

## 6. Diskussion

Im Folgenden sollen die Leitfragen der Untersuchung

- wie hoch die Validität von hausärztlichen Abrechnungsdiagnosen im Vergleich zu anderen Ländern ist,
- ob die Klassifizierung von Patienten durch ein Risikoadjustierungssystem verändert wird, wenn hausärztliche Abrechnungsdiagnosen eines Jahres herangezogen werden im Vergleich zu den im gleichen Zeitraum in den Patientenakten dokumentierten Behandlungsanlässen, und
- welche Rolle die in der Literatur beschriebenen Einflussfaktoren auf die Validität von Abrechnungsdiagnosen im hausärztlichen Bereich spielen,

anhand der vorgestellten Ergebnisse der eigenen Studie diskutiert und in Bezug zur internationalen Literatur gesetzt werden. Dazu wird zunächst auf die methodisch bedingten Einschränkungen der Studie eingegangen, und es werden mögliche Auswirkungen auf die Übertragbarkeit der Ergebnisse aufgezeigt. Anschließend werden die in Kapitel 3 im Rahmen der Leitfragen aufgestellten Hypothesen diskutiert. Im letzten Abschnitt werden aus den Ergebnissen der Untersuchung Maßnahmen abgeleitet, die möglicherweise dazu geeignet wären, die Validität von hausärztlichen Abrechnungsdiagnosen zu erhöhen.

### Diskussion der Methode

#### 6.1.1 Auswahl der Praxen

Die Art der Stichprobenziehung der Studienpraxen resultierte in einer Auswahl von Hausärzten, denen der Umgang mit einem EDV-System vermutlich vertrauter war als anderen Praxen. Da dies auch für diejenigen Ärzte zutreffen müsste, die Papierakten verwendeten, könnte hierin eine mögliche Erklärung für den geringen Unterschied zwischen den beiden Gruppen bezüglich des Einflusses auf die Diagnosenvalidität liegen. Nach Abzug der stichprobenneutralen Ausfälle resultierte eine nur geringe Responserate von 7%. Da die Inhalte der Untersuchung für die Ärzte sensible Daten betrafen, erscheint es möglich, dass die Studienpraxen sich zur Mitarbeit bereit erklärten, weil sie der Meinung waren, die Qualität ihrer Patientendokumentation sei hierfür ausreichend gut. Es ist daher wahrscheinlich, dass eine Positivauswahl in Bezug auf Dokumentationsqualität und EDV-Nutzung vorliegt, die nicht den durchschnittlichen Verhältnissen in Hausarztpraxen entspricht. Da aus Zeitgründen

keine Nonresponderanalyse durchgeführt werden konnte, kann diese Frage nicht abschließend beantwortet werden.

### **6.1.2 Erhebungsmethode**

Die Untersuchung von TRAUTNER et al. (2005) hatte vor allem Probleme bei der Fortschreibung chronischer Diagnosen gezeigt. Aus diesem Grund sollte ein Schwerpunkt der Studie auf der Analyse chronischer Erkrankungen liegen. Die Art der Stichprobenziehung für die Patienten führte dazu, dass die Stichprobe einen höheren Anteil älterer Patienten mit chronischen Krankheiten enthielt, als dem Durchschnitt in den untersuchten Praxen entsprochen hätte. Sie repräsentierte insofern nicht die „normale“ Patientenklientel des einzelnen Hausarztes, sondern die Patienten mit dem höchsten Ressourcenverbrauch. Dieser Gruppe würde bei einer (prospektiven) Risikoadjustierung jedoch ein größeres „Gewicht“ zukommen.

Ein weiteres Problem in diesem Zusammenhang stellte die Tatsache dar, dass für die Berechnung der Validitätsparameter aus Datenschutzgründen nicht die realen, von den Ärzten im Untersuchungszeitraum angegebenen Abrechnungsdiagnosen und EBM-Ziffern der KV Berlin verwendet werden konnten, sondern auf die Daten in den Abrechnungsdateien der Ärzte zurückgegriffen werden musste. Trotz der Verifizierung der Abrechnungsdiagnosen durch jeden einzelnen Hausarzt erscheint es denkbar, dass bei der Zusammenstellung und Übermittlung an die KV in Einzelfällen Diagnosen zusätzlich angegeben wurden bzw. fehlten, was die Diagnosenvalidität sowohl erhöht als auch erniedrigt haben könnte. Andererseits wurden alle Abrechnungsdiagnosen im Zweifelsfall als korrekt gewertet, wie z.B. inaktive „Dauerdiagnosen“, wenn sie in pathogenetischen Zusammenhang mit einem aktiven Behandlungsanlass zu stehen schienen. Durch dieses Vorgehen könnte die Validität eher zu hoch geschätzt worden sein. Bezüglich der ermittelten Kostendaten für die Patienten ist es möglich, dass durch Budgetüberschreitungen, Einschränkungen der Abrechenbarkeit einzelner Ziffern und einen niedrigeren Punktwert als 0,05 € im Untersuchungszeitraum letztendlich geringere Kosten resultierten als errechnet. Da jedoch die tatsächlich entstehenden Kosten pro Patient approximiert werden sollten und diese Umstände außerdem für alle Patienten gleichermaßen gelten mussten, erschien dieses Problem weniger relevant. Die Kassenärztliche Bundesvereinigung hat in einer eigenen Studie einen ähnlichen Ansatz für die Berechnung der Behandlungskosten gewählt (VON STILLFRIED und RYLL 2004). In diesem Zusammenhang muss darauf hingewiesen werden, dass für die Klassifikation von Patienten mittels eines Risikoadjustierungssystems in der Realität nicht nur hausärztliche

Diagnosen herangezogen würden, sondern Abrechnungsdiagnosen aller (Fach-)Ärzte, die der KV für einen Patienten im Abrechnungszeitraum vorlägen. Je nach Validität dieser Diagnosen könnte sich die Übereinstimmung zwischen Behandlungsanlässen und Abrechnungsdiagnosen dadurch sowohl erhöhen als auch erniedrigen. In jedem Fall ist es wahrscheinlich, dass eine solche Erweiterung der Datengrundlage zu einer veränderten Einschätzung des Ressourcenverbrauchs von Patienten im ACG-System geführt hätte.

Ein weiteres Problem betraf die Identifikation der Behandlungsanlässe anhand der Patientenakten und die Codierung nach dem ICD-10-System. Da sie nur durch einen Untersucher erfolgte, könnte dadurch ein entsprechender Bias vorliegen. Daher wurden die ermittelten und codierten Behandlungsanlässe in zwei Praxen vollständig und in den übrigen für eine sechzigprozentige Stichprobe von den Ärzten selbst anhand der Patientenakte noch einmal verifiziert und gegebenenfalls korrigiert. Da sich in allen Praxen eine Übereinstimmung von 74%-99% zwischen Hausärzten und Untersucher gezeigt hatte, wurde davon ausgegangen, dass die identifizierten Behandlungsanlässe die im Untersuchungszeitraum behandelten Erkrankungen der Patienten in der Mehrzahl der Fälle korrekt abbilden müssten.

### **6.1.3 Repräsentativität der Ergebnisse**

Bei der vorliegenden Untersuchung handelt es sich um eine eher explorativ angelegte Studie mit dem Ziel, den Prozess der Generierung und Codierung von hausärztlichen Abrechnungsdiagnosen sowie Einflussfaktoren auf die Validität dieser Daten zu untersuchen, diesbezügliche Probleme zu identifizieren und Maßnahmen abzuleiten, welche die Diagnosenvalidität erhöhen könnten. Die Ergebnisse beziehen sich auf eine relativ kleine Anzahl von überwiegend älteren und chronisch kranken Patienten aus wenigen, bezüglich der Qualität der Dokumentation und EDV-Kompetenz vermutlich positiv selektierten Hausarztpraxen in einem großstädtischen KV-Bezirk. Die Repräsentativität der Ergebnisse erscheint aufgrund der selektierten Patientenklientel und der geringen Stichprobenzahl insofern begrenzt. Andererseits entsprach die Verteilung des Geschlechts, die Krankenkassenzugehörigkeit und die Art der häufigsten hausärztlichen Diagnosen (Hypertonie, Hyperlipidämie, Rückenschmerzen, KHK und Diabetes mellitus) weitgehend der des ZI-Panels als repräsentativer Stichprobe von 77.500 im ersten Quartal 2003 behandelten Patienten aus 60 Hausarztpraxen in der KV-Region Nordrhein (Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland 2003a). Altersverteilung und Prävalenz dieser Erkrankungen zeigten dagegen einen Rechtsverschiebung zu höheren

Altersklassen. Dementsprechend fand sich im Vergleich zum ZI-Panel ein um 19% höherer Anteil an Patienten mit dem Kassenstatus „Rentner“. Die Werte für die Prävalenz der genannten Erkrankungen waren mit denen höherer Altersklassen in nationaler Studien vergleichbar (GERSTE und GUTSCHMIDT 2006, WOLF-MAIER et al. 2003, THEFELD 1999, THAMM 1999). In der DETECT-Studie wurde eine repräsentative Zufallsauswahl von 55.518 Patienten aus 3795 hausärztlichen Praxen in Deutschland über ein Jahr untersucht (WITTCHEN et al. 2005). 37% der Patienten wiesen eine Hypertonie (eigene Studie 41%), 30% eine Hyperlipoproteinämie (eigene Studie 36%), 15% einen Diabetes mellitus (eigene Studie 18%) und 12% eine KHK auf (eigene Studie 21%). Alle untersuchten Erkrankungen nahmen mit dem Alter zu: Während in der Altersgruppe der 46-65jährigen beispielsweise 10% eine KHK aufwiesen, stieg die Zahl bei den über 66jährigen auf 27%. Die Befunde weisen darauf hin, dass die Studienpatienten entsprechend der Art der Stichprobenziehung die Gruppe der älteren Patienten mit chronischen Erkrankungen in der Hausarztpraxis relativ gut repräsentieren. Da außerdem eine statistische Fallzahlberechnung durchgeführt wurde, kann davon ausgegangen werden, dass die Ergebnisse der Studie für die genannte Gruppe Allgemeingültigkeit haben dürften. Die vorliegenden Ergebnisse lassen jedoch keine Aussage darüber zu, ob die Morbidität der Patientenklientel von Hausarztpraxen anhand ambulanter Diagnosen so abgebildet wird, dass entsprechende Regelleistungsvolumina berechnet werden können, weil ausschließlich hausärztliche Abrechnungsdiagnosen untersucht wurden. Die Beantwortung dieser Frage muss daher zukünftigen, größer angelegten Studien vorbehalten bleiben, die den realen Querschnitt der Patienten von hausärztlichen Praxen abbilden, fachärztliche Diagnosen einbeziehen und unterschiedliche KV-Bezirke, sowie Stadt- und Landpraxen z.B. mittels der Methode der Clusterrandomisierung einbeziehen.

## **Diskussion der Ergebnisse**

### ***6.1.4 Validität hausärztlicher Abrechnungsdiagnosen***

***Hypothese 1: Die generelle Übereinstimmung zwischen Abrechnungsdiagnosen und Behandlungsanlässen ist eher gering, da z.B. auch Diagnosen als Abrechnungsdiagnosen angegeben werden, für die keine Behandlungsleistungen erbracht wurden.***

In der Untersuchung war von der Hypothese ausgegangen worden, dass die Übereinstimmung von Abrechnungsdiagnosen und Behandlungsanlässen insgesamt gering sei, da vor allem die Angabe von im Untersuchungszeitraum nicht behandelten Diagnosen („Dauerdiagnosen“) die Validität von Abrechnungsdiagnosen durch ein „Overreporting“ vermindern würde. Dieser Verdacht wurde dadurch unterstützt, dass in Deutschland eine höhere Anzahl von ca. zehn

Abrechnungsdiagnosen pro Patient vorlag als in anderen Ländern (LAUTERBACH und LÜNGEN 2005). Anhand der internationalen Literatur war dagegen das Problem identifiziert worden, dass viele Behandlungsanlässe nicht als Abrechnungsdiagnosen angegeben werden („Underreporting“).

In der vorliegenden Studie fanden sich im Jahr 2003 im Mittel 6,1 Abrechnungsdiagnosen pro Patient. Im Vergleich dazu wiesen die Patienten im ZI-Panel im ersten Quartal 2003 durchschnittlich 4,6 Diagnosen auf, wobei der Altersdurchschnitt im Panel mit 47 Jahren niedriger lag als in der eigenen Untersuchung (55 Jahre). Studien aus Kanada (REID et al. 2001) zählten 3,6 (Provinz Manitoba) bzw. 4,6 (Provinz British Columbia) Diagnosen pro Patient und Jahr, während in Spanien bei einem Durchschnittsalter von 43 Jahren im Mittel 2,3 Diagnosen festgestellt wurden (ORUETA et al. 2006). Allerdings sollten bei einem solchen Vergleich Unterschiede in der Datengrundlage beachtet werden. In der spanischen Studie wurde eine episodenzugeordnete Dokumentation verwendet, d.h. alle Behandlungsanlässe für ein Gesundheitsproblem wurden immer der gleichen Episode zugeordnet. Pro Jahr wurden nur diejenigen Episoden als Diagnosen codiert, die in dieser Zeit behandelt wurden. In Kanada wie auch in Deutschland werden Hausärzte nach Einzelleistungen vergütet. Jede abgerechnete Leistung erfordert die Angabe einer Diagnose als Begründung. Die größere Anzahl an Diagnosen pro Patient in Kanada und Deutschland im Vergleich zu Spanien könnte daher in den genannten Unterschieden im Dokumentations- und Abrechnungsverfahren begründet liegen.

Die generelle Übereinstimmung von Abrechnungsdiagnosen und Behandlungsanlässen in der vorliegenden Untersuchung betrug 40,1%. STEINWACHS et al. (1998) fanden eine Übereinstimmung zwischen Patientenakte und Abrechnungen von 82%, wobei lediglich die Diagnosen Diabetes, Hypertonie, Asthma, Otitis media, Schwangerschaft und Kindervorsorge untersucht wurden. Bei HORNER et al. (1991) ergab sich eine Übereinstimmungsrate von 60% zwischen Abrechnungsdiagnosen und Behandlungsanlässen. Durch die Verwendung von größeren ICD-Kategorien ließ sich die Übereinstimmung auf 78% erhöhen. In 42% der Fälle lag ein Behandlungsanlass ohne Abrechnungsdiagnose vor („Underreporting“), die umgekehrte Situation („Overreporting“) fand sich nur in 3% der Fälle und in 27% war eine Abrechnungsdiagnose inkorrekt.

In der eigenen Studie lagen die Ursachen für die hohe Rate von Abrechnungsdiagnosen, die nicht mit einem Behandlungsanlass übereinstimmten (59,9%), in 29,7% der Fälle darin begründet, dass Behandlungsanlässe von den Ärzten nicht als Abrechnungsdiagnosen codiert worden waren („Underreporting“). Demgegenüber erklärte die Angabe von Diagnosen, die im Untersuchungszeitraum keine Behandlungsleistung erfordert hatten („Overreporting“), 19,0% der nicht-übereinstimmenden Fälle. In den verbleibenden 11,2% der Fälle waren Diagnosen inkorrekt codiert worden. Das Ergebnis ließ sich nicht dadurch verbessern, dass statt vier- und fünfstelliger dreistellige ICD-10-Codes verwendet wurden. Am häufigsten fehlten Maßnahmen wie Vorsorgeuntersuchungen, Impfungen und psychosoziale Beratung, gefolgt von unspezifischen Symptomen (Bauchschmerz, Schwindel, Gangstörung, Hyperglycämie) und endokrinen Erkrankungen (Hyperlipoproteinämie, Hyperuricämie, Schilddrüsenerkrankung; insgesamt 48,3% aller fehlenden Diagnosen). Bei den 19,0% zusätzlichen Diagnosen („Overreporting“) handelte es sich vor allem um Wirbelsäulen- und Gelenkbeschwerden, Venenerkrankungen, Hypertonie, Hyperlipoproteinämie, Hyperuricämie, Adipositas und psychische Erkrankungen (entsprach 52,3% aller zusätzlichen Diagnosen). Zum Vergleich war die Übereinstimmung in der Studie von HORNER et al. (1991) bei endokrinen und psychischen Krankheiten, Vorsorgeuntersuchungen, sozialen Problemen und Symptomen am geringsten (entsprach 45% der nicht-übereinstimmenden Diagnosen). Diagnosen mit der höchsten Übereinstimmung stammten in der eigenen Untersuchung am häufigsten aus den ICD-10-Kapiteln Herz-Kreislaufsystem (I; 69% korrekte Diagnosen), endokrines System (E; 60% korrekte Diagnosen) und Muskel- und Skelettsystem (M; 53% korrekte Diagnosen).

Da die generelle Übereinstimmung zwischen Abrechnungsdiagnosen und Behandlungsanlässen mit 40,1% als gering anzusehen ist, konnte der erste Teil der untersuchten Hypothese verifiziert werden. Der zweite Teil der Hypothese bestätigte sich jedoch nicht. Die Ergebnisse der vorliegenden Untersuchung deuten darauf hin, dass trotz der Vielzahl von Abrechnungsdiagnosen pro Patient ein „Underreporting“ der im Zeitraum eines Jahres behandelten Erkrankungen vorliegt, d.h. auch in Deutschland scheint die mangelnde Vollständigkeit von Abrechnungsdiagnosen ein Problem zu sein, das die Validität von Abrechnungsdiagnosen stärker beeinflusst als die Angabe von Dauerdiagnosen („Overreporting“).

***Hypothese 2: Es bestehen Defizite in der Vollständigkeit und Korrektheit von häufigen hausärztlichen Abrechnungsdiagnosen.***

Die Diagnosen Hypertonie, Hyperlipidämie, Rückenschmerzen, KHK und Diabetes mellitus stellen vermutlich die häufigsten Erkrankungen des hausärztlichen Behandlungsspektrums in Deutschland dar (Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland 2006a, 2006b, 2003a, 2003b). Daher wurden für diese Diagnosen sowie für akute Atemwegserkrankungen und Impfungen Kappa-Werte, Sensitivität und positiv prädiktiver Wert (PPW) als Maß für die Übereinstimmung, Vollständigkeit und Korrektheit von Abrechnungsdiagnosen berechnet. Da der PPW von der Prävalenz einer Erkrankung abhängig ist, wurde diese anhand der Behandlungsanlässe in der Patientenakte als „Goldstandard“ ermittelt. Für sechs der Diagnosen wurden die Ergebnisse der Studienpatienten mit internationalen Studien zur Diagnosenvalidität aus dem ambulanten Bereich verglichen (für Impfungen existierten leider keine Vergleichswerte). Da es sich nur um drei Untersuchungen handelte, wurden zusätzlich Ergebnisse aus dem stationären Bereich herangezogen. Stichprobenumfang und Prävalenz der einzelnen Erkrankungen variierten in den verschiedenen Untersuchungen erheblich (siehe Tabelle 43).

**Tabelle 43: Stichprobenumfang und Prävalenz für sechs häufige Erkrankungen des hausärztlichen Behandlungsspektrums in Deutschland in der eigenen Untersuchung und in den vorliegenden internationalen Studien**

(HT=Hypertonie, HLP=Hyperlipoproteinämie, RS=Rückenschmerzen, KHK= Koronare Herzkrankheit, DM=Diabetes mellitus, AWI= akuter Atemwegsinfekt; rot=eigene Untersuchung, blau=Studien aus dem ambulanten (primärärztlichen) Bereich, schwarz=Studien aus dem stationären Bereich, - =nicht untersucht)

Studie	n	HT	HLP	RS	KHK	DM	AWI
eigene Untersuchung	250	41%	36%	30%	21%	18%	36%
Wilchesky et al. 2004	14980	52%	-	-	-	18%	-
Fowles et al. 1998	213	23%	-	18%	13%	10%	-
Fowles et al. 1995	1596	35%	9%	-	10%	14%	13%
Lee et al. 2005	1808	-	-	-	-	33%	-
Quan et al. 2002	1200	-	-	-	-	9%	-
Humphries et al. 2000	817	41%	41%	-	-	18%	-
Kieszak et al. 1999	1945	-	-	-	-	20%	-
Jollis et al. 1993	12937	51%	31%	-	-	19%	-
Kashner et al. 1998	414	7%	1%	-	4%	4%	-
Malenka et al. 1994	485	-	-	-	-	8%	-

Die Art der Stichprobenziehung, das daraus resultierende mittlere Alter der Patienten, der Untersuchungszeitraum und vor allem die Methodik für den Vergleich der Diagnosen sind als wichtige Einflussfaktoren bei der Interpretation der Ergebnisse zu berücksichtigen. Beispielsweise existierte in der Studie von FOWLES et al. (1995) die Vorschrift, dass eine

Erkrankung nur dann als in der Patientenakte vorhanden gezählt werden durfte, wenn eine entsprechende Diagnose vom behandelnden Arzt eingetragen worden war. Indirekte klinische Hinweise auf eine Erkrankung (z.B. Laborwerte oder Medikamentenverordnungen) wurden nicht berücksichtigt. In der Studie von FOWLES et al. (1998) wurden dagegen auch Abrechnungsdiagnosen als übereinstimmend gewertet, für die in der Patientenakte nur indirekte Hinweise existierten. WILCHESKY et al. (2004) verglichen Angaben zu Erkrankungen aus einem Fünfjahreszeitraum mit Abrechnungsdaten aus einem Jahr, während FOWLES et al. (1998) zunächst einen Zweijahreszeitraum analysierten, jedoch später Angaben in den Patientenakten aus fünf vorangegangenen Jahren benutzten, um Abrechnungsdiagnosen als vorhanden zu verifizieren, die im Untersuchungszeitraum nicht behandelt worden waren. In diesem Zusammenhang wird deutlich, dass genau differenziert werden muss, unter welcher Fragestellung Diagnosen in den zu vergleichenden Studien analysiert wurden: Wurde untersucht, welche Erkrankungen ein Patient generell aufwies oder wurden nur diejenigen Erkrankungen betrachtet, die im Untersuchungszeitraum behandelt worden waren? Durch diese Unterschiede in der Perspektive könnten die Prävalenz und der PPW der untersuchten Erkrankungen höher, die Sensitivität dagegen niedriger eingeschätzt worden sein als in der Studie von FOWLES et al. (1995), in der Abrechnungsdaten und Patientenakten aus einem Jahr untersucht wurden.

Die vorliegende Studie ist aufgrund des höheren Altersdurchschnitts und der Kontaktrate der Studienpatienten am ehesten mit der Untersuchung von WILCHESKY et al. (2004) zu vergleichen. Wie in der Studie von FOWLES et al. (1995) wurden nur Diagnosen im Untersuchungszeitraum einbezogen, es zählten jedoch auch indirekte Hinweise auf vorhandene Diagnosen in der Patientenakte (siehe Tabelle 44).

**Tabelle 44: Wichtige Charakteristika internationaler Studien zur Diagnosenvalidität im primärärztlichen Bereich im Vergleich zur vorliegenden Untersuchung**

Studie	Alter der Patienten in Jahren	Vergleichszeitraum		methodische Besonderheiten
		Abrechnung	Akte	
eigene Untersuchung	57 (+/-18)	1 Jahr	1 Jahr	Arztkontakt im Vorquartal und im Untersuchungszeitraum indirekte Diagnoseangaben in der Patientenakte zählen mehrere ICD-10-Codes entsprechen einer Diagnose
Wilchesky et al. 2004	über 66	1Jahr	5 Jahre	mindestens zwei Arztkontakte im Untersuchungszeitraum mehrere ICD-9-Codes entsprechen einer Diagnose
Fowles et al. 1998	18-64 (50%) >65 (50%)	2 Jahre	2 Jahre (5 Jahre)	indirekte Diagnoseangaben in der Patientenakte zählen
Fowles et al. 1995	keine Angabe	1 Jahr	1 Jahr	nur direkte Diagnoseangaben in der Patientenakte zählen

Die Diagnosen Diabetes mellitus und Hypertonie waren am häufigsten untersucht worden und wiesen die höchste Übereinstimmung, Vollständigkeit und Korrektheit auf. Die Prävalenz dieser Erkrankungen war mit den Ergebnissen der Stichprobe vergleichbar. Anhand der vorliegenden Untersuchung kann davon ausgegangen werden, dass die Abrechnungsdiagnose Diabetes mellitus in 86% mit einem entsprechenden Behandlungsanlass übereinstimmt. In 85% der Fälle weist ein Patient mit der Diagnose Diabetes in der Patientenakte diese auch als Abrechnungsdiagnose auf, und in 89% der Fälle kann davon ausgegangen werden, dass bei einem Patienten mit der Abrechnungsdiagnose Diabetes Hinweise auf das Vorliegen dieser Erkrankung in der Patientenakte existieren, was die internationalen Ergebnissen bestätigt. Die Diagnose Hypertonie wies in der eigenen Studie einen höheren Kappa-Wert, aber ähnliche Werte für Vollständigkeit und Korrektheit auf als in den vergleichbaren Untersuchungen. Bei allen analysierten Erkrankungen lag die Spezifität mit 99% im Rahmen der internationalen Ergebnisse. Ein Vergleich der Übereinstimmung, Vollständigkeit und Korrektheit für die übrigen Erkrankungen gestaltete sich aufgrund der Heterogenität und der unterschiedlichen Perspektiven (behandelte versus vorhandene Erkrankungen) in den internationalen Studien schwierig. Außerdem waren einige Diagnosen (z.B. Rückenschmerzen und akute

---

Atemwegsinfektionen) selten untersucht worden, so dass nur wenige Ergebnisse vorlagen. Aus dem ambulanten Bereich stand für die meisten Diagnosen nur die Studie von FOWLES et al. (1995) für einen Vergleich zur Verfügung, bei der die angewandte Methodik zu den insgesamt niedrigen Werten für Prävalenz, Kappa, Sensitivität und PPW geführt haben könnte. Dagegen zeigte die Untersuchung von FOWLES et al. (1998) trotz niedrigerer Prävalenz der Erkrankungen als in der vorliegenden Studie ähnlich gute Ergebnisse für den PPW bei den Diagnosen Rückenschmerzen und KHK, während die Sensitivität etwas geringer war (siehe Tabelle 45).

**Tabelle 45: Kappa, Sensitivität, Spezifität und positiv prädiktiver Wert (PPW) für sechs häufige Erkrankungen des hausärztlichen Behandlungsspektrums in Deutschland im Vergleich zu internationalen Studienergebnissen**

(HT=Hypertonie, HLP=Hyperlipoproteinämie, RS=Rückenschmerzen, KHK= Koronare Herzkrankheit, DM=Diabetes mellitus, AWI= akuter Atemwegsinfekt; rot=eigene Untersuchung, blau=ambulanter (primärärztlicher) Bereich, schwarz=stationärer Bereich, -=nicht untersucht; Kappa<0=schlechte Übereinstimmung („poor agreement“), Kappa 0-0,20=geringe Übereinstimmung („slight agreement“), Kappa 0,21-0,40=mäßige Übereinstimmung („fair agreement“), Kappa zwischen 0,41-0,60=mittlere Übereinstimmung („moderate agreement“), Kappa 0,61-0,80=gute Übereinstimmung („substantial agreement“) Kappa>0,81=sehr gute Übereinstimmung („almost perfect agreement“) jenseits des Zufalls (LANDIS und KOCH 1977)

	Studie	HT	HLP	RS	KHK	DM	AWI
<b>Kappa</b>	eigene Untersuchung	0,86	0,56	0,71	0,76	0,86	0,70
	Fowles et al. 1995	0,45	0,25		0,27	0,73	0,32
	Quan et al. 2002	-	-	-	-	0,74	-
	Humphries et al. 2000	0,58	0,31	-	-	0,83	-
	Kashner et al. 1998	0,66	0,80	-	0,64	0,82	-
	Malenka et al. 1994	-	-	-	-	0,86	-
	Jollis et al. 1993	0,56	0,31	-	-	0,83	-
<b>Sensitivität</b>	eigene Untersuchung	81%	49%	74%	71%	85%	58%
	Wilchesky et al. 2004	61%	-	-	-	52%	-
	Fowles et al. 1998	90%	-	53%	64%	100%	-
	Fowles et al. 1995	65%	22%	-	33%	76%	25%
	Lee et al. 2005	-	-	-	-	81%	-
	Quan et al. 2002	-	-	-	-	82%	-
	Humphries et al. 2000	60%	32%	-	-	79%	-
	Kieszak et al. 1999	-	-	-	-	68%	-
Jollis et al. 1993	65%	36%	-	-	83%	-	
<b>Positiv prädiktiver Wert (PPW)</b>	eigene Untersuchung	93%	76%	71%	82%	89%	91%
	Fowles et al. 1995	64%	46%	-	35%	76%	73%
	Fowles et al. 1998	91%		76%	78%	100%	
	Lee et al. 2005	-	-	-	-	96%	-
	Quan et al. 2002	-	-	-	-	71%	-
	Humphries et al. 2000	88%	84%	-	-	93%	-
<b>Spezifität</b>	eigene Untersuchung	99%	99%	99%	99%	99%	99%
	Wilchesky et al. 2004	88%	-	-	-	98%	-
	Fowles et al. 1998	98%	-	97%	97%	100%	-
	Fowles et al. 1995	80%	96%	-	90%	86%	96%
	Lee et al. 2005	-	-	-	-	98%	-
	Quan et al. 2002	-	-	-	-	97%	-
	Humphries et al. 2000	95%	96%	-	-	99%	-
	Jollis et al. 1993	91%	91%	-	-	98%	-

Trotz der geringen Werte für die Gesamtstichprobe war die Übereinstimmung zwischen Abrechnungsdiagnosen und Behandlungsanlässen bei den meisten der untersuchten Erkrankungen hoch. Dieses Ergebnis liegt unter anderem darin begründet, dass die

untersuchten Diagnosen aus den ICD-10-Kapiteln mit dem höchsten Anteil an korrekten Abrechnungsdiagnosen stammten. Die beiden Diagnosen Hyperlipoproteinämie und akuter Atemwegsinfekt, die eine niedrigere Übereinstimmung jenseits des Zufalls aufwiesen als die übrigen Erkrankungen, zeigten eine gute Korrektheit bei einer tendenziell geringeren Vollständigkeit („Underreporting“). Die Hypothese, dass häufige hausärztliche Abrechnungsdiagnosen eine mangelnde Vollständigkeit und Korrektheit aufweisen, konnte insgesamt somit nicht bestätigt werden. Allerdings ist dabei zu berücksichtigen, dass die analysierten ICD-10-Codes aufgrund der kleinen Stichprobe zu größeren Kategorien ähnlich der ICD-9-Einteilung zusammengefasst werden mussten, was die Übereinstimmung im Vergleich zu einer drei- oder vierstelligen Codierung erhöht hat.

In Bezug auf die Einführung eines Patientenklassifikationssystems könnte ein solches „Underreporting“ von Erkrankungen möglicherweise zu einer Unterschätzung der Morbidität und damit des Kostenaufwands von Patienten führen. Dabei ist von Bedeutung, zu welchem Zweck Abrechnungsdiagnosen vom Arzt dokumentiert werden, nämlich um durchgeführte Untersuchungen und therapeutische Maßnahmen zu begründen. In diesem Zusammenhang sollen nur die in einem bestimmten Zeitraum (z.B. einem Quartal oder einem Jahr) behandelten Erkrankungen eines Patienten einbezogen werden. Bedingt durch die langfristige Betreuung werden sich vor allem in der hausärztlichen Patientenakte Angaben zu vielen anderen aktiven und inaktiven Erkrankungen eines Patienten finden, die jedoch im Untersuchungszeitraum nicht behandlungsbedürftig waren. Wenn jedoch die umfassende Morbidität eines Patienten inklusive aller dauerhaft vorliegenden Gesundheitsprobleme, wie sie in der Akte dokumentiert ist, mit einem Morbiditätsausschnitt der in einem Jahr behandelten Diagnosen verglichen wird, folgt daraus zwangsläufig das Ergebnis eines „Underreportings“ bzw. eine niedrige Sensitivität.

Im Hinblick auf eine Risikoadjustierung stellt sich dabei die Frage, welche Daten für eine möglichst hohe Prognosegüte des Systems benötigt werden: Sollte die Morbidität eines Patienten möglichst umfassend mit allen vorliegenden Diagnosen dargestellt werden oder sollten nur die in einem Jahr behandlungsbedürftigen und damit Kosten verursachenden Diagnosen herangezogen werden? Eine mögliche Antwort auf diese Frage dürfte unter anderem davon abhängen, ob eine Risikoadjustierung zeitgleich (concurrent) oder prospektiv erfolgen soll. In einem zeitgleichen System werden vor allem die im betrachteten Zeitraum behandelten Erkrankungen berücksichtigt. In einem prospektiven Modell fallen chronische

Erkrankungen stärker ins Gewicht. Diagnosen, die im Vorjahr keine Behandlungsleistung erfordert haben, könnten im nächsten Jahr wieder aktiv werden und Kosten verursachen. Daher erscheint es für eine prospektive Risikoadjustierung eher sinnvoll, dass inaktive Erkrankungen berücksichtigt werden, bei denen ein substantielles Risiko besteht, dass sie in Zukunft eine erneute (möglicherweise kostenintensive) Behandlung erfordern.

***Hypothese 3: Behandlungsanlässe, die schwerwiegende chronische Erkrankungen betreffen, werden eher als Abrechnungsdiagnosen codiert als akute leichte Krankheiten.***

In der vorliegenden Untersuchung konnten für die mangelnde Übereinstimmung von Abrechnungsdiagnosen und Behandlungsanlässen insgesamt folgende Ursachen identifiziert werden:

1. ICD-10-Codes waren auch für im Abrechnungszeitraum nicht behandlungsbedürftige Erkrankungen angegeben worden („Overreporting“). Bei solchen Erkrankungen handelte es sich meist um chronische Diagnosen, die in über der Hälfte der Fälle in den Folgequartalen wieder eine Behandlungsleistung erfordern könnten, und die daher als latente Probleme eingeschätzt wurden.
2. Vor allem Verdachtsdiagnosen und nur laborchemisch nachgewiesene Diagnosen waren nicht als Abrechnungsdiagnosen angegeben worden, obwohl sie eine Behandlungsleistung erfordern hatten.
3. Für Maßnahmen aus dem Kapitel Z, wie Gesundheits- und Krebsvorsorgeuntersuchungen sowie Impfungen, wurden häufig keine ICD-10-Schlüssel, sondern nur Abrechnungsziffern angegeben.
4. Die Ätiologie einer Erkrankung war unklar.
5. Der Schweregrad einer Erkrankung war unklar.
6. Eine Erkrankung war zu spezifisch oder zu unspezifisch codiert worden.

In der internationalen Literatur fanden sich in einigen Studien aus dem stationären Bereich Hinweise darauf, dass die Sensitivität von Abrechnungsdiagnosen sowohl für Haupt- als auch für Nebendiagnosen bei schweren Erkrankungen und Prozeduren höher ist als bei leichten, akuten Krankheiten. In der eigenen Untersuchung war die Vollständigkeit der chronischen Erkrankungen Diabetes, Hypertonie und KHK höher als bei akuten Atemwegsinfekten. Die akute Diagnose Rückenschmerzen zeigte jedoch ebenfalls eine hohe Vollständigkeit, während die Sensitivität bei der chronischen Diagnose Hyperlipoproteinämie niedrig war, was daran lag, dass die meisten Fälle nur laborchemisch festgestellt, aber nicht medikamentös behandelt

worden waren. Diese Ergebnisse legen den Schluss nahe, dass die Angabe einer Erkrankung als Abrechnungsdiagnose im ambulanten Bereich weniger von der Schwere der Krankheit als von dem Aufwand der notwendigen medizinischen Therapiemaßnahmen abhängig sein könnte. Es ist denkbar, dass Erkrankungen dann seltener dokumentiert und verschlüsselt werden, wenn sie mit wenig Zeitaufwand zu behandeln sind, wie beispielsweise akute Atemwegsinfekte, wenn sie nur unspezifische Maßnahmen erfordern, wie ein Beratungsgespräch, oder wenn eine Krankheit zwar im Rahmen einer diagnostischen Untersuchung erkannt wird, aber keiner medikamentösen Therapie bedarf. Andererseits fiel bei der Morbiditätsbeschreibung von Patienten im ACG-System auf, dass anhand von Behandlungsanlässen eine höhere Anzahl von Patienten in niedrigere Morbiditätskategorien gruppiert worden waren als anhand ihrer Abrechnungsdiagnosen. Dieses Ergebnis würde für die eingangs aufgestellte Hypothese sprechen, dass akute Diagnosen mit einem geringeren Schweregrad seltener als Abrechnungsdiagnosen codiert werden.

#### ***6.1.5 Mögliche Auswirkungen der Validität hausärztlicher Abrechnungsdiagnosen auf die Morbiditätsklassifikation und auf die Schätzung des Ressourcenverbrauchs von Patienten im ACG-System***

***Hypothese 4: Die Morbiditätsklassifikation von Patienten im ACG-System fällt unterschiedlich aus, wenn Abrechnungsdiagnosen bzw. Behandlungsanlässe berücksichtigt werden.***

Sowohl anhand von Abrechnungsdiagnosen als auch anhand von Behandlungsanlässen ließen sich ca. 75% der Patienten 10 ACG-Kategorien zuordnen. Dieses Ergebnis findet sich auch in anderen europäischen Studien aus dem primärärztlichen Bereich: In Spanien waren 66% der Patienten, in Schweden bis zu 80% der Patienten in 11 bzw. 10 ACG-Klassen gruppiert worden (CARLSSON et al. 2006, CARLSSON et al. 2004, CARLSSON et al. 2002, ORUETA et al. 1999, JUNCOSA et al. 1999). In diesen Untersuchungen an repräsentativen hausärztlichen Patientenkollektiven fanden sich vor allem akute Erkrankungen, Kombinationen von zwei bis drei verschiedenen Krankheiten, rezidivierende und stabile chronische Erkrankungen sowie präventive und administrative Behandlungsleistungen, während Kombinationen von mehr als vier Krankheiten seltener vorkamen. DUCKETT und AGIUS (2002) untersuchten eine Selektion von „High users“ in allgemeinmedizinischen Praxen in Australien. Diese Patienten zeichneten sich durch ein hohes Alter (>68 Jahre), dem Vorliegen chronischer Erkrankungen und eine hohe Kontaktrate mit den Praxen aus und schienen daher mit der Patientenauswahl in der vorliegenden Untersuchung vergleichbar. Sowohl in der australischen Stichprobe als auch bei den eigenen Studienpatienten fanden sich

überwiegend ACG-Klassen, die eine höhere Morbidität anzeigten (2-3, 4-5 und 6-9 ADG-Kombinationen). Die übrigen beobachteten Kategorien (akute leichte und schwere, rezidivierende sowie chronische stabile Erkrankungen) waren zwar vorhanden, kamen aber seltener vor. Insofern entsprach die ACG-Klassifizierung der eigenen Studienpatienten insgesamt der von „High users“ (siehe Tabelle 46). Diesen Eindruck bestätigte der Vergleich mit den im ACG-System voreingestellten Werten eines amerikanischen Patientenkollektivs, anhand dessen unadjustierte Kostengewichte für die eigene Stichprobe berechnet worden waren: Die Patienten in der vorliegenden Untersuchung wiesen im Vergleich zu der repräsentativen amerikanischen Stichprobe einen dreimal höheren Ressourcenverbrauch auf.

**Tabelle 46: Häufigste ACG-Kategorien für hausärztliche Patienten in internationalen Studien (ENGSTROM et al. 2006, CARLSSON et al. 2006, DUCKETT und AGIUS 2002, CARLSSON et al. 2002, REID et al. 2001, ORUETA et al. 1999, JUNCOSA et al. 1999)**  
(ACG=Adjusted Clinical Group, ADG=Aggregated Diagnosis Group)

Rang	Häufigste ACG-Kategorien		Häufigste ACG-Kategorien „high user“	
1	300	Acute minor	4100	2-3 other ADG-Combinations
2	4100	2-3 other ADG-combinations	4400	4-5 other ADG-Combinations
3	500	Recurrent	4900	6-9 other ADG-Combinations
4	400	Acute major	900	Chronic medical stable
5	900	Chronic medical stable	800	Chronic medical unstable
6	1800	Acute minor, acute major	300	Acute minor
7	1600	Preventive/administrative	400	Acute major
8	2100	Acute minor, recurrent	2100	Acute minor, recurrent
9	2300	Acute minor, chronic medical stable	2300	Acute minor, chronic medical stable
10	4400	4-5 other ADG-combinations	1800	Acute minor, acute major

Die generelle Übereinstimmung der Zuordnung von Patienten zu den genannten 10 ACG-Gruppen anhand von Abrechnungsdiagnosen bzw. Behandlungsanlässen war mit 30% gering. Kappa-Werte für die fünf häufigsten Kategorien bewegten sich zwischen 0,07 und 0,27. Während die Spezifität hoch war, blieb die Sensitivität unter 50%, was die Ergebnisse der kanadischen Studie von WILCHESKY et al. (2004) bestätigte, in der die Übereinstimmung

auf der Ebene von ADGs analysiert wurde. Der PPW bewegte sich zwischen 13% und 44%. Nur die Kategorie 300 („Acute minor“) für leichte, kurz dauernde Erkrankungen wies einen PPW von 100% auf.

Anhand von Abrechnungsdiagnosen fanden sich 13% weniger Patienten mit akuten leichten Krankheiten bzw. nur einer chronischen Erkrankung (Morbiditätsgruppe 1) als anhand von Abrechnungsdiagnosen. Die Hälfte dieser Patienten war stattdessen in die ACG-Kategorie 4100 der Morbiditätsgruppe 2 eingestuft worden. 7% der Patienten mit 6-9 bzw. über 10 ADG-Kombinationen (Morbiditätsgruppe 3 und 4) wären anhand ihrer Behandlungsanlässe Kategorien mit 2-3 oder 4-5 ADG-Kombinationen (Morbiditätsgruppe 2) zugeordnet worden. Anhand von Abrechnungsdiagnosen waren dagegen 12% der Patienten in eine höhere Morbiditätskategorie gruppiert worden (anstatt einer chronischen Erkrankung 2-3 ADG-Kombinationen, anstatt 2-3 ADGs 4-5 bzw. 6-9 ADG-Kombinationen). In einer schwedischen Studie zur Morbidität von Patienten im primärärztlichen Bereich stellten CARLSSON et al. (2006) fest, dass das ACG-System sensibel auf die Vollständigkeit und Korrektheit von Diagnosen reagierte: Eine höhere Vollständigkeit von Diagnosen resultierte in einer höheren Zahl von komplexen ACG-Klassen. Die Analyse der fehlenden („untercodierten“) Abrechnungsdiagnosen in der eigenen Untersuchung hatte ergeben, dass in ca. 50% der Fälle akute Diagnosen (z.B. Symptome) und Maßnahmen nicht codiert worden waren, was offensichtlich zu einer geringeren Besetzung von ACG-Kategorien für leichte, akute Erkrankungen geführt hatte. In einem Drittel der Fälle waren außerdem chronische Erkrankungen wie Hyperlipoproteinämie, Hyperuricämie, KHK, Hypertonie und chronische muskuloskelettale Erkrankungen nicht als Abrechnungsdiagnosen angegeben worden, beispielsweise weil für sie nur eine Labor- oder Blutdruckkontrolle durchgeführt worden war. Die Einbeziehung dieser Diagnosen hätte in 12% der Fälle zu einer Eingruppierung in eine höhere Morbiditätskategorie geführt, wie der Vergleich mit den Behandlungsanlässen zeigte. Dagegen hatte die zusätzliche Einbeziehung von inaktiven, überwiegend chronischen „Dauerdiagnosen“ zur Folge, dass Patienten mit einer akuten Erkrankung vermehrt der ACG 4100 (2-3 other ADG combinations) bzw. Patienten mit 2-3 ADG-Kombinationen höheren Morbiditätsgruppen zugeordnet wurden. Eine zunehmende Anzahl an codierten Abrechnungsdiagnosen bzw. Behandlungsanlässen führte also auch in dieser Untersuchung zu einer Gruppierung von Patienten in komplexere ACG-Kategorien.

In diesem Zusammenhang hatten REID et al. (2001) in einer Studie mit kanadischen Hausarztpatienten beobachtet, dass die Zahl und Schwere der ADGs mit der Höhe des Ressourcenverbrauchs korrelierte. Da jede Diagnose einer ADG-Klasse zugeordnet wird, hat die Zahl der Diagnosen ebenfalls einen Einfluss auf die Schätzung der Behandlungskosten von Patienten. Entsprechend der beschriebenen Einteilung der eigenen Studienpatienten in ACG-Kategorien ergab deren Zusammenfassung zu Ressourcenverbrauchsklassen auf Basis von amerikanischen und eigenen Kostendaten einen geringeren Anteil von Patienten mit niedrigen und hohen Kosten sowie einen höheren Anteil an Patienten mit mittlerem und sehr hohem Ressourcenverbrauch, wenn Abrechnungsdiagnosen zugrunde gelegt wurden. Der Vergleich der Kostengewichte auf der groben Ebene der Morbiditätsgruppen ergab, dass der Ressourcenverbrauch der Patienten anhand von Abrechnungsdiagnosen insgesamt um 6% überschätzt wurde.

Um das Ausmaß einer Über- oder Unterschätzung genauer zu analysieren, war für einzelne Kostenquintile ermittelt worden, in welcher Höhe die erwarteten Behandlungskosten von den tatsächlichen Kosten der Patienten abwichen. Dabei zeigte sich, dass die Behandlungskosten von Patienten im niedrigen Kostenbereich generell eher überschätzt und im hohen Bereich eher unterschätzt wurden, ein bekannter mathematischer Effekt der „Regression zur Mitte“ (ZWINGMANN und WIRTZ 2005), der von einem Patientenklassifikationssystem so gering wie möglich gehalten werden sollte. Bei Systemen mit aktuarischem Ansatz, die jeden Patienten in eine Zelle einordnen, ist dieser Effekt aufgrund der höheren Aggregation von Risikofaktoren im allgemeinen stärker ausgeprägt als bei Regressionsmodellen (VON STILLFRIED und RYLL 2004). Anhand der Einteilung der Patienten in ACG-Kategorien wurden die durchschnittlichen Behandlungskosten für die niedrigste bzw. höchste Kostenkategorie auf der Basis von Abrechnungsdiagnosen um 200% überschätzt bzw. um 33% unterschätzt. Anhand von Behandlungsanlässen lagen die selben Werte mit 100% bzw. 22% deutlich niedriger. Über alle Kostenquintile ergab die Schätzung der durchschnittlichen Behandlungskosten anhand von Behandlungsanlässen bessere Ergebnisse als anhand von Abrechnungsdiagnosen, d.h. das Predictive Ratio lag näher an dem angestrebten Wert 1.

Zusätzlich wurde der  $R^2$ -Wert bestimmt, der ein Maß für die Fähigkeit eines Patientenklassifikationssystems darstellt, die Varianz der Behandlungskosten von Patienten zu erklären. Da die Validität von Parametern wie  $R^2$  und Predictive Ratio in hohem Maße von der Größe der Stichprobe abhängig ist, und die Besetzung der einzelnen ACG-Klassen

aufgrund der kleinen Anzahl der Studienpatienten zum Teil sehr gering war, fand sich eine hohe Varianz der Behandlungskosten, was eine korrekte Vorhersage vor allem in den höheren Kostenklassen entsprechend erschwerte (DUCKETT und AGIUS 2002). Aus diesem Grund wurden auf der Ebene der ACG-Klassifizierung nur die Kostenquintile verglichen, während Kostengewichte und  $R^2$ -Werte auf der Ebene der vier Morbiditätsgruppen berechnet wurden. Der  $R^2$ -Wert betrug in der vorliegenden Untersuchung 26% für das auf Abrechnungsdiagnosen und 30% für das auf Behandlungsanlässen beruhende Morbiditätsgruppen-Modell, sowie 19% bzw. 32% für das amerikanische RUB-Modell. In einer deutschen Studie der KBV mit einer repräsentativen Stichprobe von 7,6 Mio Patienten und Diagnosedaten aus dem 3. Quartal 2001 bis zum 2. Quartal 2003 wurde ein  $R^2$  von 26% für das ACG-Modell und 22% für RUBs errechnet (VON STILLFRIED und RYLL 2004). In einer Untersuchung mit Versichertendaten der AOK aus den Jahren 2001 und 2002 ergaben sich Werte von 8% bis 37%, je nachdem, ob Gesamtkosten mit oder ohne Kappungsgrenzen betrachtet, bzw. Sachausgaben einbezogen wurden (DRÄTHER et al. 2006).

Im Vergleich zu diesen Ergebnissen war der  $R^2$ -Wert in der australischen Studie mit 19% eher gering, während er in US-amerikanischen, kanadischen und schwedischen Untersuchungen zwischen 38% und 43% lag. In den spanischen Studien betrug er dagegen bis zu 53% (ENGSTROM et al. 2006, DUCKETT und AGIUS 2002, REID et al. 2001, ORUETA et al. 1999, JUNCOSA et al. 1999, WEINER et al. 1991). Die Höhe des  $R^2$ -Wertes wird von der Ausprägung und Validität der abhängigen (z.B. Behandlungskosten, Zahl der Arztbesuche) und unabhängigen (z.B. Geschlecht, Alter, Diagnosen) Variablen beeinflusst. So benutzten schwedische, kanadische und deutsche Autoren beispielsweise Kappungsgrenzen in unterschiedlicher Höhe für besonders teure Patienten. In einigen Studien wurden diejenigen Patienten von der Analyse ausgeschlossen, die im Untersuchungszeitraum keinen Arztkontakt aufwiesen („Non-user“), in anderen wurden alle Patienten einbezogen. In den spanischen Analysen lagen keine Kostendaten vor, sondern als Parameter für die Schätzung des Ressourcenverbrauchs wurde die Anzahl der Arztbesuche benutzt. In der deutschen Studie waren nicht nur hausärztliche Diagnosen, sondern die Diagnosen aller Ärzte einbezogen worden, welche die Patienten im Laufe eines Jahres ambulant behandelt hatten. In der australischen Untersuchung stammten die ambulanten Diagnosedaten nicht aus den Abrechnungsunterlagen eines Jahres, sondern sie waren zu Studienzwecken erhoben worden, indem ein Arzt die Patienten im Rahmen einer Eingangsuntersuchung einmalig über ihre Erkrankungen befragt hatte. Auch in den spanischen Studien wurden keine

Abrechnungsdiagnosen verwendet, sondern die behandelten Erkrankungen waren während der Konsultationen prospektiv in Form von Episoden von den Hausärzten notiert und codiert worden. Die Dokumentation von Diagnosen als Episoden entspricht exakt der Konzeption des ACG-Systems, das jede Diagnose eines Patienten nur einmal im Beobachtungszeitraum berücksichtigt. Die hohe Vorhersagekraft des ACG-Modells im spanischen Primärarztssystem könnte daher darin begründet liegen, dass die Validität der verwendeten Diagnosen durch die Art und Form der verwendeten Dokumentation besonders hoch war. Eine solche Erklärung wird durch die Ergebnisse der vorliegenden Studie gestützt, in der beide Gütemaße für Patientenklassifikationsmodelle (Predictive Ratio und  $R^2$ ) bessere Ergebnisse zeigten, wenn Behandlungsanlässe anstatt Abrechnungsdiagnosen für eine Klassifizierung der Patienten herangezogen wurden. In der eigenen Untersuchung war die Erklärungskraft des Morbiditätsgruppen-Modells aufgrund der gröberen Einteilung insgesamt geringer als in den übrigen europäischen Studien.

Zusammenfassend ist zu sagen, dass sich aus den dargestellten Ergebnissen sowohl auf Basis einer umfangreichen amerikanischen Population als auch anhand der eigenen Stichprobe Hinweise darauf ergeben, dass die Morbiditätsklassifikation und die Schätzung des Ressourcenverbrauchs durch das ACG-System vor allem für ältere und chronisch kranke Patienten signifikant höher ausfallen könnte, wenn für eine solche Klassifizierung nach ICD-10-codierte Abrechnungsdiagnosen im Vergleich zu den Behandlungsanlässen eines Jahres herangezogen werden. Insofern konnte die eingangs aufgestellte Hypothese verifiziert werden. Da die vorliegende Stichprobe jedoch sehr klein war und ausschließlich hausärztliche Diagnosen berücksichtigt wurden, müsste diesen Hinweisen in einer umfangreicheren, repräsentativen Untersuchung unter Einbeziehung aller (auch fachärztlicher) Diagnosen der Patienten weiter nachgegangen werden.

### ***6.1.6 Potentielle Einflussfaktoren auf die Validität von Abrechnungsdiagnosen***

#### **6.1.6.1 Eignung des Codierungssystems**

***Hypothese 5: Mit dem in Deutschland verwendeten ICD-10-GM-Codierungssystem können häufige hausärztliche Behandlungsanlässe (z.B. Symptome, präventive und psychosoziale Beratung) nicht korrekt und vollständig verschlüsselt werden.***

***Hypothese 6: Aufgrund der vielen möglichen Synonyme für einen Behandlungsanlass entstehen falsche Codes.***

***Hypothese 7: Mit dem ICD-10-GM-Codierungssystem verschlüsselte Abrechnungsdiagnosen unterscheiden sich von den in der Akte dokumentierten Behandlungsanlässen hinsichtlich des klinischen Inhalts.***

Von den 1748 identifizierten Behandlungsanlässen waren 1108 von den Ärzten mit Hilfe des ICD-10-Systems als Abrechnungsdiagnosen codiert worden. Nur für 15 in der Patientenakte als Diagnosen dokumentierte Behandlungsanlässe lag kein ICD-10-Code vor (0,8%). Dieses Ergebnis legt den Schluss nahe, dass Behandlungsanlässe mittels des ICD-10-Systems auch codiert worden waren, wenn sie als Textdiagnosen vorlagen. Dagegen waren die meisten der 625 uncodierten Behandlungsanlässe nicht explizit als Diagnosen in der Patientenakte dokumentiert worden. Andererseits betrafen fehlende Diagnosen am häufigsten Maßnahmen (ICD-10-Kapitel Z) und unspezifische Symptome (Kapitel R), was darauf hindeuten könnte, dass Hausärzte Schwierigkeiten haben, solche Behandlungsanlässe als ICD-10-Textdiagnosen und -Codes inhaltlich abzubilden. In der Untersuchung wurde jedoch deutlich, dass sich prinzipiell jedem Behandlungsanlass ein passender ICD-10-Code zuordnen ließ. Die Hypothese, dass häufige hausärztliche Behandlungsanlässe im ICD-10-System nicht korrekt und vollständig verschlüsselt werden können, wurde somit falsifiziert.

Allerdings ergaben sich - vermutlich aufgrund der pathogenetischen Ausrichtung und der Detailliertheit des ICD-10-Systems - Schwierigkeiten, eher unspezifische hausärztliche Behandlungsanlässe einem ICD-10-Code zuzuordnen, der Kenntnisse über die Ursache der Erkrankung voraussetzte. Dieses Problem war in 41% der Fälle die Ursache für eine inkorrekte Abrechnungsdiagnose. Beispielweise waren die Symptome „Übelkeit“ und „Diarrhoe“ häufig als „infektiöse Darmerkrankung“ (ICD-10-Kapitel A) anstatt als „nichtinfektiöse Diarrhoe“ (Kapitel K) oder als entsprechendes „Symptom“ (Kapitel R) codiert worden, obwohl kein Stuhltest vorlag, der eine infektiöse Genese der Erkrankung belegt hätte. Da die Codierungsvorschrift existiert, dass Diagnosen bezüglich ihrer Ätiologie so spezifisch wie möglich zu verschlüsseln sind, wurde der Behandlungsanlass Schwindel als Störung der Vestibularfunktion, unspezifische Bauchschmerzen als Gastritis, Sodbrennen als Refluxoesophagitis und Schmerzen im Fußgelenk als V.a. Borreliose codiert. Da alle Studienpraxen den in ihrer Praxis-EDV enthaltenen ICD-10-Thesaurus verwendeten, beeinflusste die Auswahl der darin für einen Behandlungsanlass angezeigten Codes die letztendliche Zuordnung einer Diagnose. Im ICD-10-GM-Browser des Zentralinstituts für die kassenärztliche Versorgung ([http://www.zi-berlin.de/Zi\\_ICD10Browser/zi\\_icd\\_10\\_browser.htm](http://www.zi-berlin.de/Zi_ICD10Browser/zi_icd_10_browser.htm), Stand 30.08.06) fanden sich für den

Behandlungsanlass Schwindel 42 verschiedene ICD-Codes. Ein weiteres Beispiel in diesem Zusammenhang stellt die Diagnose Diabetes mellitus dar. Bei dieser Erkrankung unterscheidet das ICD-10-System zwischen primär insulinabhängigem Typ I-Diabetes (E10) und nicht primär insulinabhängigen Typ II-Diabetes (E11). Außerdem existiert der Code E14 für einen nicht näher bezeichneten Diabetes. Diese Schlüsselnummer soll jedoch nicht verwendet werden, wenn der Patient in die Kategorie E10 oder E11 fällt. Das Auftreten von diabetischen Komplikationen soll durch eine Ziffer an vierter Stelle des Codes gekennzeichnet und die Folgeerkrankungen Polyneuropathie, Nephropathie und Retinopathie zusätzlich noch mit einer so genannten Kreuz-Stern-Schlüsselnummer versehen werden. Um für einen Patienten mit Typ II-Diabetes und einer Polyneuropathie den korrekten Code aus einer Thesaurusliste auszuwählen, sind also einerseits profunde Kenntnisse der Codierungsregeln notwendig und andererseits ist dieser Prozess zeitaufwendig. Daher wurde der Einfachheit halber in der Hälfte der Fälle von den Ärzten der Code E14 anstatt E11 verwendet. Diese Vielfalt der Codierungsmöglichkeiten ist für den hausärztlichen Bereich wenig sinnvoll und hat die Übereinstimmung von Behandlungsanlässen und Abrechnungsdiagnosen in der vorliegenden Studie vermindert. Insofern trifft die Hypothese zu, dass die vielen möglichen Synonyme im ICD-10-System zu einer Entstehung falscher Codes beitragen.

IEZZONI (2003) hatte angemerkt, dass an der Validität des klinischen Inhalts von zu administrativen Zwecken erhobenen Diagnosedaten immer wieder Zweifel geäußert werden. Daher wurde geprüft, inwieweit die zu Abrechnungszwecken generierten ICD-10-Codes inhaltlich mit den aus der Akte ersichtlichen Behandlungsanlässen übereinstimmen. Die Vielfalt der möglichen Diagnosencodes für einen Behandlungsanlass, der sich ätiologisch nicht eindeutig zuordnen ließ, führte bei 41% der inkorrekt codierten Abrechnungsdiagnosen dazu, dass der angegebene ICD-10-Code eine andere Erkrankung beschrieb als der ursprüngliche Behandlungsanlass. Da diese Diagnosengruppe jedoch nur 11% der Gesamtstichprobe ausmachte, ergab sich insgesamt eine hohe Übereinstimmung des klinischen Inhalts von ICD-10-codierten Abrechnungsdiagnosen und Behandlungsanlässe (96%). Die diesbezügliche Hypothese wurde somit falsifiziert.

#### **6.1.6.2 Codierungs- und Abrechnungsvorschriften**

***Hypothese 8: Wenn Abrechnungsdiagnosen als G (gesichert), V (Verdacht auf), A (Ausschluss von) und Z (Zustand nach) gekennzeichnet werden, erhöht sich die Übereinstimmungsrate zwischen Abrechnungsdiagnosen und Behandlungsanlässen.***

Bei der Analyse der Eignung der derzeitigen Diagnosenzusätze für die Anwendung im Rahmen eines Patientenklassifikationssystems stellte sich zunächst die Frage nach der Validität der Zusätze selbst. GERSTE und GUTSCHMIDT (2006) verglichen die Verwendung von Diagnosenzusätzen bei Diabetikern in den Jahren 2001 und 2004. Während sich im Jahr 2001 in 93% der Fälle die Angabe einer gesicherten Diagnose fand, sank dieser Anteil in 2004 auf 85%. Gleichzeitig stieg die Angabe von Verdachts- und Ausschlussdiagnosen von 2% im Jahr 2001 auf 9% in 2004 und von als „Zustand nach“ gekennzeichneten Erkrankungen im gleichen Zeitraum von 1% auf 4%. Die vorliegende Studie mit Daten aus dem Jahr 2003 ergab einen Anteil von 93% gesicherten, 4% Verdachts- und Ausschluss-, sowie 3% „Zustand nach“-Diagnosen und zeigte damit ähnliche Werte, obwohl unterschiedliche akute und chronische Erkrankungen einbezogen wurden.

Anhand eines Vergleichs mit den Behandlungsanlässen wurde die von GERSTE und GUTSCHMIDT (2006) aufgeworfene Frage untersucht, ob der geringe Anteil von Diagnosen mit dem Zusatz V, A und Z den tatsächlichen Häufigkeiten entspricht. Der mittels Patientenakte verifizierbare Anteil an gesicherten Diagnosen in der Stichprobe lag bei 89%. Als „Zustand nach“ einzuschätzende Diagnosen waren dagegen mehr als doppelt so häufig in den Akten zu finden wie von den Ärzten angegeben (7% versus 3%). Bei der Analyse einzelner Diagnosencodes stellte sich heraus, dass Krebserkrankungen und Infektionen, wie Hepatitis B und Polio, aber auch Herzinfarkt, Bandscheibenprolaps, Hüft-TEP, Herzklappenersatz, verschiedene Operationen und Frakturen den Ärzten offenbar besondere Schwierigkeiten boten, sich für den Diagnosenzusatz G oder Z zu entscheiden. In der Mehrzahl der Fälle wurden diese Diagnosen als gesichert gekennzeichnet. Eine Erklärung für diesen Befund könnte darin bestehen, dass in diesem Zusammenhang unterschiedliche Perspektiven denkbar sind. Wenn Ärzte Erkrankungen als „gesicherte“ Diagnosen angeben, verstehen sie darunter möglicherweise, dass eine Diagnose zweifelsfrei gestellt wurde. Ein solches Kriterium trifft jedoch ebenfalls auf einen in der Vergangenheit erlittenen Herzinfarkt zu, der sich mit aufwändigen Untersuchungen bestätigen ließ. Als „Zustand nach“ sollen dagegen symptomlose Erkrankungen gekennzeichnet werden, deren Behandlung abgeschlossen ist. Wenn jedoch ein Patient beispielsweise nach der Implantation einer Endoprothese der Hüfte wegen andauernder Hüftbeschwerden in Behandlung ist, oder ein Herzschrittmacher implantiert wurde, der regelmäßig überprüft werden muss, stehen diese Behandlungsanlässe in direktem Zusammenhang mit der erfolgten Operation. Aus dieser Perspektive könnte die Therapie für die codierte Diagnose „Hüft-TEP“ als noch nicht

abgeschlossen und das Kennzeichen „Z“ daher als inadäquat erscheinen. Diese Beispiele legen nahe, dass sich mit dem derzeitigen System in einigen Fällen Schwierigkeiten bei der klaren Zuordnung von Diagnosenzusätzen ergeben könnten, da es Raum für verschiedene Interpretationen bietet. Es ist zu vermuten, dass Ärzte in Zweifelsfällen den Diagnosenzusatz „G“ wählen, was den überproportionalen Anteil dieses Zusatzes erklären würde. Insgesamt wurde die Anzahl der gesicherten Diagnosen zu Ungunsten der als „Zustand nach“ klassifizierten Erkrankungen um 4% überschätzt, während ein Anteil von 4% Verdachts- und Ausschlussdiagnosen anhand der Akten bestätigt werden konnte.

Die Einführung von Diagnosenzusätzen soll es ermöglichen, anhand von Abrechnungsdiagnosen tatsächlich vorliegende behandlungsbedürftige Erkrankungen zu identifizieren und diese von Verdachts- und Ausschlussdiagnosen sowie symptomlosen Krankheiten zu unterscheiden. So beobachteten TRAUTNER et al. (2005), dass für ein Viertel der von ihm untersuchten Patienten mit chronischen Diagnosen diese nicht über mehrere Quartale fortgeschrieben wurden. Er vermutete, dass es sich bei diesen Erkrankungen um Verdachts- bzw. Ausschlussdiagnosen handeln könnte. Mit der Möglichkeit, solche Abrechnungsdiagnosen identifizieren und gegebenenfalls von weiteren Analysen ausschließen zu können, erhofft man sich eine Erhöhung der Validität insbesondere auch im Hinblick auf die Einführung eines diagnosenbasierten Patientenklassifikationssystems. Da Risikoadjustierungssysteme den Ressourcenverbrauch von Patienten anhand von im Abrechnungszeitraum behandelten Erkrankungen einschätzen, könnten geeignete Diagnosenzusätze in diesem Zusammenhang eine wichtige Rolle spielen. Die bisherigen Zusätze wurden aus dem US-amerikanischen System übernommen, in dem die Abrechnungsvorschrift gilt, dass für jede einzelne Behandlungsleistung eine ICD-Diagnose angegeben werden muss. Da in Deutschland Abrechnungsdiagnosen derzeit ohne Datums- und Leistungsbezug für den gesamten Abrechnungszeitraum kumuliert an die KVen übermittelt werden, besteht in diesem System zusätzlich die Gefahr, dass inaktive Diagnosen in den Dateien enthalten sind. Daher stellt sich die Frage, ob die verwendeten Zusätze beispielsweise auch dazu geeignet wären, Dauerdiagnosen zu identifizieren. In der vorliegenden Untersuchung wurden 410 nicht behandelte „Dauerdiagnosen“ identifiziert. 11% dieser zusätzlichen Diagnosen entsprachen einem symptomlosen Zustand, der mit dem Buchstaben „Z“ zu kennzeichnen wäre. Weitere 2% waren Verdachts- bzw. Ausschlussdiagnosen. Bei 87% handelte es sich um gesicherte akute oder chronische Diagnosen. Anhand der Zusätze Z, V und A wäre es also lediglich möglich gewesen, 13% der

nicht behandlungsbedürftigen Diagnosen zu erkennen und auszuschließen. Die Hypothese, dass die Kennzeichnung von Abrechnungsdiagnosen als G, V, A oder Z die Übereinstimmungsrate zwischen Abrechnungsdiagnosen und Behandlungsanlässen erhöht, konnte daher nicht bestätigt werden. Sollen Verdachts- und Ausschlussdiagnosen in Zukunft bei einer Risikoadjustierung gänzlich unberücksichtigt bleiben, ergibt sich darüber hinaus das Problem, dass solche Diagnosen zwar nicht langfristig behandlungsbedürftig sind, die Ausschlussdiagnostik selbst jedoch im Abrechnungszeitraum Kosten verursacht hat. Bei einer zeitgleichen („concurrent“) Risikoadjustierung könnte es daher sinnvoll sein, diese Diagnosen trotzdem einzubeziehen. Ansonsten erscheint es möglich, dass durch ein solches Vorgehen der Ressourcenverbrauch für Erkrankungen insgesamt unterschätzt wird.

Würde anstatt der derzeitigen Diagnosenzusätze eine - bei der problemorientierten episodenzugehörigen Dokumentation übliche - Unterscheidung in aktive und inaktive Erkrankungen benutzt, hätten in der vorliegenden Untersuchung 97% der zusätzlichen Diagnosen als inaktiv identifiziert und von einer Risikoadjustierung ausgeschlossen werden können. Eine weitere Unterscheidung von inaktiven Diagnosen in latente bzw. langfristig inaktive Gesundheitsprobleme würde darüber hinaus die Möglichkeit bieten, den Status von derzeit nicht behandlungsbedürftige Erkrankungen bezüglich der Wahrscheinlichkeit eines zukünftigen Ressourcenverbrauchs zu beschreiben.

### **6.1.6.3 Qualität und Struktur der Dokumentation**

*Hypothese 9: Es existiert ein positiver Zusammenhang zwischen Qualität der Dokumentation und der Validität von Abrechnungsdaten.*

*Hypothese 10: Durch eine episodenzugehörige Darstellung können Behandlungsanlässe von Patienten vollständiger und korrekter abgebildet werden als durch eine chronologische Dokumentation.*

SURJAN (1999) stellte die Hypothese auf, dass die Qualität der Patientenakte von entscheidender Bedeutung für die Validität der auf Basis der dort enthaltenen Informationen codierten ICD-10-Diagnosen sein müsste. Daher wurde in der vorliegenden Untersuchung geprüft, ob sich ein höherer Anteil an fehlenden, inkorrekten und zusätzlichen Abrechnungsdiagnosen in Praxen nachweisen ließ, deren Dokumentationsqualität als eher schlecht eingeschätzt worden war. Es zeigte sich, dass lediglich die Zahl fehlender Diagnosen in diesen Praxen um 2% höher lag als bei den Ärzten mit guter Dokumentationsqualität. Dieser Unterschied war nicht signifikant. Ein positiver Zusammenhang zwischen Validität von Abrechnungsdiagnosen und der Dokumentationsqualität im Sinne der aufgestellten

Hypothese konnte somit nicht bestätigt werden. Dieses Ergebnis erscheint zunächst überraschend. Allerdings fanden LOGAN et al. (2001) in Bezug auf die Vollständigkeit der Angaben in der Patientenakte bei verschiedenen Ärzten eine hohe Variabilität, d.h. manche Ärzte dokumentierten sehr umfangreich, während andere nur einen „Minimalstandard“ notierten. Dabei wiesen beide Gruppen jedoch eine hohe Korrektheit der Angaben auf. Diese Befunde deuten darauf hin, dass zwischen Umfang der Dokumentation und Korrektheit der enthaltenen Daten möglicherweise kein direkter Zusammenhang besteht. Die Entwicklung des Messinstrumentes für die Dokumentationsqualität erfolgte anhand des Electronic Medical Record Tools (EMRAT). Das EMRAT setzt für die Beurteilung von Patientenakten die Anwendung des SOAP-Schemas als wünschenswertem Standard voraus. Wie die Studie von KÖRNER et al. (2004) schon gezeigt hatte, existiert in deutschen Hausarztpraxen jedoch ausschließlich eine chronologische Dokumentation. Da also auch Praxen, deren Dokumentationsqualität als gut eingeschätzt worden war, ihre Akten nicht nach dem SOAP-Schema führten, unterschieden sich die Praxen weniger in Inhalt und Struktur als hauptsächlich im Umfang der Eintragungen in die Patientenakte. Es ist daher wahrscheinlich, dass es mit Hilfe des entwickelten Instrumentes nicht möglich war, die geringen strukturellen und inhaltlichen Unterschiede in den Akten zu messen. Stattdessen resultierten die Ergebnisse in erster Linie aus einem Vergleich des Dokumentationsumfangs, der sowohl unter den eher gut als auch unter den eher schlecht dokumentierenden Praxen entsprechend variierte und darüber hinaus eher keine Korrelation mit der Korrektheit von Abrechnungsdiagnosen gezeigt hatte. Um einen potentiellen Zusammenhang zwischen Dokumentationsqualität und Validität von Abrechnungsdiagnosen abschließend zu prüfen, sollte eine Vergleichsgruppe herangezogen werden, die nach dem SOAP-Schema dokumentiert.

Auch die Prüfung der zweiten in diesem Zusammenhang aufgestellten Hypothese wurde dadurch erschwert, dass die Studienpraxen nicht nach dem Episodenprinzip dokumentierten und ein direkter Vergleich zwischen einer chronologischen und episodenzugeordneten Dokumentation daher nicht möglich war. Die aus diesem Grund vorgenommene Analyse von chronologisch strukturierten elektronischen Patientenakten zeigte, dass es zwar möglich ist, Behandlungsanlässe mittels Buchstabencodes zu kennzeichnen und dabei auch längerfristig bestehende von akuten Erkrankungen zu unterscheiden. In automatisch generierte Zusammenstellungen von Erkrankungen wurden jedoch alle jemals mit einem solchen Buchstabencode versehenen Diagnosen einbezogen. Nur anhand des Datums und des Buchstabencodes ließ sich nicht entscheiden, welchen Behandlungsstatus eine Erkrankung

hatte. Außerdem wurden Diagnosen, die zur gleichen Krankheitsentität gehörten, in einigen Fällen als separate Erkrankungen verschlüsselt. Durch die Verwendung einer episodenzugeordneten Dokumentationsstruktur würde sichergestellt, dass solche Diagnosen der gleichen Episode zugeordnet werden. Durch die Unterscheidung von aktiven und inaktiven, abgeschlossenen Episoden und Gesundheitsproblemen von bleibender Bedeutung könnten im Untersuchungs- oder Abrechnungszeitraum behandelte Erkrankungen zweifelsfrei erkannt und inaktive Episoden ausgeschlossen werden. Durch die insgesamt korrektere Darstellung von Behandlungsanlässen als Episoden konnte die Übereinstimmung mit den Abrechnungsdiagnosen in der Gesamtstichprobe um 9% erhöht werden. Es ist zu vermuten, dass eine episodenzugeordnete Struktur in Verbindung mit einer problemorientierten Dokumentation nach dem SOAP-Schema Behandlungsanlässe auch vollständiger abbilden und damit die Zahl der fehlenden Abrechnungsdiagnosen vermindert werden könnte. Wenn jedes Gesundheitsproblem des Patienten und jede diagnostische bzw. Therapiemaßnahme einer neuen oder einer bestehenden Episode zugeordnet würde, dürfte sich die Wahrscheinlichkeit erhöhen, dass der Arzt alle Behandlungsanlässe eines Patienten vollständig erfasst. Um die eingangs aufgestellte Hypothese jedoch endgültig zu verifizieren bzw. zu falsifizieren zu können, müsste in einer entsprechenden Interventionsstudie prospektiv geprüft werden, ob sich der Anteil behandelter Erkrankungen, die nicht als Abrechnungsdiagnosen angegeben wurden, durch eine episodenzugeordnete Dokumentation tatsächlich vermindern lässt. Außerdem sollte analysiert werden, welche Auswirkungen eine solche Dokumentationsstruktur auf die Validität von Abrechnungsdiagnosen für den Einsatz in einem Risikoadjustierungsmodell hätte. Eine solche Studie wäre in Deutschland demnächst mit den Daten des CONTENT-Projektes durchführbar (KÖRNER et al. 2005, LAUX et al. 2005). Die Implementierung eines ICPC-2E-basierten, episodenzugeordneten Dokumentationssystems in deutschen Hausarztpraxen wird im Rahmen dieses vom BMBF geförderten Projektes der Sektion Allgemeinmedizin und Versorgungsforschung der Universität Heidelberg in Zusammenarbeit mit dem Institut für Allgemeinmedizin der Universität Frankfurt durchgeführt.

#### **6.1.6.4 Nutzung einer elektronischen Patientenakte**

*Hypothese 11: Praxen mit einer reinen elektronischen Patientenakte weisen eine höhere Validität von Abrechnungsdaten auf als Praxen, die ihre EDV überwiegend zu administrativen Zwecken nutzen.*

In der vorliegenden Untersuchung wurde die Hypothese untersucht, ob die Verwendung einer elektronischen Patientenakte die Validität von Abrechnungsdiagnosen im Vergleich zu Ärzten

erhöht, die Patientenkontakte in Papierform dokumentieren und ihre Praxis-EDV nur für administrative Zwecke (Formularwesen, Abrechnung) nutzen. Dabei wurde davon ausgegangen, dass die vielfältigen Möglichkeiten der Eintragung und Organisation von Daten in einer elektronischen Patientenakte es erleichtern, Behandlungsanlässe korrekt zu erfassen und als zutreffende Abrechnungsdiagnosen zu codieren. Der Vergleich der Validität von Abrechnungsdiagnosen zwischen Ärzten, die eine elektronische bzw. eine Papierakte verwendeten, zeigte in letzterer Gruppe einen um 11,4% höheren Anteil an fehlenden Diagnosen („Underreporting“). Dagegen wiesen Praxen mit elektronischen Akten eine fast doppelt so hohe Zahl von inaktiven „Dauerdiagnosen“ auf („Overreporting“). Dieser Unterschied war im Chi-Quadrat-Test signifikant ( $p < 0,001$ ). Für dieses Ergebnis liegt folgende Begründung nahe: Mittels des Messinstrumentes für die Dokumentationsqualität wurden Ärzte als eher gute „Dokumentierer“ beurteilt, deren Akten in den verschiedenen Kategorien des SOAP-Schemas jeweils mindestens einen Eintrag enthielten. Unterschiede zwischen den Gruppen basierten in erster Linie auf dem Vorhandensein oder Fehlen von subjektiven Beschwerden des Patienten (S) und objektiven Befunden (O), wie Laborwerten und Überweisungsergebnissen. Neben den Diagnosen (A) wurden vor allem diese Angaben herangezogen, um Aufschluss über die relevanten Behandlungsanlässe im Untersuchungszeitraum zu erhalten. Je mehr Daten in diesen Kategorien vorhanden waren, desto höher war die Wahrscheinlichkeit, dass sich nicht zu jedem auf diese Weise generierten Behandlungsanlass eine Abrechnungsdiagnose finden ließ, was dann in einer fehlenden Diagnose resultierte. Eine mögliche Erklärung für den hohen Anteil an zusätzlichen inaktiven „Dauerdiagnosen“ in Praxen mit elektronischer Patientenakte stellt dagegen die Tatsache dar, dass Praxis-EDV-Systeme eine in ein Diagnosenfeld der Akte eingegebene Erkrankung automatisch in eine separate Datei transferieren können. In dieser Datei wurden in den meisten Studienpraxen alle jemals für den Patienten notierten Diagnosen gespeichert, so dass aktive und inaktive Diagnosen nebeneinander vorlagen und in vielen Fällen in dieser Form komplett in die Abrechnung übernommen wurden. Die Zahl der korrekten Diagnosen lag in Praxen mit Papierakten lediglich um ca. 2% höher als in Praxen mit einer elektronischen Patientenakte, was auf den geringeren Anteil an zusätzlichen Diagnosen zurückzuführen war. Damit ließ sich insgesamt kein signifikanter Zusammenhang zwischen dem Grad der EDV-Nutzung und der Übereinstimmung von Abrechnungsdiagnosen und Behandlungsanlässen nachweisen ( $p=0,2$ ); die entsprechende Hypothese wurde somit falsifiziert.

Es stellte sich dagegen heraus, dass ein signifikanter Zusammenhang zwischen EDV-Nutzung und Dokumentationsqualität bestand ( $p < 0,001$ ): 75% der Praxen mit eher schlechter Dokumentationsqualität benutzten eine elektronische Patientenakte, während die eher gut dokumentierenden Ärzten je zur Hälfte eine Papier- und elektronische Akte verwendeten. Umgekehrt wiesen 76% der Praxen mit Papierakten eine eher gute Dokumentationsqualität auf, während sich bei den EDV-Nutzern je zur Hälfte eine eher gute bzw. schlechte Dokumentation fand. Da mit Hilfe des Instrumentes zur Beurteilung der Dokumentationsqualität vermutlich in erster Linie der Umfang der Dokumentation gemessen wurde, bedeutete die Feststellung einer eher guten Qualität das Vorliegen einer ausführlicheren Dokumentation. Entgegen der ursprünglichen Annahme schienen Praxen mit Papierakten insgesamt also eine höhere Zahl an relevanten Informationen über den Patienten schriftlich festzuhalten als Praxen, die eine elektronische Akte nutzten.

Sowohl die bisher dargestellten internationalen Studien als auch die Ergebnisse der eigenen Untersuchung weisen darauf hin, dass die Validität von Diagnosedaten aus Patientenakten höher ist und der klinische Inhalt realistischer abgebildet wird, als anhand von Abrechnungsdaten. Daher wäre die Nutzung dieser Datenquelle für eine Risikoadjustierung zu präferieren. Das größte Hindernis lag bisher in dem Zeit- und Kostenaufwand, den die manuelle Aufarbeitung von Patientenakten benötigte. Mit der zunehmenden Verbreitung von elektronischen Patientenakten bietet sich prinzipiell die Chance, Diagnosen relativ schnell und unkompliziert aus den Akten zu extrahieren. Damit die relevanten Informationen bei derartigen Abfragen zuverlässig zu finden sind, müsste jedoch eine Standardisierung der Dokumentation vorgenommen werden. JOLLIS et al. (1993) kritisierten in ihrer Untersuchung zur Qualität von Patientenakten das Fehlen einer standardisierten Dokumentation sowohl in der elektronischen Form als auch in Papierakten. Auch in Deutschland kamen HUMMERS-PRADIER et al. (2003) zu dem Schluss, dass die routinemäßige Nutzung von Daten aus dem Praxis-EDV-System von Hausärzten für Versorgungsforschung und Qualitätssicherungszwecke erst dann sinnvoll ist, wenn eine einheitliche Datenstruktur vorliegt. SZECSENYI und BROGE (1999) schlugen vor, eine elektronische Patientenakte nicht chronologisch, sondern problemorientiert mit Hilfe des SOAP-Schemas zu organisieren. Vor allem im hausärztlichen Bereich bietet die problemorientierte Dokumentation (WEED 1968) eine gute Möglichkeit, die durch den langen Betreuungszeitraum oft umfangreichen Patienteninformationen sinnvoll und übersichtlich zu strukturieren.

## **Maßnahmen zur Erhöhung der Validität von Abrechnungsdiagnosen**

Abrechnungsdiagnosen und Behandlungsanlässe wiesen in der vorliegenden Untersuchung eine geringe Übereinstimmung von 40% auf. Anhand der dargestellten Analyse möglicher Ursachen für dieses Ergebnis sollen im Folgenden Maßnahmen abgeleitet werden, die möglicherweise dazu geeignet sein könnten, die Diagnosenvalidität zu erhöhen.

### **6.1.7 Maßnahmen zur Erhöhung der Vollständigkeit von Abrechnungsdiagnosen**

Als Hauptproblem der Diagnosenvalidität in den untersuchten hausärztlichen Praxen hat sich ein „Underreporting“ vor allem akuter Diagnosen mit geringem Schweregrad und Symptomcharakter, chronischer Diagnosen, die keine medikamentöse Therapie erfordern, sowie von Maßnahmen, wie beispielsweise Vorsorgeuntersuchungen oder Impfungen, herausgestellt. In Anbetracht der Erfahrungen bei der Einführung von DRGs im stationären Bereich kann argumentiert werden, dass die Vollständigkeit der Diagnosendokumentation auch im ambulanten Bereich zunehmen wird, wenn die Vergütung der Ärzte von der Anzahl und Schwere der dokumentierten Diagnosen abhängt. Dabei besteht jedoch prinzipiell die Gefahr, dass von den Ärzten lediglich eine größere Anzahl an Abrechnungsdiagnosen im Sinne eines „Upcodings“ angegeben wird, ohne dass dadurch die Validität dieser Diagnosen zunimmt.

#### **6.1.7.1 Einführung einer problemorientierten Dokumentation als verbindlichen Dokumentationsstandard**

Es erscheint daher sinnvoll, die bisherige Form der ärztlichen Dokumentation so zu strukturieren, dass die tatsächlich aufgetretenen Behandlungsanlässe möglichst einfach und zeitsparend, dabei aber korrekt und nachvollziehbar abgebildet werden können. Die Untersuchung hat gezeigt, dass Praxen, die eine elektronische Patientenakte verwenden, eine höhere Vollständigkeit und eine geringere Korrektheit von Abrechnungsdiagnosen aufweisen als Praxen, die Arzt-Patienten-Konsultationen in einer Papierakte festhalten. Vermutlich fördert eine elektronische Dokumentation einerseits das schriftliche Festhalten von Behandlungsanlässen, und vereinfacht andererseits die automatische Speicherung von Diagnosen in einer separaten Datei. Unabhängig von ihrer Behandlungsbedürftigkeit wurden dann allerdings sämtliche in dieser Datei aufgeführten Diagnosen für die Abrechnung übermittelt, was die Korrektheit der Diagnosen verminderte. In den Praxen, die Papierakten verwendeten, fanden sich häufig Hinweise auf Behandlungsanlässe in der Beschreibung von Symptomen der Patienten, ohne dass diese explizit als Diagnosen dokumentiert oder codiert worden waren. Die verbindliche Einführung einer problemorientierten Dokumentation nach

dem SOAP-Schema würde eine Struktur vorgeben, die eine stärkere Standardisierung von Patientenakten zur Folge hätte. Dadurch könnte die Wahrscheinlichkeit erhöht werden, dass für jeden Patienten Beratungsanlass, Untersuchungsergebnisse, (Verdachts-) Diagnose und die veranlassten Maßnahmen notiert würden.

### **6.1.8 Maßnahmen zur Erhöhung der Korrektheit von Abrechnungsdiagnosen**

#### **6.1.8.1 Einführung einer episodenzusätzlichen Dokumentationsstruktur und der Diagnosenzusätze „aktiv/inaktiv“**

Durch die Einführung einer episodenzusätzlichen anstatt der bisherigen chronologischen Dokumentation könnte das Ausmaß des zweiten Problems bei der Validität von Abrechnungsdiagnosen, ein „Overreporting“ inaktiver Dauerdiagnosen, vermutlich erheblich reduziert werden. Wenn jede Episode als offen oder geschlossen klassifiziert, schwerwiegende Gesundheitsprobleme in einer Problemliste zusammengefasst und als aktiv, inaktiv oder latentes Problem gekennzeichnet würden, könnten inaktive bzw. geschlossene Episoden jederzeit identifiziert und gegebenenfalls von einer Abrechnung ausgeschlossen werden. Die korrekte Anwendung der derzeit verwendeten Diagnosenzusätze G, V, A und Z gestaltete sich in den untersuchten Hausarztpraxen in vielen Fällen dagegen schwierig, da die Regeln für eine Zuordnung zu Diagnosen mehrdeutig schienen und daher von den Ärzten unterschiedlich aufgefasst wurden. Außerdem war in der vorliegenden Studie der Anteil von Verdachts- und Ausschlussdiagnosen eher gering, so dass eine Identifikation solcher Erkrankungen nur einen geringen Einfluss auf die Diagnosenvalidität zeigte. Wenn eine problemorientierte, episodenzusätzliche Dokumentation in eine elektronische Patientenakte integriert würde, könnten Hausärzte vor allem bei multimorbiden Patienten eine wesentlich bessere Übersicht über die einzelnen aktiven und inaktiven Erkrankungen, ihre Therapie und mögliche Interaktionen gewinnen bzw. behalten. Außerdem wäre es möglich, Diagnosedaten sowohl für Vergütungs- und Forschungs-, als auch für Qualitätssicherungszwecke einfach und schnell bereitzustellen.

#### **6.1.8.2 Einführung eines für den hausärztlichen Bereich geeigneten Codierungssystems mit geringerer Verschlüsselungstiefe**

Mit dem Codierungssystem ICD-10 konnten prinzipiell alle hausärztlichen Behandlungsanlässe verschlüsselt werden. Es boten sich jedoch Anhaltspunkte dafür, dass die ätiologische Ausrichtung sowie die hohe Komplexität des ICD-10-Systems und seiner Codierungsregeln dazu geführt haben könnte, dass in ca. 11% der Fälle Abrechnungsdiagnosen inkorrekt verschlüsselt worden waren. Durch die Zusammenfassung

einzelner ICD-10-Codes zu größeren Kategorien ähnlich dem ICD-9-System konnten für die im hausärztlichen Bereich am häufigsten behandelten Erkrankungen relativ hohe Werte für Korrektheit und Vollständigkeit erreicht werden. Es ist daher wahrscheinlich, dass ein speziell auf den hausärztlichen Bereich zugeschnittenes Codierungssystem die Validität von Abrechnungsdiagnosen zusätzlich erhöhen könnte. In diesem Zusammenhang würde sich das ICPC-2E-System anbieten, das in vielen Ländern bereits eingesetzt wird und aufgrund seiner Struktur mit einer problemorientierten, episodensbasierten Dokumentation kompatibel ist.

### **Schlussfolgerungen bezüglich des Einsatzes eines Risikoadjustierungssystems im hausärztlichen Bereich**

In der vorliegenden Studie spiegelte sich die geringe Validität der Abrechnungsdiagnosen der Studienpatienten in einer ebenfalls eher geringen Validität ihrer ACG-Klassifikation wider. Zahl und Schwere der angegebenen Erkrankungen bestimmten die Zuordnung zu den verschiedenen Kategorien des Risikoadjustierungsmodells. Die Kombination aus einem „Underreporting“ akuter Diagnosen mit geringem Schweregrad bzw. von Vorsorgeuntersuchungen und Impfungen einerseits und einem „Overreporting“ chronischer Diagnosen andererseits führte zu einer Überschätzung der Morbidität der Studienpatienten anhand ihrer Abrechnungsdiagnosen. Dabei ist allerdings zu berücksichtigen, dass in der vorliegenden Untersuchung lediglich hausärztliche Diagnosen einbezogen wurden, während den KVen für eine Risikoklassifizierung von Patienten in der Realität die Diagnosen aller (Fach-) Ärzte zur Verfügung stünden, die im jeweiligen Abrechnungszeitraum an der Behandlung beteiligt gewesen sind. Für die risikoadjustierte Berechnung der Gesamtvergütung, die prospektiv erfolgen soll, hat sich die KBV inzwischen für das Patientenklassifikationsmodell HCC der Firma DxCG entschieden, ein Regressionsmodell, das Zuschlüsse vor allem für komplexe Erkrankungen beinhaltet. Dazu stellten CUMMING et al. (2002) fest, dass für die Abbildung hausärztlicher Behandlungsanlässe ein Patientenklassifikationssystem passender sein könnte, das nicht nur schwere, chronische Erkrankungen einbezieht. In Modellen, die akute Krankheiten und Symptome mit geringem Schweregrad von einer Klassifikation ausschließen, könnte ein großer Teil der allgemeinmedizinischen Arbeit unabgebildet bleiben. CARLSSON et al. (2002) merkten an, dass das ACG-System für den hausärztlichen Bereich möglicherweise besonders geeignet sei, da es nicht in epidemiologischer Weise die Verteilung von Erkrankungen beschreibt, sondern die Morbidität auf Patientenebene abbildet. Die Ergebnisse der vorliegenden Studie deuten darauf hin, dass sich hausärztliche Behandlungsanlässe aufgrund ihrer relativen Homogenität in eine überschaubare Anzahl von Morbiditätsklassen gruppieren lassen, so dass ein

---

kompliziertes Regressions- bzw. Zuschlagsmodell in diesem Bereich möglicherweise entbehrlich sein könnte. Für eine Risikoadjustierung auf der Ebene hausärztlicher Praxen erscheint es grundsätzlich sinnvoll, Patienten und nicht Diagnosen als Grundlage für eine Morbiditätsbeschreibung heranzuziehen. Es ist wahrscheinlich, dass sich der Ressourcenverbrauch von multimorbiden Patienten anders darstellt als die Summe der durchschnittlichen Behandlungskosten für die einzelnen Diagnosen, die diese Multimorbidität bedingen.