

» Übereinstimmung einer zeitgleichen Bewertung von Fehlbelegung mit deren retrospektiver Beurteilung anhand von Patientenakten

Zusammenfassung: Hintergrund und Ziele: Mit der Einführung einer deutschen Adaptierung des „Appropriate Evaluation Protocols“ zur Beurteilung von Fehlbelegung in Krankenhäusern waren Ärzte und Klinikmanager besorgt, dass eine rückwirkende (retrospektive) Beurteilung auf der Basis von Krankenakten das Ausmaß der Fehlbelegung zu hoch einschätzen könnte. Ziel der vorliegenden Untersuchung ist es, die Übereinstimmung zwischen einer zeitgleichen und retrospektiven Bewertung bei einer Fehlbelegungsprüfung zu untersuchen.

Methodik: In einer chirurgischen Abteilung eines akademischen Lehrkrankenhauses wurde die Notwendigkeit von 42 Neuaufnahmen am Aufnahmetag vor Ort bewertet. Ferner wurde die Notwendigkeit von einzelnen Behandlungstagen von 25 vollstationären Patienten in der Klinik geprüft. Diese Ergebnisse wurden mit einer aktenbasierten Bewertung nach drei Monaten verglichen.

Ergebnisse: Die Übereinstimmungen zwischen zeitgleicher und retrospektiver Untersuchung innerhalb derselben Prüfer waren für Aufnahmen 86% (95% Konfidenzintervall 75%–96%) und für einzelne Behandlungstage 96% (88%–100%). Die Kappa-Werte von 0,85 und 0,78 liegen in einem Bereich der als „exzellent“ beschrieben wird. Ein Vergleich zwischen zwei unterschiedlichen Gutachtern – einer prüfte zeitgleich, ein anderer retrospektiv – ergab Übereinstimmungen von 86% (Aufnahmen) und 90% (Behandlungstage). Der Anteil nicht notwendiger Krankenhausbehandlungen ist bei beiden Bewertungsverfahren gleich.

Schlussfolgerung: Diese Ergebnisse widersprechen der Befürchtung, dass eine retrospektive Bewertung von Fehlbelegung nicht möglich ist, da entscheidende klinische Informationen nicht berücksichtigt würden oder Untersuchungsergebnisse vorlägen, die an diesem Behandlungstag noch nicht zur Verfügung standen. Die Wahrscheinlichkeit von Fehleinschätzungen durch eine retrospektive Beurteilung liegt jedoch innerhalb der unvermeidbaren Streubreite des Prüfverfahrens. Die Übertragbarkeit der Ergebnisse ist durch die relativ kleine Stichprobe chirurgischer Krankenhausbehandlungen begrenzt. Weitere Studien sind notwendig, um diese Ergebnisse in einem breiteren Patientenspektrum zu bestätigen.

S. Schneeweiss^{1,2}, O. Sangha^{1,3}, M. Wildner¹, H. Siebert⁴, M. Hübner⁵, J. Witte⁶

¹ Bayerischer Forschungsverbund Public Health und Institut für Med. Informationsverarbeitung, Biometrie und Epidemiologie

² Div. of Pharmacoepidemiology and Pharmacoeconomics, Dept. of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

³ Department of Health Policy and Management, Harvard School of Public Health

⁴ Medizinischer Dienst der Krankenversicherung in Bayern, Augsburg

⁵ Medizinischer Dienst der Krankenversicherung in Hessen, Oberursel

⁶ Klinik für Allgemein- und Viszeralchirurgie, Zentralklinikum Augsburg

Agreement of Concurrent and Retrospective Assessment of the Appropriateness of Hospital Utilization: Background and purpose: Since the availability of a German adaptation of the "appropriateness Evaluation Protocol" (AEP) for hospital utilization review, physicians and health care managers were concerned about a potential overestimation of the extent of inappropriate hospital use in a retrospective assessment. The objective was to assess the agreement of concurrent and retrospective assessment of inappropriate hospital use.

Methods: The appropriateness of 42 admissions to surgical wards in a teaching hospital was tested concurrent to the admission. In 25 hospitalized patients the appropriateness of hospital days was assessed. Results were compared with the retrospective testing of the same patients after three months.

Results: Agreement between concurrent and retrospective assessment within the same reviewer was 86% of admissions (95% CI=75%–96%) and 96% of hospital days (88%–100%). Kappa showed values of 0.85 and 0.78, which can be described as excellent. A comparison of two different reviewers, one testing concurrently, the other retrospectively showed agreements of 86% and 90%. The proportion of inappropriate days was the same in both designs.

Conclusions: The results refute prior concerns that a retrospective assessment of the appropriateness of hospital use is impossible because in a retrospective design important clinical information would remain unconsidered or clinical information not yet available at that time would be used for the evaluation. The possibility of an incorrect assessment due to a retrospective design is within the variability inherent to the instrument. Despite the strong agreement, the small sample of surgical patient limits the generalizability of results.

Key words: Appropriateness of Hospital Utilization – Reproducibility – Concurrent Design – Retrospective Design

Hintergrund und Ziele

Es wird kontrovers diskutiert, inwieweit die rückwirkende (retrospektive) Beurteilung der Notwendigkeit von vollstationären Aufnahmen und Behandlungen valide und reliabel anhand von Krankenunterlagen, oft Monate nach dem Aufenthalt, vorgenommen werden kann. Die vorherrschende Meinung gerade unter Krankenhausärzten und Krankenhausmanagern ist, dass nur die zeitgleiche Beurteilung realistisch

ist. Dies wird damit begründet, dass einerseits in der konkreten klinischen Situation mehr Informationen zur Verfügung stehen, als sich später in der Dokumentation finden lassen und andererseits in der Krankenakte auch Ergebnisse von zeitlich später stattgefundenen Untersuchungen gefunden werden können.

In Deutschland steht seit kurzem eine adaptierte Version des „Appropriateness Evaluation Protocols“ (AEP) zur Verfügung [1, 2]. Das Instrument bewertet die Notwendigkeit von Krankenhausaufnahmen und Krankenhausbehandlungen nach expliziten, nicht diagnosebezogenen Kriterien. Das Verfahren ist für eine retrospektive Bewertung anhand der ärztlichen und pflegerischen Dokumentation entwickelt. Die Reproduzierbarkeit des AEP bei retrospektiver Durchführung ist sowohl in Abteilungen für Innere Medizin als auch für Chirurgie mit mindestens 74%, meist aber über 80% sehr hoch [3, 4]. Die Validität des Instruments wurde ebenfalls untersucht und als hinreichend gegeben angesehen.

Eine der wenigen Gegenüberstellungen von zeitgleichen und retrospektiven Prüfungen erfolgte durch Santos-Eggimann mit einer französischen Version des AEP [5]. Ihre Studie an 155 Patienten mit 1035 Behandlungstagen von sieben operativen und nicht operativen Abteilungen eines Kreiskrankenhauses zeigte eine Übereinstimmung von 88% bzw. 86% der Beurteilung aller Behandlungstage zwischen zeitgleicher und retrospektiver Beurteilung nach einem Jahr. Dabei erhöhte sich der Anteil von nicht notwendigen Behandlungstagen durch eine retrospektive Bewertung von 10% auf 20% bzw. von 21% auf 30%.

Ziel der vorliegenden Studie ist es, festzustellen, inwieweit eine zeitgleiche Beurteilung von Krankenhausaufnahmen oder Behandlungstagen im Hinblick auf Reihbelegung anhand eines deutschen Instruments mit den Ergebnissen einer retrospektiven Untersuchung übereinstimmt.

Methoden

Notwendigkeit von vollstationären Aufnahmen

42 konsekutive Patienten einer chirurgischen Abteilung eines akademischen Lehrkrankenhauses, die innerhalb einer Woche (vierte Woche in 1999) vollstationär aufgenommen wurden, wurden von zwei chirurgischen Fachärzten des MDK Bayern hinsichtlich der Notwendigkeit der Aufnahme beurteilt. Prüfer A bewertete 13 Patienten und Prüfer B 29 Patienten. Die Prüfer hatten dabei eine Beurteilung am Aufnahmetag vorzunehmen, spätestens jedoch am folgenden Morgen anhand der Aufnahmedokumentation nach den Kriterien des deutschen AEP. Beide Prüfer hatten die Möglichkeit, über die aufnehmenden Ärzte oder das Pflegepersonal zusätzliche Informationen zur Aufnahmesituation einzuholen.

Drei Monate später haben dieselben Prüfer diese Patienten ein zweites Mal anhand der Krankengeschichte hinsichtlich der Notwendigkeit der Aufnahme geprüft. Obwohl den Prüfern jeweils die gesamte Krankenakte zur Verfügung stand, wurden diese angewiesen, nur den Aufnahmetag (+24 Stunden) als Informationsquelle für die Beurteilung heranzuziehen.

Unabhängig von dieser Prüfung hat einer der zwei Fachärzte (Prüfer A) die Patienten des zweiten Prüfers (Prüfer B) anhand der Krankenakten retrospektiv bewertet, um die retrospektive Inter-Rater-Übereinstimmung festzustellen, aber auch, um die Übereinstimmung zwischen zeitgleicher und retrospektiver Bewertung von unterschiedlichen Prüfern zu testen.

Notwendigkeit von vollstationären Behandlungstagen

Zur Überprüfung, inwieweit sich eine zeitgleiche von einer retrospektiven Beurteilung von einzelnen Behandlungstagen unterscheidet, wurde ein chirurgischer Facharzt (Prüfer B) aufgefordert, während einer definierten Woche (vierte Woche in 1999) 25 vollstationäre Patienten einer chirurgischen Station hinsichtlich der Notwendigkeit der stationären Behandlung am Prüftag zu beurteilen. Neben diesem Prüftag wurde auch die Notwendigkeit der maximal drei unmittelbar vorangehenden Behandlungstage geprüft. Nur der Prüftag ist jedoch zeitgleich geprüft worden und ist daher primäres Untersuchungsziel. Beide Prüfer konnten weitere Informationen zu den Patienten über das behandelnde ärztliche und pflegerische Personal einholen.

Drei Monate später hat derselbe Prüfer diese Patienten ein zweites Mal anhand der Krankenakten hinsichtlich der Notwendigkeit der zuvor evaluierten maximal vier Behandlungstage geprüft. Obwohl dem Prüfer die gesamte Krankenakte zur Verfügung stand, wurde dieser darauf hingewiesen, lediglich die einzelnen Behandlungstage von jedem Patienten als Informationsquelle für die Beurteilung heranzuziehen.

Unabhängig von dieser Prüfung hat ein zweiter Facharzt für Chirurgie (Prüfer A) dieselben Patienten geprüft, um die retrospektive Inter-Rater-Übereinstimmung festzustellen. Mit diesen Ergebnissen wurde zusätzlich die Übereinstimmung zwischen zeitgleicher und retrospektiver Bewertung von unterschiedlichen Prüfern getestet.

Statistische Verfahren

Die Übereinstimmung zwischen der zeitgleichen und retrospektiven Beurteilung der Notwendigkeit von vollstationären Behandlungen (Aufnahmen und Behandlungstage) wurde mit zwei Methoden berechnet: 1. Gesamtübereinstimmung, indem die Anzahl der Tage mit übereinstimmender Beurteilung beider Zeitpunkte geteilt wurde durch die Anzahl aller geprüften Tage (jeweils getrennt für Aufnahmen und Behandlungstage); 2. Übereinstimmung von Untersuchungspaaren (zeitgleich und retrospektiv) mittels Cohens Kappa-Statistik, die für den Anteil an zufälliger Übereinstimmung adjustiert [6] wurde. Wir weisen darauf hin, dass Kappa-Werte widersprüchlich niedrig ausfallen können, wenn die Prävalenz entweder notwendiger oder nicht notwendiger Behandlungen sehr gering ist [7]. Bei den Behandlungstagen berechneten wir die Übereinstimmung für jeden der maximal vier Tage getrennt. Die entsprechenden 95% Konfidenzintervalle werden berichtet. Alle Berechnungen wurden mit SAS-Software durchgeführt [8].

Ergebnisse

Notwendigkeit von Krankenhausaufnahmen

Die Untersuchung der Aufnahmen umfasste 42 Patienten. Die Stichprobe bestand zu 49% aus Frauen, im Schnitt waren die Patienten 55 Jahre alt (± 20), von 18 Jahren bis 79 Jahren, und verbrachten im Mittel 9,2 Tage ($\pm 6,9$) im Krankenhaus. Die häufigsten Diagnosen waren ICD241 (Struma, n=5), ICD550 (Abdominalhernien, n=4), ICD 560 (Ileus, n=4) und ICD540 (Appendizitis, n=3). 16% der Patienten hatten zwei oder mehr Begleiterkrankungen.

Die Übereinstimmung für Prüfer A zwischen zeitgleicher und retrospektiver Bewertung beträgt 12/13 (92%; 95% Konfidenzintervall: 78–100) mit einem Kappa von 0,85 (0,56–1). Prüfer B stimmte in 24/29 Patienten überein (83%; 69–97) mit einem Kappa von 0,70 (0,48–0,92). Für Prüfer A und B zusammengefasst betrug die Übereinstimmung 36/42 (86%; 75–96) mit einem Kappa von 0,70 (0,48–0,92).

Die Inter-Rater-Übereinstimmung zwischen Prüfer A und Prüfer B bei retrospektiven Bewertungen war 26/28 (93%; 83–100) mit einem Kappa von 0,85 (0,66–1). Eine zeitgleiche Bewertung durch Prüfer B verglichen mit einer retrospektiven Bewertung durch Prüfer A ergab eine Übereinstimmung von 25/29 (86%; 74–99) mit einem Kappa von 0,72 (0,47–0,97). Ergebnisse sind in einer Übersicht in Abb. 1 zusammengefasst.

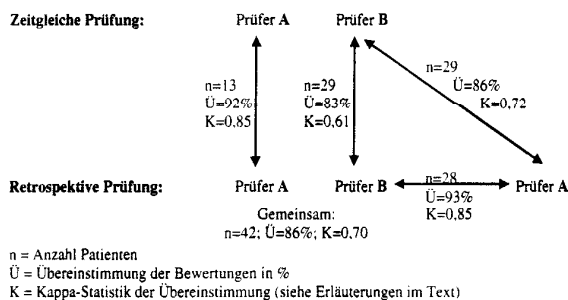


Abb. 1 Überblick über Prüfplan und Übereinstimmungen zwischen zeitgleicher und retrospektiver Prüfung: Die Bewertung der Notwendigkeit von Aufnahmen mit dem AEP an 42 vollstationären chirurgischen Patienten.

Notwendigkeit von Behandlungstagen

Die Untersuchung der Notwendigkeit von Behandlungstagen umfasste 25 Patienten. Die Stichprobe bestand zu 52% aus Frauen, im Schnitt waren die Patienten 66 Jahre alt (± 14), von 30 Jahren bis 77 Jahren, und verbrachten zum Zeitpunkt des Prüftages im Mittel bereits 14,8 Tage ($\pm 12,5$) im Krankenhaus. Die häufigsten Diagnosen waren ICD153 (Neoplasien des Kolons, n=4), ICD154 (Neoplasien des Rektums, n=3), ICD 560 (Ileus, n=3) und ICD574 (Cholelithiasis, n=3). 32% der Patienten hatten zwei oder mehr Begleiterkrankungen.

Die Übereinstimmung zwischen zeitgleicher Bewertung des Prüftages und retrospektiver Bewertung beträgt 23/24 (96%; 88–100) mit einem Kappa von 0,78 (0,36–1). Bei zeitgleicher Untersuchung wurden 3/24 der Prüftage als nicht notwendig

beurteilt (12,5%; 0–26) und bei retrospektiver Bewertung 2/24 (8%; 0–20).

Die Inter-Rater-Übereinstimmung zwischen Prüfer A und Prüfer B bei retrospektiven Bewertungen war 19/20 (95%; 85–100) mit einem Kappa von 0,64 (0–1). Eine zeitgleiche Bewertung des Prüftages durch Prüfer B verglichen mit einer retrospektiven Bewertung durch Prüfer A ergab eine Übereinstimmung von 18/20 (90%; 77–100) mit einem Kappa von 0,44 (0–1). Die Ergebnisse sind in einer Übersicht in Abb. 2 zusammengefasst.

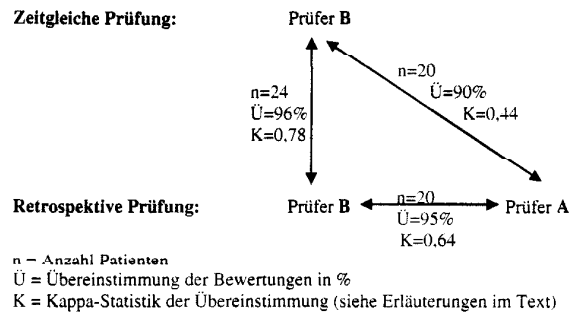


Abb. 2 Überblick über Prüfplan und Übereinstimmungen zwischen zeitgleicher Bewertung des Prüftages und dessen retrospektiver Prüfung: Die Bewertung der Notwendigkeit von einzelnen Behandlungstagen mit dem AEP an 24 vollstationären chirurgischen Patienten.

Tab. 1 zeigt, wie sich die Übereinstimmungen verschieben, wenn sich die zeitgleiche Prüfung nicht auf den Prüftag selber bezieht, sondern auf die bis zu maximal drei vorangegangenen Behandlungstage.

Diskussion

Eine Übereinstimmung von über 86% zwischen einer zeitgleichen Bewertung der Notwendigkeit von Krankenhausaufnahmen und deren retrospektiver Bewertung nach drei Monaten liegt in der gleichen Größenordnung der einzigen weiteren europäischen Studie zu dieser Fragestellung [5]. Übereinstimmungen von mehr als 95% waren bei Betrachtung von einzelnen Behandlungstagen zu beobachten. Die Kappa-Werte von 0,70 bzw. 0,64 können nach Rosner als exzellente Übereinstimmung interpretiert werden [9].

Unsere Ergebnisse weisen darauf hin, dass die Übereinstimmung zeitgleicher Untersuchungen mit retrospektiven Untersuchungen auf einem etwa gleich hohen Niveau liegt wie die Reproduzierbarkeit retrospektiver Untersuchungen untereinander. Zwei Untersuchungen an jeweils 50 Patienten zeigen ähnlich hohe Inter-Rater Reproduzierbarkeiten des AEP bei retrospektiver Anwendung [3, 4].

Bewerten zwei unterschiedliche Prüfer dieselben Patienten retrospektiv anhand von Krankenakten, finden sich etwas größere Übereinstimmungen, als wenn ein Untersucher dieselben Patienten zeitgleich prüft und der andere retrospektiv. Die Unterschiede sind gering und nicht statistisch signifikant, können jedoch ein Hinweis für die nahe liegende Hypothese sein, dass bei methodisch gleicher Durchführung, z.B. beide retrospektiv anhand von Krankenakten, die Übereinstim-

Tab. 1 Übereinstimmung der zeitgleichen Prüfung mit der retrospektiven Bewertung von Behandlungstagen in Abhängigkeit der Zeit zwischen Prüftag und zu prüfendem Behandlungstag

Geprüfter Behandlungstag ist:	Vergleich		Zeitgleich Prüfer A versus retrospektiv Prüfer B		Retrospektiv Prüfer A versus retrospektiv Prüfer B	
	Zeitgleich Prüfer A versus retrospektiv Prüfer A					
Prüftag	23/24	96% (88–100)	19/20	95% (85–100)	18/20	90% (77–100)
Prüftag -1	23/25	92% (81–100)	21/22	95% (87–100)	19/22	86% (72–100)
Prüftag -2	17/22	77% (59–95)	18/19	95% (85–100)	14/19	74% (54–93)
Prüftag -3	19/20	95% (85–100)	17/17	100%	16/17	94% (83–100)
Gesamt	82/91	90% (84–96)	75/78	96% (92–100)	67/78	86% (78–94)

mung zwischen unterschiedlichen Prüfern höher ist als bei unterschiedlichen Verfahren, z.B. der eine Prüfer zeitgleich und der zweite retrospektiv. Ein dazu widersprüchliches Ergebnis liefert jedoch die vergleichende Untersuchung von einzelnen Behandlungstagen, die dem Prüftag vorausgegangen sind. Je weiter der zu prüfende Behandlungstag von dem Prüftag entfernt ist, desto mehr muss sich der Prüfer alleine auf die Akten verlassen. Man sollte daher annehmen, dass mit zunehmender Entfernung vom Prüftag die Übereinstimmung mit der retrospektiven Prüfung nach drei Monaten größer wird, da unter ähnlicheren Bedingungen gearbeitet wurde. Dies kann jedoch nicht beobachtet werden, im Gegenteil, die Übereinstimmung scheint tendenziell zu fallen. Die Fallzahlen sind für aussagekräftige Schlussfolgerungen zu dieser spezifischen Fragestellung jedoch zu gering.

Größte Sorge von Klinikern und Krankenhausmanagern ist, dass durch die retrospektive Bewertung von Fehlbelegung wichtige klinische Informationen verloren gehen können oder Untersuchungsergebnisse in die Bewertung einfließen, die an dem entsprechenden Tag noch nicht verfügbar waren und dadurch der Anteil an Fehlbelegung zu hoch eingeschätzt wird. Die Daten der vorliegenden Studie können diese Annahme nicht unterstützen, zeigen sie doch nur geringe Unterschiede in dem Anteil an Fehlbelegung mit weit überlappenden 95% Konfidenzintervallen. Dabei reduziert sich sogar der Anteil an nicht notwendig eingeschätzten Behandlungstagen bei retrospektiver Betrachtung leicht. Obwohl die Fallzahlen dieser Studie gering sind, reichen sie wegen der sehr hohen Übereinstimmungen aus, um die primäre Fragestellung zu beantworten.

Zusammenfassend zeigen unsere Ergebnisse, dass zeitgleiche Prüfungen durch retrospektive, aktenbasierte Untersuchungen mit sehr hoher Übereinstimmung ersetzt werden können. Die Wahrscheinlichkeit von Fehleinschätzungen durch eine retrospektive Beurteilung ist gleich oder geringer als die allgemeine Streubreite des Verfahrens, die jedem Bewertungsalgorithmus zur Fehlbelegung inhärent ist. Die Übertragbarkeit der Ergebnisse ist durch die relativ kleine Stichprobe chirurgischer Krankenhausbehandlungen begrenzt. Weitere Studien sind notwendig, um diese Ergebnisse in einem breiteren Patientenspektrum zu bestätigen.

Literatur

- Sangha O, Wildner M, Schneeweiss S, Siebert H, Hebeisen H, Hübner M, Mittelstaedt G, Witte J. Fehlbelegung im Krankenhaus – Entwicklung eines standardisierten Verfahrens zur Beurteilung der Notwendigkeit von vollstationären Krankenhausbehandlungen. *Der Chirurg* 1999; 38: 201–210
- Gertman PM, Restuccia JD. The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. *Medical Care* 1981; 19: 855–71
- Schneeweiss S, Sangha O, Siebert H, Hebeisen H, Hübner M, Heinrich J, Wildner M, Witte J. Erfassung der Notwendigkeit von stationären Behandlungen in der Inneren Medizin: Reproduzierbarkeit eines deutschen Verfahrens. *Deutsche Medizinische Wochenschrift*. Manuskript zur Veröffentlichung angenommen
- Schneeweiss S, Sangha O, Siebert H, Hübner M, Fuhrmann J, Wildner M, Witte J. Reproduzierbarkeit eines deutschen Instruments zur Erfassung der Notwendigkeit von stationären Behandlungen in der Chirurgie. Manuskript zur Veröffentlichung eingereicht
- Santos-Eggimann B, Sidler M, Schopfer D, Blanc T. Comparing results of concurrent and retrospective designs in a hospital utilization review. *Int J Quality in Health Care* 1997; 9: 115–120
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure* 1960; 2: 37
- Feinstein AR, Cichetti DV. High agreement but low kappa: 1. The problem of two paradoxes. *Journal of Clinical Epidemiology* 1990; 43: 543
- SAS/STAT Software. Changes and Enhancements through Release 6.12. Cary, NC: SAS Institute Inc. 1996
- Rosner B. *Fundamentals of Biostatistics*. Wadsworth: International Publishing Group (4th ed.), 1995: 426

Dr. med. Sebastian Schneeweiss, SM, SD

Div. of Pharmacoepidemiology and
Pharmacoeconomics, Dept. of Medicine
Brigham and Women's Hospital and Harvard Medical School
221 Longwood Ave.
Boston, MA 02115
USA

E-mail: schneeweiss@post.harvard.edu