

## » Risiko-Adjustierung mit Angaben von Pflegekräften: Nutzen in epidemiologischen und ökonomischen Studien in der klinischen Forschung und im Qualitätsmanagement

Astrid Manstetten<sup>1</sup>, M. Liebetrau<sup>2</sup>, O. Sangha<sup>1,3</sup>, S. Schneeweiss<sup>4,5</sup>

<sup>1</sup> Bayerischer Forschungsverbund Public Health und Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE) der Ludwig-Maximilians-Universität, München

<sup>2</sup> Helios Klinik Blankenhain, Blankenhain

<sup>3</sup> Department of Health Policy & Management, Harvard University School of Public Health, Boston

<sup>4</sup> Department of Medicine, Harvard Medical School, Boston

<sup>5</sup> Department of Epidemiology, Harvard School of Public Health, Boston

**Zusammenfassung. Fragestellung:** Inwieweit stimmt die Beurteilung ausgewählter Krankheitsbilder zwischen Pflegepersonal und Ärzten in der internistischen Krankenhausversorgung überein und inwieweit können Angaben des Pflegepersonals in epidemiologischen und ökonomischen Studien für eine Risiko-Adjustierung eingesetzt werden? **Patienten und Methoden:** In zwei Querschnittsstudien mit je 50 internistischen Patienten wurde die Übereinstimmung zwischen Pflegekräften und Ärzten in der Bewertung von Erkrankungen des Index for CoExisting Diseases (ICED) untersucht. In der ersten Phase fand die Beurteilung zeitgleich bei Aufnahme statt, in der zweiten untersuchten die Ärzte bei Entlassung. Übereinstimmung, Kappa, Sensitivität und Spezifität, sowie Spearman-Korrelationen wurden mit deren 95% Konfidenzintervallen berechnet. **Ergebnisse:** Pflegekräfte erfassten im Durchschnitt eine geringere Anzahl von Erkrankungen pro Patient als Ärzte (2,5 vs. 3,8;  $p = 0,006$ ) und bewerteten den Schweregrad der Erkrankungen (von 0 bis 4) im Mittel geringfügig leichter als Ärzte (2,5 vs. 3,0;  $p = 0,001$ ). Die Übereinstimmung des Vorhandenseins einer Erkrankung lag in Phase 1 bei 84% (70%–90%), in Phase 2 bei 80% (67%–89%). Je differenzierter der Schweregrad einer Erkrankung nachgefragt wurde, desto geringer war die Übereinstimmung zwischen Ärzten und Pflegepersonal. Das Pflegepersonal erkannte Erkrankungen in Phase 1 mit einer Sensitivität von 0,63 und Spezifität von 0,83 unter der Annahme, dass die ärztliche Feststellung von Erkrankungen den Goldstandard darstellt. **Schlussfolgerungen:** Eine Beurteilung von (Begleit-)Erkrankungen durch pflegerisches Personal ohne eine spezielle Schulung im Sinne einer „Study Nurse“ ist für einzelne Anwendungen zum derzeitigen Zeitpunkt nicht ausreichend. Umfangreichere Studien müssen die ärztliche Variabilität in der Feststellung von Erkrankungen miteinbeziehen.

**Schlüsselwörter:** Risiko-Adjustierung – Begleiterkrankungen – Pflegekräfte – Übereinstimmung – Bias

**Risk Adjustment Using Morbidity Information Assessed by Nurses: Utility for Epidemiological and Economic Studies in Clinical Research and Quality Management. Objectives:** To what extent do nurses and physicians agree in the assessment of the prevalence and severity of medical conditions and to what extent can morbidity data provided by nurses be

used for risk adjustment in epidemiological and economic studies? **Patients and methods:** In two cross-sectional studies of 50 medical patients each, the authors studied the agreement between nurses and physicians in the assessment of conditions listed in the Index for CoExisting Diseases (ICED). Phase 1 studied the agreement when patients were evaluated by both groups at the same time. In phase 2 physicians assessed patients at discharge. **Results:** Nurses recorded on average fewer conditions (2.5 vs. 3.8;  $p = 0.006$ ) and rated them less severe (2.5 vs. 3.0 on a scale from 0 to 4;  $p = 0.001$ ). Agreement in the prevalence of conditions was 84% (70%–90%) in phase 1 and 80% (67%–89%) in phase 2. The agreement decreases continuously with an increasing number of severity categories that were to be assessed. Assuming physicians to be the gold standard, nursing information on the prevalence of conditions showed a sensitivity of 0.63 and specificity of 0.83. **Conclusions:** The assessment of medical conditions by nurses without a specific training may not be sufficient for epidemiological and economic studies in medical research. Larger studies should also consider the variability between physicians in assessing medical conditions.

**Key words:** Risk Adjustment – Comorbidity – Nurses – Agreement – Bias

### Hintergrund

Die Nicht-Vergleichbarkeit von Patienten bzw. Patientengruppen ist eine bekannte Herausforderung in der klinisch-epidemiologischen und -ökonomischen Forschung: Warum reagieren manche Patienten besser auf eine Therapie als andere und warum verursachen Patienten mit gleicher Grunderkrankung unterschiedliche Kosten?

In der klinischen Prüfung von Medikamenten wird die Vergleichbarkeit von zwei Patientengruppen durch eine zufällige Aufteilung der Patienten in die jeweiligen Vergleichsgruppen erreicht (Randomisierung). Im Mittel sind damit Patienteneigenschaften wie Geschlecht, Alter, Schweregrad der Grunderkrankung und Begleiterkrankungen zu Beginn der Intervention in beiden Gruppen gleich verteilt [1]. Randomisierte Studien sind jedoch aus einer Vielzahl von Gründen im Bereich der klinischen Forschung häufig nicht möglich [2,3]. Man bildet daher Patientengruppen oder „Schichten“ mit gleichen Eigenschaften wie z. B. gleiche Hauptdiagnose, Alter, Geschlecht, Be-

gleiterkrankungen etc., die gleichermaßen Prädiktoren für die gesundheitlichen Konsequenzen („outcomes“) bzw. Krankheitskosten sind.

Ziel einer Risiko-Adjustierung ist es, nur Patienten innerhalb Schichten mit ähnlichen Patienteneigenschaften zu vergleichen [4]. Die Ergebnisse aller Schichten werden in einem gewichteten Mittel zusammengefasst [5]. Eine effizientere Erweiterung dieses Verfahrens ist die Anpassung multivariater Regressionsmodelle (z. B. [6]).

Ein wesentlicher Bestandteil der epidemiologischen Risiko-Adjustierung ist die Erfassung von Begleiterkrankungen und deren Schweregrad. Begleiterkrankungen werden im Rahmen einer Risiko-Adjustierung als Störgrößen („confounder“) bezeichnet. Werden Confounder-Variablen durch die Erfasser fehlklassifiziert, so kann dieser Confounder nur unzureichend adjustiert werden, was wiederum eine Assoziation zwischen einer Exposition und dem Haupteffekt verzerren kann.

Eine Vielzahl von Instrumenten zur Erfassung von Begleiterkrankungen ist international verfügbar (z. B. [7,8]). Obwohl im Allgemeinen von Begleiterkrankungen gesprochen wird, erfassen diese Instrumente in der Regel alle bestehenden Erkrankungen der Patienten einschließlich der Haupterkrankung, die zur Einweisung geführt hatte und sind damit eher als Messinstrumente der gesamten Morbidität eines Patienten zu verstehen. Wir verwenden daher den allgemeineren Begriff „Erkrankungen“, der Begleiterkrankungen miteinschließt. Das Ausmaß einer Risiko-Adjustierung, das durch diese Instrumente erreicht werden kann, variiert und ist von der Patientenzusammensetzung und der jeweiligen Zielgröße einer Untersuchung abhängig [9 – 11].

Die standardisierte Erfassung und Bewertung von Erkrankungen stehen damit im Mittelpunkt aller nicht-randomisierten klinisch-epidemiologischen Studien. Insbesondere in Nordamerika wird die Erfassung von Erkrankungen im Rahmen von klinischen Studien häufig durch ausgebildetes Kodierpersonal oder durch speziell fortgebildetes Pflegepersonal („Study Nurse“) übernommen [8]. Hauptgrund dafür ist die zeitlich wie ökonomisch effizientere Abwicklung von klinischen Studien. Untersuchungen haben gezeigt, dass bei entsprechendem Training derartige Erfassungen mit Bewertungen durch ärztliches Personal vergleichbar sein können [12]. Der Einsatz von standardisierten Instrumenten zur Erfassung von Erkrankungen im Rahmen von klinischen Studien durch das Pflegepersonal ist im deutschen Sprachraum unüblich und Erfahrungen über die Validität solcher Angaben in Bezug zu ärztlichen Angaben liegen nicht vor.

Die Fragestellung der vorliegenden Arbeit war, inwieweit die Beurteilung ausgewählter Krankheitsbilder zwischen Pflegepersonal und Ärzten in der internistischen Krankenhausversorgung übereinstimmt und inwieweit Angaben des Pflegepersonals ohne spezifisches Training in der klinisch-epidemiologischen und -ökonomischen Forschung für eine Risiko-Adjustierung eingesetzt werden können?

## Patienten und Methodik

Die vorliegende Studie wurde in der Internistischen Abteilung einer Klinik der Regelversorgung mit insgesamt 195 Planbetten, davon 65 in einer Abteilung für Innere Medizin, durchgeführt.

In die Studie aufgenommen wurden alle Patienten, die konsekutiv zwischen dem 7. Juni und 28. Juli 2000 stationär auf eine der Stationen der Abteilung für Innere Medizin aufgenommen wurden. Ausgeschlossen wurden Patienten, die ausschließlich zu diagnostischen Zwecken in das Schlaflabor der Abteilung aufgenommen wurden. Die Verweildauer der Patienten war für die Aufnahme in die Studie kein Ausschlusskriterium. Die Datenerhebung endete in Phase 1 und Phase 2 jeweils mit Erreichen der angestrebten Fallzahl von 50 Patienten (siehe unten).

### Studiendesign und Durchführung

Entsprechend einem vorher festgelegten Studienprotokoll wurden die Erfassung und Bewertung von Erkrankungen durch pflegerisches und ärztliches Personal in zwei aufeinanderfolgenden Erhebungsphasen verglichen: Phase 1 umfasste die Datenerhebung durch Pflegepersonal und ärztliches Personal zum gleichen Zeitpunkt bei der Aufnahme (+ 24 h). In Phase 2 erfasste das Pflegepersonal zum Zeitpunkt der Aufnahme (+ 24 h), das ärztliche Personal erfasste die Erkrankungen für die gleichen Patienten jedoch zum Zeitpunkt der Entlassung (-24 h).

Phase 2 prüft, inwieweit während des Krankenhausaufenthalts weitere chronische Erkrankungen diagnostiziert und angegeben werden. Nachdem sich der Schweregrad einzelner Diagnosen während des Aufenthalts erheblich ändern kann berücksichtigt die Analyse von Phase 2 lediglich das Vorhandensein von Erkrankungen und nicht deren Schweregrad.

Die Bewertungen wurden durch examiniertes Pflegepersonal mit jeweils dreijähriger Ausbildung, 2 Ärzten im Praktikum und 4 Assistenzärzten ohne Facharztanerkennung durchgeführt.

Es wurde durch eine Informationsveranstaltung und Studienmanagement sichergestellt, dass beide Gruppen ihre Erhebung unabhängig voneinander durchführten und kein Datenabgleich erfolgte. Beide Gruppen erhielten als Studienmaterial dieselbe Übersichtstabelle zur Einstufung der Schweregrade der zu erfassenden Erkrankungen (siehe Anhang 2). Es wurde keine spezifische Schulung zur Bewertung des Vorhandenseins und des Schweregrades der Erkrankungen durchgeführt. Beide Berufsgruppen waren instruiert, für die Datenerfassung keine zusätzlichen inhaltlichen oder zeitlichen Ressourcen einzusetzen. Das Pflegepersonal sollte die Daten durch direkte Nachfrage beim Patienten im Rahmen der stationären Aufnahmedokumentation erheben und durch Informationen, die den Pflegekräften in der Routineversorgung zur Verfügung stehen, wie z. B. Medikationspläne, Einweisungsschein, Vorbefunde oder den Verlegungsbrief, ergänzen.

Das ärztliche Personal sollte die Daten ebenfalls im Rahmen ihres herkömmlichen Informations- und Dokumentationsmodus, d. h. unter Einbeziehung der Aktenlage und – in Phase 2 – aller während des stationären Aufenthalts gesammelten Untersuchungsergebnisse, erheben.

Um die Datenqualität vergleichbar und den Zeitaufwand gering zu halten, wurde bei der Einführung darauf hingewiesen, die Fragebogen von einer Person am Aufnahmezeitpunkt (+ 24 h) in einem einmaligen Arbeitsgang abzuschließen. Sobald ein Bogen ausgefüllt war, sollte dieser nicht weiter eingesehen oder nachgebessert werden.

Es wurde bewusst darauf verzichtet, den Einfluss des Weiterbildungsstands des Pflegepersonals und des ärztlichen Personals differenziert zu untersuchen, da bei der Datenerfassung der Austausch unter Kollegen/-innen der gleichen Berufsgruppe nicht ausgeschlossen worden war.

**Erhebungsinstrument**

Für die Datenerhebung wurde ein 1-seitiger standardisierter Fragebogen eingesetzt (Anhang 1). Neben Aufnahme- bzw. Entlassungsdatum, Alter, Geschlecht und Hauptdiagnose (ICD-10) wurden die Erkrankungen der Studienpatienten mit Beurteilung des jeweiligen Schweregrads erhoben. Die Erfassung der Erkrankungen erfolgte in Anlehnung an eine deutsche Adaptierung des ICED (Index of CoExistent Disease) [8,13]. Untersuchungen haben gezeigt, dass der ICED gemeinsam mit den Variablen Alter und Geschlecht die Variabilität von Ergebnissen der medizinischen Versorgung in unterschiedlichen Patientengruppen zu einem großen Teil erklären kann [14,15]. Lediglich die Subskala zur Bestimmung des Schweregrades von Erkrankungen (Index of Disease Severity, IDS) wurde verwendet. Dabei werden für 15 ausgewählte Erkrankungen das Vorliegen und die entsprechende Beurteilung des Schweregrads nach einer fünfstufigen Schweregradskala erfasst. Die Beurteilung richtet sich nach einer speziellen Liste, die anhand von Symptomen, Hinweiszeichen und Laborwerten die Einteilung des Schweregrades unterstützt und als Leitlinie gilt (siehe Anhang 2).

In einem zweiten Schritt wurde daraus der ICED/IDS Score gebildet, indem die jeweils höchsten Schweregrade je Patienten in einem von 0 bis 4 reichenden Score zusammengefasst wurden. Zusätzlich wurde die Gesamtsumme aller Schweregrade der 15 Krankheitsgruppen gebildet. Der maximal erreichbare Gesamtscore beträgt 60.

**Analyse**

Für eine umfassende Analyse der Übereinstimmung wurden die Kategorien zur Beurteilung des Schweregrades der Erkrankungen schrittweise zusammengefasst: Die ursprünglich 5 Ka-

tegorien (0 = keine Erkrankung, 1 = asymptomatisch, 2 = leichte Symptome, 3 = deutliche Symptome, 4 = schwere Symptome) wurden zusammengefasst in 3 (0/1,2/3,4) bzw. 2 Kategorien (0 = Erkrankung nicht vorhanden/1,2,3 oder 4 = Erkrankung vorhanden).

Das Ausmaß der Übereinstimmung zwischen beiden Berufsgruppen wurde berechnet, indem die Anzahl der Patienten mit übereinstimmender Beurteilung beider Gruppen geteilt wurde durch die Anzahl aller Studienpatienten.

Die Übereinstimmung von Untersucherpaaren wurde zusätzlich mittels Cohen's Kappa Statistik (bei 2 Kategorien) und linear gewichteten Kappas (bei 3 oder 5 Kategorien) berechnet, die beide für den Anteil an zufälliger Übereinstimmung adjustieren [16 - 18]. Wenn einzelne Schweregrade nicht von beiden Gruppen innerhalb der Studienpopulation angegeben wurden, kann keine Kappa-Statistik berechnet werden.

Für die Analyse von 2 Ausprägungen wurde außerdem Sensitivität und Spezifität für die Richtigkeit der pflegerischen Angaben berechnet unter der Annahme, dass die ärztlichen Angaben als Goldstandard zu sehen sind.

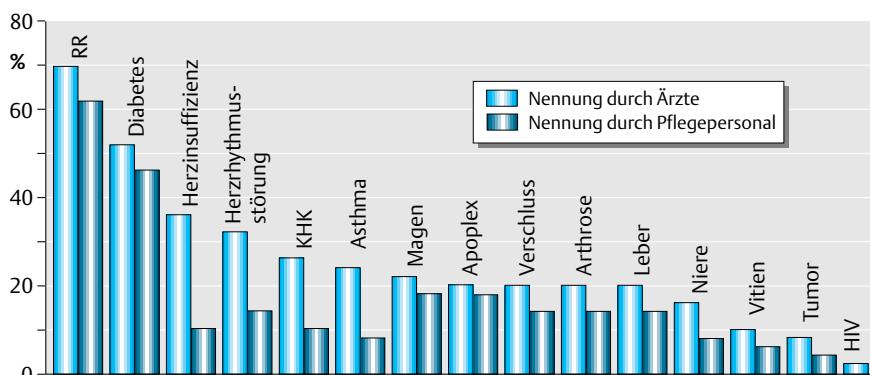
Der lineare Zusammenhang der Gesamtsummen der Schweregrade zwischen Pflegekräften und Ärzten wurde mit dem nichtparametrischen Spearman Korrelationskoeffizienten quantifiziert, da bei Betrachtung der Gesamtsummen nicht von normalverteilten Daten ausgegangen werden konnte.

Unterschiede in der mittleren Anzahl von genannten Erkrankungen und Schweregraden wurden mit dem nicht-parametrischen Mann-Whitney-Test geprüft [19].

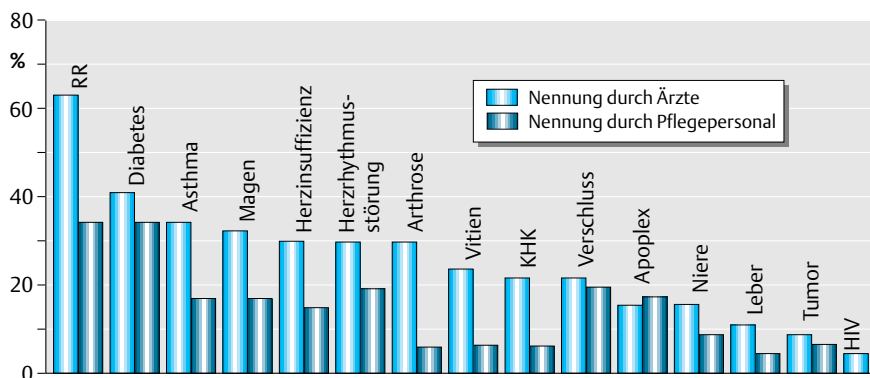
Für alle Schätzer wurden 95%-Konfidenzintervalle (KI) angegeben. Alle Berechnungen wurden mit der SAS Software durchgeführt [20].

**Ergebnisse**

Die organisatorische Einführung des Fragebogens in die klinische Versorgung war unproblematisch. Eine Informationsveranstaltung nahm 60 Minuten in Anspruch. Das Durchschnittsalter der Stichprobe von Phase 1 (n = 50) lag bei 64,2 Jahren (± 17,34), 54% der Studienteilnehmer waren weiblich. In Phase 2 reduzierte sich die angestrebte Stichprobengröße von n = 50 wegen zwei Todesfällen und zwei Verlegungen auf 46 Patienten, da keine abschließenden ärztlichen Untersuchungen



**Abb. 1** Prävalenz von Erkrankungen nach Angabe von Ärzten und Pflegepersonal zum gleichen Zeitpunkt (Phase 1, n = 50).



**Abb. 2** Prävalenz von Erkrankungen nach Angabe von Ärzten und Pflegepersonal zu verschiedenen Zeitpunkten (Phase 2, n = 46).

**Tab. 1** Übereinstimmung (Ü) der Beurteilung von Erkrankungen in der internistischen Versorgung zwischen Pflegekräften und Ärzten zum gleichen Zeitpunkt in Phase 1. Übereinstimmungen und 95% Konfidenzintervalle (KI) wurden für 5, 3 und 2 Kategorien\* dargestellt

Erkrankung	Anzahl d. Beurteilungspaare	5 Kategorien (0/1/2/3/4)*		3 Kategorien (0/1 + 2/3 + 4)*		2 Kategorien (0/1)*			
		Ü in % (95% KI)	gew. Kappa (95% KI)	Ü in % (95% KI)	gew. Kappa (95% KI)	Ü in % (95% KI)	Kappa (95% KI)	Sensitivität	Spezifität
KHK	50	76 (63–86)	0,23 (0–0,47)	78 (65–87)	0,24 (0–0,49)	80 (67–89)	0,35 (0,06–0,64)	0,80	0,80
Herzinsuffizienz	50	68 (54–79)	–	72 (58–83)	0,38 (0,13–0,64)	74 (60–84)	0,33 (0,10–0,56)	1	0,71
Herzrhythmusstörung	50	70 (56–81)	–	70 (56–81)	0,12 (0–0,32)	70 (56–81)	0,19 (0–0,46)	0,58	0,72
andere Herzkrankung	50	86 (74–93)	–	86 (74–93)	–	88 (76–94)	0,19 (0–0,60)	0,33	0,91
Bluthochdruck	50	44 (31–58)	0,4 (0,21–0,58)	58 (44–71)	0,43 (0,23–0,64)	76 (63–86)	0,47 (0,22–0,72)	0,87	0,58
Schlaganfall	50	78 (65–87)	0,4 (0,10–0,69)	80 (67–89)	0,45 (0,15–0,76)	82 (69–90)	0,42 (0,10–0,73)	0,55	0,88
Periphere art. Verschlusskrankheit	50	74 (60–84)	0,22 (0–0,5)	74 (60–84)	0,2 (0–0,49)	78 (65–87)	0,23 (0–0,55)	0,43	0,84
Diabetes	50	76 (63–86)	0,8 (0,70–0,91)	84 (71–92)	0,83 (0,72–0,94)	94 (84–98)	0,88 (0,75–1,01)	1	0,88
Chronische Atemwegserkrankung	50	82 (69–90)	–	82 (69–90)	–	84 (71–92)	0,43 (0,13–0,73)	1	0,83
Arthrose	50	74 (60–84)	0,24 (0–0,57)	74 (60–84)	0,2 (0–0,49)	78 (65–87)	0,23 (0,20–0,80)	0,43	0,86
Magen-Darm-Erkrankung	50	72 (58–83)	–	80 (67–89)	0,46 (0,18–0,74)	84 (71–92)	0,5 (0,19–0,82)	0,43	0,81
Lebererkrankung	50	78 (65–87)	–	80 (67–89)	0,35 (0,10–0,59)	86 (74–93)	0,51 (0,08–0,80)	0,71	0,88
Nierenerkrankung	50	84 (71–92)	–	90 (79–96)	0,45 (0,10–0,80)	88 (76–94)	0,44 (0,08–0,80)	0,75	0,90
Bösartiger Tumor	50	90 (76–96)	–	90 (79–96)	–	92 (81–97)	0,3 (0–0,79)	0,50	0,94
HIV/AIDS	50	98 (90–100)	–	98 (90–100)	–	98 (90–100)	–	0,02	0,98

\* Die maximal 5 Kategorien des ICD wurden in zusätzlichen Analysen zu 3 bzw. 2 Kategorien zusammengefasst: 5 Kategorien: 0 = keine Erkrankung, 1 = asymptomatisch, 2 = leichte Symptomatik, 3 = deutliche Symptomatik, 4 = schwere Symptomatik, 3 Kategorien: 0/1 + 2/3 + 4; 2 Kategorien: 0/1: 0 = Begleiterkrankung trifft nicht zu, 1 = Begleiterkrankung trifft zu

durchgeführt werden konnten. Der Anteil an Frauen lag in Phase 2 bei 58%, das Durchschnittsalter betrug 65,4 Jahre ( $\pm 17,75$ ) und die mittlere Verweildauer 11,3 Tagen ( $\pm 6,37$ ).

Pflegekräfte erfassten im Durchschnitt eine geringere Anzahl von Erkrankungen pro Patient als Ärzte (2,5 vs. 3,8;  $p = 0,006$ ), wenn diese Beurteilung zum gleichen Zeitpunkt stattfand (siehe Abb. 1). Beurteilten die Ärzte bei Entlassung, so vergrößerte sich dieser Unterschied leicht (2,2 vs. 3,8;  $p < 0,001$ ; siehe Abb. 2). Pflegekräfte bewerteten den Schweregrad der Erkrankungen im Mittel geringfügig leichter als Ärzte (Zeitgleich: 2,5 vs. 3,0;  $p = 0,001$ ; Phase 2: 2,3 vs. 3,1;  $p < 0,001$ ).

Übereinstimmungen von weniger als 80% zeigten sich für die häufigsten kardiovaskulären Krankheitsbilder, wie Herzinsuffizienz, Herzrhythmusstörungen und Bluthochdruck. Die Übereinstimmung lag in Phase 1 (Beurteilung des Schweregrades zum gleichen Zeitpunkt) für 5 Kategorien bei 77% (95% Konfidenzintervall 68%–98%) für 3 Kategorien bei 80% (70%–90%) und für 2 Kategorien bei 84% (70%–90%) (Tab. 1). In Phase 2 (Beurteilung der Übereinstimmung zu unterschiedlichen

Zeitpunkten) lag das Mittel für 5 Kategorien bei 72% (58%–83%), für 3 Kategorien 75% (61%–85%) und 2 Kategorien bei 80% (67%–89%) (Tab. 2).

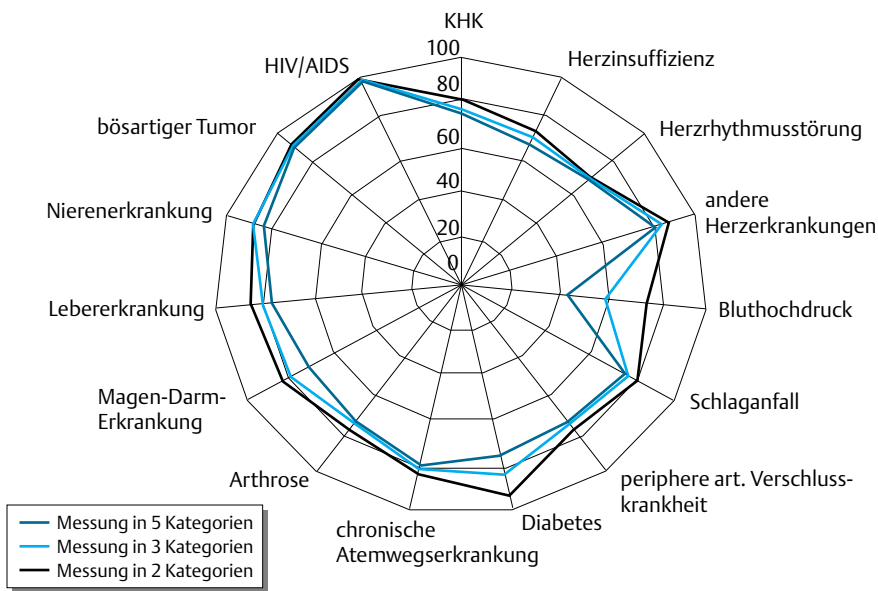
Insgesamt nahm das Maß der Übereinstimmung mit Reduzierung der Anzahl der Kategorien zu, d. h. je weniger differenziert der Schweregrad einer Erkrankung nachgefragt wurde, umso höher war die Übereinstimmung zwischen Ärzten und Pflegepersonal. Deutliche Verschlechterungen der Übereinstimmung zeigten sich für Bluthochdruck und Diabetes, mit zunehmender Differenzierung des Schweregrades (siehe Abb. 3 und 4).

Das Pflegepersonal erkannte Erkrankungen in Phase 1 mit einer Sensitivität von 0,63 und Spezifität von 0,83, sofern angenommen wird, dass die ärztliche Feststellung von Erkrankungen den Goldstandard darstellt.

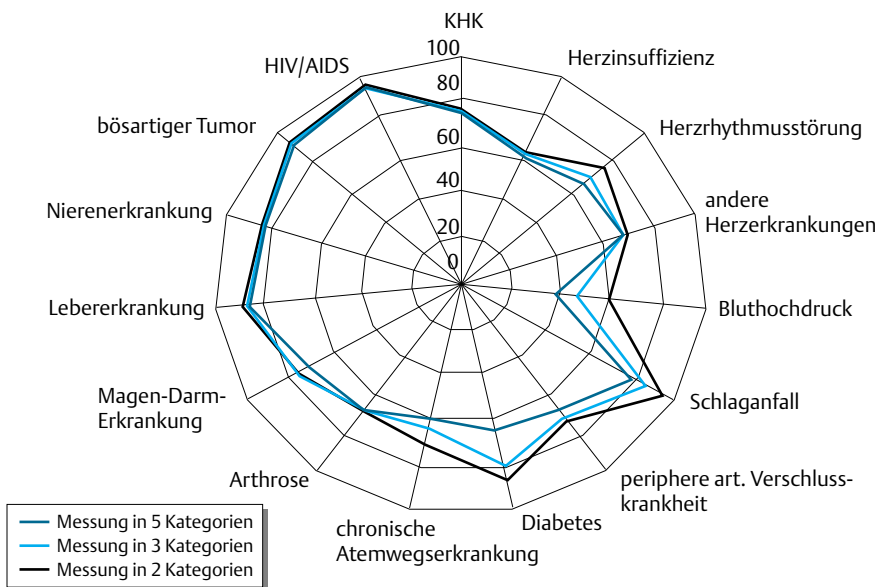
Bildet man einen Komorbiditäts-Score (Grade 0 bis 4) über alle Erkrankungen, so zeigte sich in Phase 1 eine Übereinstimmung von 44% (31%–58%). Für Phase 2 ergab sich 33% Übereinstimmung (21%–47%).

**Tab. 2** Übereinstimmung (Ü) der Beurteilung von Erkrankungen in der internistischen Versorgung zwischen Pflegekräften und Ärzten zu verschiedenen Zeitpunkten in Phase 2. Übereinstimmungen und 95% Konfidenzintervalle (KI) wurden ausschließlich für 2 Kategorien (Erkrankung vorhanden/nicht vorhanden) dargestellt

Erkrankung	Anzahl d. Beurteilungs-paare	2 Kategorien (Erkrankung vorhanden/nicht vorhanden)			
		Ü in % (95% KI)	Kappa (95% KI)	Sensitivität	Spezifität
KHK	46	76 (62–86)	0,06 (0–0,33)	0,33	0,79
Herzinsuffizienz	46	63 (49–75)	0 (0–0,24)	0,29	0,69
Herzrhythmusstörung	46	76 (62–86)	0,37 (0,08–0,67)	0,66	0,78
andere Herzerkrankung	46	70 (56–81)	0	0	0,74
Bluthochdruck	46	59 (45–72)	0,24 (0–0,47)	0,81	0,46
Schlaganfall	46	93 (82–98)	0,76 (0,5–1)	0,75	0,97
Periphere art. Verschluss-krankheit	46	72 (58–83)	0,14 (0–0,46)	0,33	0,81
Diabetes	46	85 (72–93)	0,68 (0,46–0,90)	0,88	0,83
Chronische Atemwegs-erkrankung	46	70 (56–81)	0,24 (0–0,52)	0,63	0,71
Arthrose	46	67 (53–79)	0,01 (0–0,22)	0,43	0,7
Magen-Darm-Erkrankung	46	76 (62–86)	0,38 (0,10–0,66)	0,75	0,76
Lebererkrankung	46	89 (77–95)	0,24 (0–0,68)	0,5	0,9
Nierenerkrankung	46	85 (72–93)	0,28 (0–0,67)	0,5	0,88
Bösartiger Tumor	46	93 (82–98)	0,54 (0,08–0,99)	0,66	0,95
HIV/AIDS	46	96 (86–99)	– –	0	1



**Abb. 3** Ausmaß der Übereinstimmung der unabhängigen Beurteilung von internistischen Erkrankungen durch Pflegekräfte und Ärzte zum gleichen Zeitpunkt in Prozent (n = 50).



**Abb. 4** Ausmaß der Übereinstimmung der unabhängigen Beurteilung von internistischen Erkrankungen durch Pflegekräfte und Ärzte zu verschiedenen Zeitpunkten in Prozent (n = 46).

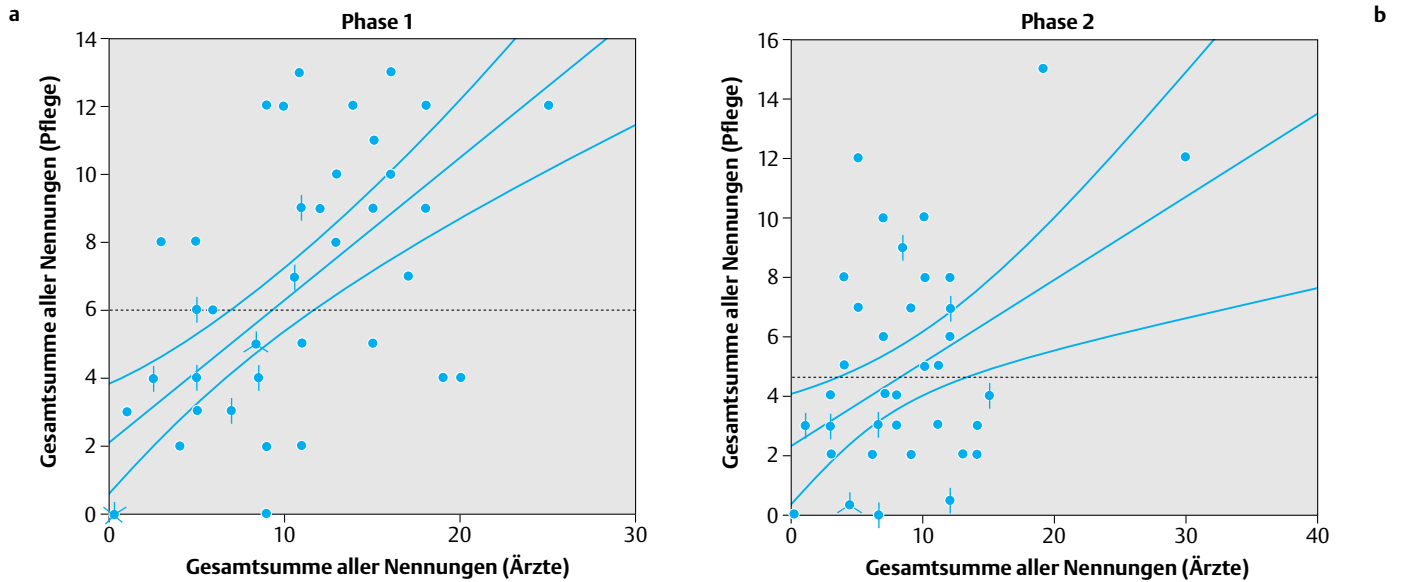
Für den linearen Zusammenhang der Gesamtsummen der Schweregrade aller Erkrankungen ergaben sich Spearman Korrelationskoeffizienten in Phase 1 von  $r = 0,65$  (0,45–0,79) und in Phase 2 von  $r = 0,29$  (0–0,54, siehe Abb. 5).

## Diskussion

In einem Vergleich der Bewertung von Erkrankungen anhand eines standardisierten Erfassungsinstrumentes durch examinierte Pflegekräfte und Ärzte konnte gezeigt werden, dass die Übereinstimmung über ein Spektrum von 15 Erkrankungen insgesamt mit 84% (Phase 1) und 80% (Phase 2) relativ hoch erschienen. Es muss jedoch darauf hingewiesen werden, dass Übereinstimmungen immer zu hohen Werten tendieren, wenn, wie in den vorliegenden Daten, die Prävalenz von Erkrankungen niedrig und die Spezifität der Beurteilung hoch ist [21]. Für einzelne Diagnosen, insbesondere kardiovaskuläre Erkrankungen, fielen diese Übereinstimmungen niedriger aus. Dabei identifizierte das Pflegepersonal systematisch weniger

Erkrankungen und bewertete Erkrankungen als weniger schwerwiegend als die ärztlichen Kollegen. Die Kappa-Statistiken fallen für viele der Diagnosen niedriger als erwartet aus. Dieses liegt an der numerischen Eigenschaft von Kappa-Statistiken, niedrige Werte zu liefern, wenn die Prävalenz einzelner Schweregrade sehr gering ist bzw. einzelne Diagnosen in dem Patientenkollektiv kaum vorkamen, z. B. HIV-Infektionen [22].

Wie erwartet, waren die Übereinstimmungen bei der Einteilung der Erkrankungen in nur 2 Kategorien (vorhanden/nicht vorhanden) systematisch höher als bei einer Differenzierung nach Schweregrad. Dies ist numerisch auf die zunehmende Variationsmöglichkeit zurückzuführen, zeigt jedoch auch den möglicherweise begrenzten Gewinn von mehr als zwei Schweregradeinteilungen. Insbesondere im Falle von Bluthochdruck und Diabetes, deren Beurteilung des Vorhandenseins gut und sehr gut übereinstimmte, verringerte sich die Übereinstimmung erheblich. Dies könnte durch eine unvollständige Darstellung der Symptomatik durch Patienten oder mangelndes



**Abb. 5** Linearer Zusammenhang zwischen ICED-Gesamtsummen nach Beurteilung durch Pflegekräfte und Ärzte.

Detailverständnis der Pflegekräfte erklärt werden. Alternativ könnten sich Pflegekräfte genauer an die Definitionen der Schweregradeinteilung des ICED gehalten haben, Ärzte jedoch eher eine Globalbegutachtung entsprechend den vorgegebenen Kategorien „leichte, deutliche und schwere Symptomatik“ abgegeben haben.

Werden pflegerische Beurteilungen bei Aufnahme mit ärztlichen Beurteilungen bei Entlassung verglichen, so verringert sich die Übereinstimmung geringfügig von 84% auf 80%. Wir gehen davon aus, dass diese Reduktion in der Übereinstimmung auf die zusätzlichen, während des stationären Aufenthalts gewonnenen Informationen zurückzuführen ist, die Ärzten zur Verfügung standen. Es ist unwahrscheinlich, dass es sich bei den zusätzlich angegebenen Erkrankungen um Komplikationen handelte, da nur chronische Erkrankungen abgefragt wurden, die bei Aufnahme möglicherweise nicht diagnostiziert waren.

Welche Konsequenzen ergeben sich aus diesen Ergebnissen für die klinisch-epidemiologische und klinisch-ökonomische Forschung? Der Haupteffekt kann durch fehlerklassifizierte Confounder in Abhängigkeit mehrerer Faktoren verzerrt sein: Im Allgemeinen gilt, dass mit zunehmendem Ausmaß einer zufälligen Fehlklassifikation des Confounders, mit zunehmendem Ausmaß der Confounder-Outcome-Beziehung und mit zunehmender Prävalenz des Confounders die Verzerrung des Haupteffekts zunimmt [23]. Tab. 3 stellt das Ausmaß der möglichen Verzerrung durch missklassifizierte Confounder-Variablen dar [24]. Grau schattiert ist ein Szenario, das etwa unseren Daten entspricht: 60% Sensitivität, 90% Spezifität, 10% Prävalenz der Erkrankungen und unter der Annahme, dass die Begleiterkrankungen die Wahrscheinlichkeit des Auftretens eines Studien-outcomes verdoppeln (Odds ratio, OR = 2,0). Dies würde zu einer Verzerrung des Haupteffekts von 21% (OR = 1,21) führen. Eine Verbesserung der Sensitivität der pflegerischen Daten auf 90% würde die Verzerrung auf 16% reduzieren.

**Tab. 3** Auswirkung einer nicht-differenziellen Fehlklassifikation in der Bestimmung von (Begleit-)Erkrankungen auf eine Assoziation zwischen Exposition und Zielgröße (outcome). Es wurde angenommen, dass kein Zusammenhang zwischen Exposition und Studien-Outcome besteht, d. h.  $OR_{EO} = 1.0$

Sensitivität für (Begleit-) Erkrankung	Spezifität für (Begleit-) Erkrankung	Prävalenz der (Begleit-) Erkrankung	Geschätztes Odds Ratio* für die Assoziation zwischen Begleiterkrankung und Studien-outcome			
			1,5	2,0	5,0	10,0
0,6	0,6	0,1	1,14	1,25	1,62	1,84
0,6	0,6	0,5	1,18	1,33	1,94	2,53
0,6	0,9	0,1	1,12	1,21	1,51	1,70
0,6	0,9	0,5	1,14	1,25	1,68	2,10
0,6	0,99	0,1	1,07	1,13	1,32	1,43
0,6	0,99	0,5	1,11	1,20	1,54	1,87
0,9	0,9	0,1	1,09	1,16	1,38	1,52
0,9	0,9	0,5	1,06	1,12	1,31	1,51

\* Unter Annahme des tatsächlichen Odds Ratio ( $OR_{EO}$ ) von 1,0 ergibt sich eine Verzerrung aus der Differenz mit dem geschätzten  $OR_{EO}$ , z. B.  $1,21 - 1,00 = 0,21$  oder 21%. Tabelle nach Kelsey et al. [24].

In Abhängigkeit von der Fragestellung und der Messgenauigkeit weiterer Confounder ist eine Verzerrung von 21 % in manchen Fällen durch die Einsparungen bei der Datengewinnung nicht zu vertreten. Ausgiebigeres Training von Pflegekräften im Sinne von „Study Nurses“ kann jedoch eine lohnende Investition sein, wenn sich die Sensitivität der Bewertung verbessern lässt und das Personal für mehrere Studien oder große Stichproben eingesetzt werden kann.

Die Korrektur von verzerrten Haupteffekten ist sowohl in einfachen Fällen mit nur einer Confoundervariable [25] als auch in multivariaten Modellen unter Verwendung von Informationen analog dieser Validierungsstudie und einer Rückgewichtung möglich [26,27]. Jedoch sind diese Verfahren aufwändig und die Schätzer der Haupteffekte verlieren durch die zusätzliche Schätzungenauigkeit in der Validierungsstudie an Präzision.

Eine wesentliche Einschränkung dieses Pilotprojekts ist die mäßig große Fallzahl. In dem für internistische Patienten relativ jungen Patientenkollektiv ist das Auftreten von einigen Erkrankungen zu gering, um stabile Schätzer der Übereinstimmung zu erhalten. Dennoch ist in diesen Fällen meist ein klarer Trend zu erkennen.

Die Ergebnisse sind nur eingeschränkt auf Spezialabteilungen übertragbar. Unklar bleibt aufgrund unserer Daten auch das Ausmaß der Variabilität in der Feststellung von Erkrankungen innerhalb der Ärzte. Diese wird in der Literatur häufig als hoch beschrieben [28]. Für die Abschätzung der Auswirkungen von

Fehlklassifikation von Erkrankungen in der klinisch-epidemiologischen und -ökonomischen Forschung wurde jedoch davon ausgegangen, dass die ärztliche Einschätzung den Goldstandard darstellt. Da dies vermutlich nicht realistisch ist, könnte sich die tatsächliche Übereinstimmung besser darstellen.

Füllen Patienten selber einen Fragebogen zu ihren bestehenden Erkrankungen aus, so haben wissenschaftliche Untersuchungen in den USA gezeigt, dass ein solches Instrument – ursprünglich für die medizinischen Berufe entwickelt [7] – mit guter Übereinstimmung beantwortet werden kann [29]. Es ist derzeit jedoch noch unklar, inwieweit sich diese Ergebnisse auf Deutschland übertragen lassen.

Zusammenfassend zeichnet sich ab, dass eine Beurteilung von (Begleit-)Erkrankungen durch pflegerisches Personal ohne eine spezielle Schulung im Sinne einer „Study Nurse“ für einzelne Anwendungen zum jetzigen Zeitpunkt nicht ausreichend ist. Umfangreichere Studien müssen die ärztliche Variabilität in der Feststellung von Erkrankungen miteinbeziehen.

### Danksagung

Wir danken allen Pflegekräften und Ärzten der Helios Klinik Blankenhain für ihre aufgeschlossene Unterstützung dieser Studie. Dr. Schneeweiss wurde unterstützt durch die Deutsche Forschungsgemeinschaft (SCHN 527/3 & SCHN 527/4), U.S. Agency for Health Care Research and Quality (RO3HSO9855, RO1HS10881) und dem AOK Bundesverband.

**Anhang 1:** Standardisierte Erfassung des ICED

Patienten-Klebe-Etikett

! Bitte unbedingt abtrennen, bevor der Fragebogen die Klinik verlässt !



<b>Fragebogen 1</b>	<b>Studie zur Übereinstimmung der Erfassung von (Begleit-)Erkrankungen (Teil 1)</b> Einstufung durch Pflegepersonal	Code-Nr. (1-50)
-------------------------	--	--------------------

Aufnahme (+ 24 Stunden)

- 1. Geschlecht  männlich  weiblich
- 2. Geburtsdatum (TT. MM. JJ) □□. □□. 19
- 3. Aufnahmedatum (TT. MM. JJ) □□. □□. 20
- 4. Hauptdiagnose (ICD-10, vierstellig) .□□□.□.

5. Welche der folgenden (Begleit-)Erkrankungen treffen für den Patienten zu ?  
(Mehrfachnennungen möglich. Nicht gekreuzte Krankheiten bedeuten, dass Erkrankung nicht besteht)

(Begleit-)Erkrankungen:	<b>Symptomatik</b>			
	asymptomatisch (keine Behandlung)	leicht (behandlungs- bedürftig)	deutlich (unter Behandlung)	schwer (akute Behandlung notwendig)
KHK, Angina pectoris, Herzinfarkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Herzinsuffizienz	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Herzrhythmusstörungen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
andere Herzerkrankungen (Vitien)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bluthochdruck	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Schlaganfall	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
periphere art. Verschlusskrankheit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Diabetes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
chronische Atemwegserkrankungen/Asthma	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Arthrose, Arthritis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Magen-Darm-Erkrankung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lebererkrankung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nierenerkrankung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
bösartiger Tumor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
HIV/AIDS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Anhang 2:** Studienhilfe zum ICED

Erkrankung	Schweregrad			
	1	2	3	4
<b>Herzerkrankungen</b>				
KHK, Angina pectoris, Herzinfarkt (Myokardinfarkt MI)	asymptomatisch mit ischämischen EKG-Abnormalitäten, leichte Angina bei starker körperl. Belastung (NYHA I–II)	Anamnese eines MI oder Bypasses ohne Residualeffekt, minimale Herzinsuff., Angina oder Atemnot bei Aktivitäten des täglichen Lebens (ein Stockwerk), NYHA II	akuter MI in den vergangenen 6 Monaten, mittlere bis schwere Herzinsuff., Atemnot in Ruhe, NYHA III und IV	Koma, Herzstillstand, akutes Lungenödem, akuter MI, kompliziert durch kardiogenen Schock
Herzinsuffizienz	Anamnese mit einzelner Episode einer Herzinsuff., einfach zu beherrschen, ohne Komplikationen	leichtes Fußödem, leichte Atemnot bei Belastung und/oder in Ruhe, Anamnese mit mehrfachen Episoden von Herzinsuffizienz, derzeit medikamentös eingestellt	entsprechend der kardiovaskulären Erkrankung (mäßige bis ausgeprägte Atemnot unter Belastung, mäßige bis ausgeprägte Fußödeme, Herzvergrößerung, chron. Müdigkeit)	entsprechend der kardiovaskulären Erkrankung (Koma, Herzstillstand)
primäre Herzrhythmusstörungen	asymptomatisch, keine Medikation erforderlich, nur EKG Veränderungen	geringe Symptome, mit Medikamenten oder Schrittmacher behandelt	ausgeprägte Symptome wie Schwindel oder Synkopen infolge Arrhythmie oder Blockierungen im Reizleitungssystem	primärer Herzstillstand, nicht durch ischämische Herzkrankheit oder Vitien verursacht
andere Herz-erkrankungen (Vitien, Kardiitis, Kardiomyopathie)*	asymptomatisch, keine Medikation erforderlich, nur EKG- oder Echo-Veränderungen, keine Herzgeräusche	mit Medikamenten stabilisiert, leichte bis mäßig ausgeprägte Atemnot bei starker Belastung, minimale Ödeme, NYHA I–II	geringe bis minimale Belastbarkeit, Atemnot in Ruhe, Rechts Herzinsuffizienz, akute Endokarditis, zerebrale Beeinträchtigung o. Embolie, Herzinsuffizienz, akuter Infarkt, NYHA III–IV	Koma, Herzstillstand
Bluthochdruck	diagnostizierter Bluthochdruck, asymptomatisch, derzeit ohne Medikamente	Blutdruck kleiner 160/100 unter medikamentöser Therapie	Blutdruck höher als 160/100 trotz medikamentöser Therapie, keine zentralnervösen Erscheinungen, keine hypertensive Krise	hypertensive Krise oder Koma, nicht im Zusammenhang mit anderen Erkrankungen
<b>Zerebrale und periphere Gefäßerkrankungen</b>				
zerebral vaskuläre Ereignisse (ischämisch und hämorrhagisch)	Anamnese mit TIA ohne bleibende Ausfälle	Anamnese mit Schlaganfall ohne bleibende Ausfälle, oder mit geringer Parästhesie/Ataxie, mehr als 2 TIAs, bekanntes Aneurysma oder teilweiser Verschluss eines Hirn-/Karotisgefäßes ohne Symptome	Anamnese mit Schlaganfall mit Hemiplegie, Paraplegie oder Tetraplegie, akute subarachnoidale Blutung, häufige TIAs	Koma in Zusammenhang mit Schlaganfall
periphere Gefäßerkrankungen (venös und arteriell)	Anamnese einer Thrombophlebitis ohne bleibenden Effekt, peripherer Gefäßbypass ohne Symptome, Ödeme ohne Verschluss	Claudicatio intermittens infolge peripherer Gefäßerkrankung	Lungenembolie in der Vorgeschichte, ausgeprägte Ödeme infolge venöser Abflussstörung, ischämischer Ulkus oder Gangrän, Ruheschmerz durch periphere Gefäßerkrankung	Gangrän mit Septik, Schock

## Anhang 2: Fortsetzung

Erkrankung	Schweregrad			
	1	2	3	4
<b>Weitere häufige Erkrankungen</b>				
Diabetes mellitus	laborchemisch nachweisbarer aber nicht behandelter Diabetes	medikamentös oder diätetisch kontrollierter Diabetes (BZ < 300)	nicht kontrollierter Diabetes (BZ > 300) oder Diabetes mit Komplikation (Neuropathie, Nephropathie, Retinopathie, Gangrän etc.)	diabetisches Koma, Schock, terminales Nierenversagen, schwere KHK
chron. obstrukt. Atemwegs-erkrankungen/ Bronchitis/ Asthma/ Emphysem	chronischer Husten, derzeit ohne Medikamente, körperliche Untersuchung und Röntgen normal	morgendlicher Husten mit Auswurf, leichte Atemnot bei schwerer Belastung, FEV <sub>1</sub> : 60–80%	Atemnot in Ruhe, FEV <sub>1</sub> < 60%, wiederholte Infektionen der Atemwege vor der Krankenhausaufnahme	beatmungspflichtig
Arthrose/Arthritis	leichte bis mäßig ausgeprägte Arthritis	Hüft- oder Knieendoprothese innerhalb der letzten 3 Monate, leichte pulmonale und kardiale Beteiligung einer rheumatoïden Erkrankung	ausgeprägte Deformitäten, chronische Schmerzen, ausgepr. Einschränkung der Beweglichkeit, ausgepr. pulmonale und kardiale Beteiligung	akute, ausgeprägte Vaskulitis mit Fieber, Nierenversagen, Angiitis der Hirngefäße, veränderter Bewusstseinszustand
Magen-Darm-Erkrankungen	Anamnese eines Gastroduodenalulkus innerhalb des vergangenen Jahres, leicht symptomatische Gastritis oder Divertikulitis, wiederkehrendes Colon irritabile	aktives Ulkus derzeit medikamentös behandelt, behandelte Divertikulitis, Reflux Ösophagitis, Polypenentfernung innerhalb des letzten Monats, leicht ausgeprägte Colitis ulcerosa	alle GI-Krankheiten, die zu Perforation, Blutung, Obstruktion, Peritonitis oder Fistelbildung führen, z. B.: Ulkus, Divertikulitis, Appendizitis, Colitis ulcerosa, Hiatushernie mit Anämie, Striktur oder Aspirationspneumonie	Bakteriämie, Septik, Schock
Leber- und Gallengangserkrankungen	Hepatitis vor mehr als 12 Monaten, leichte, asymptomatische Zirrhose	Obstruktion der Gallengänge, Hepatitis innerhalb der letzten 12 Monate, unkompliz. akute Hepatitis (viral, toxisch, medikament.)	chronisch persistierende Hepatitis, chronisch aktive Hepatitis, Portaler Hochdruck, Lebervenen-thrombose	Leberversagen (Bewusstseinsstörungen, schneller Anstieg der Leberenzyme, Aszites, Koma)
Nierenerkrankungen	vergangene unkomplizierte Erkrankungen: akuter unkompl. Harnwegsinfekt, unkomplizierte Nephritis vor weniger als 3 Monaten, Harnsteinleiden vor weniger als 6 Monaten	akute unkomplizierte Erkrankungen: Akute Nephritis, Harnsteinleiden, leicht ausgeprägte Nierenarterienstenose, chronische Harnwegsinfekte	akute komplizierte Erkrankungen: (Serumnitrat > 40, Crea > 3), obstruktive Uropathie, Nierenversagen, Enzephalopathie, mittlere bis ausgepräg. Nierenarterienstenose, Z.n. Nierentransplantation	terminales Nierenversagen, Koma
bösartige Tumoren (ohne Basaliom der Haut)	Anamnese eines Tumors, letzte Behandlung liegt mehr als 5 Jahre zurück	Anamnese eines Tumors, letzte Behandlung liegt länger als 1 Jahr, aber kürzer als 5 Jahre zurück	akute Tumorerkrankung oder letzte Tumorthherapie innerhalb der letzten 12 Monate	terminale Tumorerkrankung
HIV/AIDS	asymptomatische HIV Infektion	akute unkomplizierte Erkrankungen: Lymphknotenschwellung, Fieber, Dermatitis, Diarrhö, Husten, Pneumonie, Gewichtsverlust	akute komplizierte Erkrankungen: PCP <sup>2</sup> , CMV <sup>3</sup> Enteritis, Fieber >1 Monat, Diarrhö >1 Monat, Enzephalitis, Virusinfektionen, Kachexie	terminales Stadium

<sup>1</sup> Ist das zugrunde liegende Leiden eine ischämische Herzerkrankung, dann bitte die KHK-Kriterien verwenden.

<sup>2</sup> PCP: Pneumocystitis-carinii-Pneumonie.

<sup>3</sup> CMV: Cytomegalie-Virus.

## Literatur

- 1 Unnebrink K, Pritsch M. Grundprinzipien klinischer Therapiestudien – was, wie und warum? *Med Klin* 1999; 94: 458–464
- 2 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312: 1215–1218
- 3 Feinstein AR. Problems of randomized trials. In: Abel U, Koch A. *Nonrandomized comparative clinical studies*. Düsseldorf: Symposion Publishing 1998
- 4 Schneeweiss S, Sangha O. Weiterentwicklung von Krankenhausbetriebsvergleichen: Wie wichtig ist Risiko-Adjustierung für den Krankenhausvergleich. In: Sieben G, Litsch M (Hrsg). *Krankenhausbetriebsvergleiche*. Berlin: Springer 2000: 131–147
- 5 Mantel N, Haenszel N. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; 22: 719–748
- 6 Koch A, Windeler J. Statistische Modelle und die Auswertung von Beobachtungsstudien. *Med Klin* 1999; 94: 587–593
- 7 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis* 1987; 40: 373–383
- 8 Greenfield S, Blanco DM, Elashoff RM, Ganz PA. Patterns of care related to age of breast cancer patients. *JAMA* 1987; 257: 2766–2770
- 9 Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000; 29: 891–898
- 10 Iezzoni LI. *Risk adjustment for measuring healthcare outcomes*, 2nd ed. Chicago: Health Administration Press 1997
- 11 Schneeweiss S, Seeger J, Maclure M, Wang P, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol* 2001: im Druck
- 12 Kaku K, Gilbert FI, Sachs R. Comparison of health appraisals by nurses and physicians. *PublicHealth Reports* 1970; 85: 1042–1046
- 13 Schneeweiss S, Sangha O, Manstetten A, Schlottmann N, Liebetrau M, Hartman W, Walter B, Schultz W, Paeger A, Euler P, Eichenlaub A, Leber WD, Rauh G. Identifikation von medizinischen Indikatoren für Ergebnisqualität in der internistischen Krankenhausversorgung: Ergebnisse der QMK Pilotstudie. *Gesundheitsökonomie und Qualitätsmanagement* 2000; 5: 173–182
- 14 Greenfield S, Apolone G, McNeil BJ, Cleary PD. The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement. *Comorbidity and outcomes after hip replacement*. *Medical Care* 1993; 31: 141–154
- 15 Bennett CL, Greenfield S, Aronow H, Ganz P, Vogelzang NJ, Elashoff RM. Patterns of care related to age of men with prostate cancer. *Cancer* 1991; 67: 2633–2640
- 16 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure* 1960; 20: 37–46
- 17 Ciccetti DV, Allison T. A new procedure for assessing the reliability of scoring EEG sleep recordings. *Am J EEG Technol* 1971; 11: 101–109
- 18 Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996; 7: 199–202
- 19 Daniel WW. *Biostatistics: A foundation for analysis in health sciences*, 5th ed. New York: Wiley 1991: 592–595
- 20 SAS/STAT Software. *Changes and Enhancements through Release 6.12*. Cary, NC: SAS Institute Inc. 1996
- 21 Syklo M, Nieto FJ. *Epidemiology beyond the basics*. Gaithersburg, MD, Aspen: 2000: 373
- 22 Feinstein AR, Cichetti DV. High agreement but low kappa: 1. The problem of two paradoxes. *J Clin Epidemiol* 1990; 43: 543
- 23 Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; 112: 564–569
- 24 Kelsey JL, Whittemore AS, Evans AS, Thompson WD. *Methods in observational epidemiology*. New York: Oxford University Press 1996: 354
- 25 Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. *Int J Epidemiol* 1983; 12: 93–97
- 26 Rosner B, Spiegelman DL, Willet WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990; 116: 114–122
- 27 Spiegelman DL, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an „alloyed gold standard“. *Am J Epidemiol* 1997; 145: 184–196
- 28 Elmore JG, Feinstein AR. A bibliography of publications on observer variability (Final installment). *J Clin Epidemiol* 1992; 45: 567–580
- 29 Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care* 1996; 34: 73–84

Dr. med. S. Schneeweiss, Sc. D.

Department of Medicine  
Div. of Pharmacoepidemiology and Pharmacoeconomics  
Brigham and Women's Hospital and Harvard Medical School  
221 Longwood Ave (BLI-341)  
Boston, MA 02115  
E-mail: schneeweiss@post.harvard.edu