

## 2. Hintergrund und Stand der Forschung

### Risikoadjustierungssysteme

#### 2.1.1 Was ist Risikoadjustierung?

Der Begriff „Risikoadjustierung“ wird in verschiedenen Zusammenhängen verwendet und bezieht sich allgemein auf eine Situation, in der Gruppen hinsichtlich eines bestimmten Ergebnisses miteinander verglichen werden sollen. Wenn sich diese Gruppen bezüglich verschiedener Faktoren unterscheiden, die das Ergebnis beeinflussen könnten, müssen diese Einflussfaktoren berücksichtigt und kontrolliert werden, damit keine unzutreffenden Schlussfolgerungen gezogen werden.

#### 2.1.2 Risikoadjustierung im Gesundheitswesen

In der Gesundheitsökonomie wird eine Risikoadjustierung beispielsweise im Zusammenhang mit pauschalen Vergütungsmodellen für Leistungserbringer angewendet. Da Patienten je nach ihrem Gesundheitsstatus unterschiedlichen Behandlungsaufwand verursachen können, sollten die Pauschalen, die Leistungserbringer für die Behandlung verschiedener Patienten bekommen, entsprechend differenziert ausgestaltet sein. Für „kränkere“ Patienten sollten höhere, für „gesündere“ Patienten niedrigere Vergütungen gezahlt werden. Ohne eine solche Risikoadjustierung würden Anreize für eine Risikoselektion von Seiten der Leistungserbringer geschaffen, da eine Behandlung von Patienten, deren Behandlungsaufwand die Vergütungspauschale übersteigt, einen finanziellen Nachteil bedeuten würde.

Im Rahmen der Qualitätssicherung im stationären und ambulanten Bereich sollen verschiedene Leistungserbringer bezüglich ihrer Behandlungsergebnisse miteinander verglichen werden. Bei einem solchen Vergleich spielt der generelle Gesundheitsstatus der Patienten und die Schwere der behandelten Erkrankungen eine entscheidende Rolle, denn Patienten können auch dann unterschiedliche Behandlungsergebnisse aufweisen, wenn Behandlungsleistungen mit gleicher Qualität erbracht wurden. Die Therapie chronisch kranker, älterer und multimorbider Patienten ist im Allgemeinen kostenintensiver, und Komplikations- und Mortalitätsraten sind höher als bei jüngeren Patienten mit akuten Erkrankungen. Ärzte und Kliniken, die viele ältere, multimorbide Patienten betreuen, werden daher im Qualitäts- und Preiswettbewerb vermeintlich schlechter abschneiden als Konkurrenten mit einer weniger aufwendig zu behandelnden Patientenklientel. Um entscheiden zu können, ob tatsächlich ein erhöhter Ressourcenverbrauch bzw.

Qualitätsprobleme vorliegen, müssen zunächst gleiche Ausgangsbedingungen zwischen Leistungserbringern im Hinblick auf die zu vergleichenden Ergebnisse („Outcome“) geschaffen werden. Dazu müssen Faktoren, die das Outcome beeinflussen können, möglichst vollständig identifiziert und kontrolliert werden.

### 2.1.3 Methodischer Ansatz und Wirkprinzipien von Risikoadjustierungssystemen

#### 2.1.3.1 Das Konzept des „Confounding“

Aus der epidemiologischen Forschung ist das Konzept des „Confounding“ bekannt. „Confounder“ wirken als Störgrößen, die den Zusammenhang zwischen einer Exposition und dem Auftreten einer Erkrankung beeinflussen. Dadurch kann der Anschein einer in Wahrheit nicht vorhandenen kausalen Beziehung entstehen. Für einen solchen Einflussfaktor müssen zwei Voraussetzungen vorliegen (GORDIS 2001):

1. Er ist als Risikofaktor für das Auftreten der Erkrankung bekannt.
2. Er steht in Zusammenhang mit der Exposition, ist aber keine Folge davon.

Eigenschaften des Patienten, wie z.B. hohes Alter oder Multimorbidität, stellen Risikofaktoren für ein schlechteres Behandlungsergebnis bzw. einen erhöhten Ressourcenverbrauch dar, und werden mit einiger Wahrscheinlichkeit aufwendigere Behandlungsleistungen erforderlich machen. Gleichzeitig sind solche Faktoren in der Regel keine Folge der Behandlungsleistung. In diesem Sinne können Alter oder Komorbidität ebenfalls als Confounder betrachtet und behandelt werden (siehe Abbildung 1).

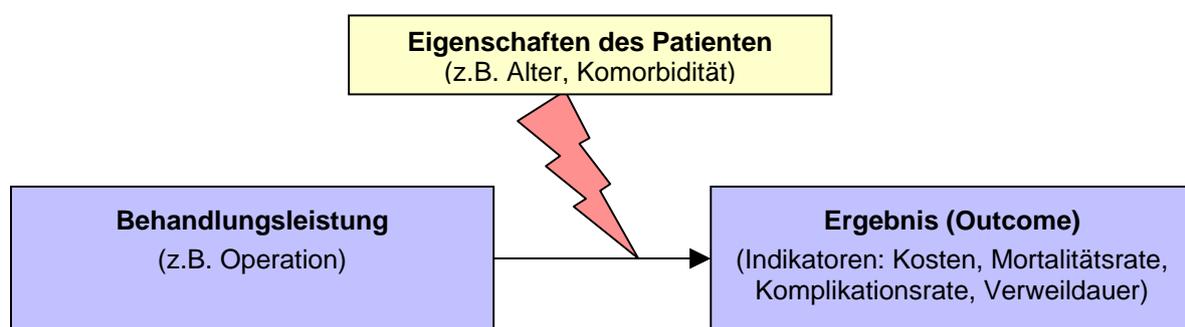


Abbildung 1: Zusammenhang zwischen Behandlungsleistung und Ergebnis und „Confounding“-Effekte

Werden die Effekte potentieller Confounder bei der Durchführung einer Studie nicht berücksichtigt, können unzulässige Schlussfolgerungen bei der Interpretation von

Studienergebnisse resultieren. Analog zu einer fälschlich angenommenen Ursache-Wirkungsbeziehung zwischen Exposition und Erkrankung könnte beim Vergleich verschiedener Leistungserbringer beispielsweise der Eindruck entstehen, es bestünde ein unmittelbarer Zusammenhang zwischen der Durchführung einer Behandlungsleistung und ihren Kosten bzw. dem Behandlungsergebnis; Unterschiede im Outcome wären somit eine direkte Folge unterschiedlich durchgeführter Behandlungsleistungen. Dadurch würde die Kosten- und Behandlungseffektivität von Leistungserbringern möglicherweise unzutreffend eingeschätzt. Aus den genannten Gründen sollten bekannte und potentielle Confounder mittels geeigneter Verfahren kontrolliert werden.

### **2.1.3.2 Methoden der Kontrolle von Confoundern in der epidemiologischen Forschung**

Die Auswahl eines Verfahrens zur Kontrolle von Confoundern hängt einerseits davon ab, ob bekannt ist, wie viele und welche Einflussfaktoren das Studienergebnis beeinflussen können und andererseits, ob es möglich ist, diese schon beim Studiendesign zu berücksichtigen. Ist ein Confounder bekannt, kann beispielsweise bei Fall-Kontrollstudien das „Matched Pairs“-Prinzip angewendet werden. Dabei wird zu jedem exponierten Fall eine Kontrolle gesucht, die hinsichtlich des Einflussfaktors die gleichen Merkmale aufweist, wodurch sein Effekt nivelliert wird. Ist dagegen nicht bekannt, welche Confounder den untersuchten Zusammenhang möglicherweise beeinflussen, können Probanden mittels eines Randomisierungsverfahrens zufällig auf die Expositions- und Kontrollgruppe verteilt werden. Eine solche „Randomisierung erhöht die Wahrscheinlichkeit, dass die Gruppen nicht nur im Hinblick auf uns bekannte Variablen vergleichbar sind, sondern auch hinsichtlich uns unbekannter Größen, die wir weder erkennen noch messen können“ (GORDIS 2001), die aber vielleicht einen Einfluss auf das Outcome haben.

Wenn es nicht möglich ist, Confounding-Effekte bei der Datenerhebung zu kontrollieren, können sie bei der Auswertung berücksichtigt werden. Dazu stehen im wesentlichen zwei Methoden zur Verfügung: Die Stratifizierung der Daten und die Anwendung statistischer Regressionsmodelle (ROTHMAN 2002). Sind die relevanten Einflussfaktoren bekannt und nicht zu zahlreich, kann die Auswertung stratifiziert erfolgen, d.h. die Ergebnisse werden beispielsweise nach Altersgruppen und Geschlecht getrennt betrachtet. Sollen viele Einflussfaktoren gleichzeitig einbezogen werden, wird diese Art der Analyse allerdings sehr aufwendig und es ist ein großes Untersuchungskollektiv nötig, da sich die gesamte Stichprobe auf immer mehr einzelne „Zellen“ aufteilt und die Besetzung der einzelnen Zellen sehr klein werden kann (KREIENBROCK und SCHACH 1995). In dieser Situation kann der jeweilige

Einfluss verschiedener Confounder effizienter mittels multivariater statistischer Verfahren wie der Regressionsanalyse berücksichtigt werden (ROTHMAN 2002, KREIENBROCK und SCHACH 1995). Die Regressionsanalyse zählt zu den Strukturen-Prüfenden Verfahren. Mit Hilfe von Regressionsmodellen können sowohl Ursachenanalysen als auch Wirkungsprognosen oder Zeitreihenanalysen durchgeführt werden. Bei einer Ursachenanalyse wird ein vermuteter Zusammenhang zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen überprüft und quantifiziert. Bei der Erstellung einer Wirkungsprognose wird die Frage untersucht, auf welche Weise sich die abhängige Variable mit einer Änderung der unabhängigen Variablen verändert, und mit einer Zeitreihenanalyse kann analysiert werden, ob und wie sich die abhängige Variable im Zeitverlauf ändert, d.h. man schätzt Werte der abhängige Variable für die Zukunft (Trendprognose). Durch seine Flexibilität ist die Methode eines der am häufigsten angewendeten multivariaten Verfahren (BACKHAUS et al. 1996).

In der epidemiologischen Forschung werden Regressionsmodelle überwiegend in zwei Fällen eingesetzt:

- um kausale Zusammenhänge zwischen einer Exposition und dem Auftreten einer Erkrankung zu überprüfen, und
- um Individuen bzw. Bevölkerungsgruppen mit erhöhtem Risiko für das Auftreten einer bestimmten Krankheit zu identifizieren (Prognosemodell).

In ersterem Fall soll der Einfluss potentieller Confounder mit Hilfe des Verfahrens einzeln analysiert und quantifiziert werden, so dass der „wahre“ Effekt der Exposition isoliert werden kann. Bei der Konzipierung von medizinischen Prognosemodellen werden Informationen über bekannte Risikofaktoren einer Erkrankung (z.B. Alter, Geschlecht, Blutdruck- und Cholesterinwerte, Raucherstatus für koronare Herzkrankheit) in das Regressionsmodell einbezogen und die Wahrscheinlichkeit für einen Patienten oder eine Bevölkerungsgruppe berechnet, in einem definierten Zeitraum (z.B. in den nächsten 10 Jahren) an dieser Krankheit zu erkranken oder zu versterben (ROTHMAN 2002).

### **2.1.3.3 Methoden der Kontrolle von Confoundern in der Outcomeforschung**

Outcomeforschung als Teilgebiet der Versorgungsforschung analysiert die Effektivität und das Resultat verschiedener Versorgungsmaßnahmen unter Alltagsbedingungen (PFAFF 2005). Bei einem Vergleich verschiedener Leistungserbringer können nur die Behandlungsergebnisse der von diesen Ärzten und Kliniken behandelten Patienten untersucht

werden. Dabei kann nicht davon ausgegangen werden, dass sich Patienten mit ihren spezifischen Charakteristika zufällig auf bestimmte Leistungserbringer verteilen, sondern es ist davon auszugehen, dass insbesondere in spezialisierten Kliniken oder Arztpraxen diesbezüglich eine hochselektierte Auswahl an Patienten vorliegt. Insofern ist eine Kontrolle von Confoundern in diesem Bereich eher nicht durch ein entsprechendes Studiendesign (z.B. durch eine Randomisierung von Studienteilnehmern) erreichbar.

Um dennoch einen fairen Vergleich von Leistungserbringern zu ermöglichen, müssen Faktoren, die das Behandlungsergebnis beeinflussen könnten, daher meist bei der Datenauswertung berücksichtigt werden. Sind diese Faktoren bekannt, nicht sehr zahlreich und ist die Datenmenge genügend umfangreich, kann eine Stratifizierung der Daten (z.B. nach Alter und Geschlecht) erfolgen. Sollen jedoch viele Confounder gleichzeitig einbezogen werden, kann ihr jeweiliger Einfluss auf das Outcome in einem Regressionsmodell einzeln überprüft und quantifiziert werden. Es handelt sich dabei um ein Prognosemodell, mit dessen Hilfe Patientengruppen mit erhöhtem Ressourcenverbrauch oder ungünstigem Behandlungsergebnis identifiziert bzw. Ressourcenverbrauch oder Behandlungsergebnis für einzelne Patienten prognostiziert werden können. Ein solches Regressionsmodell sagt im Prinzip voraus, welchen Wert das untersuchte Behandlungsergebnis annähme, wenn mögliche Confounder „einberechnet“ würden, d.h. es wird ein „erwartetes“ Ergebnis berechnet, das mit dem realen Outcome verglichen werden kann. Der Unterschied zwischen dem erwarteten und dem tatsächlichen Wert stellt den Anteil des Ergebnisses dar, der nicht durch die in das Modell einbezogenen Einflussfaktoren erklärt wird. Weist ein Krankenhaus beispielsweise trotz Berücksichtigung relevanter Confounder höhere Mortalitätsraten auf als andere Kliniken, könnte diese Tatsache einen Hinweis auf echte Qualitätsmängel bedeuten, da sich die Ergebnisse nicht durch eine unterschiedliche Patientenklientel erklären lassen.

#### ***2.1.4 Prinzip der Entwicklung von Risikoadjustierungssystemen***

Die Konstruktion eines Prognosemodells für eine gegebene Fragestellung erfolgt in mehreren Schritten. Als Voraussetzung für die Identifizierung und Auswahl der für die Fragestellung relevanten Risikofaktoren sollten zunächst die folgenden vier Fragen beantwortet werden (IEZZONI 2003):

1. In Bezug auf welches Ergebnis (Outcome) soll für potentielle Risikofaktoren adjustiert werden?
2. In welchem Zeitraum soll für potentielle Risikofaktoren adjustiert werden?
3. In welcher Population soll für potentielle Risikofaktoren adjustiert werden?

4. Warum/zu welchem Zweck soll für potentielle Risikofaktoren adjustiert werden?

Die am häufigsten untersuchten **Ergebnisse** (Outcomes) sind Höhe des Ressourcenverbrauchs (Kosten, Verweildauer), Qualität der Behandlungsergebnisse (Mortalitätsrate, Komplikationsrate, funktioneller Status) und patientenbezogene Ergebnisse (z.B. Lebensqualität). Der **Untersuchungszeitraum** kann entweder eine Behandlungsepisode umfassen, wie z.B. einen Krankenhausaufenthalt oder die Dauer der Behandlung einer bestimmten Erkrankung, oder auch einen definierten Zeitraum, z.B. wenn die Höhe der Behandlungskosten pro Jahr evaluiert werden soll. Für die Auswahl geeigneter Risikofaktoren zur Risikoadjustierung eines Outcomes ist die zu untersuchende **Patientenpopulation** besonders relevant. So spielen bei geriatrischen Patienten andere Faktoren eine Rolle, die das Behandlungsergebnis beeinflussen können, als bei psychisch Kranken oder behinderten Menschen. Das **Ziel** von Evaluationen, die einer Risikoadjustierung bedürfen, ist einerseits der Vergleich verschiedener Leistungserbringer in Bezug auf Kosteneffizienz und Behandlungsqualität. Andererseits sollen auf diese Weise angemessene Vergütungspauschalen für Ärzte und Krankenhäuser festgelegt werden, wie es z.B. mit der Einführung des Systems der Diagnosis Related Groups (DRGs) beabsichtigt war. Außerdem sollen risikoadjustierte Vergütungsmodelle Risikoselektion von Seiten der Leistungserbringer verhindern, indem sie einen finanziellen Ausgleich für die Behandlung multimorbider, alter und chronisch kranker Patienten bieten.

Sind potentielle Confounder für ein bestimmtes Outcome noch nicht hinreichend bekannt, muss zunächst versucht werden, anhand von Literaturrecherchen und Expertenbefragungen mögliche Einflussfaktoren zu identifizieren, für die ein klinischer Zusammenhang mit dem Ergebnis nachvollziehbar ist. Potentielle Einflussfaktoren lassen sich in die folgenden Kategorien einteilen (IEZZONI 2003):

- Demographische Faktoren (Alter, Geschlecht, Ethnizität)
- Klinische Faktoren (Akute klinische Stabilität, Hauptdiagnose, Komorbidität, funktionaler Status, psychischer und kognitiver Status)
- Sozioökonomische Faktoren (Familiensituation, Bildung, Einkommen, Beruf, Wohnungssituation, kultureller Hintergrund)
- Gesundheitsbezogenes Verhalten (Alkohol, Rauchen, Drogenkonsum, Sexualverhalten, Ernährung und Übergewicht)

- Einstellungen und Präferenzen (Gesundheitsstatus, Lebensqualität, Präferenzen und Erwartungen an das Gesundheitssystem)

Im nächsten Schritt wird anhand einer genügend großen Studienpopulation untersucht, ob statistisch ein signifikanter Zusammenhang zwischen den ausgewählten Einflussfaktoren und dem zu untersuchenden Outcome nachweisbar ist (z.B. mittels eines Chi-Quadrattests). Diese Variablen werden in ein Regressionsmodell eingeschlossen (IEZZONI 2003). Durch die Berechnung der Regressionskoeffizienten, die ein Maß für den Erklärungswert der Variablen in Bezug auf das Outcome darstellen, kann der Einfluss einzelner Faktoren quantifiziert und als Gewicht oder Risikoscore ausgedrückt werden. Die Gewichte bzw. Scores für die einzelnen Variablen können addiert und daraus ein Maß für das Gesamtrisiko jedes Patienten gebildet werden. Um die Güte bzw. den Erklärungswert des Regressionsmodells zu prüfen (d.h. wie gut reale Unterschiede von Patienten bezüglich eines bestimmten Outcomes abgebildet werden), sollte die Vorhersagekraft des anhand einer bestimmten Studienpopulation entwickelten Modells in einem weiteren Schritt an einer anderen Stichprobe getestet werden. Um die Datenanalyse zu vereinfachen, wird der ursprüngliche Datensatz meist von vornherein in eine Entwicklungs- und Validierungsstichprobe geteilt (IEZZONI 2003).

### **2.1.5 Risikoadjustierungssysteme im Bereich der Qualitätssicherung**

Der international am häufigsten verwendete Outcomeparameter in der stationären Qualitätssicherung ist die Mortalitätsrate. Dabei wird meist die Frühmortalität in einem Zeitraum bis zu 30 Tage nach Klinikaufnahme gemessen. Als Risikofaktoren, für die in mehreren Studien ein Zusammenhang mit der Mortalitätsrate nachgewiesen werden konnte, haben sich Alter, Geschlecht, Haupt- und Nebendiagnosen sowie die akute klinische Stabilität des Patienten herausgestellt (z.B. IEZZONI 2003, CHARLSON et al. 1987, KNAUS et al. 1981). Im Folgenden sollen exemplarisch drei international verbreitete Risikoadjustierungssysteme beschrieben werden, die im Bereich der Qualitätssicherung Anwendung finden (INOUYE et al. 2003). Der in den USA entwickelte **Charlson Comorbidity Index** ist ein Beispiel für ein System, das ursprünglich für den Einsatz in Langzeittherapiestudien konzipiert wurde und den Einfluss der Komorbidität eines Patienten in Bezug auf sein Sterberisiko innerhalb eines bestimmten Zeitraums berücksichtigt. Das **APACHE III**-System wurde entwickelt, um die 30-Tage-Mortalität von Patienten auf Intensivstationen zu messen und sinnvoll vergleichen zu können und beinhaltet in der Hauptsache Indikatoren der akuten klinischen Stabilität. Der **EUROScore**, auf dem das

deutsche System des Koronarchirurgie-Score (KCH) basiert, wurde für die Adjustierung der 30-Tage-Mortalität nach kardiochirurgischen Operationen in Europa entwickelt und bezieht patienten- und operationsbezogene Risikofaktoren ein.

### **2.1.5.1 Der Charlson Comorbidity Index**

CHARLSON et al. (1987) suchten nach einer Methode, den Effekt von Begleiterkrankungen auf die Mortalität in Therapiestudien zu quantifizieren. Patienten mit Nebendiagnosen, die das Sterberisiko erhöhen, konnten aufgrund ihres Einflusses auf das Outcome bis dahin nicht in therapeutische Wirksamkeitsstudien eingeschlossen werden. Die Autoren untersuchten anhand eines Kollektivs von 604 internistischen Krankenhauspatienten den Einfluss von Nebendiagnosen auf die 1-Jahres-Überlebensrate mittels des Cox Proportional Hazard Modells (COX 1972). Bei dieser Methode handelt es sich um eine Regressionsanalyse, mit dem z.B. das Ausmaß eines Therapieeffekts in Bezug auf die Überlebenszeit unter Berücksichtigung anderer Einflussgrößen geschätzt werden kann (ZIEGLER et al. 2004). Für jeden Patienten dokumentierten die Autoren bei der Aufnahme Alter, Art und Schweregrad der Hauptdiagnose und ermittelten nach Entlassung sämtliche vorliegenden Begleiterkrankungen und ihre Schweregrade aus den Krankenakten. Für das Vorliegen einer bestimmten Nebendiagnose berechneten sie anschließend sowohl das rohe als auch das für Aufnahmediagnose, Schweregrad und übrige Nebendiagnosen adjustierte relative Sterberisiko. Anhand der adjustierten Werte wurde jeder dieser Nebendiagnosen ein Gewicht zugeordnet. Die addierten Gewichte der verschiedenen Nebendiagnosen wurden dann zu einem Index zusammengefasst, der sowohl die Zahl als auch den Schweregrad der Begleiterkrankungen jedes Patienten abbildete. Die Validierung dieses Komorbiditäts-Indexes erfolgte an einer historischen Kohorte von 685 Brustkrebspatientinnen, für die Daten über die 10-Jahres-Mortalität vorlagen. Dabei zeigte sich, dass bei Nachbeobachtungszeiträumen über fünf Jahren auch das Alter einen unabhängigen prädiktiven Faktor für die Mortalität darstellt, so dass für die Anwendung in Langzeitstudien ein modifizierter Index empfohlen wird, der Alter und Komorbidität additiv berücksichtigt. Seit seiner auf Angaben aus Krankenakten basierenden Entwicklung wurde der Charlson Comorbidity Index an ICD-9-verschlüsselte Diagnosedaten angepasst, so dass für eine Risikoadjustierung administrative Daten genutzt werden können (ROMANO et al. 1993, DEYO et al. 1992). Im Jahre 1992 machte die Einführung der ICD-10 durch die WHO eine Überarbeitung und Anpassung an die neue Verschlüsselungssystematik notwendig, der sich verschiedene Autoren angenommen haben (Übersicht bei QUAN et al. 2005). Außerdem wurde die für eine Berechnung des Indexes herangezogene Datenbasis inzwischen um ambulante Diagnosedaten erweitert, die

ausschließlich oder in Verbindung mit stationären Diagnosen genutzt werden können (KLABUNDE et al. 2000). Da der Charlson Comorbidity Index für die Untersuchung verschiedenster Fragestellungen geeignet ist, insbesondere aber für die Prognose der Mortalität von Patienten über längere Zeiträume (KLABUNDE et al. 2002), hat er international weite Verbreitung gefunden.

#### **2.1.5.2 Das APACHE III-System**

Das Acute Physiology Age Chronic Health Evaluation (APACHE)-System wurde konzipiert, um das Mortalitätsrisiko von Patienten auf Intensivstationen mittels eines Scores zu quantifizieren. Die erste Version wurde 1981 von KNAUS et al. entwickelt und enthielt 34 akute physiologische Parameter, aus denen sich ein Wert bilden ließ, mittels dessen das Sterberisiko eines Patienten berechnet werden konnte. Im Jahr 1985 folgte eine erweiterte APACHE II-Version, die zusätzlich Angaben über die Diagnose enthielt, die zur Aufnahme auf die Intensivstation geführt hatte, und im Jahr 1991 stellten die Autoren das nochmals umfangreich überarbeitete und anhand der Daten von 17.440 Intensivpatienten aus 40 Krankenhäusern validierte APACHE III-System vor (KNAUS et al. 1991). Für die Berechnung des APACHE III-Scores werden von den ursprünglichen 34 nur noch 17 physiologische Parameter herangezogen, wie z.B. Puls, Blutdruckwerte, Temperatur, Atemfrequenz, Sauerstoffpartialdruck und verschiedene Laborwerte, sowie der neurologische Status gemäß der Glasgow Coma Scale. Dafür sind Alter des Patienten und Begleiterkrankungen einbezogen worden; die im APACHE II-System eingeführten Diagnoseklassen (Major Disease Categories) wurden zahlenmäßig erweitert und in nicht-operative und postoperative Kategorien unterschieden. Zusätzlich soll der Ort angegeben werden, von dem aus der Patient auf die Intensivstation eingewiesen wurde. Die Abschätzung des Mortalitätsrisikos erfolgt anhand einer Gleichung, die neben dem APACHE-Score einen Wert für die Zuweisungsdiagnose und den Ort der Zuweisung additiv berücksichtigt. Auf diese Weise kann das APACHE III-System dazu eingesetzt werden, das Mortalitätsrisiko einzelner Patienten schon am Aufnahmetag zu schätzen. Gleichzeitig stellt der APACHE-Score ein gutes Werkzeug für die Risikostratifizierung verschiedener Patientengruppen dar.

#### **2.1.5.3 Der EuroScore und der KCH**

Ein Beispiel für einen europäischen Risikoscore stellt der EuroSCORE dar, ein Risikostratifizierungsmodell für den Vergleich von Frühmortalitätsraten in der Kardiochirurgie, das anhand der Behandlungsdaten von 19.000 herzchirurgischen Patienten in acht europäischen Ländern aus dem Jahr 1995 entwickelt wurde (NASHEF et al. 1999). Der

EuroSCORE enthält patientenbezogene Faktoren (Alter, Geschlecht, Begleiterkrankungen, vorangegangene Herzoperationen, Höhe des Serumkreatinins, präoperative Komplikationen), kardiale Faktoren (instabile Angina pectoris, linksventrikuläre Dysfunktion, Zustand nach Herzinfarkt, pulmonale Hypertension) und operationsbezogene Faktoren (Notfall-OP, Einbeziehung der thorakalen Aorta, Septumruptur nach Herzinfarkt), die je nach ihrem Einfluss auf die Mortalität mit einem Score zwischen 1 und 4 bewertet werden.

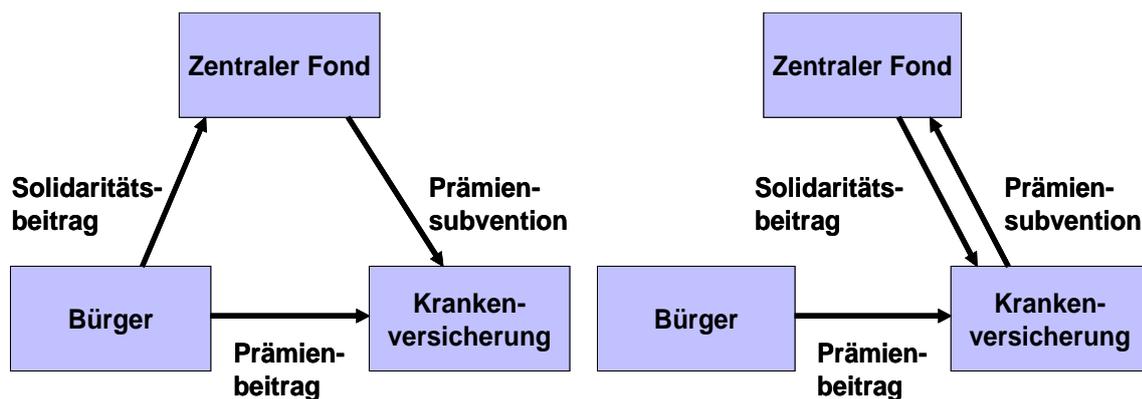
Im Rahmen der externen Qualitätssicherung im stationären Bereich wurde in Deutschland auf der Basis des EuroSCORE von der Bundesgeschäftsstelle Qualitätssicherung (BQS) ein eigenes Risikoadjustierungssystem für herzchirurgische Patienten entwickelt, der Koronarchirurgie-Score (KCH). Mit Hilfe eines logistischen Regressionsmodells wird aus 17 Risikofaktoren, für die ein statistisch signifikanter Einfluss auf die Frühmortalität nachgewiesen werden konnte, die erwartete 30-Tage-Letalität für herzchirurgische Patienten in den untersuchten Kliniken errechnet und mit der beobachteten Letalität verglichen. Um risikoadjustierte Werte pro Klinik zu erhalten, wird der Quotient aus erwarteter und beobachteter Letalität gebildet und mit der beobachteten Gesamtletalität aller Patienten aus allen Krankenhäuser multipliziert. Das Ergebnis repräsentiert die Letalität, die das untersuchte Krankenhaus aufweisen würde, wenn die Risikostruktur seiner Patientenklientel derjenigen aller anderen Kliniken entspräche (MOHR et al. 2005).

### ***2.1.6 Risikoadjustierungssysteme im Bereich der Vergütung und Kostenevaluation***

Im Bereich der Vergütung dient eine Risikoadjustierung dem Zweck, die Ziele von Gesundheitsreformen eines Landes durch entsprechende finanzielle Anreize zu unterstützen. In Ländern mit einem staatlichen, durch Steuern finanzierten Gesundheitssystem, wie z.B. Großbritannien, Spanien oder Schweden, beinhalten diese Ziele vor allem, Gleichheit („equity“) im Zugang zum Gesundheitssystem zu gewährleisten, d.h. jeder Bürger soll unabhängig von Einkommen, sozialem Status oder Region genau diejenigen Gesundheitsleistungen bekommen, die er benötigt (RICE und SMITH 2001). Der Bedarf an solchen Gesundheitsleistungen kann regional stark variieren, so dass sich die Frage ergibt, nach welchem Schlüssel die erhobenen Steuergelder auf die unterschiedlichen Regionen des Landes verteilt werden sollen, um überall eine bedarfsgerechte Versorgung der Bevölkerung sicherzustellen.

In Ländern mit einem Krankenversicherungssystem (neben Deutschland z.B. die Niederlande, Schweiz, Belgien, Israel) zahlt jeder Versicherte einen Beitrag nach seinen finanziellen Möglichkeiten und erhält dafür alle Leistungen, die er im Krankheitsfall benötigt

(Solidaritätsprinzip). Da sich der Bedarf an Gesundheitsleistungen von Versicherten der verschiedenen Krankenkassen stark unterscheiden kann, die Beiträge sich jedoch nach den wirtschaftlichen Möglichkeiten richten, sind Kassen mit überdurchschnittlich vielen Geringverdienern mit schlechtem Gesundheitsstatus im Nachteil. Damit das Solidaritätsprinzip aufrechterhalten werden kann, ist ein finanzieller Ausgleich zwischen den Versicherern nötig. Dazu existiert in diesen Ländern ein gemeinsamer Ausgleichsfond. In den Niederlanden, Belgien und Israel sind die Beiträge unterteilt in einen Prämienanteil, der an die Versicherung gezahlt wird, und einen einkommensabhängigen Solidaritätsanteil, den die Versicherten selbst in den Ausgleichsfond einzahlen. In der Schweiz und in Deutschland werden einkommensabhängige Beiträge an die Versicherer entrichtet, die dann über einen Risikostrukturausgleich unter den Krankenkassen verteilt werden (VAN DE VEN et al. 2003) (siehe Abbildung 2).



**Abbildung 2: Zwei Varianten von Ausgleichszahlungen in Gesundheitssystemen mit Versicherungsprinzip (aus: VAN DE VEN et al. 2003)**

Sowohl die Verteilung staatlicher Steuergelder an regionale Behörden als auch die Ausgleichszahlungen für die einzelnen Versicherungen erfolgen normalerweise als Pauschalen (engl. „Capitation“). Ein „Capitation payment“ ist definiert als „eine Pro-Kopf-Zahlung, die für jeden Einwohner einer Region bzw. für jeden Versicherten einer Krankenkasse in das Budget der regionalen Behörde bzw. Versicherung eingeht, um die Gesundheitsleistungen für dieses Mitglied in einem bestimmten Zeitraum zu finanzieren“ (RICE und SMITH 2001, S.82). Um den Anreiz für eine Risikoselektion durch die Versicherungsunternehmen zu vermindern, erfolgt die Berechnung der Höhe solcher Pauschalen im Allgemeinen risikoadjustiert. In den meisten Ländern enthalten entsprechende Risikoadjustierungssysteme die Variablen Alter und Geschlecht. In Großbritannien,

Skandinavien, Belgien, der Schweiz, den USA, Kanada, Australien und Neuseeland werden diese Risikofaktoren durch weitere demographische, sozioökonomische und geographische Indikatoren, aber auch durch aggregierte gesundheitsbezogene Indikatoren wie Mortalitätsraten ergänzt (siehe Übersicht bei RICE und SMITH 2001).

Um ein weiteres Ziel von Gesundheitsreformen, nämlich eine höhere Effizienz der Gesundheitssysteme zu erreichen, sollen Krankenkassen bzw. lokale Behörden stärker als bisher das Krankheitsrisiko ihrer Mitglieder tragen. Damit dieses Risiko kalkulierbar wird, muss die Risikoadjustierung von Pro-Kopf-Pauschalen so gestaltet sein, dass die zu erwartenden Kosten möglichst genau abgebildet werden. Dazu ist es notwendig, neben demographischen und aggregierten Krankheitsdaten Informationen über die Morbidität auf individueller Ebene heranzuziehen. Mit Hilfe eines Patientenklassifikationssystems können anhand von Morbiditätsindikatoren (z.B. Krankheitsdiagnosen aus dem stationären und/oder ambulanten Bereich, Medikamentenverordnungen, Prozeduren) und der Kosten für Gesundheitsleistungen aus dem Vorjahr die erwarteten Kosten pro Mitglied für das Folgejahr geschätzt werden. Viele Länder prüfen derzeit den Einsatz von Patientenklassifikationssystemen für die Berechnung von Pro-Kopf-Zahlungen oder haben bereits ein solches System eingeführt. Um Anreize für Leistungserbringer zu schaffen, Patienten kosteneffektiv zu behandeln bzw. möglichst gesund zu erhalten, kann die Vergütung im ambulanten und stationären Bereich ebenfalls mittels Kopfpauschalen erfolgen. Diese Zahlungen sollten genauso nach der Morbidität der Patienten adjustiert werden, wie Pro-Kopf-Pauschalen im Rahmen eines Risikostrukturausgleichs, da ansonsten auch hier die Gefahr einer Risikoselektion von Seiten der Leistungserbringer entstehen kann. Im Folgenden soll zunächst die historische Entwicklung und das Prinzip von Patientenklassifikationssystemen dargestellt und an zwei Modell-Beispielen verdeutlicht werden. Im Anschluss wird der aktuelle Stand der internationalen Verbreitung solcher Systeme sowie die Anwendung von Risikoadjustierungssystemen in Deutschland dargestellt.

#### **2.1.6.1 Patientenklassifikationssysteme**

In den USA wurde die Notwendigkeit einer adäquaten Risikoadjustierung von Pro-Kopf-Zahlungen seit den achtziger Jahren unter anderem im Zusammenhang mit dem von der Health Care Financing Administration (HCFA, inzwischen umbenannt in „Centers for Medicare and Medicaid Services“, CMS) angebotenen Medicare-Programm diskutiert, das die Gesundheitsversorgung für US-Bürger über 65 Jahre sicherstellt. Aus Kostengründen versuchte die HCFA schon in den siebziger Jahren, die Behandlung seiner Mitglieder durch

so genannte Health Maintenance Organizations (HMOs) zu fördern, die mittels Capitation payments vergütet wurden. Solche Pauschalen wurden bis zum Ende der neunziger Jahre nach geographischer Region und demographischen Faktoren adjustiert, wie z.B. Alter, Geschlecht, Sozialhilfe- („welfare“) und institutionellem Status (Altenheimbewohner oder nicht). In dieser Situation entstand einerseits die Sorge, dass Medicare-Patienten, für die ein hoher Behandlungsaufwand vermutet wurde, von den marktwirtschaftlich orientierten HMOs aufgrund der undifferenzierten Vergütungspauschalen nicht gern angenommen würden. Andererseits hatten mehrere Studien ergeben, dass sich überwiegend gesündere Mitglieder in HMOs einschreiben ließen, für die die Pauschalen eher zu hoch angesetzt waren (MORGAN et al. 1997, ELLIS et al. 1996). Durch die Einbeziehung des Gesundheitsstatus als Risikofaktor sollten der erhöhte Behandlungsaufwand und die Kosten für die ältere, oft multimorbide Patientenklientel von Medicare bei der Berechnung von Kopfpauschalen berücksichtigt und dadurch eine Risikoselektion durch die HMOs verhindert werden (MCCLURE 1984).

Forscher der Boston University entwickelten daraufhin Mitte der achtziger Jahre im Auftrag der HCFA ein erstes Patientenklassifikationssystem, das Diagnostic Cost Group-Modell (DCG) (ASH et al. 1989). Da der HCFA zu diesem Zeitpunkt als relevante Risikofaktoren nur Diagnosen von stationären Aufenthalten ihrer Mitglieder vorlagen, basierte das Modell ausschließlich auf diesen Diagnosen. In den neunziger Jahren wurden DCGs weiter verfeinert, an verschiedene Populationen angepasst und mit der steigenden Verfügbarkeit von Daten aus dem ambulanten und dem Arzneimittelbereich durch weitere Risikofaktoren ergänzt. Eine Modellfamilie entstand, die inzwischen neben den ursprünglichen PIP (Principal Inpatient)-DCGs die auf ambulanten oder stationären Diagnosen basierenden Hierarchical Condition Categories (DCG/HCC) und die Medikamentenverordnungen einbeziehenden RxGroups umfasst. Je nach den zur Verfügung stehenden Daten können die einzelnen Modelle kombiniert werden, um die Vorhersagekraft zu erhöhen. Seit Mitte der neunziger Jahre wurden in den USA von verschiedenen Institutionen weitere Klassifikationssysteme entwickelt, die sich in der Zahl und Kombination der einbezogenen Risikofaktoren, in Bezug auf die Zielgruppe und den betrachteten Zeitraum unterscheiden (siehe Übersicht bei RESCHKE et al. 2004).

Allgemein soll mit Hilfe eines Patientenklassifikationssystems versucht werden, den Behandlungsaufwand und damit die Kosten für Versicherte in einem bestimmten Zeitraum zu schätzen. Dabei unterscheidet man einen so genannten Zellen- und einen Regressionsansatz.

Beim Zellenansatz werden Patienten anhand ihrer in Bezug auf das Outcome „Kosten“ relevanten Risikomerkmale zu möglichst homogenen Gruppen zusammengefasst (Stratifizierung), d.h. jeder Versicherte wird genau einer Gruppe (Zelle) zugeordnet. Für jede Risikogruppe werden dann die durchschnittlichen Pro-Kopf-Ausgaben ermittelt. Beim Regressionsansatz erfolgt die Schätzung der erwarteten Ausgaben aufgrund von Regressionskoeffizienten. „Die Regressionskoeffizienten stammen aus Regressionsmodellen, mit denen die Ausgaben eines Versicherten durch eine große Anzahl von Risikomerkmale (z.B. Behandlungsanlass nach ICD-Code gruppiert) als den unabhängigen Variablen erklärt werden“ (RESCHKE et al. 2004, S. 34). Der Regressionsansatz wird auch Zuschlagsansatz genannt, weil sich die Ausgabenberechnung für jeden Versicherten dabei aus einem Grundbetrag gemäß Alter und Geschlecht und verschiedenen Zusatzbeträgen für andere Risikomerkmale, wie beispielsweise Diagnosen, zusammensetzt. Außerdem ist wichtig, welcher Zeitraum der Leistungsanspruchnahme betrachtet werden soll. Werden die betrachteten Morbiditätsindikatoren herangezogen, um Ausgaben für einen kommenden Zeitraum zu prognostizieren, handelt es sich um ein prospektives Modell, sollen sie die derzeitige Inanspruchnahme erklären, spricht man von einem retrospektiven (engl. concurrent) Modell. Beispielhaft sollen das auf ambulanten Diagnosen basierende ACG- und das für ambulante und stationäre Diagnosen geeignete HCC-System näher beschrieben werden.

#### ***2.1.6.1.1 Das System der Adjusted Clinical Groups (ACG)***

Internationale Verbreitung hat insbesondere das an der Johns Hopkins Universität in Baltimore entwickelte ACG-System gefunden (WEINER et al. 1991). In diesem System wird jede Krankheitsdiagnose des Patienten anhand der Krankheitsdauer, -schwere und -ätiologie, der Diagnosesicherheit, sowie der Wahrscheinlichkeit, dass aufwendige Spezialbehandlungen nötig werden, in 32 so genannte „Aggregated Diagnosis Groups“ (ADG) eingeteilt. Da jeder Patient theoretisch zwischen 0 und 32 Diagnosen aufweisen kann und sich hierdurch 4,3 Mrd. möglicher Kombinationen für eine Gruppierung ergeben, wurden bei der Entwicklung des Systems diese 32 ADGs gemäß der Krankheitsdauer zu 12 Gruppen zusammengefasst (CADGs). Für diese 12 Kategorien wurden die am häufigsten vorkommenden Diagnosenkombinationen mit einer möglichst geringen Ausgabenvarianz berechnet und daraus 18 so genannte „Major Ambulatory Categories“ (MACs) gebildet, die jeweils einer ACG entsprechen. 7 andere MACs, die eine größere Varianz bezüglich der Ausgaben aufwiesen, wurden für die Bildung einer ACG mit den Parametern Alter, Geschlecht und Zahl der ADGs kombiniert.

Die Zuordnung eines Patienten zu einer der 53 bis 93 ACG-Morbiditätsklassen erfolgt anhand seiner Krankheitsdiagnosen gemäß MAC bzw. ADGs, Alter, Geschlecht und zusätzlichen Informationen, sofern diese vorliegen (z.B. Geburt erfolgt/nicht erfolgt bei der ACG für Schwangerschaft). Das ACG-System ist also dazu geeignet, Kostenaspekte von Multimorbidität zu berücksichtigen.

#### ***2.1.6.1.2 Das System der Hierarchical Condition Categories (HCC)***

Dieses System ging nach mehreren Revisionen aus dem auf einem Zellenansatz basierenden Diagnostic Cost Group (DCG)-Modell hervor welches den Nachteil aufwies, dass sich Multimorbidität ungenügend abbilden ließ, da jeweils nur die kostenintensivste Diagnose für die DCG-Klassifizierung Berücksichtigung fand. Im HCC-System wird jeder Patient anhand seiner Krankheitsdiagnosen in eine von 781 Diagnostic Groups eingeordnet (DxGroups), die anhand von klinisch medizinischen Gesichtspunkten und Kostenaspekten zu 184 Condition Categories (CCs) zusammengefasst werden, welche in sich hierarchisch gegliedert sind (ELLIS et al. 1996). Jeder HCC ist je nach Schwere ein bestimmtes Kostengewicht zugeordnet. Für die Klassifizierung eines Patienten findet neben den Parametern Alter und Geschlecht die in der Hierarchie am höchsten stehenden und damit kostenintensivsten Condition Categories Eingang. Weist der Patient mehrere „gleichrangige“ Erkrankungen im Sinne einer Multimorbidität auf, können diese im HCC-Modell in Form mehrerer Zuschläge berücksichtigt werden (POPE et al. 2000).

#### **2.1.6.2 Anwendung risikoadjustierter Vergütungs- und Kostenausgleichssysteme im internationalen Kontext**

Im Folgenden soll ein Überblick über diejenigen Patientenklassifikationssysteme gegeben werden, die derzeit international zu Vergütungszwecken eingesetzt werden (siehe Tabelle 1). Die meisten Anwendungserfahrungen mit Patientenklassifikationssystemen liegen in den **USA** vor, wo fast alle verfügbaren Systeme auch entwickelt wurden.

**Tabelle 1: Überblick über Patientenklassifikationssysteme, für die internationale Anwendungserfahrungen vorliegen**  
(HMO=Health Maintenance Organization)

Abkürzung	Name	Entwickler	Anwendungszweck	Einsatz durch
DCG	Diagnostic Cost Groups	Ash et al. Boston University, Fa. DxCG	Berechnung von Kopfpauschalen für Versicherte und Budgets für HMOs und andere Leistungserbringer	U.S Department of Veterans Affairs, Kaiser Permanente, Pacific Business Group on Health, USA
PIP-DCG	Principal Inpatient Diagnostic Cost Groups	Pope et al. Boston University, Fa. DxCG	Berechnung von Kopfpauschalen für Versicherte und Medicare-Budgets für HMOs	Medicare (2000-2004), USA
DCG/HCC	Diagnostic Cost Groups mit Hierarchical Condition Categories	Pope et al. Boston University, Fa. DxCG	Berechnung von Kopfpauschalen für Versicherte und Medicare-Budgets für HMOs	Medicare (seit 2004), USA
RxGroups		Zhao et al. Boston University, Fa. DxCG	Berechnung von Kopfpauschalen für Versicherte und Medicare-Budgets für HMOs	Medicare (seit 2004), USA
CPDS	Chronic Illness and Disability System	Kronick et al. University of California San Diego	Berechnung von Kopfpauschalen für Versicherte und Medicaid-Budgets für HMOs	Medicaid in Colorado, Oregon, Tennessee, Delaware, New Jersey, Michigan, Utah, Washington, Pennsylvania, Virginia, USA
Medicaid Rx	Medicaid Risk Groups	Gilmer et al.	Berechnung von Kopfpauschalen für Versicherte und Medicaid-Budgets für HMOs	Medicaid in Kalifornien, USA
RxRisk		Fishman et al. Center for Health Studies Group Health Cooperative, Seattle	Berechnung von Kopfpauschalen für Patienten und Budgets für Leistungserbringer der Cooperative	Group Health Cooperative of Puget Sound, Seattle, USA
ACG	Adjusted Clinical Groups	Weiner et al. Johns Hopkins University, Baltimore	USA: Berechnung von Kopfpauschalen für Versicherte und Budgets für HMOs und andere Leistungserbringer Kanada: Erstellung von Morbiditäts- und Kostenprofilen für Hausarztpraxen Spanien: Berechnung von Budgets für primärärztliche Gesundheitszentren Schweden: Bestimmung der Morbidität von hausärztlichen Patienten für Qualitätsmanagement	Medicaid und kommerzielle Versicherer, USA Primary Care in British Columbia, Kanada Baskischer Gesundheitsdienst Osakidetza, Spanien Primärärztlicher Bereich, Schweden
PCG	Pharmacy Cost Groups	Lamers et al. Universität Rotterdam	Risikostruktur- ausgleich	Niederlande

Seit dem Jahr 2000 benutzte die CMS für sein Programm „Medicare+Choice“ zunächst die auf stationären Diagnosen basierenden Diagnostic Cost Groups (PIP-DCGs), seit 2004 mit der Verfügbarkeit von ambulanten Diagnosen und Medikamentenverordnungen inzwischen das DCG/HCC-Modell und RxGroups, um risikoadjustierte Pro-Kopf-Zahlungen für seine Mitglieder festzulegen und daraus die Höhe von Medicare-Budgets für HMOs zu berechnen (ZHAO et al. 2005). Auch das U.S. Department of Veterans Affairs sowie viele kommerzielle Versicherungsunternehmen und Arbeitgeber, z.B. Kaiser Permanente und die Pacific Business Group on Health, verwenden dieses oder andere Modelle der Firma DxCG. Innerhalb der Medicaid-Programme wird derzeit von zehn Staaten (Colorado, Oregon, Tennessee, Delaware, New Jersey, Michigan, Utah, Washington, Pennsylvania und Virginia) das an der University of California, San Diego entwickelte Chronic Illness und Disability Payment System (CDPS) (KRONICK et al. 1996) für eine Risikoadjustierung benutzt, während in Maryland und Minnesota zu diesem Zweck das ACG-System eingesetzt wird (WEINER et al. 1998). In Kalifornien wird außerdem ein speziell für Medicaid entwickeltes, auf Medikamentenverordnungen basierendes Modell (Medicaid Rx) erprobt (GILMER et al. 2001). Ein ähnliches System, RxRisk (FISHMAN et al. 2003), verwendet die Group Health Cooperative of Puget Sound für die Vergütung ihrer Leistungserbringer (RESCHKE et al. 2004). Das ACG-Modell wird neben seinem Einsatz als Risikoadjustierungssystem innerhalb der Medicaid-Programme in einigen Staaten in den USA auch von kommerziellen Versicherungsanbietern verwendet und hat international die weiteste Verbreitung gefunden. Nach mehreren Validierungsstudien (REID et al. 2002, REID et al. 2001, VERHULST et al. 2001, REID et al. 1999) wird es derzeit in **Kanada** in der Provinz British Columbia dazu benutzt, Morbiditäts- und Kostenprofile für allgemeinmedizinische Praxen zu erstellen. Weitere Provinzen evaluieren das System im Moment (ABRAMS 2001).

In Europa hat der baskische nationale Gesundheitsdienst Osakidetza das ACG-Modell nach gründlicher Prüfung (ORUETA et al. 1999, JUNCOSA et al. 1999) in **Spanien** erstmals eingeführt, um die Allokation von Ressourcen für primärärztliche Gesundheitszentren morbiditätsgerecht zu gestalten. In **Schweden** wurde die Übertragbarkeit des ACG-Systems auf das dortige primärärztliche System untersucht (CARLSSON et al. 2002). Dabei stand jedoch weniger die Frage der Ressourcenallokation als die Bestimmung der Krankheitslast zur Identifikation von sinnvollen Qualitätsmanagementmaßnahmen im primärärztlichen Bereich im Mittelpunkt.

Schweden und die übrigen skandinavischen Länder haben nationale Gesundheitssysteme, die durch Steuereinnahmen finanziert werden. Regionale Gesundheitsbehörden sind für die Erhebung der Einkommenssteuer und die Organisation der Gesundheitsversorgung verantwortlich. Früher wurden diese regional erhobenen Steuergelder direkt an die Leistungserbringer in derselben Region verteilt, und zwar in Höhe der in einem vorangegangenen Zeitraum verursachten Kosten unter Berücksichtigung der Inflation. Wie viele andere Länder sah sich jedoch auch das schwedische Gesundheitssystem mit Effizienzproblemen, steigenden Kosten und in der Folge mit einer zunehmenden Steuerlast für alle Bürger konfrontiert. Gesundheitsreformen in den neunziger Jahren sollten auf regionaler und lokaler Ebene Wettbewerb etablieren, z.B. durch die Möglichkeit der regionalen Gesundheitsbehörden, auf lokaler Ebene Verträge mit einzelnen Leistungserbringern abzuschließen. Parallel zu dieser Entwicklung wurden immer mehr Krankenhäuser, Primärversorger und andere Leistungserbringer privatisiert, die inzwischen als Mitbewerber um Verträge in Konkurrenz zu den öffentlich finanzierten Gesundheitszentren treten und für eine größere Wahlfreiheit der Bürger sorgen. In diesem Zusammenhang ergab sich die Notwendigkeit der Einführung eines differenzierten Risikoadjustierungssystems, mit dessen Hilfe die lokalen Budgets so berechnet werden können, dass sie möglichst genau dem tatsächlich vorhandenen Bedarf an Gesundheitsleistungen entsprechen. Da auf lokaler Ebene zum Teil nur geringe Patientenzahlen vorlagen, wäre ein statistisches Modell, das auf aggregierten Daten basiert, zu instabil gewesen. Daher wurde ein kompliziertes Matrixsystem zunächst in der Hauptstadt Stockholm erprobt und später landesweit eingeführt, das demographische und sozioökonomische Indikatoren (Alter, Familienstand, Wohnsituation, Beruf, Beschäftigungsstatus) auf individueller Ebene zur Adjustierung benutzt und auf tatsächlichen Kosten basiert (ANDERSSON et al. 2000, DIDERICHSEN et al. 1997). Dadurch soll der höhere Ressourcenverbrauch von Patienten mit niedrigem sozioökonomischem Status berücksichtigt werden, so dass in dieser Hinsicht schlechter gestellte Stadtteile einen größeren Budgetanteil bekommen.

**Großbritannien** hat ebenfalls ein zentral gesteuertes nationales Gesundheitssystem, das durch Steuergelder finanziert wird. Hier nahm jedoch die Regierung die Steuerung des Gesundheitswesens und die Verteilung der Gelder auf die einzelnen Regionen vor. Da diesem System immer wieder Ineffizienz und Ungerechtigkeit beim Zugang von Bürgern zu Gesundheitsleistungen vorgeworfen wurde, hat Großbritannien seit Beginn der neunziger

Jahre Maßnahmen für eine Dezentralisierung der Gesundheitsversorgung eingeleitet (RICE und SMITH 2001). In dem Maße, wie die finanzielle Verantwortung auf lokale Behörden übertragen wurde, wuchs der Bedarf an differenzierten Risikoadjustierungssystemen. So war ursprünglich die standardisierte Mortalitätsrate (SMR) in den verschiedenen Regionen der Parameter der Wahl, um die Allokation von Ressourcen zu steuern. Seit 1995 wird ein von Wissenschaftlern der Universität York entwickeltes, auf aggregierten Daten der lokalen Ebene beruhendes Risikostratifizierungsmodell verwendet (CARR-HILL et al. 1994), das neben SMR, Alter und Geschlecht auch eine Art standardisierter Morbiditätsrate und verschiedene sozioökonomische Faktoren einbezieht (DIDERICHSEN et al. 1997). Inzwischen wurde immer wieder Kritik am derzeitigen System laut, insbesondere in Bezug auf die ausschließliche Nutzung aggregierter Daten (ASTHANA et al. 2004), so dass die Einbeziehung von individuellen Indikatoren des Gesundheitsstatus und die Möglichkeit der Einführung eines Patientenklassifikationssystems geprüft wird (MAJEED et al. 2001a, MAJEED et al. 2001b).

In den **Niederlanden** wurde auf der Basis des auf Medikamentenverordnungen basierenden Chronic Disease Scores (VON KORFF et al. 1992) ein eigenes Klassifikationssystem, die Pharmacy Cost Groups (PCG), entwickelt und im Jahr 2002 implementiert (LAMERS und VAN VLIET 2004, LAMERS 1999a, LAMERS 1998). In der **Schweiz** wurde die Einführung eines prospektiven Systems vorgeschlagen, bei dem entweder das Vorliegen eines stationären Aufenthaltes im Vorjahr eine höhere Pauschale im Folgejahr bedingt, oder in dem stationäre Diagnosen als Risikofaktoren zusätzlich zu den derzeitigen Indikatoren Alter, Geschlecht und Region herangezogen werden (BECK et al. 2003). Die entsprechenden Studienergebnisse wurden bisher jedoch nicht umgesetzt, eine Adjustierung im Rahmen des Risikostrukturausgleichs erfolgt weiterhin nach Alter, Geschlecht und Region. In **Belgien** existiert ein kompliziertes Risikoadjustierungssystem, mit dessen Hilfe aus dem Gesamtbudget für Gesundheitsausgaben vom nationalen Institut für die Kranken- und Invaliditätsversicherung (RIZIV/INAMI) prospektive Teilbudgets für die sechs großen Krankenversicherungen berechnet werden. Dazu werden die Versicherten aller Krankenkassen zunächst nach soziodemographischen Kriterien in eine Gruppe der Angestellten und der Selbständigen geteilt. Innerhalb dieser Gruppen erfolgt eine weitere Unterteilung in Beschäftigte („active population“), Behinderte („invalids“), Rentner, Witwen und Waisen. Jede dieser Untergruppen wird wiederum nach weiteren Kriterien, wie z.B. Alter, Mortalität, Einkommen, Zahl abhängiger Personen im Haushalt und Wohnsituation

unterschieden. Mittels eines Regressionsmodells können die erwarteten Kosten für jede Gruppe berechnet werden. Anhand der Zahl der Versicherten, die jede Krankenkasse in jeder dieser Gruppen aufweist, wird den Versicherungen dann ein risikoadjustiertes Budget zugewiesen, das allerdings nur einen Teil der Ausgaben decken soll und durch ein retrospektiv festzulegendes, an den tatsächlichen Leistungsausgaben der Krankenkassen orientiertes Budget ergänzt wird. Derzeit werden nur aggregierte Daten als Indikatoren benutzt. Es werden jedoch repräsentative Studien mit individuellen Daten durchgeführt, um das Risikoadjustierungssystem weiter zu entwickeln und eine bessere Vorhersagekraft zu erreichen. (SCHOKKAERT und VAN DE VOORDE 2005, SCHOKKAERT und VAN DE VOORDE 2003, SCHOKKAERT et al. 1998).

Außerhalb Europas hat auch **Israel** ein nach europäischem Muster organisiertes Krankenversicherungssystem mit einem Risikostrukturausgleich. Eine Risikoadjustierung findet derzeit nur nach Alter statt; die Einführung von geographischen Indikatoren und Gesundheitsstatus als zusätzlichen Risikofaktoren wird jedoch diskutiert (SHMUELI et al. 2003).

### **2.1.6.3 Anwendung risikoadjustierter Vergütungs- und Kostenausgleichssysteme in Deutschland**

#### ***2.1.6.3.1 Risikoadjustierter Finanzausgleich in der gesetzlichen***

##### ***Krankenversicherung: Risikostrukturausgleich und morbiditätsorientierter RSA (Morbi-RSA)***

Ein Beispiel für den Einsatz eines Risikoadjustierungssystems in Deutschland stellt der Risikostrukturausgleich (RSA) der gesetzlichen Krankenkassen dar. Der RSA wurde gleichzeitig mit der Möglichkeit der freien Krankenkassenwahl für Versicherte eingeführt und dient dem Finanzausgleich unter den Krankenkassen. Um Kassenarten mit einem hohen Anteil an einkommenschwachen Versicherten, Rentnern und beitragsfrei mitversicherten Familienangehörigen im Wettbewerb nicht zu benachteiligen, werden alle gesetzlich Versicherten nach folgenden Kriterien klassifiziert:

- Alter
- Geschlecht
- Erwerbsunfähigkeitsstatus
- Krankengeldanspruch
- Teilnahme an einem Disease Management-Programm

Alle GKV-Versicherten werden anhand dieser Kriterien stratifiziert und für jede Gruppe die durchschnittlich zu erwartenden Kosten als so genannte standardisierte Leistungsausgaben berechnet. Auf diese Weise können für jede Kassenart die aufgrund ihrer Versichertenklientel erwarteten Ausgaben prognostiziert und mit den realen Kosten verglichen werden. Kassen, die aufgrund ihrer günstigen Versichertenstruktur mit ihren beobachteten Ausgaben unter den standardisierte Leistungsausgaben liegen, müssen die Differenz als Ausgleichszahlung an Kassen leisten, die überdurchschnittliche Kosten zu verzeichnen haben. Auf diese Weise sollen Anreize zur Risikoselektion junger, gesunder Versicherter für die Krankenkassen vermindert und ein Wettbewerb zwischen den Kassen gefördert werden, der kosteneffizientes Handeln belohnt. Leider hat sich gezeigt, dass das Instrument des RSA in seiner derzeitigen Form nicht ausreicht, um unter den Kassen für vergleichbare Beiträge zu sorgen, da die Morbidität der Versicherten im RSA bisher nicht berücksichtigt wird. Daher trat zum 01.01.2002 ein Gesetz zur Neuregelung des Risikostrukturausgleichs in Kraft, das eine schrittweise Umstrukturierung des RSA in Richtung einer direkten Umverteilung nach Morbiditätskriterien bis zum Jahr 2007 vorsieht. Um Versicherte gemäß ihrer Morbidität stratifizieren zu können, muss ein entsprechendes Patientenklassifikationssystem eingeführt werden. Inzwischen wurde der Bundesregierung ein Gutachten vorgelegt, das den Einsatz eines Systems empfiehlt, welches die krankheitsbedingten Ausgaben der Versicherten anhand von Informationen über ihre Medikamentenverordnungen und stationäre Diagnosen prognostiziert (LAUTERBACH und LÜNGEN 2005, RESCHKE et al. 2004). Es handelt sich um ein Produkt aus der in Boston inzwischen von der Firma DxCG vertriebenen DCG-Modellfamilie, den RxGroups+Inpatient-HCC.

#### ***2.1.6.3.2 Risikoadjustierte Fallpauschalen in der stationären Vergütung: Diagnosis Related Groups (DRGs)***

Das System der Diagnosis Related Groups (DRG) wurde im Jahr 1983 in den USA erstmalig für Mitglieder der staatlichen Medicare-Organisation (HCFA) für die Vergütung von Krankenhäusern eingeführt. Es handelt sich um ein Klassifikationssystem, das stationäre Behandlungsfälle nach Alter, Geschlecht, Gesundheitsstatus und durchgeführten Prozeduren in möglichst homogene Gruppen (Diagnosis Related Groups) stratifiziert, für die dann eine angemessene Vergütungspauschale festgelegt werden kann (VLADECK 1984). Seit seiner Einführung hat das DRG-System internationale Verbreitung gefunden. Die ursprüngliche amerikanische Version wurde weiterentwickelt, und es entstanden länderspezifische Formen, wie z.B. das australische AR(Australian Refined)-DRG-System, das für die deutschen G(German Refined)-DRGs modifiziert wurde.

Auch in Deutschland werden seit Beendigung der Übergangsphase im Jahr 2004 sämtliche Krankenhäuser mittels DRG vergütet. Mit Hilfe des Systems soll der Behandlungsaufwand bzw. Ressourcenverbrauch jedes Patienten sowohl im operativen als auch im nicht-operativen Bereich möglichst vollständig abgebildet und monetär bewertet werden. Dazu stratifiziert ein Computerprogramm, ein so genannter „Grouper“, die Behandlungsfälle einer Klinik retrospektiv anhand von

- durchgeführter Prozedur
- Behandlungsdiagnose
- Alter und Geschlecht des Patienten
- Evtl. gleichzeitig vorliegenden Komorbiditäten
- Komplikationen der Therapie und Beatmungsdauer
- Geburtsgewicht (im Bereich der Geburtshilfe).

Jedem Behandlungsfall wird eine bestimmte DRG zugewiesen, die mit einem relativen Gewicht bewertet ist. Für die erstmalige Berechnung dieser Relativgewichte wurden Behandlungsleistungen und -kosten in einer repräsentativen Stichprobe deutscher Kliniken fallbezogen dokumentiert und vom Institut für das Entgeltsystem im Krankenhaus (*InEK GmbH*; gegründet von den Spitzenverbänden der Krankenkassen, dem Verband der Privaten Krankenversicherung und der Deutsche Krankenhausgesellschaft) ausgewertet. Zur weiteren Pflege des DRG-Systems übermitteln die teilnehmenden Krankenhäuser fallbezogene Daten aller DRG-relevanten Fälle kontinuierlich an das Institut, so dass die Gültigkeit der Relativgewichte jährlich überprüft und gegebenenfalls angepasst werden kann. Mittelfristig sollen alle Krankenhäuser in die Datenlieferung eingeschlossen werden.

Um den Fallerlös für das Krankenhaus zu berechnen, wird das Relativgewicht für einen Behandlungsfall mit dem so genannten Basiswert (base rate) multipliziert. Zunächst erfolgte die Berechnung des Basisfallwertes für jedes Krankenhaus noch individuell. Dabei wurde in Verhandlungen zwischen Krankenhaus und Kostenträgern ein Budget für das Krankenhaus (Gesamtbetrag) sowie eine Leistungsplanung (Aufstellung der geplanten DRGs) vereinbart. Der Basisfallwert berechnete sich aus dem Gesamtbetrag dividiert durch die Summe der vereinbarten Bewertungsrelationen der DRGs. Seit dem 01.01.2005 soll dieses Budgetsystem in einer Konvergenzphase schrittweise von einem leistungsorientierten Preissystem abgelöst werden, bei dem bis zum Jahr 2009 ein einheitlicher Basisfallwert für alle Krankenhäuser eines Bundeslandes und somit ein einheitlicher Preis für gleiche Leistungen gilt.

### **2.1.6.3 Risikoadjustierte Vergütung im ambulanten Bereich:**

#### ***Morbiditätsorientierte Regelleistungsvolumina (Morbi-RLV)***

Die vertragsärztliche Vergütung im ambulanten Bereich soll in Deutschland zukünftig grundlegend umgestaltet werden. Ab dem Jahr 2007 sollen anstelle der bisherigen Gesamtvergütung regional zu berechnende und in Verträgen zwischen KVen und Krankenkassen festzulegende Leistungsmengen (so genannte morbiditätsorientierte Regelleistungsvolumina) treten, deren Höhe die Morbidität von Versicherten der verschiedenen Krankenkassenarten widerspiegelt. Auf diese Weise soll das Morbiditätsrisiko zurück zu den Krankenkassen verlagert und strukturelle Bedingungen geschaffen werden, unter denen „das Geld den Versicherten folgt“ (VON STILLFRIED 2003, S.9). Im Rahmen des Gesundheitssystemmodernisierungsgesetzes (GMG) wurden im Jahr 2004 § 85 a und b in das Sozialgesetzbuch V eingefügt, die genaue Vorgaben für das Verfahren zur Ermittlung der Morbidität von Versicherten enthalten. So ist in § 85 a Absatz 5 Satz 2 festgelegt, dass „zur Bestimmung der Morbiditätsstruktur [...] diagnosebezogene Risikoklassen für Versicherte mit vergleichbarem Behandlungsbedarf nach einem international anerkannten, zur Anwendung in der vertragsärztlichen Versorgung geeigneten Klassifikationsverfahren [gebildet werden sollen]; Grundlage hierfür sind die vertragsärztlichen Behandlungsdiagnosen gemäß § 295 Abs. 1 Satz 2“ (Sozialgesetzbuch Fünftes Buch - Gesetzliche Krankenversicherung 2004). Im Rahmen des Gutachtens zur Einführung des Morbi-RSA wurde eine ausführliche Übersicht über die meisten der existierenden Klassifikationsverfahren erstellt, inklusive der jeweiligen Vor- und Nachteile, datentechnischen Voraussetzungen und der Frage der Übertragbarkeit auf deutsche Verhältnisse. Da im SGB V ambulante Behandlungsdiagnosen als Indikatoren für Morbidität festgeschrieben wurden, ließ sich der Kreis der für die ambulante Vergütung in Frage kommenden Systeme deutlich einengen. Breitere Anwendungserfahrungen auch in anderen Ländern als den USA haben Adjusted Clinical Groups (ACG) und die DCG-Modellfamilie. International bisher keine Verwendung finden das Chronic Illness und Disability Payment System (CDPS), die Clinical Risk Groups (CRG) der Firma 3M (HUGHES et al. 2004), das von der RAND Corporation entwickelte Clinically Detailed Risk Information System for Cost (CD-RISC) (KAPUR et al. 2003) und die zeitlich auf Krankheitsepisoden basierenden Episode Risk Groups (ERG) der Firma Symmetry Health Data Systems (GOLDFIELD et al. 1999). Im Dezember 2005 beschloss der Erweiterte Bewertungsausschuss von KBV und Spitzenverbänden der Krankenkassen, zukünftig das HCC-System der Firma DxCG für die Berechnung morbiditätsadjustierter RLV zu verwenden (VON STILLFRIED 2006).

### 2.1.7 Validität von Risikoadjustierungssystemen

Soll ein Risikoadjustierungssystem seinen Zweck erfüllen, sinnvolle Vergleiche von Leistungserbringern mit unterschiedlichen Ausgangsbedingungen zu ermöglichen, ist eine ausreichende Validität der mit Hilfe des Modells ermittelten Ergebnisse von zentraler Bedeutung. Die Validität risikoadjustierter Ergebnisse wurde in zahlreichen internationalen Studien an unterschiedlichen Populationen untersucht (z.B. ZHAO et al. 2005, PETERSEN et al. 2005, IEZZONI 2003, CUMMING et al. 2002, CARLSSON et al. 2002, REID et al. 2002, ROSEN et al. 2001, ASH et al. 2001, ASH und SHWARTZ 1999, JUNCOSA et al. 1999, REID et al. 1999, LAMERS 1999b, ASH und BYRNE-LOGAN 1998, POPE et al. 1998, DUNN et al. 1996, FOWLES et al. 1996). Im Vorfeld der Einführung von Morbi-RSA und Morbi-RLV wurden auch in Deutschland einige Studien mit verschiedenen Patientenklassifikationsmodellen durchgeführt (DRÄTHER et al. 2006, VON MAYDELL und SCHULTE 2005, RESCHKE et al. 2004, VON STILLFRIED und RYLL 2004, STAFFELDT 2004). In den meisten Untersuchungen stand dabei die Analyse der Vorhersagekraft eines Modells in Bezug auf das Outcome im Vordergrund.

Neben dieser Dimension der prognostischen Validität eines Risikoadjustierungssystems existieren jedoch noch andere relevante Validitätsdimensionen (IEZZONI 2003), die meist weniger Beachtung finden:

- Augenscheinliche Validität („Face validity“): Scheint das Risikoadjustierungsmodell die medizinisch relevanten Faktoren korrekt zu messen? (z.B. entspricht das Ergebnis dem klinischen Zustand des Patienten?)
- Inhaltsvalidität („Content validity“): Der Inhalt der Test-Items erfasst das zu messende Konstrukt in seiner wichtigsten Aspekte erschöpfend (z.B. wurden alle relevanten Risikofaktoren in des Modell einbezogen? Fehlen für den zu messenden Zusammenhang wichtige Faktoren?)
- Konstruktvalidität („Construct validity“): Aus dem zu messenden Zielkonstrukt können Hypothesen abgeleitet werden, die anhand der Testwerte bestätigt werden können. Testwerte fallen so aus, wie es die aus der Theorie abgeleiteten Hypothesen vorgeben (z.B. sind die empirisch determinierten Risikofaktoren mit denjenigen Faktoren identisch, die ein Expertenpanel für dieses Modell ausgewählt hätte?)
- Prognostische Validität („Predictive validity“): Der Testwert sagt das spätere Verhalten korrekt vorher (z.B. wie gut kann das Outcome mit Hilfe des Modells erklärt/vorhergesagt werden? Wie hoch ist der statistische Erklärungswert ( $R^2$ )?)

- Kausale Validität („Attributional validity“): Inwieweit kann das Outcome tatsächlich der erfolgten Behandlungsleistung zugeschrieben werden? (z.B. inwieweit repräsentieren risikoadjustierte Ergebnisse die „wahren“ Kosten bzw. die „wahre“ Qualität und wie hoch ist das „Residuum“?)

Bezüglich der Dimension der Inhaltsvalidität werden Risikofaktoren für ein Risikoadjustierungsmodell beispielsweise nicht nur nach dem vermuteten Einfluss eines Faktors auf das interessierende Outcome, sondern vor allem in Abhängigkeit der zur Verfügung stehenden Daten ausgewählt. So entschieden sich die CMS im Jahr 2000 z.B. hauptsächlich deshalb für ein auf stationären Diagnosen basierendes Modell, weil dies zu diesem Zeitpunkt die einzigen potentiellen Risikofaktoren waren, die für ihre Versicherten relativ leicht verfügbar waren (KLABUNDE et al. 2002). Die Auswahl von Risikoindikatoren sollte jedoch in erster Linie nach dem medizinisch nachvollziehbaren Zusammenhang zwischen Risikofaktor und Outcome erfolgen. Andernfalls werden Aussagekraft und Validität der Ergebnisse eines Risikoadjustierungssystems auch bei guter statistischer Vorhersagekraft (z.B. hoher  $R^2$ -Wert) begrenzt sein (IEZZONI 2003).

Eine weitere entscheidende Voraussetzung für die Validität risikoadjustierter Ergebnisse über die Erfüllung der genannten Kriterien hinaus ist die Validität der Daten, die den ausgewählten Risikofaktoren zugrunde liegen. Mit anderen Worten, Funktion und Prognosegüte von Risikoadjustierungsmodellen hängen maßgeblich davon ab, ob die für die statistischen Berechnungen herangezogenen Werte für die einzelnen Risikofaktoren ausreichend korrekt und vollständig sind

### **Validität für Risikoadjustierungssysteme verwendeter Daten**

Untersuchungen zur Validität der einem Risikoadjustierungssystem zugrunde liegenden Daten wurden bisher weitaus seltener durchgeführt als Studien zur Vorhersagekraft der Modelle an sich (WELCH 2002). Solche Analysen sind jedoch wichtig, um eine Aussage darüber treffen zu können, ob Unterschiede im Outcome bei verschiedenen Leistungserbringen, wie z.B. Mortalitätsraten oder Höhe der Kosten, möglicherweise auf Unterschiede in der Qualität der zur Risikoadjustierung verwendeten Behandlungsdaten zurückzuführen sind. Beispielsweise haben verschiedene Studien mit dem ACG-System eine Variabilität der Vorhersagekraft unter ähnlichen Studienbedingungen gezeigt, die nicht erklärt werden konnte (WILCHESKY et al. 2004). Im folgenden Abschnitt soll daher die Validität von Daten analysiert werden, die üblicherweise als Risikofaktoren in Risikoadjustierungsmodellen eingesetzt werden.

Dazu werden zunächst mögliche Datenquellen und ihre Vor- und Nachteile vorgestellt. Die meisten Risikoadjustierungssysteme basieren auf administrativen Daten aus dem ambulanten und stationären Bereich. In Deutschland werden solche Daten zurzeit im Rahmen des DRG-Systems benutzt; zukünftig sollen sie auch die Grundlage für eine Berechnung von Morbi-RSA und Morbi-RLV bilden. Daher wird anschließend ein Überblick über die internationale Studienlage zur Validität administrativer Daten gegeben. Den internationalen Ergebnissen werden die wenigen deutschen Studien zu diesem Thema gegenüber gestellt. In einer anschließenden Zusammenfassung werden die Gemeinsamkeiten und Unterschiede der internationalen und deutschen Forschungsergebnisse und ihre möglichen Ursachen diskutiert.

### ***2.1.8 Datenquellen für Risikofaktoren***

Eine de novo Erhebung von Kosten- und Behandlungsdaten ist teuer, weshalb sowohl bei der Entwicklung als auch bei der Anwendung eines Risikoadjustierungsmodells fast immer zu anderen Zwecken erhobene Daten genutzt werden. Im wesentlichen stehen folgende Datenquellen zur Verfügung (IEZZONI 2003):

- Administrative Daten (Abrechnungsdaten aus dem ambulanten und stationären Bereich, Arzneimittelverordnungen, etc.)
- Daten aus Patientenakten (stationär und ambulant)
- Daten aus Patientenbefragungen

**Administrative Daten** haben den Vorteil, dass sie im Allgemeinen in codierter Form maschinenlesbar zur Verfügung stehen (z.B. ICD-9- bzw. ICD-10-codierte Diagnosen, Medikamentenverordnungen als ATC-Codes, Prozedurencodes für Operationen) und daher auch in großen Mengen schnell und einfach per EDV zu verarbeiten sind. Andererseits werden sie überwiegend zu Abrechnungszwecken erhoben, so dass an der Validität des klinischen Inhalts immer wieder Zweifel geäußert werden. Trotz dieser Kritik überwiegen für die meisten Kostenträger die Vorteile, und inzwischen werden sowohl im Bereich der Vergütung und Kostenevaluation als auch im Bereich der Qualitätssicherung in den meisten Fällen administrative Daten zur Risikoadjustierung benutzt.

**Patientenakten** sind Datenquellen, die wesentlich mehr Informationen enthalten als administrative Daten, und die das Krankheitsgeschehen eines Patienten umfassend und medizinisch plausibel darstellen können. Allerdings müssen die interessierenden Risikofaktoren meist mühsam von Hand extrahiert und in eine statistisch auswertbare Form gebracht werden. Mit der zunehmenden Nutzung von EDV-basierten

Dokumentationssystemen in Kliniken und Praxen steigen die Möglichkeiten, die gewünschten Patientendaten automatisiert in größeren Mengen aus den Akten zusammenzustellen. Ein Problem in diesem Zusammenhang stellt jedoch das Fehlen einer standardisierten Dokumentationsform dar (JOLLIS et al. 1993). Sowohl in eine „Papierakte“ als auch in die elektronische Patientendokumentation trägt jeder Arzt ein, was ihm individuell wichtig erscheint. Dabei existiert kein allgemein konsentierter Inhalt und auch die Reihenfolge der Einträge bzw. die in der EDV benutzten Felder und ihre Bezeichnungen wechseln je nach Computersystem und Vorlieben des Praxisinhabers oder Krankenhauses. Darüber hinaus werden Papier- und EDV-Patientenakte oft parallel verwendet, wobei auch hier individuell bzw. klinikspezifisch sehr unterschiedlich ist, welche Informationen in welchem Aktenteil dokumentiert werden (MIKKELSEN und AASLY 2001).

**Daten aus Patientenbefragungen** werden seltener im Rahmen eines Risikoadjustierungsmodells benutzt. Aufgrund der Subjektivität der Ergebnisse und potentieller Probleme bei der Datenerhebung, wie z.B. Recall-Bias oder Non-Responder-Bias, ist der Einsatz dieser Datenquellen bisher limitiert. Am häufigsten wird der vom Patienten selbst bewertete Gesundheitszustand („Self-reported health status“) als Risikofaktor einbezogen (IEZZONI 2003).

### ***2.1.9 Validität von Abrechnungsdaten - Internationaler Stand der Forschung***

Die meisten der vorliegenden Studien stammen aus dem stationären Bereich, während der ambulante Bereich eher unterrepräsentiert ist (WILCHESKY et al. 2004, KLABUNDE et al. 2002, WELCH 2002, FOWLES et al. 1995). Einige Untersuchungen analysieren die potentiellen Folgen einer mangelnden Validität von administrativen Daten in Bezug auf die Ergebnisse einer Risikoadjustierung (z.B. Veränderungen der Eingruppierung in eine bestimmte DRG oder Unterschiede in der 30-Tage-Mortalität). Im ambulanten Bereich steht überwiegend die Frage der Validität von Diagnosen im Mittelpunkt, die von Leistungserbringern im Rahmen der Abrechnung an die Kostenträger übermittelt wurden. Dabei wird untersucht, ob ein Patient an einer bestimmten Erkrankung leidet oder nicht.

Methodisch wird die Validität von Abrechnungsdaten geprüft, indem in einem bestimmten Zeitraum zu Abrechnungszwecken erhobene Patienteninformationen mit Angaben aus anderen Datenquellen verglichen werden. In einigen Studien dienen Angaben der Patienten zu ihren Erkrankungen als Goldstandard (z.B. RECTOR et al. 2004), meist werden jedoch klinische Befunde in der Krankenakte als Vergleich benutzt. In einer extrem aufwendigen

Untersuchung wurde die Validität beider Verfahren gleichzeitig analysiert. Dazu beobachteten die Untersucher konsekutive Konsultationen von insgesamt ca. 4500 Patienten bei 138 verschiedenen Hausärzten und verglichen den Goldstandard „Beobachtung“ mit der ärztlichen Dokumentation in den Patientenakten sowie mit dem Ergebnis einer schriftlichen Befragung der Patienten über die Konsultation (STANGE et al. 1998).

IEZZONI (1997) empfiehlt folgende Dimensionen für die Untersuchung der Validität von Abrechnungsdiagnosen:

- Klinischer Inhalt (z.B. inwieweit bilden Diagnosencodes den klinischen Zustand des Patienten ab?),
- Korrektheit der Kodierung (z.B. hat der Patient wirklich die Erkrankungen, für die Diagnosencodes angegeben sind?),
- Vollständigkeit der vorliegenden Diagnosencodes (z.B. wurden alle Erkrankungen verschlüsselt, die der Patient hat?).

Als Messgrößen werden die Kriterien für die Beurteilung der Güte diagnostischer Tests benutzt:

- Sensitivität und Spezifität,
- positiv prädiktiver Wert (PPW) und negativ prädiktiver Wert (NPW),
- Kappa-Wert.

Dabei stellt die Sensitivität ein Maß für die Vollständigkeit von Abrechnungsdaten dar, während der PPW ihre Korrektheit quantifiziert. In diesem Zusammenhang bedeutet eine hohe *Sensitivität*, dass eine Diagnose oder Prozedur, die in der Patientenakte dokumentiert wurde, in der Mehrzahl der Fälle auch in den Abrechnungsunterlagen aufgeführt ist. Eine hohe *Spezifität* bedeutet, dass eine in den Akten nicht dokumentierte Diagnose oder Prozedur mit großer Wahrscheinlichkeit auch nicht in den Abrechnungsdaten erscheint. Der *positiv prädiktive Wert* beantwortet die Frage, wie viele der in den Abrechnungsdaten aufgeführten Diagnosen, Prozeduren, etc. sich auch in den Patientenakten wieder finden. Der *negativ prädiktive Wert* beschreibt, in welchen Maße davon ausgegangen werden kann, dass das Fehlen einer Diagnose oder Prozedur in den Abrechnungsunterlagen bedeutet, dass ein Patient diese Erkrankung tatsächlich nicht hat bzw. die Prozedur nicht durchgeführt wurde, soweit dies aus den Akten hervorgeht. Der *Kappa-Koeffizient* (COHEN 1960) ist ein Maß für das Verhältnis zwischen der Anzahl der Fälle, in denen Daten in Patientenakten und Abrechnungsdaten übereinstimmen und der Anzahl der Fälle, in denen sie übereinstimmen

könnten, wenn man die zufällige Übereinstimmung berücksichtigt. Bei einem Vergleich von Behandlungsanlässen und Abrechnungsdiagnosen ist der Kappa-Wert künstlich erhöht, weil es viele Patienten gibt, die nicht unter einer bestimmten Erkrankung leiden und diese Diagnose daher weder in der Akte noch in der Abrechnung aufweisen (FOWLES et al. 1995). Der Koeffizient kann einen theoretischen Wert von 1 bis -1 annehmen. Nach LANDIS und KOCH (1977) zeigt ein Wert unter 0 eine schlechte Übereinstimmung an („poor agreement“). Werte bis 0,20 werden als geringe Übereinstimmung („slight agreement“), Werte zwischen 0,21 und 0,40 als mäßige Übereinstimmung („fair agreement“), Werte zwischen 0,41 und 0,60 als mittlere Übereinstimmung („moderate agreement“), Werte zwischen 0,61 und 0,80 als gute Übereinstimmung („substantial agreement“) und Werte ab 0,81 als sehr gute Übereinstimmung („almost perfect agreement“) jenseits des Zufalls eingeschätzt.

### **2.1.9.1 Datenvalidität im stationären Bereich**

In den vorliegenden Studien aus dem stationären Bereich wird im Rahmen der Qualitätssicherung die Validität der Angaben zu Hauptdiagnosen und Begleiterkrankungen als Risikofaktoren für die 30-Tage-Mortalität untersucht. Die meisten Ergebnisse zeigen, dass für eine Risikoadjustierung wichtige Begleiterkrankungen (z.B. die im Charlson Comorbidity Index enthaltenen Komorbiditäten) in administrativen Daten seltener dokumentiert sind als in den Patientenakten (KLABUNDE et al. 2002, KIESZAK et al. 1999, MALENKA et al. 1994), was Auswirkungen auf die Berechnung der 30-Tage-Mortalität haben kann. In einer Studie zum Vorliegen von prognostisch ungünstigen Faktoren bei Herzkatheterpatienten konnte beispielsweise in weniger als der Hälfte der untersuchten Fälle in den Abrechnungsdaten eine Angabe über solche Faktoren gefunden werden (JOLLIS et al. 1993). Eine kanadische Studie verglich die Dokumentation von Hauptdiagnose und Begleiterkrankungen von Patienten mit Herzinsuffizienz in Patientenakten und Abrechnungsunterlagen und berechnete Auswirkungen auf die 30-Tage-Mortalität. Die administrativen Daten zeigten einen hohen PPW für das Vorliegen der Hauptdiagnose. Für Komorbiditäten wurde eine hohe Spezifität und PPW bei geringer Sensitivität sowie eine geringere Prävalenz als in den Akten nachgewiesen. Sensitivität der Abrechnungsdaten und schlechtere Vorhersagewerte für die 30-Tage-Mortalität konnten jedoch durch die Einbeziehung von Daten aus dem Vorjahr verbessert werden (LEE et al. 2005). Zwei andere Untersuchungen aus Kanada (QUAN et al. 2002, HUMPHRIES et al. 2000) verglichen ebenfalls die im Charlson Comorbidity Index enthaltenen 17 Begleiterkrankungen in Patientenakten und Abrechnungsunterlagen. Insgesamt wurde die Prävalenz von Begleiterkrankungen anhand der Abrechnungsdaten in Abhängigkeit von der Diagnose in den

genannten Studien unterschätzt. Vor allem asymptomatische Erkrankungen waren in vielen Fällen nicht aufgeführt. Spezifität und NPW zeigten für alle Erkrankungen hohe, Sensitivität und PPW für COPD, Herzinsuffizienz, periphere Gefäßerkrankungen eher niedrige Werte (50%-62%). Hypertonie, Hyperlipidämie und die Angabe einer zerebrovaskulären Erkrankung in der Anamnese zeigten einen hohen PPW (70%-90%) bei niedriger Sensitivität (ca. 60%, 32% und 43%). Für die Diagnose Diabetes mellitus waren die Werte für Sensitivität, Spezifität, PPW, NPW und Kappa gleichermaßen hoch (79%-99%). Die Berechnung des Charlson Comorbidity Index auf Basis von Abrechnungsunterlagen im Vergleich mit den Patientenakten zeigte jedoch keine unterschiedlichen Werte. Im Unterschied zu diesem Ergebnis wurden in zwei amerikanischen Studien risikoadjustierte Mortalitätsraten von kardiochirurgischen und allgemeinchirurgischen Patienten anhand von administrativen und klinischen Daten berechnet. Die Vorhersagekraft des auf administrativen Daten basierenden Risikoadjustierungsmodells war in beiden Studien geringer als die des Modells mit aus der Patientenakte abgeleiteten Risikofaktoren (GORDON et al. 2005, GERACI et al. 2005). KASHNER (1998) untersuchten die Validität der administrativen Datenbasis des Departments of Veterans Affairs (VA) anhand einer nationalen Zufallsstichprobe von 383 stationären und 428 ambulanten Patienten aus 152 VA-Behandlungszentren. Verglichen wurden demographische Variablen, stationäre Haupt- und Nebendiagnosen, Häufigkeit der Inanspruchnahme der verschiedenen stationären und ambulanten Behandlungseinrichtungen im Zeitraum eines Quartals. Die größte Übereinstimmung zeigten demographische Informationen (Kappa 0,92), gefolgt von stationären Haupt- und Nebendiagnosen (Kappa 0,75 bzw. 0,62), während für die Angabe der Inanspruchnahme stationärer und ambulanter Behandlungseinrichtungen niedrigere Werte ermittelt wurden (Kappa 0,50 bzw. 0,46). In dieser Studie war die in den Abrechnungsunterlagen angegebene Prävalenz für einige Diagnosen (z.B. koronare Herzkrankheit (KHK), Hypertonie und Diabetes mellitus) zwischen 19% und 39% höher, als aus den Patientenakten zu entnehmen war.

Im Rahmen der Vergütung und Kostenevaluation wurde in vielen Studien die Validität der für Patienten dokumentierten Diagnosen und Prozeduren im Zusammenhang mit dem DRG-System analysiert. Eine Studie aus der Einführungsphase dieses Systems in US-amerikanischen Krankenhäusern ergab, dass auf der Basis von administrativen Daten ermittelte DRGs in über 50% der Fälle nicht mit der anhand der Patientenakte festgelegten DRG übereinstimmten. Als Ursache wurde unter anderem genannt, dass Hauptdiagnose und

Prozedur falsch dokumentiert wurden, Begleiterkrankungen und Prozeduren bei den Abrechnungsdiagnosen oft nicht angegeben waren und Alter und Geschlecht fehlten oder falsch sein konnten (JOHNSON und APPEL 1984). Im Rahmen der nationalen DRG-Validierungsstudie in den USA wurden im Zeitraum von Oktober 1984 bis März 1985 7050 Patientenakten aus 239 Kliniken analysiert. Dabei zeigte sich eine Rate von 20,8% fehlerhaften DRG-Eingruppierungen, die zu einem Anstieg des Casemix-Index um 1,9% führten (HSIA et al. 1988). In einer weiteren Untersuchung im Jahr 1988 konnte jedoch nachgewiesen werden, dass die Zahl falscher DRGs inzwischen auf 14,8% gesunken und kein Einfluss auf den Fallschweregrad nachweisbar war (HSIA et al. 1992). Da in diesen beiden Studien ausschließlich die resultierenden DRGs verglichen worden waren, analysierten FISHER et al. (1992) anhand des Datenmaterials aus der nationalen DRG-Validierungsstudie die Validität der Kodierung einzelner Diagnosen und Prozeduren. Dabei stellte sich heraus, dass Sensitivität und PPW für die untersuchten Prozeduren- und Diagnosencodes höher waren, wenn es sich um schwerwiegende Eingriffe und Krankheitsbilder handelte. Eine neuere Studie untersuchte die Validität von 35 häufig vorkommenden Prozeduren und kam zu ähnlichen Ergebnissen. Während schwerwiegende operative Eingriffe eine hohe Sensitivität von über 69% hatten, lag sie bei über der Hälfte der kleinen Eingriffe unter 50%, was auf die unterschiedliche Dokumentationsrate zurückgeführt wurde: Die Dokumentation kleiner Eingriffe auf Stationsebene wurde vom Personal eher vergessen (QUAN et al. 2004). Im Vorfeld der Einführung von DRGs in Deutschland fanden LÜNGEN und LAUTERBACH (2001) bei einer systematischen Literaturrecherche in *Medline* in den Jahren 1980 bis 2000 33 Studien aus neun Ländern (USA, Kanada, Australien, Norwegen Schweden, Großbritannien, Frankreich, Italien, Deutschland), die eine quantitative Analyse der Kodierqualität von Diagnosen und Prozeduren im stationären Bereich zum Inhalt hatten. In einer Metaanalyse ermittelten sie eine ungewichtete Fehlerrate über alle Studien von durchschnittlich 23%, die in 18% der Fälle zu einer Änderung der DRG geführt hatte.

Die Analyse der vorliegenden internationalen Studien zur Datenvalidität im stationären Bereich hat gezeigt, dass die Abrechnungsunterlagen bezüglich der angegebenen Hauptdiagnose in hohem Maße korrekt sind, das heißt, ein „Overreporting“ bzw. „Upcoding“ scheint inzwischen selten zu sein. In Bezug auf die Angabe von Begleiterkrankungen weisen die Studienergebnisse durch eine eher niedrige Sensitivität vor allem bei asymptomatischen Erkrankungen auf ein „Underreporting“ hin. Vollständigkeit und Korrektheit von Abrechnungsdiagnosen variieren je nach Erkrankung (z.B. hohe Werte beider Dimensionen

bei Diabetes mellitus; gute Korrektheit aber geringe Vollständigkeit bei Hypertonie und Hyperlipidämie). Je schwerwiegender die Diagnose, desto höher sind Vollständigkeit und Korrektheit der Abrechnungsdaten. Die vorgestellten Ergebnisse legen die Schlussfolgerung nahe, dass Prävalenzangaben vor allem für Krankheiten mit leichterem und phasenweise asymptomatischem Verlauf auf der Basis von Abrechnungsdiagnosen vermutlich zu niedrig geschätzt werden. Die vorliegenden Untersuchungen lassen jedoch keine klare Aussage darüber zu, welche Auswirkungen ein „Underreporting“ bestimmter Diagnosen auf die Klassifizierung von Patientengruppen in einem Risikoadjustierungsmodell hat.

### **2.1.9.2 Datenvalidität im ambulanten Bereich**

Einige Studien zur Validität von administrativen Daten im ambulanten Bereich analysieren relevanten Diagnosen und Prozeduren für ausgewählte Indikationen. So untersuchten KATZ et al. (1997) Sensitivität und PPW für die Dokumentation von verschiedenen rheumatologischen Erkrankungen in Abrechnungsunterlagen und Patientenakten von Rheumatologen. Die Autoren fanden eine hohe Sensitivität und PPW für rheumatoide Arthritis, systemischen Lupus erythematodes (SLE) sowie für Gelenkpunktionen, während die Diagnose Fibromyalgie eine eher niedrige Sensitivität und PPW aufwies. Bei der Diagnose Arthrose waren Sensitivität und PPW umso höher, je unspezifischer die ICD-9-codierten Angaben sein durften (z.B. ohne Angabe der Lokalisation). COOPER et al. (1999) analysierten in einer Studie mit ähnlichem Design die Sensitivität von Diagnosen und Prozeduren aus ambulanten und stationären Abrechnungsunterlagen für die Identifizierung von Krebsfällen bei Medicare-Patienten. Als Goldstandard diente in dieser Untersuchung ein gemeinsames Krebsregister des National Cancer Institutes und der CMS. Es zeigte sich, dass die Sensitivität von Abrechnungsunterlagen für die Identifikation von Krebsfällen durch die Einbeziehung ambulanter Informationen von 77% auf 91% erhöht werden konnte. Außerdem nahm die Sensitivität mit der Länge des Beobachtungszeitraums zu. NEWCOMER et al. (1999) untersuchten in einem ähnlichen Design, wie viele von 5379 in einer Datenbank für Demenzkranke aufgeführten Patienten während eines Ein-Jahreszeitraums auch anhand ihrer Medicare-Abrechnungsdiagnosen als demenzkrank identifiziert worden wären. Der Datenvergleich ergab, dass lediglich 31% der Patienten in einem Jahr die Abrechnungsdiagnose Demenz aufwies. Der Anteil der Patienten mit dieser Abrechnungsdiagnose stieg jedoch mit der Zahl der Beobachtungsjahre an. Das gleiche Muster fand sich auch bei anderen chronischen Erkrankungen. Die Autoren kamen zu dem Schluss, dass die Wahrscheinlichkeit, Erkrankungen anhand von Abrechnungsdiagnosen zu entdecken, entscheidend vom Beobachtungszeitraum abhängt. Je kürzer dieser Zeitraum,

desto mehr unterscheidet sich die ermittelte Prävalenz von der „wahren“ Prävalenz. Als ein zweites Problem in diesem Zusammenhang identifizierten die Autoren die Nutzung von Abrechnungsdaten zur Konzeption und Anwendung von Risikoadjustierungsmodellen: Da hier oft nur Abrechnungsdaten eines Jahres verwendet werden, ist die Gefahr hoch, dass die Prävalenz einer Erkrankung in einer Population unterschätzt wird. Dadurch könnten die durchschnittlichen Kosten für die Behandlung einer Diagnose überschätzt werden, da chronische Erkrankungen nur dann in den Abrechnungsdiagnosen auftauchen, wenn sie der Behandlung bedürfen.

Andere Studien vergleichen die Übereinstimmung von Diagnoseangaben für ein größeres Spektrum chronischer Erkrankungen. So berechneten FOWLES und LAWTHERS (1995) anhand von 1927 Medicare-Patienten aus 91 hausärztlichen Praxen in Maryland, USA, für den Zeitraum eines Jahres die prozentuale Übereinstimmung und Kappa-Werte für 27 verschiedene Diagnosen und durchgeführte Prozeduren in Akten und Abrechnungsunterlagen. Sie stellten fest, dass die meisten der in den Abrechnungsunterlagen aufgeführten Diagnosen auch in den Patientenakten zu finden waren (hoher PPW), andererseits fehlten in den Abrechnungsunterlagen Diagnosen, die aus den Patientenakten zu entnehmen waren (eher niedrige Sensitivität). Die höchste Übereinstimmung zeigte sich bei den Diagnosen Diabetes mellitus (Kappa 0,73), Herzinsuffizienz (Kappa 0,60), COPD (Kappa 0,56) und Hypertonie (Kappa 0,45). Rheumatoide Arthritis und TIA wiesen Kappa-Werte zwischen 0,44 und 0,40 auf, während alle anderen untersuchten Diagnosen eine schlechte Übereinstimmung aufwiesen (Kappa < 0,40). Labortests und andere Prozeduren zeigten insgesamt eine schlechtere Übereinstimmung als Diagnosen. In einer anderen Studie wurden die Akten und Abrechnungsunterlagen von 213 Patienten aus einer Gruppenpraxis mit einem von den Patienten selbst ausgefüllten Fragebogen zum Gesundheitsstatus verglichen. Für 17 chronische Diagnosen wurde die Übereinstimmung zwischen den Datenquellen berechnet, wobei die Patientenakte als Goldstandard diente (FOWLES et al. 1998). Beim Vergleich der Abrechnungsdiagnosen mit den Diagnosen in der Patientenakte fand sich eine hohe Spezifität bei geringerer Sensitivität, sowie eine starke Abhängigkeit dieser Parameter von der Art der Diagnose. Während Asthma, COPD, Diabetes, Hypertonie, Krebserkrankungen und Sehprobleme eine hohe Sensitivität von 80% und darüber aufwiesen, lag die Sensitivität von psychischen-, Rücken- und Gelenkproblemen, sowie Herz-, Nieren-, Darm-, Nerven- und Muskelproblemen zwischen 53% und 68%. Mit 50% bzw. 20% war die Sensitivität der Diagnosen Lähmung und Apoplex sowie Alkohol- und Drogenabusus besonders niedrig.

STEINWACHS et al. (1998) prüften anhand von sechs Aufgreifdiagnosen (Diabetes, Hypertonie, Asthma, Otitis media, Schwangerschaft und Kindervorsorge), welcher Anteil von innerhalb eines Jahres abgerechneten Arztbesuchen für 2407 Medicaid-Patienten in Maryland auch in der Patientenakte dokumentiert war und inwieweit Besuchsdatum und Diagnoseangaben übereinstimmten. Insgesamt fanden sich 82% der in den Abrechnungsunterlagen aufgeführten Arztbesuche mit gleichem Datum und gleicher Diagnose in den Akten wieder. Während Diagnosen und Leistungen für Patienten mit 1-3 Arztbesuchen („low utilization“) seltener abgerechnet wurden als aus den Akten hervorging, lagen für Patienten mit über 10 Arztbesuchen häufiger Abrechnungsunterlagen vor, für die kein Akteneintrag vorhanden war. Insgesamt lag die Zahl der in den Akten dokumentierten Besuche um 2,6% höher als in den Abrechnungsunterlagen. In einem anderen Studienansatz untersuchten HORNER et al. (1991), welche Art der Konsultation, Diagnosen und durchgeführte Prozeduren innerhalb eines zweiwöchigen Zeitraums in den Akten von 1140 Patienten dokumentiert worden waren, die in dieser Zeit eine allgemeinmedizinische Gruppenpraxis aufgesucht hatten. Diese Informationen verglichen sie mit den Abrechnungsunterlagen. Die Autoren stellten fest, dass die Übereinstimmung für die Art der Konsultation und die Zahl der behandelten Diagnosen mit einem Kappa-Wert von 0,29 bzw. 0,28 gering war. Nur in 60% der Fälle entsprach eine der im Untersuchungszeitraum gelisteten Abrechnungsdiagnose einem Behandlungsanlass in der Patientenakte. In 27% der Fälle konnte für eine Abrechnungsdiagnose kein Behandlungsanlass gefunden werden. Wenn gröbere ICD-9-Kategorien verwendet wurden, konnte die Übereinstimmung um 10%-15% erhöht werden. 45% der Diagnosen mit einer schlechten Übereinstimmung fanden sich vor allem in folgenden ICD-9-Kategorien: Endokrine und psychische Erkrankungen, Symptome und Maßnahmen, wie Vorsorgeuntersuchungen und Beratung bei sozialen Problemen. Die anderen Kategorien machten nur 5%-7% der Fehlerrate aus. Die Autoren kamen zu dem Schluss, dass die Verwendbarkeit von ambulanten Abrechnungsdiagnosen für Vergütungs- und Studienzwecke aufgrund ihrer schlechten Qualität limitiert sei.

In einer kanadischen Studie untersuchten WILCHESKY et al. (2004) Sensitivität und Spezifität von 14 ausgewählten Diagnosen, die bei der Medikation älterer Patienten berücksichtigt werden sollten. Außerdem analysierten sie die im Charlson Comorbidity Index enthaltenen Diagnosen, sowie diejenigen Diagnosencodes, die im Patientenklassifikationssystem ACG für eine Risikoadjustierung herangezogen werden. Dazu wurde eine zufällig ausgewählte Kohorte von 14.980 geriatrische Patienten von 110

Hausärzten aus einer anderen Studie benutzt, für die Abrechnungsdaten aller behandelnden Ärzte aus einem Kalenderjahr vorlagen. Die Abrechnungsdiagnosen wurden mit den Diagnoseeinträgen in den hausärztlichen Patientenakten verglichen, die innerhalb eines Zeitraums von fünf Jahren vor Studienbeginn dokumentiert worden waren. Die Spezifität von Diagnosen, die bei der Medikation geriatrischer Patienten berücksichtigt werden sollten, war hoch; ebenso die Spezifität der Diagnosen aus dem Charlson Comorbidity Index. Die Sensitivität war im Vergleich zu den US-amerikanischen Untersuchungen jedoch insgesamt eher niedrig. Die höchste Sensitivität zeigten Hypertonie und Diabetes (61% und 52%), gefolgt von COPD, Asthma und Herzinsuffizienz (24%-30%). Die Werte für alle anderen Diagnosen lagen unter 20%, verbesserten sich jedoch, wenn die Abrechnungsdiagnosen aller Ärzte einbezogen wurden, die den Patienten im Untersuchungszeitraum behandelt hatten. Dieser Befund spricht dafür, dass hausärztliche Patientenakten in Kanada Angaben zu sämtlichen Erkrankungen eines Patienten enthalten, nicht nur zu den vom Hausarzt behandelten Diagnosen. Um die Sensitivität und Spezifität ambulanter Abrechnungsdiagnosen für eine Risikoadjustierung mittels des ACG-Systems zu prüfen, wurden die vorliegenden Diagnosen für jeden Patienten in verschiedene ADG-Klassen gruppiert. Anschließend wurde die Klassifizierung jedes Patienten aufgrund der Diagnosen in den Akten mit der Einstufung anhand der Abrechnungsdiagnosen verglichen. Außer der Gruppe „chronic medical-stable“, für die ein Wert von 40% berechnet wurde, war die Spezifität hoch. Die gleiche Kategorie zeigte andererseits eine hohe Sensitivität (82%), während die Sensitivität in den anderen Kategorien sehr variabel und insgesamt eher niedrig war. Dazu merkten die Autoren an, dass die niedrige Sensitivität in ihrer Studie möglicherweise durch das Studiendesign mitbedingt gewesen sein könnte: Da für die Auswahl der Diagnoseinformationen in den Akten ein Fünfjahreszeitraum gewählt wurde und diese mit Abrechnungsdiagnosen aus nur einem Jahr verglichen wurden, könnten viele der Diagnosen in den Akten asymptomatisch gewesen sein und im Abrechnungszeitraum keine Behandlung erfordert haben. In diesem Fall wären solche Diagnosen als falsch negativ eingestuft worden und hätten die Werte für die Sensitivität erniedrigt. Die Autoren stellten fest, dass eine Risikoadjustierung auf der Basis von hausärztlichen Abrechnungsdiagnosen die wahre Morbidität von Patienten erheblich unterschätzen könnte und schlugen folgende Maßnahmen zur Erhöhung der Sensitivität von Abrechnungsdiagnosen vor:

- Einbeziehung von Abrechnungsdiagnosen aus mehreren Jahren,
- Einbeziehung von Abrechnungsdiagnosen von mehreren ambulanten Behandlern,

- Einbeziehung von Abrechnungsdiagnosen aus anderen Behandlungszusammenhängen (stationäre Diagnosen, Medikamentenverordnungen)

Die vorliegenden Studien zur Validität von Abrechnungsdiagnosen aus dem ambulanten Bereich zeigen ähnliche Probleme wie im stationären Sektor. Auch hier liegt international eher ein „Underreporting“, d.h. eine mangelnde Vollständigkeit von Abrechnungsdiagnosen im Vergleich mit der Patientenakte vor. Dabei zeigen sich jedoch Unterschiede zwischen einzelnen Arztgruppen. Abrechnungsdiagnosen in fachärztlichen Praxen weisen eine höhere Übereinstimmung mit der Patientenakte auf als in hausärztlichen Praxen. Die untersuchte Verschlüsselungsebene spielt ebenfalls eine Rolle für die Diagnosenvalidität: Je unspezifischer ein Diagnosencode sein darf, desto höher ist die Übereinstimmung zwischen Abrechnungsdiagnosen und Angaben in der Patientenakte. Analog zum stationären Bereich findet sich auch bezüglich der Validität ambulanter Abrechnungsdiagnosen eine Diagnosenspezifität. Während Angaben zu Diabetes mellitus und Hypertonie sowie zum Vorliegen einer Krebserkrankung überwiegend vollständig und korrekt sind, werden vor allem im hausärztlichen Bereich häufige Behandlungsanlässe, wie psychische Erkrankungen, verschiedene Symptome, Vorsorgeleistungen und psychosoziale Beratung zwar in der Patientenakte dokumentiert, im allgemeinen jedoch selten als Abrechnungsdiagnosen genannt. Dadurch könnten für einige Erkrankungen zu niedrige Prävalenzangaben resultieren, insbesondere wenn nur Abrechnungsdiagnosen eines Jahres, eines Behandlungssektors oder einer Arztgruppe als Basis herangezogen werden. In Bezug auf die Einführung eines Patientenklassifikationssystems könnte ein solches „Underreporting“ von Erkrankungen insbesondere im hausärztlichen Bereich einerseits zu einer Unterschätzung der Morbidität von Patienten führen. Andererseits würden die Behandlungskosten für eine Erkrankung aufgrund einer zu niedrig angegebenen Prävalenz insgesamt als zu hoch eingeschätzt werden.

#### ***2.1.10 Studien zur Datenvalidität in Deutschland***

Inzwischen liegen auch in Deutschland Erfahrungen mit der Validität von Patienteninformationen im Zusammenhang mit der Einführung des DRG-Systems vor. In zwei Studien zur Kodierqualität wurden Diagnosen, Prozeduren und Angaben zu anderen, für die DRG-Eingruppierung relevanten Parametern anhand der Patientenakten überprüft und die jeweils resultierende DRG-Einstufung verglichen. Beide Studien kamen zu sehr unterschiedlichen Ergebnissen, was mit der Tatsache begründet wurde, dass das in der früheren Untersuchung verwendete australische DRG-System noch nicht gut genug an deutsche Verhältnisse angepasst war. Während die ältere Studie nur in ca. 50% der Fälle eine

korrekte Angabe für die Hauptdiagnose und in bis zu 37% der Fälle Unterkodierungen bzw. in 60% Überkodierungen feststellte, fand sich in der neueren Untersuchung in 65% der Fälle eine korrekte Hauptdiagnose und in 9% der Fälle Unterkodierungen und in 34% Überkodierungen. Durch eine Korrektur von Diagnosen- und Prozedurenschlüsseln änderte sich in 22% der Fälle die Einstufung in eine DRG, was insgesamt zu einer Verminderung des Casemix-Index (Fallschweregrad) von 0,84 auf 0,81 führte, während der Fallschweregrad in der älteren Studie nach korrekter DRG-Gruppierung um 6,9% angestiegen war (KLAUS et al. 2005, DIRSCHEDEL et al. 2003).

Die Validität von Abrechnungsdaten aus dem ambulanten Bereich wurde bisher kaum untersucht. Dem Zentralinstitut für die Kassenärztliche Versorgung liegen für verschiedene KV-Bereiche ICD-10-codierte Abrechnungsdaten vor. In der Vergangenheit wurde versucht, aus diesen Daten die Prävalenz einzelner Erkrankungen in Deutschland zu schätzen (HEUER et al. 2003, KEREK-BODDEN et al. 2000). Da Abrechnungsdaten aufgrund der bisher geltenden gesetzlichen Bestimmungen nicht personenbezogen vorlagen, war ein Vergleich mit Patientenakten nicht möglich. HUMMERS-PRADIER et al. (2003) analysierten den Inhalt von Patientenakten bezüglich der Qualität der Medikation von Asthmapatienten, verglichen die Angaben zu dieser und anderen Erkrankungen jedoch nicht mit den ICD-10-codierten Abrechnungsdiagnosen. In einer Studie mit ähnlicher Thematik untersuchten SCHNEIDER et al. (2005) den Zusammenhang zwischen der Therapie von 857 Asthma- und COPD-Patienten und den ICD-10-codierten Abrechnungsdiagnosen. In einer logistischen Regressionsanalyse fanden die Autoren einen stärkeren Einfluss der Ergebnisse diagnostischer Tests auf die Behandlungsmaßnahmen als der angegebenen ICD-10-Diagnosen. Sie folgerten daraus, dass ICD-10-Codes möglicherweise nicht immer den wirklichen Zustand des Patienten widerspiegeln, und dass die Qualität der medizinischen Versorgung für Asthma- und COPD-Patienten besser sein könnte, als Studien auf der Basis von ICD-10-codierten Abrechnungsdiagnosen vermuten lassen. Im Rahmen ihres Gutachtens für die Einführung des Morbi-RSA prüften LAUTERBACH und LÜNGEN (2005) die Eignung ambulanter Diagnosen als Risikofaktoren. In ihrem Datensatz, einer repräsentativen Versichertenstichprobe mit Personenbezug, fiel auf, dass die Dichte der ambulant erworbenen Diagnosen pro Konsultation in Deutschland im Unterschied zu anderen Ländern extrem hoch war: Der Durchschnitt lag bei 10 Diagnosen pro Patient. Lediglich 8% wiesen nur eine Diagnose, aber 61% bis zu sechs und immerhin 2% über 30 Diagnosen auf. Die zweite Studie, die sich mit der Validität von ambulanten Abrechnungsdiagnosen in Bezug auf den Einsatz

als Risikofaktoren in einem Patientenklassifikationssystem beschäftigt, wurde von TRAUTNER et al. (2005) durchgeführt. Die Autoren analysierten anhand der ICD-10-codierten Abrechnungsdiagnosen von insgesamt ca. 1,95 Millionen Patienten aller niedersächsischen Vertragsärzte aus sechs Quartalen der Jahre 2002 und 2003, wie viel Prozent der häufigsten 16 akuten und 28 chronischen Diagnosen im Laufe des Untersuchungszeitraums fortgeschrieben wurden. Sie stellten fest, dass sich lediglich zwischen 0,5% (Grippe) und 9,3% (Glaukomverdacht) der akuten Diagnosen über alle Quartale hinweg wieder fanden. Bei den chronischen Diagnosen wurden zwei Gruppen unterschieden: Diejenigen Patienten, die die Angabe der Erkrankung schon im letzten Quartal 2001 aufwiesen und Patienten, die zu Beginn des Untersuchungszeitraums diagnostiziert wurden. In der ersten Gruppe fand sich je nach Erkrankung eine hohe Fortschreibungsrate von 66% (Skoliose) bis 85% (Hypertonie). In der zweiten Gruppe war die Rate dagegen erheblich niedriger und lag zwischen 19% (COPD) und 53% (Hypertonie), wobei ein massiver Abfall der Werte im Folgequartal nach Diagnosestellung auffiel und die Rate im weiteren Verlauf eher konstant blieb. Dieser Befund lässt vermuten, dass es sich möglicherweise um Ausschlussdiagnosen handelte; weiterführende Schlüsse konnten aus dem vorhandenen Datensatz jedoch nicht gezogen werden.

### ***2.1.11 Zusammenfassung der Studienergebnisse zur Validität von Abrechnungsdaten***

Die meisten der vorgestellten Studien zur Validität von Informationen über Diagnosen und Prozeduren aus administrativen Datenquellen im stationären und ambulanten Bereich stammen aus den USA und Kanada. Insgesamt zeigen diese Untersuchungen eine hohe Spezifität und NPW, d.h., diejenigen Patienten, die nicht unter einer bestimmten Erkrankung leiden, können anhand von Abrechnungsdaten wahrscheinlich relativ zuverlässig identifiziert werden. Auch der PPW ist relativ hoch, d.h., wenn eine Abrechnungsdiagnose vorliegt, weist der Patient in den meisten Fällen auch die Erkrankung auf (wenig falsch positive Diagnosen; „Overreporting“ durch Leistungserbringer eher gering). Die Sensitivität, d.h. die Fähigkeit, Patienten mit einer bestimmten Erkrankung anhand von Abrechnungsdiagnosen richtig zu identifizieren, ist dagegen niedriger (viele falsch negative Diagnosen, „Underreporting“ der Leistungserbringer). Informationen über Risikofaktoren und nicht aktuell behandlungsbedürftige Erkrankungen sind in den Abrechnungsunterlagen unterrepräsentiert. Die Sensitivität ist krankheitsspezifisch stark variabel und zeigt eine Abhängigkeit von der Schwere der Erkrankung bzw. Prozedur. Die Einbeziehung von Abrechnungsdiagnosen aus anderen Behandlungszusammenhängen (z.B. Haus- und Fachärzte, stationärer Bereich, Medikamentenverordnungen) und die Ausdehnung des untersuchten Zeitraums konnte in

Studien aus den USA und Kanada die Sensitivität erhöhen. Die Ergebnisse der wenigen deutschen Untersuchungen zur Validität von administrativen Daten für den Einsatz in Risikoadjustierungssystemen zeigen im Vergleich zur internationalen Studienlage einige Abweichungen und Besonderheiten. Die Validität von Abrechnungsdaten, die im Rahmen des DRG-Systems erhoben wurden, war geringer als in den US-amerikanischen Studien, ein „Overreporting“ deutlich häufiger. Im ambulanten Bereich fiel eine extrem hohe Zahl von bis zu 30 angegebenen Abrechnungsdiagnosen pro Patient sowie eine wechselnde Fortschreibung vor allem chronischer Diagnosen auf.

Die Frage, ob die Validität von Abrechnungsdaten im Vergleich zu den Angaben in der Patientenakte ausreichend ist, um als Risikofaktoren in einem Risikoadjustierungsmodell eingesetzt zu werden, wird in der gesichteten Literatur kontrovers diskutiert. Bezüglich der Berechnung eines Krankheitsindex, wie dem Charlson Comorbidity Index (CCI), kommen einige Studien zu dem Schluss, dass Akten die Morbidität eines Patienten vollständiger abbilden können, da sie z.B. auch Informationen über Erkrankungen und Risikofaktoren enthalten, die im Untersuchungszeitraum nicht behandlungsbedürftig waren. Dies gilt jedoch vermutlich nur, wenn die Akten aller an der Behandlung des Patienten beteiligten Ärzte eingesehen werden (DE GROOT et al. 2003). Andere Untersuchungen konnten keine Unterschiede zwischen auf Basis von Patientenakten bzw. Abrechnungsdiagnosen berechnetem CCI feststellen. Die Angaben zu Hauptdiagnosen und Prozeduren für die DRG-Klassifizierung werden in den USA inzwischen als hinreichend zuverlässig eingeschätzt, während sich in Deutschland noch Defizite im Sinne eines „Upcodings“ zeigen. Aufgrund der international eher niedrigen Sensitivität mancher Abrechnungsdiagnosen im ambulanten Bereich wurde das Problem deutlich, dass Patientenklassifikationssysteme die Prävalenz bestimmter Erkrankungen in einer Population unterschätzen und damit die durchschnittlichen Kosten für die Behandlung einer Diagnose überschätzen könnten. Da im Allgemeinen nur Diagnosen eines Jahres für die Klassifizierung von Patienten einbezogen werden, könnten beispielsweise chronische Erkrankungen in den Abrechnungsdiagnosen fehlen, wenn sie in diesem Zeitraum nicht behandlungsbedürftig waren. Für Deutschland liegt dagegen der Befund vor, dass viele chronische Diagnosen über mehrere Quartale fortgeschrieben werden. Unklar ist derzeit jedoch, ob eine solche Fortschreibung gleichbedeutend mit einer guten Vollständigkeit (=hohen Sensitivität) von Abrechnungsdiagnosen im Vergleich mit der Patientenakte ist.

### **Potentielle Einflussfaktoren auf die Validität von Abrechnungsdaten**

Die internationale Studienlage zeigt als Hauptproblem, dass vor allem leichtere, nicht-chronische Erkrankungen und unspezifische Symptome selten als Abrechnungsdiagnosen angegeben werden. In diesem Zusammenhang stellt sich die Frage, welche Ursachen dieser Befund haben könnte, und welche Faktoren generell die Validität von Abrechnungsdiagnosen beeinflussen könnten.

#### ***2.1.12 Codierungssystem und Codierungsprozess***

Einen möglichen Einflussfaktor auf die Validität von Abrechnungsdiagnosen könnte die Tatsache darstellen, dass sie Informationen über den Patienten in verschlüsselter Form enthalten. Beispielsweise könnte das verwendete Codierungssystem einen Einfluss darauf haben, wie gut sich die Morbidität von Patienten abbilden lässt. Das in Europa benutzte ICD-10-System unterscheidet sich vom US-amerikanischen ICD-9-Codierungssystem unter anderem bezüglich der Gliederung und der Möglichkeit, Erkrankungen spezifisch zu verschlüsseln. Die größeren Kategorien des ICD-9-Systems könnten die Morbidität von Patienten verändert darstellen im Vergleich zu Ländern, die das ICD-10-System benutzen. Andererseits erhöhte die Verwendung größerer Verschlüsselungskategorien in Studien (z.B. HORNER et al. 1991) die Übereinstimmung zwischen Abrechnungsdiagnosen und den in der Patientenakte angegebenen Diagnosen.

In beiden ICD-Systemen existieren für eine bestimmte Erkrankung viele Synonyme mit unterschiedlichen Diagnoseschlüsseln, so dass eine hohe Variabilität der Zuordnung möglich ist (HORN BROOK et al. 1998). So fand SURJAN (1999) in einer Literaturübersicht zur Validität von Diagnoseschlüsseln beispielsweise Fehlerraten von 10% bei dreistelliger und 25%-30% bei vierstelliger (auch EDV-gestützter) Verschlüsselung nach ICD-10.

Für den stationären Bereich schlug SURJAN (1999) einen theoretischen Ansatz vor, um den Prozess der Generierung von ICD-Codes und seine Fehlerquellen systematisch zu erfassen. Dazu gliederte er den Codierungsprozess in folgende Schritte:

1. Aus den Angaben in der Patientenakte wird ein Entlassungsbrief formuliert, der die relevanten Behandlungsdiagnosen enthalten sollte.
2. Im nächsten Schritt werden anhand der diagnostischen Angaben die zugrunde liegenden Erkrankungen identifiziert.
3. Anschließend werden die Erkrankungen des Patienten ICD-Kategorien zugeordnet und zu einzelnen ICD-Codes transformiert. Dieser Schritt

beschreibt die eigentliche Codierung von Behandlungsanlässen und wird inzwischen meist computergestützt durchgeführt.

4. Zum Abschluss des Prozesses werden die codierten Erkrankungen als primäre Behandlungsanlässe, Komorbiditäten und Komplikationen klassifiziert.

Bei jedem der genannten Schritte können Fehler auftreten. Im Entlassungsbrief könnten klinische Diagnosen aus der Patientenakte übersehen worden sein, vor allem, wenn es sich um Begleiterkrankungen handelt, die keine Relevanz für die abzurechnende Behandlung hatten. Im nächsten Schritt der diagnostischen Zuordnung unterscheidet der Autor zwischen einer Diagnose, die das klinische Wissen über einen Patienten subsumiert und zur Rechtfertigung bestimmter therapeutischer Maßnahmen dient, und einer Erkrankung, die die pathogenetische Ursache unterschiedlicher Diagnosen darstellt. Nur Erkrankungen und nicht Diagnosen sollten einer Codierung zugeführt werden, da die Klassifizierungssystematik des ICD-Systems auf pathogenetischen Ursachen von Erkrankungen basiert. Wird dem Unterschied zwischen Diagnose und Erkrankung keine Beachtung geschenkt, könnten beispielsweise mehrere Symptome einer Erkrankung als Diagnosen einzeln codiert werden. Während die Angabe eines Codes für die Erkrankung korrekt gewesen wäre, entstünden auf diese Weise mehrere Codes für verschiedene Diagnosen. Auch bei der Zuordnung zu den ICD-Codes könnten Probleme entstehen. Wenn eine klinische Diagnosenangabe z.B. sehr spezifisch ist, steigt die Wahrscheinlichkeit, dass sie einer anderen, allgemeineren Kategorie zugeordnet wird. In diesem Zusammenhang ist von Bedeutung, wie und von wem diese Zuordnung vorgenommen wird, ob eine Klinik beispielsweise medizinische Dokumentare oder ein Codierprogramm für die Codierung einsetzt, oder ob Ärzte neben ihrer sonstigen Arbeit Diagnosen und Prozeduren codieren müssen (LINCZAK et al. 2004, LINCZAK et al. 2003).

### ***2.1.13 Länderspezifische Codierungs- und Abrechnungsvorschriften***

Sowohl bei der Codierung als auch beim letzten Schritt der Klassifizierung von Erkrankungen für Abrechnungszwecke spielen länderspezifische Codierungsregeln und Abrechnungsvorschriften ebenfalls eine wichtige Rolle für die Validität von Abrechnungsdaten. Beispielsweise vermuteten RECTOR et al. (2004) für die USA, dass Diagnosen, die in den Abrechnungsunterlagen, aber nicht in der Patientenakte erscheinen, Ausschlussdiagnosen darstellen könnten. In Deutschland zogen TRAUTNER et al. (2005) in ihrer Studie die gleiche Erklärung für das Verschwinden von chronischen Erkrankungen im Folgequartal nach Diagnosestellung in Betracht.

In den USA gilt die Abrechnungsvorschrift, dass jede für einen Patienten durchgeführten Maßnahme in den Abrechnungsunterlagen anhand einer Diagnose begründet werden muss. Daher nahmen RECTOR et al. (2004) an, dass Erkrankungen, die in den Abrechnungsunterlagen, aber nicht in der Akte des Patienten erschienen, als Rechtfertigung für eine abgerechnete Leistung angegeben worden waren (z.B. Labortest, Röntgen, EKG, etc.). Generell sollen in den USA nur Diagnosen in den Abrechnungsunterlagen aufgeführt werden, für die im Abrechnungszeitraum Behandlungsleistungen erbracht wurden. Diese Vorschrift macht es wahrscheinlich, dass viele „inaktive“ chronische Diagnosen nur in der Patientenakte, nicht aber in den Abrechnungsunterlagen erscheinen. Viele Organisationen geben auch EDV-kompatible Formulare vor, auf denen Ärzte und Kliniken erbrachte Leistungen mit Datum und Diagnose in codierter Form angeben müssen. Die Vollständigkeit der Diagnosen für einen multimorbiden Patienten könnte dann beispielsweise davon beeinflusst werden, wie viele Zeilen in dem Formular für Diagnoseeinträge vorgesehen sind (IEZZONI et al. 1992).

Als weitere mögliche Ursachen für Differenzen zwischen Abrechnungsunterlagen und Akten vermuteten VIRNIG und MCBEAN (2001), dass Erkrankungen oft nicht in den Abrechnungsdaten erscheinen, die mit nicht erstattungsfähigen Maßnahmen behandelt wurden. Darüber hinaus stellten sie fest, dass die Qualität derjenigen Patienteninformationen am höchsten ist, die Abrechnungszwecken dienen. IEZZONI (2003) machte ebenfalls die Beobachtung, dass Leistungserbringer besonders auf die Vollständigkeit von Daten achten, die Voraussetzung für eine Abrechnung von Leistungen sind.

#### ***2.1.14 Qualität der Dokumentation***

Bei der Vorstellung seines theoretischen Konzeptes für den Codierungsprozess vermutete SURJAN (1999), dass die Qualität der Patientenakte als Informationsgrundlage von entscheidender Bedeutung für die Validität der resultierenden ICD-Codes sein müsste. Dabei stellt sich die Frage, auf welche Weise die Qualität von Patientenakten definiert und gemessen werden sollte, und welche Qualitätsindikatoren für eine solche Messung herangezogen werden könnten. In einer Studie zur Validität von Patientenakten untersuchten ROMM und PUTNAM (1981), welche Inhalte von 55 transkribierten Konsultationen in der internistischen Poliklinik eines Krankenhauses in den USA auch in den Akten der Patienten dokumentiert worden waren. Die Autoren entwarfen dazu ein Auswertungsschema, das die Inhalte der Konsultationen in „Informationseinheiten“ bewertete und verglichen Art und Anzahl dieser Einheiten in Transkripten und Akten. Sie fanden eine hohe Übereinstimmung für den

Hauptbehandlungsanlass (92%) und die ihn begleitenden Informationen (71%). Die übrige Anamnese zeigte mit 29% eine schlechtere Übereinstimmung.

LOGAN et al. (2001) entwickelten einen Ansatz, um die Qualität der in elektronischen Patientenakten enthaltenen Daten zu messen und erprobten ihn im Rahmen einer Pilotstudie. Als Messgrößen wählten sie die Dimensionen Vollständigkeit und Korrektheit und klassifizierten vier mögliche Zustände bezüglich der Übereinstimmung bzw. Nicht-Übereinstimmungen zwischen dem gewählten Vergleichsstandard und der Patientenakte:

1. Eine Information ist in der Akte und im Vergleichsstandard richtig enthalten (ein „korrektes Element“)
2. Eine Information ist im Vergleichsstandard und in der Akte enthalten, in der Akte aber nicht richtig wiedergegeben (ein „inkorrektes Element“)
3. Eine Information ist im Vergleichsstandard, aber nicht in der Akte enthalten (ein „fehlendes Element“)
4. Eine Information ist in der Akte, aber nicht im Vergleichsstandard enthalten (ein „zusätzliches Element“)

Als Qualitätsindikatoren empfahlen die Autoren, Befunde und Inhalte von Konsultationen zusammenzustellen und anhand dieser Zusammenstellung „wichtige Informationen“ zu identifizieren, die unbedingt in der Akte vorhanden sein sollten. Dabei wurden „wichtige Informationen“ als Inhalte definiert, die ärztlicherseits als relevant eingestuft würden. Die von den Autoren durchgeführte Pilotstudie hatte zum Ziel, zwei unterschiedliche Dokumentationsmethoden miteinander zu vergleichen (Diktat mit nachfolgender Transkription und ein computergestütztes Dokumentationssystem). Dazu setzten die Autoren Videoaufzeichnungen von vier verschiedenen Konsultationen als Goldstandard ein und ließen ein Expertenteam in einem Delphi-Verfahren zunächst die relevanten Inhalte definieren. Anschließend wurde von acht Ärzten mit beiden Methoden eine Dokumentation von jeweils zwei der Konsultationen angefertigt. Die Diktat- bzw. computergestützte Dokumentation zeigte eine Vollständigkeit von 69% bzw. 68% und eine Korrektheit von 98% bzw. 93%.

### ***2.1.15 Nutzung einer elektronischen Patientenakte***

Im ambulanten und stationären Bereich findet die EDV-gestützte Patientenakte immer größere Anwendung. Daher wurde die Validität von Patienteninformationen in elektronischen Patientenakten im Vergleich zur traditionellen Papierakte in einigen Studien untersucht. HOGAN und WAGNER (1997) kamen in einem Literaturreview zunächst zu dem Schluss,

dass die vorliegenden Analysen eine zu große Variabilität der Ergebnisse aufwiesen, um die Validität von elektronischen Patientenakten einschätzen zu können, und empfahlen weitere Forschungsarbeiten zum Thema. Einige Jahre später verglichen TANG et al. (1999) Ärzte, die entweder eine Papier- oder eine elektronische Patientenakte benutzten und fanden eine bessere Vollständigkeit von Diagnose- und Medikationslisten in der elektronischen Akte.

### **Spezifische Aspekte potentieller Einflussfaktoren auf die Validität von Abrechnungsdaten im hausärztlichen Bereich**

Im primärärztlichen Bereich könnten sich durch die Besonderheiten der allgemeinmedizinischen Tätigkeit über die genannten Probleme hinaus noch andere spezielle Aspekte ergeben, die die Validität von Abrechnungsdiagnosen beeinflussen könnten.

#### ***2.1.16 Codierungssystem und Codierungsprozess***

KÖRNER et al. (2005, S. A3168) stellten fest, dass „die allgemeinmedizinische Versorgung eher durch Beratungsanlässe als durch Diagnosen bestimmt“ ist. Es stellt sich die Frage, inwieweit das auf pathogenetisch herleitbaren Diagnosen basierende ICD-10-Codierungssystem für die Allgemeinmedizin ein „geeignetes Raster“ darstellt, um „Patientenprobleme versorgungsnah abzubilden“ (ebenda). So besteht beispielsweise die Gefahr, dass bei geringfügigen Symptomen für die Codierung Diagnosen gewählt werden, die dem Symptom zwar ähnlich sind, aber einen höheren Schweregrad der Erkrankung suggerieren als eigentlich vorliegt (FISCHER und ROSSA 1995). Dadurch könnte die Morbidität von Patienten unzutreffend eingeschätzt werden. Viele Länder setzen daher speziell für den primärärztlichen Bereich konzipierte Codierungssysteme ein, wie z.B. die 1987 von der World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA) erstmals veröffentlichte International Classification of Primary Care (ICPC) (LAMBERTS et al. 1993). Das System ist in 17 organbezogene Kapitel und sieben Komponenten gegliedert, in denen neben Diagnosen Symptome, Therapien, psychosoziale Maßnahmen und andere spezifisch hausärztliche Tätigkeiten schnell und einfach codiert werden können. Inzwischen liegt die ICPC in zweiter Auflage als elektronische Version vor und verfügt über einen Thesaurus, mit dessen Hilfe eine automatisierte Übersetzung von ICPC-Codes ins ICD-10-System möglich ist (OKKES et al. 2002, OKKES et al. 2000). Ein anderes Codierungsverfahren, das vor allem bei Hausärzten in Großbritannien Anwendung findet, stellen die „Read-Codes“ dar. Die Codierungsinhalte sind denen im ICPC-System ähnlich und umfassen Angaben zum Beruf des Patienten, zu Anamnese, Diagnostik und Tests, präventiven Maßnahmen, Operationen und Prozeduren sowie zu administrativen Arbeiten (JUNIUS et al. 1995).

Da Hausarztpraxen in Deutschland im Allgemeinen nicht über medizinische Dokumentare verfügen, wird die Codierung von Behandlungsanlässen in den meisten Fällen von den Ärzten selbst erledigt. Viele Praxis-EDV-Systeme verfügen über integrierte Verschlüsselungsprogramme, die auf dem offiziellen Thesaurus des DIMDI basieren (Deutsches Institut für Dokumentation und Information (DIMDI) 2006). Eine vom Arzt eingegebene Klartextdiagnose wird von diesem Programm automatisch mit einem geeigneten ICD-10-Code versehen. Auf welche Weise dieses Verfahren die Validität von Abrechnungsdiagnosen beeinflusst, wurde bisher nicht untersucht.

### ***2.1.17 Dokumentationsstruktur und Dokumentationsstandards***

Eine weitere Besonderheit ergibt sich im Bereich der Dokumentation von Arzt-Patienten-Kontakten. Ein wesentlicher Inhalt der hausärztlichen Tätigkeit ist die langfristige Behandlung von Patienten, nicht selten über mehrere Generationen hinweg. Die Begleitung chronisch kranker und multimorbider Patienten stellt ebenfalls eine zentrale Aufgabe des Hausarztes dar. Die Art der Dokumentation von Behandlungsanlässen und Befunden sollte daher so organisiert sein, dass der Krankheitsverlauf verschiedener Erkrankungen langfristig übersichtlich abgebildet werden kann. In Deutschland und in vielen anderen Ländern werden die einzelnen hausärztlichen Konsultationen chronologisch dokumentiert. Bei jedem Besuch des Patienten werden das Datum sowie die verschiedenen Beschwerden, Untersuchungsergebnisse und ergriffenen Maßnahmen mehr oder weniger ausführlich vom Arzt notiert. Im Laufe der Zeit entstehen auf diese Weise umfangreiche Akten, aus denen der Verlauf einzelner chronischer Erkrankungen des Patienten nur mit Schwierigkeiten nachverfolgt werden kann, da die dazugehörigen Befunde an wechselnden Stellen in der Akte zu finden sind. Die Einführung elektronischer Patientenakten hat diese Struktur der Dokumentation nicht wesentlich verändert.

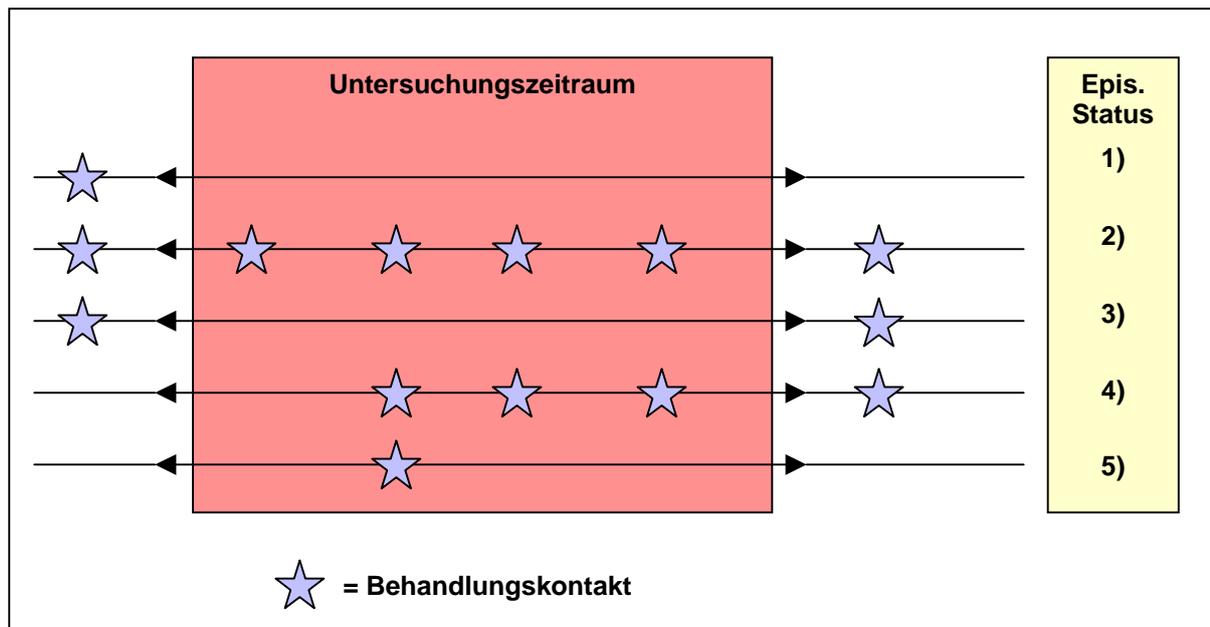
WEED (1968) schlug vor, Patientendokumentationen nicht chronologisch, sondern problemorientiert zu führen. Eine solche Dokumentationsstruktur wurde von LAMBERTS (1993) im Rahmen des Konzepts der Behandlungsepisode („Episode of care“) aufgegriffen. Das Konzept der medizinischen Behandlungsepisode wurde erstmalig von SOLON et al. (1967) im Rahmen einer Studie zur Inanspruchnahme von Gesundheitsleistungen vorgestellt. Er definierte eine solche Episode als „eine Folge einer oder mehrerer medizinischer Behandlungsleistungen für ein Individuum während einer Periode mit relativ kontinuierlichem Kontakt mit einem oder mehreren Leistungserbringern wegen eines

speziellen medizinischen Problems“ (SOLON et al. 1969, S.936). Behandlungsepisoden („Episodes of care“) fokussieren also auf die Behandlungsrealität von Gesundheitsproblemen, während Erkrankungsepisoden („Episodes of illness“) die subjektive Realität des Patienten abbilden, indem sie den Zeitraum umfassen, während dessen ein Patient unter Krankheitssymptomen leidet. Krankheitsepisoden („Episodes of disease“) dagegen beschreiben den gesamten Verlauf eines Gesundheitsproblems von seinem Beginn bis zur Heilung oder dem Tod des Patienten. Erkrankungs- und Krankheitsepisoden ziehen nicht zwangsläufig eine Inanspruchnahme von Behandlungsleistungen nach sich.

Das Konzept der Behandlungsepisode wurde von niederländischen Forschern für das Gebiet der Allgemeinmedizin übernommen. Nach LAMBERTS und HOFMANS-OKKES (1996a) beginnt eine Behandlungsepisode mit der ersten Konsultation wegen eines bestimmten Gesundheitsproblems und endet mit dem Abschluss der Behandlung wegen dieses Problems. Somit kann eine Episode z.B. bei chronischen Erkrankungen mit regelmäßigen Folgekonsultationen und medikamentöser Dauertherapie einen sehr langen Zeitraum umfassen. Andererseits kann sie auch sehr kurz sein, wie beispielsweise bei einem akuten Atemwegsinfekt, der nach der Erstuntersuchung und –therapie und evtl. noch einer weiteren Folgeuntersuchung ausgeheilt ist. In der Praxis können bei der Abgrenzung einzelner Behandlungsepisoden Schwierigkeiten auftreten. Die Entscheidung, wann eine Episode als abgeschlossen betrachtet werden kann, wird im ursprünglichen Konzept von SOLON (1967) davon abhängig gemacht, ob zwischen zwei Behandlungskontakten für das selbe Gesundheitsproblem ein Ereignis oder Zeitraum liegt, der einen Bruch in der Behandlungskontinuität bedeutet. So kann auch bei einer chronischen Erkrankung wie Diabetes mellitus eine Behandlungsepisode abgeschlossen sein, wenn der Patient nach erfolgreicher medikamentöser Einstellung und nunmehr stabiler Stoffwechsellage keine weiteren Behandlungskontakte aufweist. Wird er nach weiteren sechs Monaten wieder zur Kontrolle einbestellt, würde eine neue Episode für dieses Gesundheitsproblem eröffnet werden. Andererseits könnte die gleiche Behandlungsepisode auch als durchgehend betrachtet werden, wenn der Patient nach der medikamentösen Einstellung weiterhin regelmäßigen Kontakt zur Praxis hält, indem er beispielsweise jeden Monat ein Rezept für seine Medikamente abholt. Als epidemiologische Messgrößen unterscheiden LAMBERTS und HOFMANS-OKKES (1996a) in diesem Zusammenhang neue (inzidente) und bekannte (prävalente) Behandlungsepisoden. VAN DEN AKKER et al. (2004) definieren darüber hinaus Episoden mit besonderer Relevanz, die den Funktionsstatus von Patienten

einschränken oder einschränken könnten, als „Gesundheitsprobleme“, die jeweils den Status „aktiv“, „inaktiv“, „abgeschlossen“ oder „von bleibender Bedeutung“ aufweisen können. Inaktive Erkrankungen können zusätzlich als „latente Probleme“ gekennzeichnet werden, wenn wahrscheinlich ist, dass sie zukünftig wieder einen aktiven Status annehmen. Soll das Behandlungsgeschehen in einem bestimmten Zeitraum betrachtet werden, sind verschiedene Kombinationen dieser Ausprägungen möglich, um den Status von Episoden und Gesundheitsproblemen zu beschreiben (siehe Abbildung 3):

1. Episoden, die vor dem Untersuchungszeitraum abgeschlossen wurden (bekannte abgeschlossene Gesundheitsprobleme, z.B. Z.n. erfolgter Cholecystektomie bei Cholecystolithiasis). Solche Episoden könnten auch von bleibender Bedeutung sein.
2. Episoden, die vor dem Untersuchungszeitraum bestanden und währenddessen weiter behandelt werden (bekannte aktive Gesundheitsprobleme, z.B. bei chronischen Erkrankungen, wie Hypertonie, KHK, Diabetes mellitus).
3. Episoden, die vor dem Untersuchungszeitraum bestanden und währenddessen nicht behandlungsbedürftig sind (bekannte inaktive Gesundheitsprobleme, z.B. Skoliose, Varikosis, Osteoporose). Solche Episoden können auch als latente Probleme gekennzeichnet werden.
4. Episoden, die während des Untersuchungszeitraums erstmalig auftreten und währenddessen (und darüber hinaus) weiter behandelt werden (neue aktive Gesundheitsprobleme, wie bei Erstdiagnose eines Prostatacarcinoms).
5. Episoden, die während des Untersuchungszeitraums erstmalig auftreten und innerhalb dieses Zeitraums abgeschlossen werden (neue Episoden, die kein gravierendes Gesundheitsproblem darstellen, z.B. bei akuten Erkrankungen wie Atemwegsinfekten, Harnwegsinfekten, Gastroenteritis, Tonsillitis).



**Abbildung 3: Verschiedener Status von Episoden bezogen auf einen festgelegten Untersuchungszeitraum**

1)=bekannte abgeschlossene Episode (erster und letzter Behandlungskontakt vor dem Untersuchungszeitraum); 2)=bekannte aktive Episode (erster Behandlungskontakte vor, weitere Kontakte während und letzter Kontakt nach dem Untersuchungszeitraum); 3)=bekannte inaktive Episode (erster Behandlungskontakte vor und nächster Kontakt nach Untersuchungszeitraum); 4)=neues aktives Gesundheitsproblem (erster Behandlungskontakt im Untersuchungszeitraum und letzter Kontakt danach); 5)=neue Episode (erster und letzter Kontakt im Untersuchungszeitraum)

Inhaltlich beginnt eine Behandlungsepisode mit dem subjektiven Behandlungsanlass des Patienten („Subjective“), dem meist die Einleitung von diagnostischen Schritten („Objective“) und die Einschätzung des Arztes in Form einer Arbeitshypothese („Assessment“) folgt. Je nach Diagnose werden dann therapeutische Maßnahmen ergriffen („Plan“). Alle relevanten Gesundheitsprobleme eines Patienten sollten darüber hinaus in einer Problemliste zusammengefasst und ihr Status (aktiv/inaktiv) gekennzeichnet werden. Erfolgt die Dokumentation in der Patientenakte episodenzugeordnet nach dieser „SOAP“-Struktur, wird eine übersichtliche, problembezogene Darstellung des gesamten Behandlungsgeschehens für jede einzelne Erkrankung von Patienten möglich. Wird das SOAP-Schema in einer elektronischen Patientenakte benutzt, können unterschiedliche Erkrankungen in einzelnen Masken getrennt dargestellt und die im Krankheitsverlauf durchgeführten Maßnahmen immer wieder neu hinzugefügt werden, so dass jederzeit der aktuelle Status jedes Gesundheitsproblems für einen Patienten abgerufen werden kann. Durch die Möglichkeit der Zuordnung von Maßnahmen und Leistungsziffern zu den einzelnen Behandlungsepisoden könnten beispielsweise Aussagen zum Ressourcenverbrauch und Behandlungsaufwand von Patienten bzw. von bestimmten Erkrankungen in einer Praxis getroffen werden.

Im Gegensatz zu Deutschland, wo ein vergleichbarer Dokumentationsstandard derzeit noch fehlt, ist beispielsweise in Spanien, in den Niederlanden und in Belgien die Nutzung einer elektronischen Patientenakte zusammen mit einer episodenzugeordneten Dokumentation nach dem SOAP-Schema im primärärztlichen System allgemein akzeptiert. Das hausärztliche Codierungssystem ICPC ist speziell auf eine Dokumentation mittels SOAP-Schema zugeschnitten, deshalb wird in diesen Ländern gleichzeitig das ICPC-System zur Verschlüsselung von Behandlungsanlässen verwendet (VERBEKE 2004, ORUETA et al. 1999, LAMBERTS et al. 1993).

### ***2.1.18 Qualität der hausärztlichen Patientendokumentation***

In der Untersuchung von LOGAN et al. (2001) zur Entwicklung von Qualitätskriterien für die Inhalte elektronischer Patientenakten fiel eine große Variabilität der Werte für Vollständigkeit von Patienteninformationen je nach dokumentierendem Arzt auf (zwischen 55% und 90%), während die Werte für Korrektheit bei allen Ärzten gleichmäßig hoch waren. Diese Ergebnisse deuten darauf hin, dass bei einem Fehlen von verbindlichen Dokumentationsstandards interindividuelle Unterschiede zwischen den Praxen sowohl im Umfang als auch in der Art der über den Patienten in der Akte dokumentierten Informationen vorliegen könnten. Es erscheint denkbar, dass vollständigere Angaben in der Patientenakte wiederum eine höhere Validität von Abrechnungsdiagnosen nach sich ziehen könnten, eine solche Hypothese wurde bisher jedoch nicht empirisch untersucht.

Die „European Working Party on Quality in Family Practice“ (EQuIP), eine Arbeitsgruppe der WONCA, hat für den hausärztlichen Bereich Qualitätskriterien für die Patientendokumentation entwickelt, die international konsentiert sind. Die EQuIP legt dabei das SOAP-Schema als wünschenswerten Dokumentationsstandard zugrunde. Die von der EQuIP entwickelten Qualitätsindikatoren stellen einen Itempool dar, aus dem je nach Gesundheitssystem und Anwendungsfeld relevante Kriterien ausgewählt und als Basis für die Entwicklung eines nationalen Bewertungsinstruments benutzt werden können. Sie sind in essentielle, notwendige und wünschenswerte Items unterteilt. Darüber hinaus wurde von EQuIP ein eigenes Beurteilungsinstrument für die Qualität elektronischer Patientenakten entwickelt, das „Electronic Medical Record Audit Tool“ (EMRAT).

Die Prüfdimensionen des EMRAT umfassen:

- 1) Demographische Patientendaten
- 2) Liste der Hauptprobleme
- 3) Dynamik und Entwicklung von Ereignissen
- 4) Beratungsblatt (jeder Besuch wird in einem strukturierten Formular dokumentiert, das einer logischen SOAP-Sequenz klinischer Überlegungen folgt)
  - a) Subjektive Daten
  - b) Objektive Daten
  - c) Assessment
  - d) Plan
- 5) Medikation
- 6) Präventive Maßnahmen und Risikofaktoren
- 7) Laboruntersuchungen, Bildgebung und andere Untersuchungen
- 8) Chronische Erkrankungen - Follow-up
- 9) Familienanamnese
- 10) Informationsaustausch und Speicherung von Dokumenten

Das EMRAT wurde in Deutschland bisher in einer Studie zur Dokumentationsqualität in deutschen Hausarztpraxen getestet (KÖRNER et al. 2004). Tabelle 2 zeigt die Ergebnisse von 300 ausgewerteten Patientenakten aus 10 allgemeinmedizinischen Praxen in der Region Heidelberg. Da die möglichen Ergebnisse der Items im EMRAT entweder 5 oder 10 Punkte betragen, wurden die Mittelwerte der Teilnehmer jeweils in Prozent der möglichen Punkte angegeben.

**Tabelle 2: Ergebnisse des EMRAT von 300 Patientenakten aus 10 Hausarztpraxen in Heidelberg (aus: KÖRNER et al. 2004)**

(MW=Mittelwert, %=Prozent der maximal erreichbaren Punktzahl, SD=Standardabweichung)

Item	Min.	Max.	MW	%	SD
Ergebnisse bildgebender Verfahren in chronologischer Ordnung	10	10	10	100	0
Laborergebnisse in chronologischer Ordnung	10	10	10	100	0
Demographische Angaben	3,5	5	4,3	86	0,51
Patientenakte sicher vor unbefugtem Zugriff	5,9	8	7,2	72	0,65
der verantwortliche Arzt klar identifiziert	0	5	3,5	70	2,42
Probleme der letzten 5 J. in der Problemliste	0	10	6,5	65	3,59
Diagnose/Problem definiert	1,6	4	2,8	56	0,75
dokumentiert, was dem Pat. gesagt wurde	0	9,6	2	40	3,44
Diagnose/Problem durch Daten gestützt	1,1	2,7	1,9	38	0,6
Behandlungsplan	0,8	2,4	1,6	32	0,56
Behandlungsanlass/Patientenbeschwerden	0,1	3,1	1,4	28	1,09
Angaben zu Allergien	0,3	3,5	1,6	16	1,05
Raucher/Nichtraucher	0	9,1	1,6	16	2,82
Behandlungsdauer für Medikamente angegeben	0	3	0,3	6	0,95

Es wird deutlich, dass diejenigen Items, die formale Aspekte (chronologische Ordnung von Labor- und anderen Befunden, Datenschutz, demografische Angaben) bewerten, bessere Ergebnisse zeigen als Items zu inhaltlichen Aspekten (Behandlungsanlass definiert; dokumentiert, was dem Patienten gesagt wurde; Behandlungsdauer für Medikamente, etc.).

### **2.1.19 Nutzung einer elektronischen Patientenakte**

PRINGLE et al. (1995) untersuchten Vollständigkeit und Korrektheit von elektronischen Patientenakten in vier britischen Hausarztpraxen anhand eines Vergleichs mit den Papierakten und mit Videoaufzeichnungen von Arzt-Patienten-Kontakten. Die Autoren fanden hohe Werte für Vollständigkeit bei Erkrankungen wie Diabetes mellitus und Glaukom (97% und 92%), während Risikoverhalten sowohl in der Papier- als auch in der elektronischen Patientenakte seltener dokumentiert wurde. Nur in 52% der Fälle waren Angaben zum Raucherstatus und in 38% zu Alkoholkonsum enthalten. Im Vergleich mit der Papierakte waren in der elektronischen Patientenakte in 82% der Fälle Diagnosen und in 67% Überweisungen

vollständig dokumentiert. Besonders gut war die Übereinstimmung für Medikamentenverordnungen: In allen Fällen (100%) waren in der Papierakte enthaltene Arzneimittelverordnungen auch im Computer zu finden. Auch beim Vergleich mit den Videoaufzeichnungen bestätigte sich die hohe Validität von Diagnosen, Überweisungen und Medikamentenverordnungen. Sonstige Informationen über den Patienten fanden sich dagegen fast nur in der Papierakte und nicht in der elektronischen Patientendokumentation.

In einer Übersichtsarbeit analysierten THIRU et al. (2003) die Studienlage im primärärztlichen Bereich und fanden einen durchgängig hohen PPW für Daten in der elektronischen Patientenakte, während die Sensitivität je nach Art der untersuchten Daten variabel war. Angaben zu Medikamentenverordnungen zeigten dabei die höchste Validität. In einer weiteren Übersichtsarbeit analysierten JORDAN et al. (2004) sämtliche vorliegenden Studien aus dem hausärztlichen Bereich in Großbritannien, die die Korrektheit und Vollständigkeit der Codierung von Erkrankungen in elektronischen Patientenakten mittels Read-Codes untersuchen. Die Autoren kamen zu dem Ergebnis, dass die Validität der Diagnosencodierung von der Art der Erkrankung abhängig ist.

### ***2.1.20 Zusammenfassung der Studienergebnisse zu potentiellen Einflussfaktoren auf die Validität von Abrechnungsdaten im hausärztlichen Bereich***

Die Ergebnisse der dargestellten Studien ermöglichen Vermutungen darüber, welche Faktoren die Validität von Abrechnungsdiagnosen insbesondere im hausärztlichen Bereich beeinflussen könnten. Es wäre denkbar, dass die Validität von Abrechnungsdaten in solchen Ländern am höchsten ist, die im Rahmen der Nutzung einer elektronischen Patientenakte eine episodensbasierte, problemorientierte Dokumentation als Dokumentationsstandard vorgeben, sowie ein speziell für die Abbildung der Morbidität von hausärztlichen Patienten konzipiertes Codierungssystem verwenden.

In der Literatur finden sich nur wenige Studien zu einzelnen dieser Themenbereiche. Beispielsweise liegen bisher keine vergleichenden Analysen des potentiellen Einflusses von chronologischer und episodensbezogener problemorientierter Dokumentation auf die Validität von hausärztlichen Abrechnungsdiagnosen vor. Allerdings fielen bei der Prüfung des ACG-Systems für den Einsatz im primärärztlichen Bereich in Spanien höhere prädiktive Werte („ $R^2$ “) auf, als die Entwickler mit US-amerikanischen Daten erreichen konnten (ORUETA et al. 2006, ORUETA et al. 1999). Da dieselben Diagnosen in den meisten Patientenklassifikationssystemen pro Untersuchungszeitraum nur einmal berücksichtigt werden, entspräche das Konzept der Behandlungsepisode eher den Vorgaben des Systems als

---

die Dokumentation von Diagnosen für jede Konsultation, wie es im amerikanischen „fee-for-service“-System vorgeschrieben ist. Die hohe Vorhersagekraft des ACG-Modells im spanischen Primärarztssystem könnte daher darin begründet liegen, dass die Validität der verwendeten Diagnosen durch eine episodenzugehörige Dokumentation besonders hoch war.

Die genannten potentiellen Einflussfaktoren auf die Validität von Abrechnungsdiagnosen wurden bisher weder im Zusammenhang noch speziell im Hinblick auf die Einführung eines diagnosenbasierten Patientenklassifikationssystems analysiert. In der im Folgenden vorgestellten empirischen Untersuchung sollte daher einerseits die Validität von Abrechnungsdiagnosen in hausärztlichen Praxen ermittelt und andererseits die genannten potentiellen Einflussfaktoren auf die Diagnosenvalidität im Kontext der Einführung eines diagnosenbasierten Risikoadjustierungssystems untersucht werden.