

Erfassung der Notwendigkeit von stationären Behandlungen in der Inneren Medizin

Reproduzierbarkeit eines für Deutschland adaptierten Verfahrens

S. Schneeweiss^{1,2}, O. Sangha^{1,3}, H. Siebert⁴, H. Hebeisen⁴, M. Hübner⁵, R. Heinrich⁶, M. Wildner¹, J. Witte⁷

¹ Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie der Ludwig-Maximilians-Universität, München

² Department of Medicine, Harvard Medical School, und Department of Epidemiology, Harvard School of Public Health, Boston

³ Department of Health Policy & Management, Harvard University School of Public Health, Boston

⁴ Medizinischer Dienst der Krankenversicherung in Bayern, Augsburg

⁵ Medizinischer Dienst der Krankenversicherung in Hessen, Oberursel

⁶ Abteilung Innere Medizin 5, Krankenhaus Neuperlach, München

⁷ Klinik für Allgemein- und Viszeralchirurgie, Zentralklinikum Augsburg, Augsburg

Grundproblematik und Fragestellung: Die Prüfung der Angemessenheit von Krankenhausaufnahmen und -behandlungen, auch Fehlbelegungsprüfung genannt, hat in den vergangenen Jahren in Deutschland zunehmend an Aktualität gewonnen. Prüfungen durch den Medizinischen Dienst der Spitzenverbände (MDS) und die Medizinischen Dienste der Krankenkassen (MDK) in den einzelnen Ländern zeigten deutlich den Bedarf an einem standardisierten, validen und reproduzierbaren Instrument zur Beurteilung der Angemessenheit von Krankenhausaufenthalten. Ziel der vorliegenden Studie war es, eine deutschsprachige Adaptierung eines solchen Instruments, basierend auf dem »Appropriateness Evaluation Protocol« (AEP), bezüglich dessen Reproduzierbarkeit zu testen.

Patienten und Methodik: Aus allen 2317 Aufnahmen des Jahres 1997 eines hessischen Lehrkrankenhauses auf die Abteilung für Innere Medizin wurden 52 Patienten für die Untersuchung der Inter-rater-Reliabilität (54% Frauen, Alter: 66 Jahre \pm 18) und 49 Patienten für die Intra-rater-Reliabilität (53% Frauen, Alter: 61 Jahre \pm 20) zufällig ausgewählt. Es wurden Gesamtübereinstimmungen, spezifische Übereinstimmungen und Kappa-Statistiken für Krankenhausaufnahmen und einzelne Krankenhaustage mit den jeweiligen 95%-Konfidenzintervallen berechnet.

Ergebnisse: Das adaptierte AEP zeigte eine Inter-rater-Übereinstimmung von 92% (88–96%) der Aufnahmen und 76% (73–80%) aller Behandlungstage und eine Intra-rater-Übereinstimmung von 96% (88–100%) der Aufnahmen und 93% (91–95%) der Behandlungstage. Diese hohe Übereinstimmung ist unabhängig von der Länge der Liegezeit und unabhängig vom Anteil angemessener oder nicht angemessener Krankenhaustage.

Folgerung: Ein Standardinstrument mit bekannten metrischen Eigenschaften sollte dem internen Qualitätsmanagement von Krankenhäusern dienen, um sich auf die Herausforderungen eines sich konsolidierenden Marktes vorzubereiten. Das AEP kann dieses Standardinstrument sein.

Nachdem ein Großteil der Gesundheitsausgaben im Krankenhausbereich entstehen, werden zunehmend Anstrengungen unternommen, Wirtschaftlichkeitsreserven in diesem Bereich zu mobilisieren. Dabei haben die Prüfung der Angemessenheit von Krankenhausaufnahmen und -behandlungen, auch Fehlbelegungsprüfung genannt, in den vergangenen Jahren

Evaluation of the appropriateness of hospital care in internal medicine: Reliability of a German adaptation of the »Appropriateness Evaluation Protocol«.

Background and objectives: The evaluation of the appropriateness of hospital admissions and hospital stays has become an increasingly important issue in Germany. Evaluations by the German Physicians Review Organizations (MDS, MDK) demonstrated the need for a standardized, valid and reliable assessment tool for inappropriate hospital use. Objective of this study was to test the reliability of a German adaptation of the »Appropriateness Evaluation Protocol« (AEP).

Patients and Methods: From the 2317 admissions to medical wards of a teaching hospital in Hesse in 1997 we randomly selected 52 patients to test the inter-rater-reliability (54% female, mean age=66 years \pm 18). Another 49 patients were randomly selected to test the intra-rater-reliability of the AEP (53% female, mean age=61 years \pm 20). We estimated general agreement, specific agreement and Kappa statistics for the agreement of the evaluation of hospital admissions and hospital days. 95% confidence intervals were reported.

Results: The German adaptation of the AEP showed an inter-rater-agreement of 92% (88–96%) for hospital admissions and 76% (73–80%) for hospital days. Correspondingly, we observed an intra-rater-agreement of 96% (88–100%) for hospital admissions and 93% (91–95%) for hospital days. The high agreement was independent of the length of hospital stay and independent of the proportion of inappropriate hospital days.

Conclusion: A standard instrument for the assessment of the appropriateness of hospital care with known metric properties may be useful for quality management in hospitals and provide competitive advantages in a consolidating health care market. The AEP could be such an instrument.

in Deutschland zunehmend an Aktualität gewonnen. Die Verringerung unnötiger Krankenhausaufenthalte kann nicht nur Kosten reduzieren, sondern gleichzeitig die Qualität der Versorgung verbessern (5). Die Diskussion in Deutschland erlebte einen ersten Gipfel, nachdem der Medizinische Dienst der Spitzenverbände (MDS) im Frühjahr 1997 ein Gutachten veröffentlichte, welches die Prüfung von 63665 Patienten aus 11 Bundesländern zusammenfasste. Das Vorhaben hatte nach § 275a SGB V Modellcharakter und basierte auf einer freiwilli-

gen Teilnahme von je einem Krankenhaus jeder Versorgungsstufe in allen Bundesländern. Obschon das seinerzeit verwendete Verfahren erhebliche Probleme aufwies, haben die regionalen medizinischen Dienste der Krankenversicherungen (MDK) ihre Prüfung von Krankenhäusern fortgesetzt. Die vorrangigen Probleme bei den aktuellen Verfahren liegen in deren zweifelhafter und ungeprüfter Reproduzierbarkeit (Reliabilität) und Gültigkeit (Validität). Entsprechend lassen Ergebnisse dieser Prüfungen erheblichen Spielraum für alternative Erklärungen zu.

In einigen Ländern wurde über zum Teil unangemessenen Krankenhausaufenthalte von erheblichem Umfang berichtet (1, 2, 3, 13, 14, 18, 22, 28, 34), aber nur wenigen Häusern ist es gelungen, diesen Anteil auf der Basis solcher Studien und Feedback-Mechanismen tatsächlich zu reduzieren (18, 20, 33).

Ziel der vorliegenden Studie war es, die Reproduzierbarkeit der adaptierten deutschen Version des »Appropriateness Evaluation Protocol« zur Beurteilung der Angemessenheit von Krankenhausaufenthalten im Verlauf stationärer internistischer Krankenhausaufenthalte zu untersuchen.

Patienten und Methodik

Appropriateness Evaluation Protocol (AEP)

Die Methodik, Angemessenheit von Krankenhausaufenthalten zu beurteilen, basiert auf dem »Appropriateness Evaluation Protocol«. Das AEP, ursprünglich von Gertman und Restuccia entwickelt (13), wurde später modifiziert, um sowohl Krankenhausaufnahmen, als auch einzelne Krankenhaustage bezüglich ihrer Angemessenheit zu beurteilen (21). Das Instrument wurde weltweit in einer großen Zahl von Krankenhäusern mit einer positiven Bewertung seines Nutzwertes angewandt (2, 16, 18, 19, 21, 22). Methodische Studien zum AEP konzentrierten sich auf die Beurteilung der Reproduzierbarkeit (Reliabilität) und Gültigkeit (Validität) (13, 15, 16, 22, 30, 31) sowie die Anpassung des ursprünglichen Instruments an Krankenhäuser mit unterschiedlichen Versorgungsaufträgen, Abteilungen oder Diagnosegruppen (8, 12, 17).

Das AEP besteht aus einer Liste diagnoseunabhängiger Kriterien der medizinischen Angemessenheit, um einzelne Tage eines Krankenhausaufenthalts zu beurteilen. Die Kriterien betreffen medizinische Leistungen und Prozeduren, die pflegerische Versorgung und den Gesundheitszustand des Patienten, welche eine enge stationäre Überwachung notwendig machen. Wird ein Tag als unangemessen bewertet, d.h. können in den Krankenhausakten keine Hinweise gefunden werden, die mindestens einem der Kriterien zuzuordnen sind, dann wird einer Liste von Gründen eine Beschreibung der Faktoren entnommen, die potenziell für die medizinische Unangemessenheit des stationären Aufenthalts verantwortlich sind. Analog wird die medizinische Angemessenheit von Krankenhausaufnahmen bewertet. Das adaptierte Instrument ist publiziert (23, 24).

Das ursprüngliche Instrument wurde von einem Arzt mit englischer Muttersprache und von einem Arzt mit deutscher Muttersprache vom Englischen ins Deutsche übersetzt. Alle Kriterien, deren Erläuterungen (Manual) und die Liste von Gründen wurden in mehrfachen Sitzungen eines Expertengremiums bestehend aus erfahrenen Medizinern, Vertretern von Fach- und Berufsverbänden, Prüfern und Methodikern

detailliert diskutiert und gegebenenfalls geändert. Änderungen an der englischen Originalversion wurden in Betracht gezogen, wenn es Unterschiede in der deutschen Versorgungspraxis gab, wenn einzelne Leistungen in Deutschland nicht erbracht werden oder wenn neuere Entwicklungen der medizinischen Versorgung in dem ursprünglichen AEP noch nicht berücksichtigt waren.

Das adaptierte Instrument wurde in eine computergestützte Form gebracht, die es den Prüfern ermöglicht, das gesamte Manual abzufragen, über Pulldown-Menüs und Hilfe-Texte die Auswertung von Patientenakten zu strukturieren und die Datenqualität durch automatisierte Plausibilitätsprüfungen direkt bei Eingabe in die Datenbank zu verbessern.

Studiendesign und Patientenstichprobe

Die Reliabilitätsstudie wurde in einem 450-Betten-Lehrkrankenhaus in Frankfurt durchgeführt. Alle Patienten (n=2317), die im Jahr 1997 auf die Abteilung für Innere Medizin aufgenommen worden waren, wurden für die retrospektive Analyse mit dem AEP in Betracht gezogen. Patienten, die innerhalb dieses Jahres mehrfach aufgenommen worden waren, konnten mehrfach in der Studie berücksichtigt werden. Patienten mit privater Krankenversicherung oder Sozialhilfe (n=193) wurden wegen eines fehlenden Prüfauftrages des MDK von der Untersuchung ausgeschlossen.

Die einzelnen Patienten einer zufällig ausgewählten Stichprobe von 278 Patienten, stratifiziert nach Alter (< oder \geq Median), Liegezeit (< oder \geq Median) und Diagnose (14 häufigsten Diagnosen plus eine Gruppe »andere Diagnosen«) wurden mit der Entlassungsstatistik des Krankenhauses identifiziert. Aus dieser Stichprobe wurden 38 Patienten ausgeschlossen, da ihre Akten nicht auffindbar oder unvollständig waren oder sie nachträglich als Sozialhilfeempfänger kodiert wurden. Von jedem der verbleibenden 240 Patienten wurde der gesamte Krankenhausaufenthalt inklusive Aufnahme und Entlassungstag für die Beurteilung der Notwendigkeit der Krankenhausbehandlung berücksichtigt. Für die Untersuchungen zur Reproduzierbarkeit wurden zufällig 52 Patienten für die Inter-rater-Reliabilität (Übereinstimmung zwischen zwei Beurteilungen desselben Patienten durch zwei unterschiedliche Gutachter) und 49 Patienten für die Intra-rater-Reliabilität (Übereinstimmung zwischen zwei Beurteilungen desselben Patienten durch einen Gutachter zu unterschiedlichen Zeitpunkten) ausgewählt. Kein Patient war in beiden Stichproben vertreten.

Drei Fachärzte für Innere Medizin des MDK Hessen führten die erste Untersuchung durch. Die Fachärzte wurden explizit angeleitet, ihre Beurteilung über die Notwendigkeit einer Krankenhausaufnahme ausschließlich anhand des Aktenmaterials des Aufnahmetages und bis zu 24 Stunden nach Aufnahme abzugeben. Die Beurteilung eines einzelnen Krankenhaustages wurde ausschließlich auf den jeweiligen Tag basierte und nicht auf den gesamten Verlauf. 3 Wochen später führten zwei der Internisten (C+D) eine erneute Untersuchung der Stichprobe von 49 Patienten durch, um die Intra-rater-Reliabilität zu bestimmen. Keiner der Prüfer wusste während der ersten Prüfung, dass eine Untergruppe der Patienten ein zweites Mal untersucht werden wird. Um die Inter-rater-Reliabilität zu bestimmen, führten zwei Fachärzte für Innere Medizin des MDK Bayern eine Prüfung an der Stichprobe von 52 Patienten durch.

	Gutachter 1 (bzw. Zeitpunkt 1)	
	angemessen	nicht angemessen
Gutachter 2 (bzw. Zeitpunkt 2)	angemessen	a
	nicht angemessen	c
Gesamtübereinstimmung	$= (a+d)/(a+b+c+d)$	
spezifische Übereinstimmung (angemessener Aufenthalt)	$= a/(a+b+c+d)$	
spezifische Übereinstimmung (nicht angemessener Aufenthalt)	$= d/(a+b+c+d)$	

Tab. 1 Inter-rater- (bzw. Intra-rater-)Übereinstimmung bei der Beurteilung der Angemessenheit von Krankenhausaufenthalten.

Analyse

Die Reproduzierbarkeit zwischen Untersuchern (Inter-rater) und innerhalb eines Gutachters (Intra-rater) wurde mit drei Methoden berechnet (**Tab. 1**):

1. Gesamtübereinstimmung, indem die Anzahl der Patienten mit übereinstimmender Beurteilung beider Ärzte (a+d) geteilt wurde durch die Anzahl aller Patienten an einem spezifischen Tag bzw. bei Aufnahme (a+b+c+d).
2. Spezifische Übereinstimmung analog zur Gesamtübereinstimmung, jedoch getrennt für die Übereinstimmung bei angemessenen (a) und nicht angemessenen Tagen (d).
3. Übereinstimmung von Untersucherpaaren mittels Cohen's Kappa-Statistik, die für den Anteil an zufälliger Übereinstimmung adjustiert (7). Wir weisen darauf hin, dass Kappa widersprüchlich niedrig ausfallen kann, wenn die Prävalenz angemessener bzw. nicht angemessener Tage sehr gering ist (10).

Im Unterschied zu bisherigen internationalen Berichten zur Reproduzierbarkeit von Instrumenten zur Prüfung der Angemessenheit von Krankenhausbelegungen (13, 21, 22, 32) berechneten wir die Reliabilität für jeden einzelnen aufeinanderfolgenden Behandlungstag im Laufe einer stationären Behandlung (d.h. -Aufnahme, 1. Tag, 2. Tag, ..., Entlassung). Übereinstimmung wurde nur für die Tage 1 bis 22 (Inter-rater) bzw. 21 (Intra-rater) berechnet, die bei jeweils wenigstens acht Patienten der Stichproben während des stationären Aufenthalts vorkamen. Gesamtübereinstimmung und spezifische Übereinstimmung wurde zusätzlich über die Summe aller Krankenhaustage berechnet. Da die Reproduzierbarkeit an einem Patiententag mit der des nächsten Tages korreliert sein kann, wurden zunächst die Übereinstimmungen für jeden Patienten über die Summe der jeweiligen Krankenhaustage errechnet und diese Übereinstimmungen dann über alle Patienten gemittelt. Aus den Standardabweichungen dieser Mittel können unverzerrte Schätzer der Konfidenzintervalle berechnet werden (9). Gemittelte Kappa-Werte über den gesamten Krankenhausaufenthalt wurden analog zu Fleiss (11) berechnet, nachdem die Gleichheit aller Kappa-Werte zwischen allen Tagen vom Krankenhaustag mittels Cochran's Q Test geprüft wurde (6).

Für alle Schätzer werden 95%-Konfidenzintervalle angegeben. Alle Berechnungen wurden mit SAS-Software durchgeführt (27).

Ergebnisse

Charakteristika der Patientenstichproben

Die Inter-rater-Stichprobe bestand zu 53,8% aus Frauen, im Schnitt waren die Patienten 66 Jahre alt ($\pm 18,0$) und verbrachten im Mittel 14,2 Tage ($\pm 9,2$) im Krankenhaus. Die Intra-rater-Stichprobe bestand zu 52,9% aus Frauen, im Schnitt waren die Patienten 61 Jahre alt ($\pm 19,7$) und verbrachten im Mittel 13,6 Tage ($\pm 10,6$) im Krankenhaus.

Inter-rater-Reliabilität

Für die Inter-rater-Analyse wurden 52 internistische Patienten herangezogen. **Abb. 1** zeigt sowohl die Gesamtübereinstimmung und die spezifische Übereinstimmung als auch die Kappa-Statistiken für Aufnahme und fortlaufende Liegetage bis zum 22. Tag inklusive 95%-Konfidenzintervalle (KI). Die spezifische Übereinstimmung der angemessenen Aufenthalte, nimmt mit zunehmender Verweildauer ab, d.h. immer mehr Versorgungstage werden als nicht notwendig bewertet. Gesamtübereinstimmung zwischen zwei unterschiedlichen Gutachtern war 92% (95% KI: 88–96%) bei der Aufnahme und 76% (73–80%) für die Summe aller fortlaufenden Behandlungstage. Kappa für die Aufnahme betrug 0,31 (0–0,81) und gemittelt über alle Behandlungstage 0,42 (0,34–0,49). Die Kappa-Werte unterschieden sich nicht signifikant von Tag zu Tag ($Q_{K\alpha} 15,2, p=0,82$).

Intra-rater-Reliabilität

Intra-rater-Reliabilität wurde bei 49 Patienten und anhand von zwei Gutachtern untersucht, die 3 Wochen nach der ersten Beurteilung dieselben Patienten nochmals beurteilten. **Tab. 2** zeigt die Übereinstimmung innerhalb jeweils eines Gutachters. Intra-rater-Reproduzierbarkeit war mit 96% Gesamtübereinstimmung für Aufnahmen (88–100%) und 93% (91–95%) für Krankenhaustage sehr hoch. Entsprechend waren die Kappa-Werte 0,65 (0,2–1) und 0,82 (0,77–0,88). Beide Gutachter C und D urteilten mit jeweils gleich hoher Übereinstimmung ($Q_{K\alpha} 10,8, p=0,93$). Gutachter C beurteilte keine einzige Krankenhausaufnahme in seiner Stichprobe als unangemessen, so dass keine Kappa-Statistiken berechnet werden konnten. **Abb. 2** zeigt sowohl die Gesamtübereinstimmung und die spezifische Übereinstimmung als auch die Kappa-Statistiken für Aufnahme und fortlaufende Liegetage bis zum 21. Tag inklusive 95%-Konfidenzintervalle. Analog zur Inter-rater-Reliabilität ist auch hier eine Abnahme der spezifischen

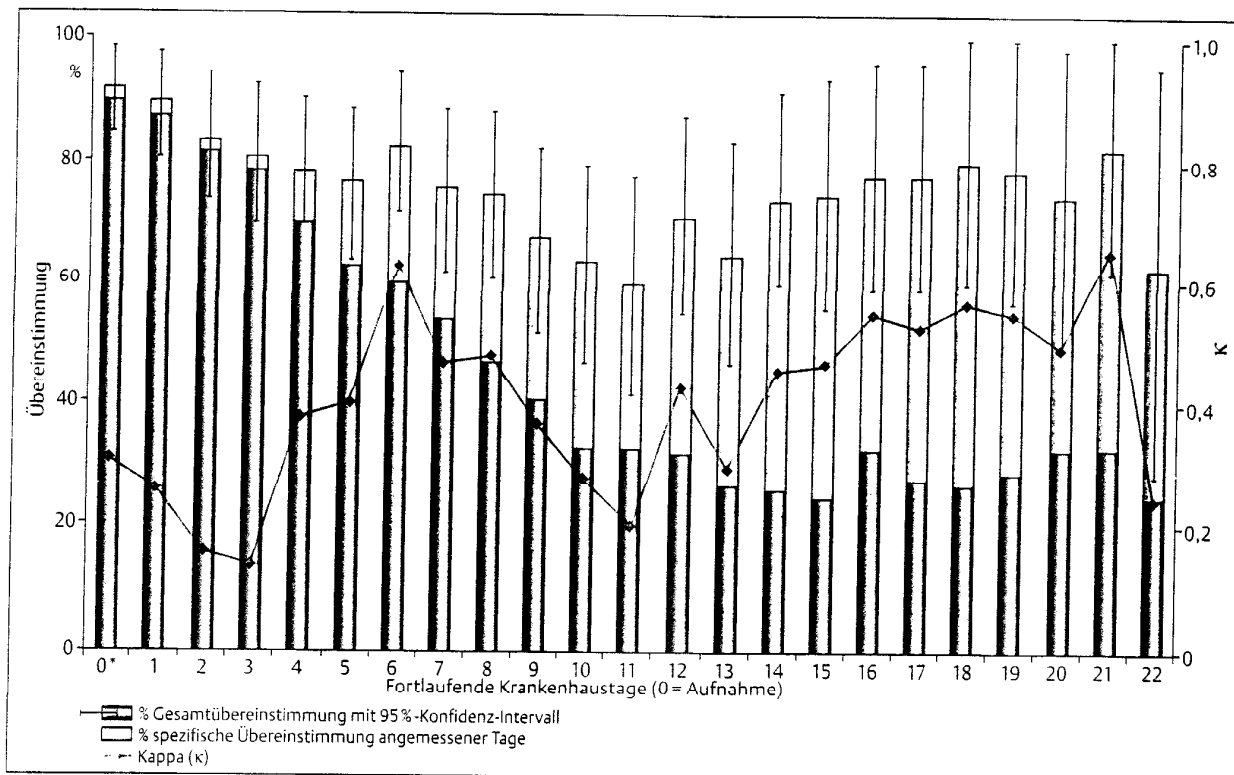


Abb. 1 Inter-rater-Reliabilität bei der Beurteilung des Krankenhausaufenthalts von 52 internistischen Patienten.

Übereinstimmung mit zunehmender Verweildauer zu beobachten.

Diskussion

Das »Appropriateness Evaluation Protocol« (AEP) ist das international am häufigsten eingesetzte Instrument zur Beurteilung der Angemessenheit stationärer Versorgung. Die Vorzüge und Grenzen des Instruments sind in seinem Ursprungsland, den USA, in einer Reihe von Publikationen gut dokumentiert. In Deutschland war bisher kein standardisiertes Instrumentarium verfügbar, dessen Reproduzierbarkeit und Validität getestet wurde und das darauf basierend allgemein akzeptiert wurde.

Die Adaptierung des AEP an die Spezifika des deutschen Sprach- und Versorgungsraums wurde mit großer Sorgfalt durchgeführt. Der Adaptierungsprozess in anderen europäischen Ländern im Rahmen eines von der Europäischen Union geförderten Projekts wurde detailliert studiert und Expertise eingeholt (2, 4, 25). Obwohl, wie in anderen Adaptierungen, einige Kriterien entfallen sind, wurden die Diagnoseunabhängigkeit und die expliziten Definitionen beibehalten.

Die Reliabilität des deutschen Instruments ist durchgehend hoch. Mit einer Inter-rater-Übereinstimmung von 92% der Aufnahmen und 76% aller Behandlungstage und einer Intra-rater-Übereinstimmung von 96% der Aufnahmen und 93% der Behandlungstage liegen die Werte gut im Bereich der internationalen Ergebnisse, in einigen Fällen besser (15). Das Instrument behält seine hohe Reliabilität unabhängig vom Anteil

Tab. 2 Intra-rater-Reliabilität bei der Beurteilung des Krankenhausaufenthalts von 49 internistischen Patienten durch zwei Gutachter (Prozent Gesamtübereinstimmung, Kappa und 95%-Konfidenzintervalle).

	Gutachter C		n	Gutachter D		Gesamt		n
	% Übereinstimmung	Kappa		% Übereinstimmung	Kappa	% Übereinstimmung	Kappa	
Aufnahme	0,96 (0,88-1)	-*	25	0,96 (0,88-1)	0,79 (0,36-1)	0,96 (0,88-1)	0,65 (0,2-1)	49
Krankenhaus-tage†	0,92 (0,88-0,95)	0,70 (0,58-0,82)	255	0,94 (0,91-0,97)	0,78 (0,69-0,88)	0,93 (0,91-0,95)	0,82 (0,77-0,88)	543

* Bei der ersten Begutachtung durch Gutachter C wurde keine Aufnahme als nicht notwendig eingestuft.

† Tage 1 bis 21 (je einschließlich), ohne Aufnahme

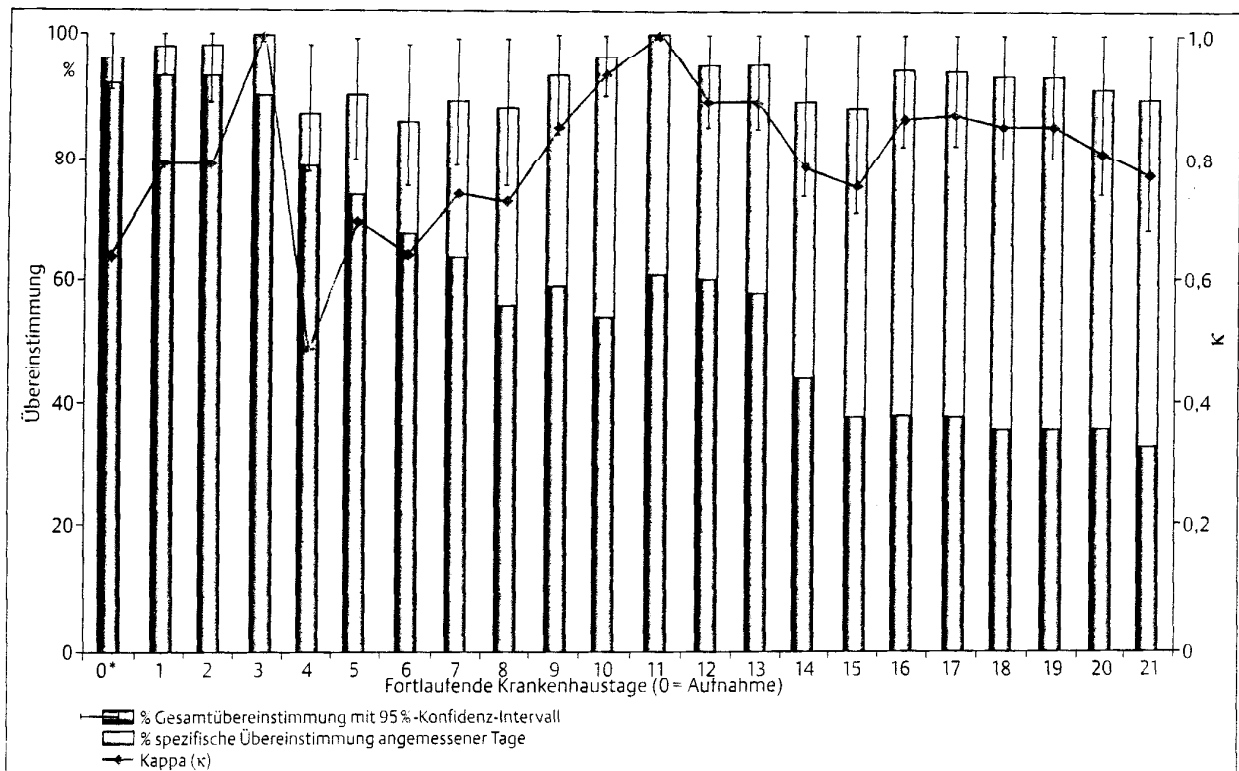


Abb. 2 Intra-rater-Reliabilität bei der Beurteilung des Krankenhausaufenthalts von 49 internistischen Patienten.

angemessener oder nicht angemessener Krankentage und unabhängig von der Liegezeit. Damit entspricht das deutsche AEP den Grundvoraussetzungen für ein standardisiertes Instrument zur Beurteilung der Angemessenheit von Krankenhausaufenthalten (31).

Das von uns vorgeschlagene longitudinale Design untersucht Krankenhausaufenthalte einzelner Patienten. Der wesentliche Vorteil einer longitudinalen Erfassung besteht darin, dass die Gutachter nicht nur einen isolierten Tag im Verlauf einer Krankenhausbehandlung betrachten, sondern sich mit dem gesamten Verlauf vertraut machen können. Ein praktischer Vorteil besteht darin, dass weniger Krankenakten angefordert werden müssen. Die Bewertungen einzelner Patiententage können jedoch untereinander hoch korreliert sein. Wird beispielsweise der Aufenthalt eines Patienten gegen Ende des Aufenthaltes als unangemessen bewertet, so werden oft alle folgenden Tage ebenfalls als unangemessen bewertet. Damit müssen komplexere Methoden zur Analyse von »Fehlbelegungsraten« herangezogen werden und die notwendige Stichprobengröße an der Gesamtzahl an Tagen, nicht Patienten, kann sich leicht erhöhen. Durch eine reine Querschnitterhebung einzelner zufällig ausgewählter Tage – wie dies in den USA häufig erfolgt, zum Beispiel durch Prüfung des jeweils letzten Tages im Quartal – kann diese Komplexität umgangen werden.

Wesentliche Kritik am AEP besteht darin, dass die Angemessenheit lediglich aktenbasiert und rückwirkend beurteilt wird, mit der Möglichkeit, dass dadurch häufiger eine Unangemessenheit attestiert wird (26). In einer eigenen Untersuchung

konnten wir eine sehr hohe Übereinstimmung zwischen zeitlicher und rückwirkender Beurteilung feststellen (28).

Es war nicht Ziel dieses Artikels, über den potenziellen Nutzwert der Fehlbelegungsprüfung wie effizienteres Krankenhausmanagement, mögliche Alternativen der Vergütungspraxis oder eine bessere Integration der vor- und nachsorgenden Institutionen sowie Pflegeeinrichtungen zu diskutieren. Hauptziel unserer Untersuchung war es vielmehr, ein international anerkanntes Instrumentarium, das nun dem deutschen Krankenhausmarkt zur Verfügung steht, bezüglich dessen metrischer Eigenschaften in deutschen Krankenhäusern zu testen.

Die Existenz von einigen zum Teil nicht öffentlich gemachten Ad-hoc-Instrumenten, deren Reproduzierbarkeit und Validität unbekannt ist, verunsichert die Anbieter stationärer Versorgung. Stattdessen sollte ein Standardinstrument mit bekannten metrischen Eigenschaften dem internen Qualitätsmanagement von Krankenhäusern dienen, um sich auf die Herausforderungen eines sich konsolidierenden Marktes proaktiv vorzubereiten. Das adaptierte AEP kann dieses Standardinstrument sein.

Danksagung: Wir bedanken uns für die umfangreiche Unterstützung durch Frau Dr. I. Wisser bei der Adaptierung des AEP.

Literatur

- 1 Alonso J, Munoz A, Anto JM. Using length of stay and inactive days in the hospital to assess appropriateness of utilization in Barcelona, Spain. *J Epidemiol Comm Health* 1996; 50: 196–201
- 2 Apolone G, Alfieri V, Braga A, Caimi V, Cestari C, Crespi V, Crosti PF, De Filippi F, Gelosa M, Lanzi E. A survey of the necessity of the hospitalization day in an Italian teaching hospital. *Quality Assurance in Health Care* 1991; 3: 1–9
- 3 Bare ML, Prat A, Lledo L, Asenjo MA, Salleras L. Appropriateness of admissions and hospitalization days in an acute-care teaching hospital. *Revue d' Epidemiologie et de Sante Publique* 1995; 43: 328–336
- 4 Bentes M, Gonsalves ML, Santos M, Pina E. Design and development of a utilization review program in Portugal. *Int J Quality Health Care* 1995; 7: 201–212
- 5 Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *New Engl J Med* 1991; 324: 370–376
- 6 Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950; 37: 256–266
- 7 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure* 1960; 20: 37
- 8 Davido A, Nicoulet I, Levy A, Lang T. Appropriateness of admission in an emergency department: reliability of assessment and causes of failure. *Quality Assurance in Health Care* 1991; 3: 227–234
- 9 Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Clarendon Press: Oxford, 1994
- 10 Feinstein AR, Cicchetti DV. High agreement but low kappa: 1. The problem of two paradoxes. *J Clin Epidemiol* 1990; 43: 543
- 11 Fleiss JL. *Statistical methods for rates and proportions*. Wiley: New York, 1981
- 12 Gloor JE, Kissoon N, Joubert GL. Appropriateness of hospitalization in a Canadian pediatric hospital. *Pediatrics* 1993; 91: 70–74
- 13 Gertman PM, Restuccia JD. The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. *Med Care* 1981; 19: 855–871
- 14 Hynes M, O'Herlihy BP, Laffoy M, Hayes C. Patients 21 days or more in an acute hospital bed: appropriateness of care. *Irish Journal of Medical Science* 1991; 160: 389–392
- 15 Inglis AL, Coast J, Gray SF, Peters TJ, Frankel SJ. Appropriateness of hospital utilization. *Med Care* 1995; 33: 952–957
- 16 Kaza S, Erdem Z, Dogrusoy S, Halici N. Reliability of a hospital utilization review method in Turkey. *Int J Quality Health Care* 1999; 10: 53–58
- 17 Kemper KJ. Medically inappropriate hospital use in a pediatric population. *New Engl J Med* 1988; 318: 1033–1037
- 18 Payne SM, Ash A, Restuccia JD. The role of feedback in reducing medically unnecessary hospital use. *Med Care* 1991; 29: AS91–106
- 19 Payne SM, Campbell D, Penzias BG, Sochowitzky E. New methods for evaluating utilization management programs. *Qual Rev Bull* 1992; 18: 340–347
- 20 Restuccia JD. The effect of concurrent feedback in reducing inappropriate hospital utilization. *Med Care* 1982; 20: 46–62
- 21 Restuccia JD, Kreger BE, Payne SM, Gertman PM, Dayno SJ, Lenhart GM. Factors affecting appropriateness of hospital use in Massachusetts. *Health Care Financing Review* 1986; 8: 47–54
- 22 Rishpon S, Lubacsh S, Epstein LM. Reliability of a method of determining the necessity for hospitalization days in Israel. *Med Care* 1986; 24: 279–282
- 23 Sangha O, Wildner M, Schneeweiss S, Siebert H, Hebeisen H, Hübner M, Bauer H, Witte J. Fehlbelegung im Krankenhaus – Entwicklung eines standardisierten Verfahrens zur Beurteilung der Notwendigkeit von vollstationären Krankenhausbehandlungen. *Chirurg* 1999; 38: 201–210
- 24 Sangha O, Wildner M, Schneeweiss S, Siebert H, Hebeisen H, Hübner M, Bauer H, Witte J. Das deutsche »Appropriateness Evaluation Protocol«. URL: <http://www.bfv-online.de/BFV/AEP>,
- 25 Santos-Eggimann B, Paccaud F, Blanc T. Medical appropriateness of hospital utilization: an overview of the Swiss experience. *Int J Quality Health Care* 1995; 7: 227–232
- 26 Santos-Eggimann B, Sidler M, Schopfer D, Blanc T. Comparing results of concurrent and retrospective designs in a hospital utilization review. *Int J Quality Health Care* 1997; 2: 115–120
- 27 SAS/STAT Software. Changes and Enhancements through Release 6.12. CSAS Institute Inc.: Cary, NC, 1996
- 28 Schneeweiss S, Sangha O, Wildner M, Siebert H, Hübner M, Witte J. Übereinstimmung einer zeitgleichen Bewertung von Fehlbelegung mit deren retrospektiver Beurteilung anhand von Patientenakten. *Das Gesundheitswesen* 2000; 62: 207–210
- 29 Siu AL, Sonnenberg FA, Manning WG, Goldberg GA, Bloomfield ES, Newhouse JP, Brook RH. Inappropriate use of hospitals in a randomized trial of health insurance plans. *New Engl J Med* 1986; 315: 1259–1266
- 30 Siu AL, Leibowitz A, Brook RH, Goldman NS, Lurie N, Newhouse JP. Use of the hospital in a randomized trial of prepaid care. *J Amer Med Ass* 1988; 259: 1343–1346
- 31 Streiner DL, Norman GR. *Health Measurement Scales*. Oxford University Press: Oxford, 1994
- 32 Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medreview Instrument, and Intensity-Severity-Discharge criteria. *Med Care* 1990; 28: 95–111
- 33 Vardi A, Modan B, Blumstein Z, Lusky A, Schiff F, Barzilay Y. A controlled intervention in reduction of redundant hospital days. *Int J Epidemiol* 1996; 25: 604–608
- 34 Winickoff RN, Restuccia JD, Finche BJ. Concurrent application of the Appropriateness Evaluation Protocol to acute admissions in Department of Veteran Affairs Medical Centers. *Med Care* 1991; 29: AS64–75

Korrespondenz

Dr. med. Oliver Sangha
 Institut für Medizinische Informationsverarbeitung
 Biometrie und Epidemiologie
 der Ludwig-Maximilians-Universität München
 Marchioninistraße 15
 81377 München
 Tel.: 089/69349-100
 Fax: 089/69349-104
 E-Mail: san@ibe.med.uni-muenchen.de