dem schweizerischen vergleichbar sind. Weiter stellt die publizierte Literatur nur einen Teil der durchgeführten Studien dar. Bei klinischen Studien ist bekannt, dass Studien mit signifikanten Resultaten eher publiziert werden als solche mit nichtsignifikanten Resultaten [6]. Dies könnte auch hier der Fall sein, was heissen würde, dass Studien mit guter Vorhersagekraft eher publiziert würden. Eine Literaturanalyse würde dann die Vorhersagekraft der Instrumente überschätzen. Eine weitere Einschränkung ist die Heterogenität der Studien. Es wurden verschiedenste Populationen untersucht, die nicht immer repräsentativ für die Gesamtbevölkerung waren. Weiter unterschieden sich die Studien in der Definition der Kosten, in den Ebenen der Validierung und weiteren Faktoren, was einen Vergleich der Resultate erschwert.

Wir haben die Vorhersagekraft von Modellen mittels R^2 beurteilt, weil diese Masszahl einfach zu interpretieren ist und bei der Beurteilung von Modellen oft verwendet wird. Leider wird das R^2 nicht nur von den Variablen im Modell, sondern auch von externen Faktoren wie beispielsweise der Studienart (Modellschätzung vs. Validierung) und der Art der Validierung (zeitgleich vs. prospektiv) beeinflusst. Es kommt hinzu, dass Personen mit sehr hohen Kosten zu einer Senkung von R^2 führen, da extreme Kosten mit den meisten Modellen unterschätzt werden. Würde ein Modell nicht wie bei der Verwendung der R^2 auf individueller Ebene beurteilt, sondern auf der Gruppenebene (die vorhergesagten Kosten für die ganze Patientengruppe werden verglichen mit den wirklich entstandenen Kosten), hätten grosse Abweichungen weniger Einfluss und das Modell würde wahrscheinlich besser abschneiden.

Erwünschte Eigenschaften von Instrumenten zur Beurteilung der Patientenstruktur

Zusätzlich zu einer möglichst hohen Vorhersagekraft sollten für den Risikoausgleich, für die Berechnung der Capitation oder für die Beurtei-

lung der Wirtschaftlichkeit von Leistungserbringern verwendete Instrumente einige weitere Eigenschaften aufweisen, welche bei der Auswahl eines Instrumentes berücksichtigt werden sollten. Der Aufwand bei der Datenerhebung sollte vertretbar sein. Die Erhebung soziodemographischer Variablen beispielsweise ist mit geringem Aufwand verbunden, während der Zeitaufwand bei auf der ICD-Codierung beruhenden diagnosebasierten Instrumenten deutlich höher ist. Instrumente können Anreize zur Manipulation schaffen. Spielen Diagnosen eine grosse Rolle, kann dies dazu verleiten, mehr oder schwerwiegendere Diagnosen zu stellen. Die Population erscheint dann kränker, als sie in Wirklichkeit ist. Instrumente können auch Anreize zu verstärkter Inanspruchnahme setzen, z.B. wenn Spitalaufenthalte wichtig sind. Schliesslich belohnen Instrumente, welche frühere Kosten beinhalten, eine ineffiziente und teure Versorgung. Die zu erhebenden Daten sollten möglichst objektiv sein. Bleibt zu viel Ermessensspielraum, besteht immer die Möglichkeit der Datenmanipulation. Die Datenerhebung sollte schliesslich möglichst reproduzierbar sein: verschiedene Personen sollten zur gleichen Beurteilung kommen. Um Manipulationen festzustellen und zu verhindern, sollte die Datenerhebung überprüfbar sein, wobei auch der Aufwand hierzu in einem vertretbaren Rahmen bleiben sollte.

Tabelle 5 zeigt eine Einschätzung dieser Eigenschaften für Instrumente, die sich auf verschiedene Daten abstützen.

Die soziodemographischen Instrumente zeichnen sich durch hohe Datenqualität und geringen Aufwand bei der Erhebung aus und setzen keine unerwünschten Anreize. Ihre Vorhersagekraft ist allerdings tief. Kostenbasierte Instrumente haben eine hohe Vorhersagekraft bei mittlerem Erhebungsaufwand, setzen aber unerwünschte Anreize. Medikamentenbasierte Instrumente zeichnen sich ebenfalls durch eine relativ hohe Vorhersagekraft aus. Der Aufwand bei der Erhebung ist mittel, die Datenqualität relativ gut, allerdings können sie dazu verleiten, die vom Modell erfass-

Tabelle 5

Zusammenfassende Einschätzung der Merkmale verschiedener Daten, die in Morbiditätsindikatoren verwendet werden (adaptiert von Cumming et al [7]).

	Soziodemographische Daten	Kosten	Medikamente	Diagnosen	Selbsteinschätzung des Gesundheitszustands
Datenqualität	Hoch	Mittel	Hoch	Mittel	Hoch
Vorhersagekraft	Tief	Hoch	Hoch	Hoch	Tief
Aufwand	Tief	Mittel	Mittel	Hoch	Tief
Anreize Inanspruchnahme	Tief/Kein	Hoch	Hoch für Medikamente	Tief/Hoch für Hospitalisationen	Tief
Anreize Codierung	Tief/Kein	Tief	Tief	Hoch	Mittel

ten Medikamente vermehrt einzusetzen. *Diagnosebasierte Instrumente* haben wie die medikamentenbasierten eine relativ hohe Vorhersagekraft bei deutlich höherem Erhebungsaufwand (ICD-Codierung). Sie können ebenfalls unerwünschte Anreize setzen, beispielsweise wenn das Modell nur Diagnosen aus Hospitalisationen enthält oder schwerwiegendere Diagnosen stark gewichtet werden. Zu bemerken ist, dass in der Schweiz ICD-Diagnosen bisher nur im stationären Bereich erhoben wurden. Die Datenqualität ist mittel und dürfte im ambulanten Bereich schlechter sein als im stationären. Modelle, die auf *subjektiven Variablen* beruhen (Einschätzung des Gesundheitszustands durch den Patienten), zeichnen sich durch einen relativ geringen Aufwand aus und setzen wenig Anreize auf Ebene der Inanspruchnahme. Allerdings ist auch ihre Vorhersagekraft gering.

In den letzten Jahren wurden zunehmend *Kombinationen aus medikamenten- und diagnosebasierten Instrumenten* untersucht. Die Kombination dieser zwei Bereiche bietet einige Vorteile. Beispielsweise kann der ambulante Bereich über Medikamente abgedeckt werden, der stationäre über Diagnosen, was den Erhebungsaufwand in der ambulanten Versorgung deutlich verringert. Diese Modelle haben eine relativ hohe Vorhersagekraft.

Schlussfolgerungen

Es existieren verschiedene Instrumente zur Beurteilung der Patientenstruktur. Vergleiche der Vorhersagekraft von verschiedenen Modellen, welche in unterschiedlichen Populationen evaluiert wurden, sind aber nur bedingt möglich, da andere Faktoren zusätzlich zu den Modellen einen Einfluss auf die R^2 haben. Studien, die mehrere Instrumente in der gleichen Population miteinander verglichen haben, sind hier hilfreicher.

Die resultierenden R^2 der verschiedenen Modelle sind insbesondere bei realitätsnaher, prospektiver Erhebung bescheiden. Ein grosser Teil der Kosten (insbesondere der Gesamtkosten) kann auch bei Berücksichtigung von Diagnosen oder Medikamenten *nicht akkurat vorhergesagt werden*. Dennoch sind Instrumente, die den Gesundheitszustand über Diagnosen oder Medikamente abbilden, den soziodemographischen Instrumenten deutlich überlegen. Welches Modell zur Dokumentation der Morbidität der versorgten Patientenpopulation am meisten Sinn macht, hängt nicht nur von dessen Vorhersagekraft ab, sondern auch von weiteren Faktoren wie dem Aufwand bei der Erhebung, Manipulierbarkeit, Überprüfbarkeit, Setzen von Anreizen usw. Bei der Auswahl eines Modelles sollten diese Aspekte ebenfalls berücksichtigt werden.

Der Aufwand ist besonders auch in der ambulanten Versorgung ein wichtiger Aspekt. Bei einigen Modellen mit guter Vorhersagekraft (insbesondere den diagnosebasierten) ist der hohe Aufwand ein Nachteil. Ein denkbarer Ansatz mit reduziertem Aufwand wäre, sich auf bestimmte, mit hohen Kosten verbundene Diagnosen zu beschränken und Patienten mit diesen Diagnosen zu identifizieren. In einer US-amerikanischen Studie [8] waren in den Gruppen mit den höchsten Kosten Personen mit Diabetes mellitus, Herzinsuffizienz, Asthma bronchiale/COPD oder einer Depression im Vergleich zur Gesamtpopulation deutlich übervertreten. Es kommen sicher noch weitere Diagnosen wie Malignome oder die HIV-Infektion für diesen Ansatz in Frage; diese wurden in der Studie aber nicht untersucht.

Zusammenfassend lässt sich sagen, dass eine grosse Zahl von Instrumenten existiert, welche die Patientenstruktur abbilden. Sie sind den rein soziodemographischen Ansätzen klar überlegen, aber auch mit diesen Instrumenten bleibt ein grosser Teil der Kosten, und insbesondere die Gesamtkosten, nicht vorhersagbar.

Danksagung

Wir danken Professor Doug Altman, Centre for Statistics in Medicine, Universität Oxford, für angelegte Diskussionen und hilfreiche Kommentare.

Literatur

- 1 Aus den Verhandlungen des FMH-Zentralvorstands. Schweiz Ärztezeitung. 2005;86(14):811-12.
- 2 Altman DG, Royston P. What do we mean by validating a prognostic model? *Statistics in Medicine*. 2000;19(4):453-73.
- 3 Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-24.
- 4 Spycher S. Der TMI – ein Indikator, der zählt. *Schweiz Ärztezeitung*. 2006;87(23):1038-45.
- 5 Reschke P, Sehlen S, Schiffhorst G, Schröder WF, Lauterbach KW, Wasem J. Klassifikationsmodelle für Versicherte im Risikostrukturausgleich. Untersuchung zur Auswahl geeigneter Gruppenbildungen, Gewichtungsfaktoren und Klassifikationsmerkmale für einen direkt morbiditätsorientierten Risikostrukturausgleich in der gesetzlichen Krankenversicherung. Endbericht; 2004.
- 6 Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. *BMJ*. 1998;316:61-6.
- 7 Cumming RB, Knutson B, Cameron BA, Derrick B. A comparative analysis of claims-based methods of health risk assessment for commercial populations. 2002.
- 8 Ash AS, Zhao Y, Ellis RP, Schlein Kramer M. Finding future high-cost cases: comparing prior cost versus diagnosis-based methods. *Health Services Research*. 2001;36(6):194-206.