

## Kapitel 11

# **Weiterentwicklung von Krankenhausbetriebsvergleichen: Wie wichtig ist Risiko-Adjustierung für den Krankenhausvergleich?**

SEBASTIAN SCHNEEWEISS, OLIVER SANGHA

Bei den gegenwärtigen Bemühungen, das Gesundheitswesen bei gleichbleibendem Qualitätsniveau effizienter zu machen, rückt das Krankenhaus zunehmend in den Mittelpunkt. Dabei üben der stetige Fortschritt in der Medizin mit neuen, teuren Technologien und Behandlungsformen sowie eine stetig alternde Bevölkerung einen zunehmenden ökonomischen Druck auf die Krankenkassen aus. Der Verpflichtung zur Beitragsstabilität steht zusätzlich eine "Deckungslücke" auf der Einnahmenseite durch die gegenwärtige Arbeitsmarktsituation gegenüber. Diese Beobachtungen und die Tatsache, dass Deutschland im internationalen Vergleich bei ähnlicher Morbiditätsstruktur hinsichtlich der Bettendichte, Liegezeiten und Auslastung von Krankenhäusern seit Jahren einen Spitzenplatz einnimmt, legitimiert das Bestreben der Krankenkassen im Bereich der stationären Versorgung Wirtschaftlichkeitsreserven zu identifizieren. In diesem Zusammenhang hat der Krankenhausbetriebsvergleich in jüngster Zeit eine zentrale Rolle eingenommen.

Vergleiche werden unter der betriebswirtschaftlichen Annahme durchgeführt, dass ein Krankenhaus mit geringeren Ausgaben pro Behandlungsergebnis wirtschaftlicher arbeitet als eines mit höheren Ausgaben. Dieser Annahme stellen Kritiker entgegen, dass sich die Leistungsempfänger hinsichtlich Morbidität, Schweregrad und weiteren Faktoren, die das Behandlungsergebnis beeinflussen, in jedem Krankenhaus unterscheiden, die Leistungen entsprechend sehr variabel erbracht werden und sich somit einem direkten Vergleich entziehen.

Herausforderungen bestehen jedoch heute nicht nur hinsichtlich der Vergleichbarkeit von Patienten, sondern auch auf der Seite der Berechnung der Kosten pro behandeltem Patient und der erbrachten Leistungen. Die gegenwärtige Krankenhausfinanzierung, sowie eine erheblich Variabilität hinsichtlich der klinikinternen Buchhaltung, machten auch in diesem Bereich einen direkten Vergleich extrem schwierig. Leistungen im Sinne von Behandlungsergebnissen werden im Vergleich zu anderen Industriezweigen äußerst rudimentär, d.h. fast ausschließlich durch die Angabe von Mortalitäts-, Morbiditäts- und Infektionsraten, erfasst. Betriebswirtschaftliche Kenngrößen, z. B. Fallkosten oder Belegungsraten, sind nicht aussagefähig, solange die zugrundeliegende Patientenpopulation nicht explizit

beschrieben werden kann. Die bloße Angabe von Alter, Geschlecht, ICD-Hauptdiagnose bzw. Prozeduren ist dabei oftmals nicht ausreichend.

Unter dem Druck, einen validen Krankenhaus- bzw. Patientenvergleich dennoch möglich zu machen, sind verschiedene Verfahren und Methoden entstanden, die in den vorangegangenen Beiträgen beschrieben wurden.

Dieses Kapitel klammert bewusst die Erfassung von Kosten, Leistungen und Ergebnissen aus und beschäftigt sich mit dem Stellenwert der "Risiko-Adjustierung" in künftigen Krankenhausvergleichen.

## 11.1 Vergleichbarkeit

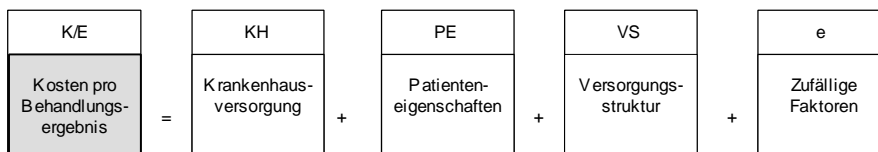
Grundsätzlich ist die Problematik der Nicht-Vergleichbarkeit von Patienten bzw. Patientenpopulationen ein bekanntes Phänomen in der Medizin. Warum reagieren manche Patienten besser auf ein Medikament als andere? Dieses hat dazu geführt, dass Medizin als Individualmedizin unter dem Einsatz sehr variabler kognitiver und materieller Ressourcen praktiziert wurde. In Zeiten ohne Kostendruck wurde eine solche Einstellung nicht herausgefordert, obwohl ein Nachweis der jeweiligen Behandlungsergebnisse fehlte.

Im Bereich der klinischen Prüfung von Medikamenten wurde das Problem methodisch gelöst. Die Vergleichbarkeit von zwei Patientengruppen wird durch strikte Ausschlusskriterien von Patienten und durch zufällige Aufteilung der Patienten in die jeweiligen Vergleichsgruppen erreicht. Im Mittel sind damit alle Patienteneigenschaften wie Geschlecht, Alter, Schweregrad der Grunderkrankung und Begleiterkrankungen in beiden Gruppen gleich. Derartige randomisierte Experimente sind jedoch im Krankenhausvergleich nicht möglich. Ersatzweise wird daher versucht, Patienteneigenschaften wie Diagnosen zu erfassen, um dann Kosten und Leistungen nur von Patientengruppen zu vergleichen, die die gleiche Diagnose haben. Je vergleichbarer Patientengruppen sein sollen, desto mehr Untergruppen oder ‚Schichten‘ mit gleichen Eigenschaften wird man bilden, z. B. Hauptdiagnose plus Alter plus Geschlecht plus Begleiterkrankungen. Eine Risiko-Adjustierung vergleicht nur Patienten der unterschiedlichen Krankenhäuser innerhalb jeder Schicht mit ähnlichen Patienteneigenschaften. Die Ergebnisse aller Schichten werden in einem gewichteten Mittel zusammengefasst.

Risiko-Adjustierung stößt an seine natürlichen Grenzen, wenn Patienteneigenschaften nicht erfasst werden oder nicht erfassbar sind. Werden solche Faktoren nicht berücksichtigt, die gleichzeitig mit Kosten oder Behandlungsergebnissen korreliert sind, so ist die Vergleichbarkeit entsprechend eingeschränkt. Zusätzlich gibt es zufällige Faktoren, die in Form von 95 % Konfidenzintervallen dargestellt werden. Sie werden in den entsprechenden Verhandlungen den beteiligten Gruppen verdeutlichen, auf welcher begrenzten empirischen Basis weitreichende Entscheidungen getroffen werden sollen.

## 11.2 Welche Faktoren beeinflussen Kosten und Behandlungsergebnisse eines Krankenhausaufenthaltes?

Die Vergleichbarkeit von Krankenhäusern wird nur durch solche Faktoren eingeschränkt, die die Kosten und Ergebnisse eines Hauses beeinflussen. Die beiden wichtigsten Faktoren neben der medizinischen Versorgung im Krankenhaus sind (1) patientenbezogenen Faktoren, wie allgemeiner Gesundheitszustand bei Aufnahme, Art und Schweregrad der Hauptdiagnose, Anzahl und Schweregrad der Begleiterkrankungen und (2) Gegebenheiten der Versorgungsstruktur wie Versorgungsstufe und Spezialabteilungen [1]. Dazu kommen die Faktoren, die als zufällige Ereignisse vorhanden sind. Wir stellen diese Konzepte in einem verallgemeinerten Versorgungsmodell dar (siehe Abbildung 11.1). Dieses Modell verdeutlicht die Wirkungen der einzelnen Komponenten auf Kosten und Ergebnisse in einer funktionalen Form. Es erleichtert, das Konzept einer Risiko-Adjustierung zu verstehen und einheitlich zu kommunizieren.



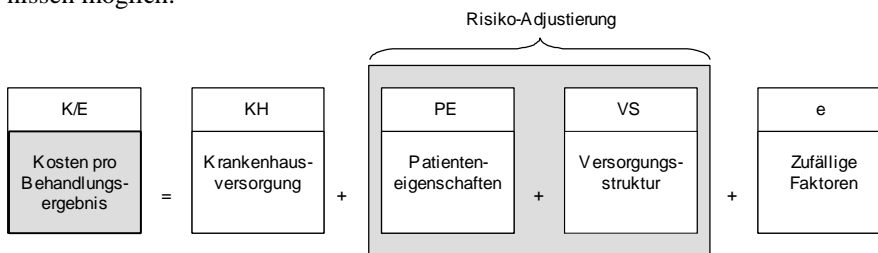
**Abb. 11.1** Das verallgemeinerte Versorgungsmodell. Kosten pro Leistung werden durch eine Funktion von drei Hauptfaktoren und zufälligen Faktoren erklärt

Das Modell verdeutlicht, dass Patienteneigenschaften, auch Patientenmix genannt, und Versorgungsstruktur zum Vergleich von Krankenhäusern ähnlich wichtige Stellungen einnehmen wie die im Krankenhaus erbrachte medizinische Versorgung. Nachdem Krankenkassen hauptsächlich an guten Behandlungsergebnissen interessiert sind und es Aufgabe der einzelnen Krankenhäuser ist, die jeweilige medizinische Versorgung zu optimieren, wird die erbrachte Krankenhausversorgung (KH) nicht explizit gemessen, sondern durch einen binären Indikator ersetzt, der das zu untersuchende Krankenhaus markiert.

## 11.3 Risiko-Adjustierung

Patienten mit gleicher Hauptdiagnose unterscheiden sich hinsichtlich Alter, Geschlecht, Schweregrad, Begleiterkrankungen, Funktionsfähigkeit und weiteren Faktoren, die den Behandlungserfolg beeinflussen können. Mit anderen Worten, diese Faktoren beeinflussen das Risiko, schlechtere Ergebnisse zu höheren Kosten bei gleich guter medizinischer Versorgung zu produzieren, als der Durchschnitt aller Patienten. Ziel einer Risiko-Adjustierung ist es, nur Patienten mit ähnlichem Risikoprofil bzw. nur Abteilungen/Krankenhäuser mit ähnlichen Patientenzusammensetzungen („Patientenmix“) zu vergleichen.

Die Risiko-Adjustierung in einem Betriebsvergleich ist also nichts anderes als der Versuch, die Effekte dieser Faktoren auf die gesundheitlichen Ergebnisse mit entsprechenden statistischen Methoden zu eliminieren oder adjustieren. Sind Patienteneigenschaften und Versorgungsstruktur hinreichend adjustiert, so ist eine direkte Gleichsetzung der Krankenhäuser mit den Kosten und Behandlungsergebnissen möglich.



**Abb. 11.2** Betriebsvergleich im verallgemeinerten Versorgungsmodell: Risiko-Adjustierung eliminiert den Effekt von Patienteneigenschaften und Versorgungsstruktur auf Kosten und Behandlungsergebnisse. Sie erlaubt damit einen direkten Vergleich von Kosten zwischen Krankenhäusern.

Unsere Darstellung der Risiko-Adjustierung in Abbildung 11.2 lässt sich direkt in ein einfaches lineares Regressionsmodell mit  $i$  Patienteneigenschaften (PE) und  $j$  Charakteristika der Versorgungsstruktur (VS) übertragen und entsprechend analysieren:

$$K/E = a + b_1 \text{Krankenhaus} + c_1 PE_1 + \dots + c_i PE_i + d_1 VS_1 + \dots + d_j VS_j + e$$

Der Regressionskoeffizient  $b_1$  ist damit ein Maß für die kostengünstiger (oder teurer) erbrachten Behandlungsergebnisse in dem untersuchten Krankenhaus im Vergleich zu allen anderen Häusern, adjustiert für Patienteneigenschaften und Versorgungsstruktur.

Sorgfältige Analysen erfordern jedoch kompliziertere Modelle, die Interaktionen zwischen den einzelnen Faktoren erlauben. Stabilere Schätzungen erhält man, wenn zufällige Schwankungen auf Patientenebene und Krankenhausebene in hierarchischen Regressionsanalysen getrennt modelliert werden [2][3].

Tabelle 11.1 zeigt Dimensionen einer weitreichenden Risiko-Adjustierung patientenbezogener Eigenschaften. Alle Faktoren sollten zum Zeitpunkt der Aufnahme erfasst werden und den Status zu diesem Zeitpunkt oder kurz davor (bei chronischen Erkrankungen länger) abbilden, um nicht durch die Behandlung oder deren Komplikationen beeinflusst zu werden. In der Praxis werden meist weniger Dimensionen herangezogen.

Wie ordnen sich bestehende Konzepte in dieses allgemeine Konzept ein?

Seit 1996 bestehen von Seiten der Sozialleistungsträger 21 Strukturmerkmale, die Krankenhäuser in Gruppen zusammenfassen sollen, die ähnliche Aufgaben erfüllen. Einteilungsmerkmale waren dabei u.a. Anzahl Fachabteilungen, Akademisches Lehrkrankenhaus, Versorgungsstufe etc.

**Tabelle 11.1** Dimensionen einer patientenbezogenen Risiko-Adjustierung im Krankenhausvergleich**Dimensionen der Risiko-Adjustierung\***

Haupterkrankung und deren Schweregrad  
 Akuter klinischer Allgemeinzustand  
 Begleiterkrankungen und deren Schweregrad  
 Vorausgehende Inanspruchnahme von Gesundheitsleistungen  
 Demographische Faktoren  
 Subjektiver Gesundheitszustand  
 Lebensstil Faktoren  
 Kognitive, seelische und soziale Funktionsfähigkeit

Vergleicht man Krankenhäuser innerhalb dieser Strukturgruppen, so wird in unserem Modell die Versorgungsstruktur zu einem gewissen Grade adjustiert, nicht jedoch die Patienteneigenschaften, da der tatsächliche Patientenmix in den zu vergleichenden Häusern unbekannt bleibt. Es kann also vorkommen, dass eine Klinik der Primärversorgung in einem Ballungsraum mit einer Uni-Klinik in der Nachbarschaft mit einer Klinik der Primärversorgung im ländlichen Raum verglichen wird. Starke Unterschiede im Patientenmix sind hier zu erwarten.

Das *WIDO-Klip*-Verfahren ist der erste Ansatz in Deutschland, Patienteninformationen zur Risiko-Adjustierung heranzuziehen. Basierend auf ICD-kodierten Diagnosen der LKA-Statistik, werden mittels Clusteranalyse jene Abteilungen aus einem Pool von Fachabteilungen identifiziert, die sich bezüglich der Verteilung der ICD-Diagnosen möglichst ähnlich sind. Kosten, Verweildauer oder andere Zielgrößen werden dann direkt zwischen den Abteilungen verglichen, die sich zusätzlich in Alters- und Geschlechtsverteilung ähnlich sind. In unserem Modell entspricht dies einer verbesserten Adjustierung des Patientenmixes, jedoch ohne Komorbidität und Schweregrad. Obwohl Diagnoseinformationen in die Bildung von Vergleichsabteilungen eingehen, ist das Verfahren nicht einer Adjustierung von patientenbezogenen Eigenschaften gleichzusetzen. Es handelt sich vielmehr um eine aggregierte Gegenüberstellung von alters-, geschlechts- und diagnoseähnlichen Abteilungen und aggregierten Kosten (Abbildung 11.3).

**Abb. 11.3** Derzeitige Reduktion auf den Vergleich von aggregierten Kosten mit einer aggregierten Risikoadjustierung

Hierzu zwei Beispiele:

1. In einem Qualitätsmanagementprojekt werden abteilungsbezogene Komplikationsraten als Maß der Ergebnisqualität verglichen.
2. In einem Betriebsvergleich werden Verweildauern als Zielgröße verglichen.

Im ersten Beispiel kann eine Abteilung mit besonders geringer Komplikationsrate identifiziert werden. Es kann jedoch nicht festgestellt werden, ob diese Komplikationen bei Personen mit schwerwiegenderen Diagnosen und mehreren Nebendiagnosen auftreten, oder relativ gesunden Patienten, die auf Grund einer schlechten medizinischen Versorgung die Komplikation erlitten haben. Diese Effekte bleiben bei dem Vergleich aggregierter Größen wegen der Entkoppelung von Prädiktoren (z. B. Diagnose) und Zielgröße unerkannt und werden als ökologischer Bias bezeichnet. Bei einem Betriebsvergleich (zweites Beispiel) hingegen ist es für die Prüfer nicht von Interesse, wie einzelne diagnoseähnliche Abteilungen zu kürzeren Liegezeiten kommen. Damit ist dieses Verfahren hier einsetzbar, jedoch nicht zum internen Qualitätsmanagement.

### **Grenzen aggregierter Verfahren**

1. Bisherige aggregierte Verfahren wie das WIdO-Klip basieren auf der Bildung von Abteilungsgruppen mit möglichst ähnlichen (homogenen) Diagnoseverteilungen. Diese Suche nach ähnlichen Abteilungen wird jedoch an Grenzen stoßen, wenn weitere Variablen wie z. B. der Schweregrad der jeweiligen Erkrankungen zur Gruppenbildung herangezogen werden. Die Anzahl der zur Überlappung zu bringenden Kategorien steigt exponentiell, und es wird sehr bald unmöglich, noch hinreichend gleiche Abteilungen zu identifizieren. Damit ist das Verfahren vermutlich bereits jetzt nahe an seiner Grenze angelangt.
2. Ein Vergleich einzelner Patientengruppen zwischen unterschiedlichen, sich nicht ähnelnden Häusern wird explizit ausgeschlossen. Dabei kann es von großer Wichtigkeit sein, z. B. die Behandlung unkomplizierter Pneumonien bzgl. der entstandenen Kosten/Liegezeiten in Lehrkrankenhäusern mit denen in Häusern der Primärversorgung zu vergleichen.
3. Ferner kann das Verfahren nur bedingt zum internen Qualitätsmanagement eingesetzt werden.

Aus diesen Gründen ist das Klip-Verfahren zwar mit Einschränkungen zum Betriebsvergleich einsetzbar, eine Weiterentwicklung einer Risiko-Adjustierung, die gleichzeitig zum Qualitätsmanagement eingesetzt werden kann, ist jedoch nur mit patientenbezogenen Verfahren erreichbar.

### **Einzelfallbewertung vs. statistische Verfahren**

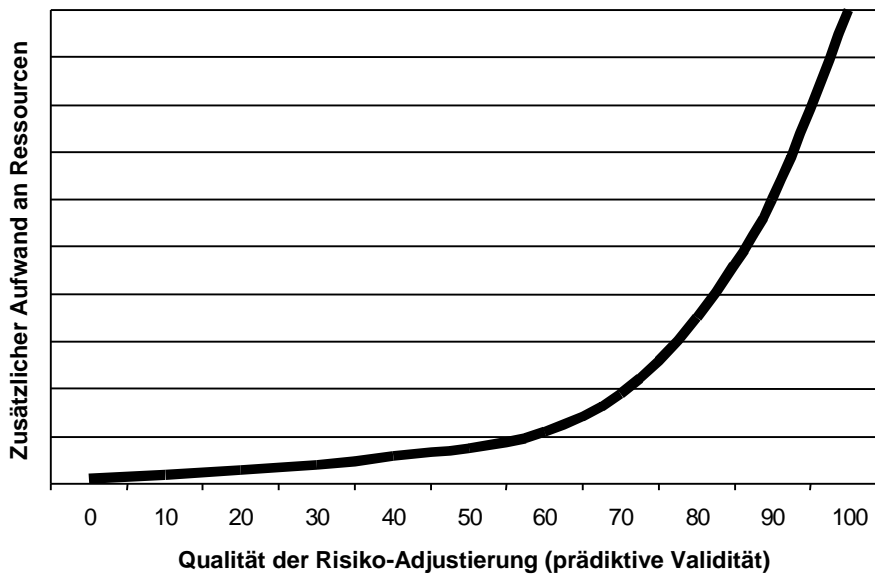
Es sei hier bemerkt, dass Risiko-Adjustierung eine statistische Methode ist, die formale Einschränkungen, Standardisierungen und Gewichtungen erfordert. Bei isolierter Betrachtung einzelner Patienten wird ein erfahrener Kliniker vermutlich eine differenziertere und bessere Einschätzung der zu erwartenden Behandlungs-

ergebnisse erstellen können, da er eine Vielzahl von z.T. schlecht erfass- und dokumentierbaren Faktoren in seinen Überlegungen berücksichtigt. Damit ist es jedoch auch unmöglich, diese Einschätzung in einer zuverlässigen und reproduzierbaren Weise abzubilden [4]. Die Stärke einer standardisierten Risiko-Adjustierung ist, dass sie möglichst objektive Vergleiche zwischen unterschiedlichen Ärzten, Häusern und Regionen zulässt und sich nicht auf subjektive Kriterien, Einzelfallberichten und Anekdoten beschränkt.

#### **11.4 Aufwand und Nutzen einer Risiko-Adjustierung**

Risiko-Adjustierung kann mit einem breiten Spektrum von Methoden durchgeführt werden, das vom einfachen Auflisten von Diagnosen bis zu komplexen Gewichtungsverfahren reicht. Die Qualität einer Risiko-Adjustierung wird gemessen an ihrer Fähigkeit, die Variabilität der gesundheitlichen Ergebnisse oder Kosten bei gleicher medizinischer Versorgung zu erklären bzw. diese für jeden Patienten vorherzusagen. Die optimale, praktisch nicht erreichbare prädiktive Validität ist dabei 100 %, d.h. Ergebnisse und Kosten jedes Patienten können aufgrund seiner Einweisungscharakteristika und bei bekannter medizinischer Versorgung bereits vollkommen vorhergesagt werden.

Die Qualität oder Vollständigkeit einer Risiko-Adjustierung hat einen Preis. Eine Vielzahl von Parametern müssen erfasst, bewertet und dokumentiert werden. Nachdem jedoch relativ grobe Parameter wie Alter, Geschlecht und ICD-Diagnosen in Kombination bereits einen nicht zu unterschätzenden Teil einer Risiko-Adjustierung darstellen, kann mit relativ geringem Mehraufwand (z. B. Anzahl der Diagnosefelder von 1 auf 5 oder sogar 25 erweitern, wie es Romano et al. empfehlen [5]) bereits sichtliche Verbesserungen erreicht werden. Um eine weitere Steigerung der prädiktiven Validität zu erreichen steigt der Aufwand jedoch zunehmend (siehe Abbildung 11.4). Dies erklärt sich dadurch, dass jede Erkrankung sehr spezifische Prädiktoren für den Behandlungserfolg hat. Da Krankenhäuser ein relativ heterogenes Spektrum von Krankheiten behandeln, müssen mehr und mehr Einzelparameter zu einer weiteren Optimierung der Risiko-Adjustierung herangezogen werden, die zum Teil mit hohem Aufwand erfasst und dokumentiert werden müssen.



**Abb. 11.4** Zunehmender Aufwand für eine Optimierung der Risiko-Adjustierung

Für unterschiedliche Anwendungen können verschiedentliche Grade der Risiko-Adjustierung ausreichen. Um eine informierte Entscheidung treffen zu können, muss daher zunächst die prädiktive Wertigkeit unterschiedlicher Verfahren bekannt sein. Danach wird der jeweilige Aufwand in Relation zum Gewinn entscheiden, welches Verfahren zum Einsatz kommt.

## 11.5 Wertigkeit unterschiedlicher Instrumente zur Risiko-Adjustierung

Dieser Abschnitt greift aus den ca. 20 existierenden Risiko-Adjustierungs-Verfahren mit noch mehr Varianten nur wenige, häufig eingesetzte Verfahren exemplarisch heraus und diskutiert deren strukturelle sowie metrische Eigenschaften. Strukturelle Eigenschaften von Risiko-Adjustierungs-Verfahren geben Auskunft über den Ressourcenaufwand und entscheiden letztlich über die Praktikabilität der einzelnen Verfahren, insbesondere, wenn sich die metrischen Eigenschaften ähnlich sind.

Wir teilen die Verfahren in solche auf, die auf administrative Daten zurückgreifen, und solche die überwiegend mit klinischen Daten arbeiten. Administrative Daten sind solche, die derzeit bereits in den Leistungsstatistiken aufgeführt werden (Diagnose(n), Alter, Geschlecht, Liegezeit, Entlassungsart und Prozeduren) oder solche, die ohne großen Mehraufwand integriert werden können, wie weitere Diagnosefelder und Diagnosetyp (Haupt, Neben und Komplikation). Klinische

Daten umfassen dagegen konkrete Ergebnisse und Bewertungen klinischer Untersuchungen. Diese Trennung entspricht den erheblichen Unterschieden in Praktikabilität und Installation der Systeme.

### **Administrative Daten**

*Diagnosis Related Groups*: DRGs [6][7] und ‚All Patient Refined‘ DRGs (APR-DRGs) [8][9] sind die am weitesten verbreiteten Systeme basierend auf administrativen Daten. DRGs wurden spezifisch für prospektive Vergütungssysteme (PVS) entwickelt. Primäres Ziel war es, Patienten in Gruppen möglichst gleichen Ressourcenverbrauchs einzuteilen. Damit ist Risiko-Adjustierung integrierter Bestandteil von DRG-Systemen. DRGs klassifizieren Patienten an Hand von 5 ICD-Entlassungsdiagnosen, 3 Prozeduren, Alter, Geschlecht und Entlassungsstatus. Die Daten werden den amerikanischen Leistungsstatistiken entnommen (UHDDS), die durch ausgebildete Kodierer von den Krankenakten nach Entlassung abstrahiert werden. Mit Computeralgorithmen („grouper“) werden die Patienten den einzelnen DRGs zugeteilt. Die weiterentwickelten APR-DRGs ordnen die Begleiterkrankungen den unterschiedlichen Schweregradstufen dabei in Abhängigkeit der Hauptdiagnose zu. Kritikpunkt der Risiko-Adjustierung ist die Einteilung des Schweregrades an Hand der zum Einsatz gekommenen Prozeduren und nicht an der Notwendigkeit der Prozeduren. Angesichts der erheblichen regionalen Variabilität in der Häufigkeit von Prozeduren [10], lässt dieses Vorgehen eine geringere Validität erwarten.

Weitere Systeme, die mit den gleichen Angaben arbeiten, sind *Patient Management Categories* (PMCs) [11][12] und *Disease Staging* (DS) [13]. Dieses Systeme kategorisieren Patienten ebenfalls in Krankheitsgruppen und entsprechende Schweregrade.

Der *Charlson Index* ist ein gewichteter Score von bis zu 17 ausgewählten Begleiterkrankungen. Er wurde ursprünglich für klinische Daten entwickelt, mittlerweile gibt es jedoch vier unterschiedliche Adaptationen für administrative Daten, die auf 3- oder 5-stelligen ICD-9 Codes basieren [14][15][16][17]. Durch seine einfache Struktur ist der Charlson Index entsprechend leicht anzuwenden und nicht mit Lizenzgebühren verbunden.

### **Klinische Daten**

Der *Computerized Severity Index* (CSI) [18][19] stellt ein umfangreiches System dar, das Patienten in 600 Diagnosen und Symptome einteilt und den Erkrankungsschweregrad an einer Vielzahl klinischer Parameter auf einer ordinalen Skala abbildet. Der CSI computeralgorithmus besitzt für jede Erkrankung eine Matrix aus Zeilen mit klinischen Parametern, z. B. EKG, Laborwerte, Röntgen, und Spalten mit den jeweiligen Schweregraden 1 bis 4. Die Felder der Matrizen sind mit den entsprechenden Kriterien für die Schweregrade ausgefüllt. Nachdem das System voll computerisiert ist, kann es leicht mit einem Krankenhausinformationssystem verbunden werden. Ein ähnlich umfangreiches System stellt das *MedisGroup System* da [20][21].

Systeme wie *APACHE III* (Acute Physiology, Age, Chronic Health Evaluation) [22], *MPM II* (Mortality Probability Model) [23], und *SAPS II* (Simplified Acute Physiology Score) [24] wurden für schwerst-krankte Patienten auf Intensivstationen entwickelt und stellen deren Schweregrad auf sehr differenzierten Skalen da. Sie unterscheiden sich im Erfassungsaufwand und damit in der Einsatzbreite. *APACHE* benötigt physiologische Parameter aus arteriellem Blut, was dessen Anwendung in der klinischen Routine außerhalb der Intensivstation erheblich einschränkt. Das *SAPS* kann als eine vereinfachte Form des *APACHE* gesehen werden [25].

Einschränkungen des Funktionsstatus können mit der *Pflegekategorisierung* nach der Bundespflegesatzverordnung oder mit dem differenzierten Instrumentarium nach Greenfield (Bestandteil des ‚Index of coexisting diseases‘, *ICED*) zusammengefasst werden. Zunehmend werden Instrumente zur Erfassung des patientenzentrierten Gesundheitszustandes in Risiko-Adjustierungsmodelle integriert. Aus der ständig wachsenden Zahl von Instrumenten stechen der *SF-36* [26] mit seiner vielfältigen Einsetzbarkeit und guten metrischen Eigenschaften und die *Dartmouth Function Charts* mit ihrer leicht verständlichen grafischen Darstellung heraus [27].

### Prädiktive Validität

Die Fähigkeit eines Instruments, die Behandlungsergebnisse oder Kosten vorherzusagen, wird üblicherweise mit der *c*-Statistic oder  $R^2$  bestimmt (Tabelle 11.3). Je näher diese Werte an 1 kommen, desto größer ist die prädiktive Validität.

**Tabelle 11.3** Bewertung der prädiktiven Validität eines Instruments zur Risiko-Adjustierung

Scalierung der Zielgröße	Maßzahl / Statistik	Bedeutung
<i>Dichotom:</i> z. B. <i>Mortalität, Wieder- einweisung, Verbesserung eines binären kli. Outco- mes wie Ödeme ja/nein</i>	<i>c</i> <i>AUC</i> = area under the ROC curve	<i>c</i> und <i>AUC</i> sind Diskriminati- onsstatistiken, die den Unter- schied zwischen vorhergesagtem binären outcomes und der tat- sächlichen Zielgröße quantifizie- ren. 0.5 bedeutet keine Vorher- sagekraft, die über den Zufall hinaus geht (man könnte also auch eine Münze werfen), 1 bedeutet eine 100 %ig korrekte Vorhersage. Numerisch gilt: $c =$ <i>AUC</i> .
<i>Kontinuierlich:</i> z. B. <i>Kosten, Verweildauer</i>	$R^2$	$R^2$ ist die erklärte Varianz in einem linearen Regressionsmo- dell. Reicht von 0 bis 1, einer 100 % Erklärung der Zielgröße.

Diese Zahlen lassen sich jedoch nur vergleichen, wenn sie aus der gleichen Studienpopulation stammen und die gleiche Zielgrößen prognostiziert wird. Leider gibt es kaum vergleichende Studien, die alle wichtigen Risiko-Adjustierungs-Verfahren umfassen und wichtige Zielgrößen untersuchen. So ist beispielsweise ein Risiko-Adjustierung nur mit Alter und Geschlecht ( $c=0.69$ ) 38 % besser als keine Risiko-Adjustierung ( $c=0.5$ ). Die Vorhersagekraft kann jedoch im Fall von akuten Herzinfarkten mit APR-DRGs ( $c=0.84$ ) um 68 % gesteigert werden (Tabelle 11.4) [28][29].

Der Anteil der durch die DRG-Risiko-Adjustierung erklärten Varianz der Verweildauer wegen Pneumonie ( $R^2=0.14$ ) verdoppelt sich annähernd im Vergleich zu dem einfachsten Modellen bestehend aus Alter und Geschlecht ( $R^2=0.08$ , Tabelle 11.5) [30].

**Tabelle 11.4** Vorhersagekraft von Krankenhaussterblichkeit (prädiktive Validität) von unterschiedlichen Risiko-Adjustierungssystemen

System	Validität ( $c$ -Statistik)*		
	Akuter Herzinfarkt	Pneumonie	Schlaganfall
MedisGroups	0.80	0.81	0.80
Charlson comorbidity	0.70	0.74	0.61
Disease Staging	0.79	0.75	0.60
PMC	0.82	0.79	0.73
APR-DRG	0.84	0.78	0.77
DRG	0.80	0.83	0.74
Alter und Geschlecht	0.69	0.67	0.60

\* ohne Risiko-Adjustierung  $c=0.5$ .

Die drei physiologisch orientierten Scores für schwerstkranke Patienten besitzen zum Zeitpunkt der Aufnahme auf eine Intensivstation eine hohe prädiktive Validität bezüglich Mortalität (APACHE:  $c=0.90$ , MPM:  $c=0.82$ , SAPS:  $c=0.86$ ). Obwohl primär für intensivpflichtige Patienten entwickelt, zeigt das SAPS hohe prädiktive Validität außerhalb von Intensivstationen ( $c=0.85$ ) [31], bei Tumor-Patienten ( $c=0.79$ ) [32], coronary care units ( $c=0.91$ ) [33][34] und stroke units.

**Tabelle 11.5** Anteil der erklärten Varianz der Krankenhausverweildauer (prädiktive Validität) von unterschiedlichen Risiko-Adjustierungssystemen in Patienten mit Pneumonie.

System	Validität ( $R^2$ ) Pneumonie
MedisGroups	0.091
Charlson comorbidity	0.085
Disease Staging	0.091
PMC	0.104
APR-DRG	0.126
DRG	0.142
Alter und Geschlecht	0.079

Zu jedem Verfahren gibt es eine Vielzahl von Untersuchungen mit einer breiten Streuung der entsprechenden Werte in Abhängigkeit von Patientenzusammensetzung, Krankenhauscharakteristika und Zielgröße. Eine generelle Beobachtung ist jedoch, dass DRG-Systeme Kosten und Ressourcenaufwand besser abschätzen, jedoch klinisch basierte Verfahren bei einem weiten Spektrum von Krankheit, Mortalität oder Morbidität (Wiedereinweisungen) etwas besser präzisieren können. Abbildung 13.5 zeigt eine grobe Rangfolge der besprochenen Verfahren bezüglich Validität und dem damit eng zusammenhängenden Ressourcenaufwand. Diese Rangfolge basiert lediglich auf amerikanischen Erfahrungen mit dem dort gängigen Kodierungsverhalten. Entsprechende Untersuchungen im deutschen Krankenhausbereich sind dringend notwendig, bevor weitreichende gesetzgeberische Maßnahmen verankert werden.

## 11.6 Positive und unerwünschte Folgen einer weitergehenden Risiko-Adjustierung

Eine Weiterentwicklung der bestehenden Verfahren des Krankenhausvergleichs auf patientenbezogene Erfassung von Kosten und Behandlungsergebnissen mit entsprechender Risiko-Adjustierung wird erhebliche Folgen auf das deutsche Krankenhaussystem haben.

1. Sie wird zu einem *faireren und besser quantifizierbaren* Betriebsvergleich führen und gleichzeitig zum internen Qualitätsmanagement beitragen [35].
2. Vor dem Hintergrund einer leistungsorientierten Vergütung werden Krankenhäuser Maßnahmen ergreifen, gute Leistungen auf wirtschaftliche Art zu erbringen. Dieses wird zu einer *zunehmenden Spezialisierung* der Krankenhäuser führen, um durch einen erhöhten Umsatz die Kosten pro Fall zu reduzieren [36][37].

3. Es ist vielfach gezeigt worden, dass diese Art von Spezialisierung mit Umsatzsteigerung ausgewählter Diagnosen und Prozeduren zu einer **besseren Behandlungsqualität** und besseren risiko-adjustierten Ergebnissen führt.
4. Sobald Krankenhäuser ihre Patientenmix besser und ohne zeitliche Verzögerung quantifizieren, ihren Kosten aus vergangenen Jahren gegenüberstellen und mit anderen Häusern vergleichen können, ist eine **vorausschauendere Budgetierung** möglich.
5. Die Einführung bzw. Erweiterung auch nur einfacher Risiko-Adjustierungs-Verfahren setzt nicht unerhebliche **Änderungen in Denkweise und Management** von Kliniken voraus. Die praktische Umsetzung kann leicht scheitern, wenn nicht genügend gezielte Qualifikation und Willen zu dieser Änderung angereichert werden kann. Hier ist eine voraussichtige Planung im Bereich der innerbetrieblichen Weiterbildung und bei der Zusammensetzung des Managements wichtig.
6. Mit der Einführung von personenbezogenen Risiko-Adjustierungs-Modellen, die durch die Versicherungen vorgegeben werden, kann jedes Krankenhaus für sich bestimmen, welche Patientengruppen in der spezifischen Situation des Hauses den größten Gewinn darstellen. In einzelnen Häusern können möglicherweise durch eine starke Gewichtung von Begleiterkrankungen gerade multimorbide Patienten mit nur mäßigen Behandlungsergebnissen nach einer Risiko-Adjustierung den größten Marktvorteil im Vergleich zu anderen Häusern darstellen. Die bessere Abschätzbarkeit von Ressourcenverbrauch, Vergütung und Gewinn pro Patient kann zu einer **Risikoselektion** mit einer bewussten Ausgrenzung weniger lukrativer Patientengruppen führen [38].
7. Krankenhäuser können sich durch dieses Systeme dazu verleitet lassen, Diagnosen und **Schweregrad aufzuwerten**, die ihnen im Krankenhausvergleich eine günstigere Position gewähren, z. B. komplizierte Erkrankung mit höherem Schweregrad und vielen Begleiterkrankungen. Dieses als ‚coding creep‘ bezeichnetes Phänomen wurde in den USA nach Einführung prospektiver Vergütungssysteme beobachtet. Es wurde geschätzt, dass bis zu einem Drittel der Zunahme eines case mix indexes zwischen 1986 und 1987 verändertem Kodierungsverhalten zuzuschreiben ist [39]. Um eine hinreichende Qualität der Risiko-Adjustierung zu gewährleisten, muss der zusätzliche Aufwand an Dokumentation und Kodierung vergütet werden und systematische Kodierfehler durch Prüfungen und entsprechende Sanktionen reduziert werden.

## 11.7 Welche Weiterentwicklungen sind notwendig, welche sind wünschenswert, und welche sind realisierbar?

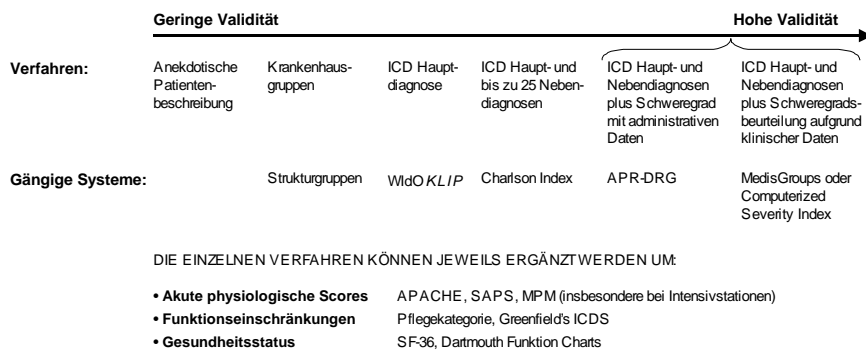
Nach den obigen Ausführungen wird deutlich, dass patientenbezogene Risiko-Adjustierungsverfahren notwendig sind. Dafür kann zunächst noch die klassische Leistungsstatistik in einer erweiterten Form herangezogen werden.

*Notwendig* erscheint dabei die möglichst rasche Erweiterung der Leistungsstatistik auf eine Hauptdiagnose und mindestens drei, idealerweise fünf und mehr Nebendiagnosen. Eine solche Erweiterung wird bei der Einführung von prospektiven Vergütungssysteme ohnehin notwendig. Ähnlich wie den erweiterten DRG-Systemen (AP-DRG, APR-DRG) könnten bestimmte Nebendiagnosen als für eine definierte Hauptdiagnose relevante Begleiterkrankungen qualifizieren und somit eine bedingte Schweregrad-Einteilung ermöglichen. Diese Erweiterung stellt keinen nennenswerten Mehraufwand für Ärzte dar. Des weiteren sollte darauf geachtet werden, dass die Verschlüsselung von Prozeduren (ICPM / OPS) akkurat erfolgt. Die Datenqualität bei der Verschlüsselung ist Grundvoraussetzung für eine valide Risiko-Adjustierung. Mit einer zunehmenden wirtschaftlichen Bedeutung von Diagnose- und Prozedurdaten in den Leistungsstatistiken, spätestens mit der Einführung prospektiver Vergütungssysteme, wird es im eigenen Interesse von Krankenhäusern liegen, diese Daten sorgfältig zu erfassen und zu kodieren.

*Wünschenswert* ist eine Erweiterung der Risiko-Adjustierung, wie sie z. B. im Rahmen prospektiver Vergütungssysteme vorgesehen ist. Der größte Vorteil von PVS liegt in Deutschland nicht nur in der Risiko-Adjustierung, sondern in der klaren Definition von erstatteten Kosten pro Patient. Diese Verlagerung der Vergütung von aggregierten (krankenhausbezogenen) zu patientenbezogenen Verfahren machen Krankenhausbetriebsvergleiche in Zukunft unnötig. Spätestens dann werden Qualitätsvergleiche in den Vordergrund treten, die von der dann etablierten personenbezogenen Risiko-Adjustierung profitieren.

*Realisierbar* sind diese Erweiterungen am leichtesten bei einem schrittweisen Ausbau der bisherigen Systeme. Auf diese Weise bleibt der jeweilige Aufwand überschaubar und die Krankenhäuser können ausreichend Expertise bilden, um die weiterentwickelten Systeme zu integrieren. Wenn man davon ausgeht, dass die LKA, entgegen unseren Empfehlungen, vorerst nicht erweitert wird, schlagen wir vor, mit diesen limitierten Daten eine personenbezogene Risiko-Adjustierung zu beginnen. Das könnte nicht nur die Risiko-Adjustierung verbessern, sondern auch Kliniken und Versicherungen auf zukünftige Denkmodelle der PVS und Qualitätsvergleiche vorbereiten.

Mit dem zögerlichen Ausbau von Krankenhausinformationssystemen werden in weiterer Zukunft PVS um mehr klinische Daten ergänzt werden können. Dies wird die prädiktive Validität erhöhen, den Vergleich von Krankenhäusern fairer machen und ultimativ die Behandlungsergebnisse transparenter machen.



**Abb. 11.5** Rangfolge unterschiedlicher R-A Verfahren bezüglich ihrer prädiktiven Validität.

## 11.8 Literaturverzeichnis

- [1] Iezzoni LI: Severity of illness measures and assessing the quality of hospital care. In Goldfield N, Nash DB (eds): Providing quality care (2<sup>nd</sup> ed.). Futures Challenges, Ann Arbor, MI, 1990.
- [2] Shwartz M, Ash AS, Iezzoni LI: Comparing outcomes across providers. In: Iezzoni LI (ed.): Risk Adjustment for Measuring Healthcare Outcomes, 2<sup>nd</sup> ed. Health Administration Press, Chicago, 1997.
- [3] Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ: Variations in the utilization of coronary angiography for elderly patients with acute myocardial infarction. An analysis using hierarchical logistic regression. Med Care 1995;33:625-42.
- [4] Richards T, Lurie N, Rogers WH, Brook RH: Measuring case mix and quality of care. Validity studies for the Graduate Medical Education Study. RAND 1987 R-3509-HHS.
- [5] Romano PS, Mark DH: Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32:81-90.
- [6] Fetter RB, Shin Y, Freeman RF, Averill RF, Thompson JD: Case mix definition by diagnosis-related groups. Med Care 1980;18 supplement:1-53.
- [7] Vladeck BC: Medicare hospital payment by diagnosis-related groups. Ann Int Med 1984;100:576-591.
- [8] Edwards N, Honemann D, Burley D, Navarro M: Refinement of the Medicare diagnosis-related groups to incorporate a measure of severity. Health Care Financing Rev 1994;16:45-64.
- [9] Goldfield N: Physician profiling and risk-adjustment. Second edition. Aspen, Gaithersburg, MD, 1996.
- [10] Ohmann C, Mueller U, Sangha O, Wildner M. Operationshäufigkeiten in Deutschland. Publikation in Vorbereitung, 1999.
- [11] Young WW, Swinkola RB, Zorn DM: The measurement of hospital case mix. Med Care 1982;20:501-512.
- [12] Young WW, Kohler S, Kowalski J: PMC patient severity scale: Derivation and validation. HSR 1994;29:367-390.
- [13] Markson LE, Nash DB, Louis DZ, Gonnella JS: Clinical outcomes management and disease staging. Eval Health Profess 1991;14:201-227.
- [14] Romano PS, Roos LL, Jollis JG: Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993;46:1075-1079.

- [15] Deyo RA, Cherkin DC, Ciol MA: Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613-619.
- [16] D'Hoore W, Sicotte C, Tilquin C: Risk Adjustment in Outcome Assessment: the Charlson Comorbidity Index. *Meth Inform Med* 1993;32:382-7.
- [17] Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA: Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol* 1996;49:273-278.
- [18] Horn SD, Sharkey PD, Buckle JM, Backofen JE, Averill RF, Horn RA: The relationship between severity of illness and length of stay and mortality. *Med Care* 1991;29:305-317.
- [19] Iezzoni LI, Daley J: A description and clinical assessment of the computerized severity index. *Quality Rev Bull* 1992;18:44-52.
- [20] Brewster AC, Karlin BG, Hyde LA, Jacobs CM, Bradburz RC, Chae ZM: MEDISGRPS: A clinically based approach to classifying hospital patients at admission. *Inquiry* 1985;22:377-387.
- [21] Steen PM, Brewster AC, Bradbury RC, Estabrook E, Young JA: Predicted probabilities of hospital death as a measure of admission severity of illness. *Inquiry* 1993;30:128-141.
- [22] Knaus et al: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36.
- [23] Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J: Mortality Probability Models (MPM II) based on an international cohort of ICU patients. *JAMA* 1993;270:2478-2486.
- [24] Le Gall JR, Lemeshow S, Saulnier F: A new simplified acute physiology score (SAPS II) based on an European/North American multicenter study. *JAMA* 1993; 270:2957-63.
- [25] Nava S. Scoring of severity in patients admitted to a respiratory intensive care unit. *Archives for Chest Disease*. 1997;52:71-72.
- [26] Ware JE, Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-481.
- [27] Landgraf JM, Nelson EC, Hays RD, Wasson JH, Kirk JW: Assessing function: Does it really make a difference? A preliminary evaluation of the acceptability and utility of the COOP function charts. In: Lipkin M. (ed.): *Functional status measurement in primary care*. Springer, New York, 1990.
- [28] Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernam YD: Judging hospitals by severity-adjusted mortality rates: The influence of the severity-adjustment method. *Am J Public Health* 1996;86:1379-1387.
- [29] Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernam YD: Severity measurement methods and judging hospital death rates. *Med Care* 1996;34:11-28.
- [30] Iezzoni LI, Shwartz M, Ash AS, Mackiernam YD: Does severity explain differences in hospital length of stay for pneumonia patients? *J Health Services Res and Policy*.
- [31] Auriant I, Vinatier I, Thaler F, Tourneur M, Loirat P: Simplified acute physiology score II for measuring severity of illness in intermediate care units. *Crit care Med* 1998;26:1368-71.
- [32] Guiguet M, Blot F, Escudier B, Antoun S, Leclercq B, Nitenberg G. Severity-of-illness scores for neutropenic cancer patients in an intensive care unit: Which is the best predictor? Do multiple assessment times improve the predictive value? *Critical Care Medicine*. 1998;26:488-93.
- [33] Schuster HP, Schuster FP, Ritschel P, Wilts S, Bodmann KF. The ability of the Simplified Acute Physiology Score (SAPS II) to predict outcome in coronary care patients. *Intensive Care Medicine*. 1997;23:1056-61.
- [34] Reina A, Vazquez G, Aguayo E, Bravo I, Colmenero M, Bravo M. Mortality discrimination in acute myocardial infarction: comparison between APACHE III and SAPS II prognosis systems. PAEEC Group. *Intensive Care Medicine*. 1997;23:326-30.
- [35] Iezzoni LI, Ayanian JZ, Bates DW, Burstin HR: Paying more fairly for Medicare capitated care. *N Engl J Med* 1998;339:1933-1938.
- [36] Farley DE, Hogan C: Case-mix specialization in the market of hospital services. *Health Services Research* 1990;25:757-781.

- 
- [37] Herzlinger R: Market driven health care. Perseus, Reading, Massachusetts, 1997.
- [38] Newhouse JP: Patients at risk: Health reform and risk adjustment. Health Affairs 1994;spring:132-146.
- [39] Carter GM, Newhouse JP, Relles DA: How much changes in the case mix index is DRG creep? The RAND Corporation, R-3826-HCFA, Santa Monica, 1990.