

## Update zum deutschen AEP: Metrische Eigenschaften und praktische Erfahrungen

S. Schneeweiss<sup>1,2</sup> und O. Sangha<sup>1,3</sup>

<sup>1</sup> Bayerischer Forschungsverbund Public Health, Ludwig-Maximilians-Universität, München

<sup>2</sup> Department of Medicine, Harvard Medical School, and Department of Epidemiology,

Harvard School of Public Health, Boston

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

### An update on the German version of the AEP: metric properties and practical experience

**Abstract.** The development of a German version of the Appropriateness Evaluation Protocol (AEP) allows for the first time the evaluation of hospital admissions and bed days in Germany. The instrument is based on international experience and has been adopted in cooperation with acknowledged members of German surgical and medical societies. The AEP showed excellent reliability in general internal medicine as well as in surgery. The validity is comparable to international studies, although further research is necessary. Approximately 90% of surgical cases could be evaluated according to the criteria of the AEP; the remaining patients were evaluated using the "override option". The majority of inappropriate care is due to poor documentation in medical records and management deficiencies during inpatient care.

**Keywords:** Appropriateness of hospital care – AEP – Reliability – Validity.

**Zusammenfassung.** Mit der Entwicklung der deutschen Version des Appropriateness Evaluation Protocol (AEP) existiert erstmals ein standardisiertes Instrument zur Erfassung der Notwendigkeit von vollstationären Krankenhausaufnahmen und -behandlungstagen. Das Instru-

ment basiert auf internationalen Erfahrungen und wurde aus dem angelsächsischen unter Mitwirkung anerkannter chirurgischer Fachvertreter an die deutschen Verhältnisse adaptiert. Dieser Bericht fasst die ersten wissenschaftlichen Ergebnisse zu den metrischen Eigenschaften und Erfahrungen mit der praktischen Anwendung des deutschen AEP zusammen. Das AEP zeigte eine sehr gute Reproduzierbarkeit sowohl in der Inneren Medizin, als auch in der Chirurgie. Die Validität des AEP entspricht internationalen Untersuchungen, wenn auch künftige nationale Erfahrungen in Weiterentwicklungen implementiert werden müssen. Das Instrument erlaubt ungefähr 90% der beurteilten Patienten anhand des vorgegebenen Kriterienkatalogs zu bewerten; für die übrigen Patienten ist die Begründung der Notwendigkeit einer vollstationären Behandlung im Klartext („Override Option“) erforderlich. Die meisten als unnötig eingestuften Krankenhaustage beruhen auf einer mangelhaften Dokumentation und organisatorischen Problemen während der stationären Behandlung.

**Schlüsselwörter:** Fehlbelegungsprüfung – AEP – Reliabilität – Validität.

Standardisierte Instrumente zur Fehlbelegungsprüfung sind international weit verbreitet und seit längerer Zeit im Einsatz [1, 3, 4]. Vor etwa einem Jahr wurde in „Der Chir-

urg“ ein deutsches Instrumentarium – das deutsche AEP (Appropriateness Evaluation Protocol) – vorgestellt [8, 9], das sich an internationalen Vorbildern orientiert und die Notwendigkeit von vollstationären Krankenhausaufnahmen und einzelnen Behandlungstagen standardisiert zu bewerten erlaubt. In der Zwischenzeit wurden konzeptionelle Arbeiten zu diesem Thema publiziert und mehrere Studien zu den metrischen Eigenschaften des AEP in der Chirurgie und inneren Medizin in Deutschland durchgeführt. In kurzer Form geben wir einen Überblick über den derzeitigen Stand der metrischen Evaluation des deutschen AEP und berichten über Erfahrungen bei der praktischen Umsetzung. Gleichzeitig stimulierte die Veröffentlichung des Instrumentariums eine lebhaftere inhaltliche Diskussion; die daraus resultierenden Vorschläge werden gegenwärtig gesammelt und in eine Adaptation des Verfahrens eingehen.

### Reproduzierbarkeit

Eine grundlegende Eigenschaft eines Instruments zur Erfassung von Fehlbelegung ist die Reproduzierbarkeit der Ergebnisse durch verschiedene Untersucher (Inter-rater-Reliabilität) oder durch dieselben Untersucher, jedoch zu unterschiedlichen Zeitpunkten (Intra-rater-Reliabilität) [6]. Kann ein Instrument oder Verfahren diese Eigenschaft

nicht erfüllen, kommt dieses einer willkürlichen Bewertung durch Untersucher gleich und eine weitere Verwendung des Instrumentariums erübrigt sich faktisch.

In Stichproben von 52 und 49 internistischen Patienten konnten sehr hohe Inter-rater-Übereinstimmungen von 92% und Intra-rater-Übereinstimmungen von 96% bei der Bewertung der Notwendigkeit von stationären Aufnahmen durch das AEP ermittelt werden. Die Übereinstimmung von einzelnen Behandlungstagen waren 76% bzw. 93% [12].

Ähnlich gute Reproduzierbarkeiten ergaben sich für chirurgische Patienten. Das AEP zeigte in Stichproben von 54 und 51 chirurgischen Patienten eine Inter-rater-Übereinstimmung von 74% der Aufnahmen und 84% aller Behandlungstage und einer Intra-rater-Übereinstimmung von 88% der Aufnahmen und 88% der Behandlungstage. Die beobachteten Übereinstimmungen waren unabhängig von der Länge der Liegezeit und unabhängig vom Anteil angemessener oder nicht angemessener Krankenhaustage [13].

Der AEP ist ein aktenbasiertes Verfahren, d.h. Untersucher beurteilen anhand der Patientenakten, ob eines der 24 diagnoseunabhängigen Kriterien für einen gerechtfertigten stationären Aufenthalt zutrifft. Ärzte und Klinikmanager waren besorgt, dass eine rückwirkende (retrospektive) Beurteilung auf der Basis von Krankenakten das Ausmaß der Fehlbelegung zu hoch einschätzen könnte [10]. Dies wird damit begründet, dass einerseits in der konkreten klinischen Situation mehr Informationen zur Verfügung stehen als dokumentiert werden und andererseits in der Krankenakte auch Ergebnisse von zeitlich später stattgefundenen Untersuchungen zu finden sind, die zum Zeitpunkt der klinischen Entscheidung nicht vorhanden waren [10]. In einer Studie an 67 chirurgischen Patienten wurde eine Übereinstimmung zwischen zeitgleicher und retrospektiver Untersuchung innerhalb derselben Prüfer von 86% für Aufnahmen und 96% für einzelne Behandlungstage mit dem deutschen AEP ermittelt. Ähnlich hohe Werte ergaben sich

bei dem Vergleich zwischen zwei unterschiedlichen Untersuchern, einer prüfte zeitgleich, ein anderer retrospektiv (86% bzw. 90%) [11].

Gemeinsam weisen diese Ergebnisse nach internationalen Standards [5] auf eine exzellente Reproduzierbarkeit des AEP hin, wenn auch die Untersuchungen an größeren Patientenkollektiven repliziert werden sollten, bevor endgültige Schlüsse gezogen werden.

### Validität

Nachdem die Reproduzierbarkeit als wichtige Voraussetzung eines Instruments nachgewiesen wurde, war zu klären, inwieweit das AEP tatsächlich das Konstrukt Fehlbelegung erfasst. In der internationalen Literatur finden sich zahlreiche Arbeiten, die trotz fehlendem Goldstandard Hinweise auf eine gute Validität des AEP geben. Während der aufwendigen Adaptierung der internationalen Originalversion [2] waren sowohl deutsche Experten der Chirurgie, Innere Medizin und Geriatrie involviert, als auch die verschiedenen Fachgesellschaften eingebunden [8]. Damit war beabsichtigt sowohl die sogenannte „Augenscheinvalidität“ und „Inhaltsvalidität“ abzusichern [6]. Ferner wurde geprüft, ob das AEP im praktischen Einsatz zu Aussagen kommt, die nach Expertenmeinung *a priori* erwartet wurde (= Konstruktvalidität) [14]. So wurden beispielsweise in multivariaten Regressionsanalysen an 250 chirurgischen und 226 internistischen Patienten bestätigt, dass die Abwesenheit von Begleiterkrankungen und ein zunehmender zeitlicher Abstand vom Aufnahmetag signifikante Prädiktoren für Fehlbelegung sind. Alter war, nach Adjustierung für Begleiterkrankungen und Liegezeit, nicht mit Fehlbelegung assoziiert [7].

Eine endgültige Bewertung der Validität des AEP nach wissenschaftliche Kriterien erfordert weitere Untersuchungen. Dennoch weisen derzeitige Ergebnisse in die Richtung eines validen Instruments. Sie erlauben jedoch derzeit noch keine Aussage über die korrekte Skalierung, d.h. ob 20% Fehlbelegung nach Auskunft des Instruments tat-

sächlich 20% entspricht oder eher 30% oder 10%. Konfidenzintervalle können die Unsicherheit des Instruments aufgrund der Stichprobenziehung quantifizieren, jedoch nicht die Fehlerbreite aufgrund derzeit nicht auszuschliessender Skalierungsdefizite.

### Handhabung von „Ausnahmen“

Bei dem Entwicklungsteam sind bereits ein Vielzahl von Kasuistiken eingegangen, wo das AEP scheinbar die Situation nicht korrekt bewerten würde und zu einer Überschätzung der tatsächlichen Fehlbelegung führen könnte. Kein Instrument dieser Welt kann allen Besonderheiten und Variationen der klinischen Medizin gerecht werden. Aus diesem Grunde wurde von Untersuchern in etwa 10% der Fälle von der sogenannten „override option“ Gebrauch gemacht, das bedeutet, Untersucher hielten den Behandlungstag für angemessen, obwohl der AEP Kriterienkatalog anzeigte, dass dies nicht der Fall wäre. Wie bereits oben dargestellt, war bei zeitgleicher Beurteilung, die einen direkten Informationsaustausch zwischen Prüfer und den aufnehmenden oder behandelnden Ärzten ermöglicht, die Beurteilung nur unwesentlich abweichend von einer rein aktenbasierten Bewertung. Dies deutet in die Richtung, dass Fachärzte mit langjähriger klinischer Erfahrung in der Lage sind, komplexe klinische Erkrankungsbilder durch Aktenstudium hinreichend gut erkennen können. Dies mag nicht für jeden Einzelfall zutreffen, jedoch für Patientengruppen und deren statistische Bewertung.

### Häufige Gründe für „Fehlbelegung“ und deren Vermeidung

Ein wesentlicher Grund von „Fehlbelegung“ lag in der mangelhaften Dokumentation in den Krankenakten. Dieses ist zum Teil bestürzend, auf der anderen Seite beruhigend, da es sich hierbei um ein modifizierbares Problem handelt. Hier bestehen erhebliche Verbesserungspotentiale, die seitens der Leistungserbrin-

ger zwingend angegangen werden sollten, da nicht nur die Vermeidung von „Fehlbelegungen“, sondern auch die Eingruppierung in künftige prospektive Entgeltsysteme von der Güte der Dokumentation abhängen. Daneben war die Einstufung von fehlbelegten Tagen hauptsächlich auf Managementfehler zurückzuführen, die wiederum auf interne Organisationsprobleme und die umliegende Versorgungsstruktur zurückzuführen waren. Interne Gründe waren überwiegend das Warten auf diagnostische Massnahmen oder auf Operationstermine. Aus ärztlicher Sicht können die Patienten nicht entlassen werden. Dennoch waren an diese Tagen keine diagnostischen/therapeutischen oder vorbereitenden Aktivitäten erfolgt, so dass dieses klassisch Managementproblem der Ablaufoptimierung zur „Fehlbelegung“ geführt hat. Verzögerte Entlassungen können sowohl auf Verzögerungen bei der „Unterbringung“ der Patienten in Rehabilitationseinrichtungen, Pflegeheimen etc. zurückgeführt werden, z.T. wurden laut AEP jedoch auch Patienten, die nach Hause entlassen wurden, unnötig lange in stationärer Behandlung belassen [7].

Zusammenfassend liegt mit dem AEP ein zuverlässiges und – nach derzeitigem Wissensstand – valides Instrument zur Bewertung der Notwendigkeit von Krankenhausaufenthalten vor. Es ist offensichtlich, dass

ein solches Verfahren Kritik induziert. Diese findet in Deutschland gegenwärtig durchaus konstruktiv statt und wird in künftige Adaptation eines solchen Verfahrens Eingang finden.

#### Literatur

1. Apolone G, Alfieri V, Braga A, Caimi V, et al (1991) A survey of the necessity of the hospitalization day in an Italian teaching hospital. *Qual Assur Health Care* 3: 1
2. Gertman PM, Restuccia JD (1981) The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. *Med Care* 19: 855
3. Hynes M, O'Herlihy BP, Laffoy M, Hayes C (1991) Patients 21 days or more in an acute hospital bed: appropriateness of care. *Ir J Med Sci* 160: 389
4. Kemper KJ (1988) Medically inappropriate hospital use in a pediatric population. *N Engl J Med* 318: 1033
5. Rosner B (1995) *Fundamentals of biostatistics*, 4<sup>th</sup> edn. International Publishing Group, Wadsworth, p 426
6. Sangha O, Wildner M (1999) Beispiele methodischer Probleme bei der Fehlbelegungsprüfung im Krankenhaus und Ansätze zu deren Behebung. *Gesundheitswesen* 61: 287
7. Sangha O, Schneeweiss S, Wildner M (1999) Fehlbelegungsprüfung im Krankenhaus. Entwicklung des deutschen Appropriateness Evaluation Protokoll (AEP). Abschlussbericht der LMU München, November 1999
8. Sangha O, Wildner M, Schneeweiss S, Siebert H, et al (1999) Fehlbelegung im Krankenhaus – Entwicklung eines standardisierten Verfahrens zur Beurteilung der Notwendigkeit von vollstationären Krankenhausbehandlungen. *Chirurg* 70: 201
9. Sangha O, Wildner M, Schneeweiss S, Siebert H, et al (2000) Die deutsche Version des Appropriateness Evaluation Protocol (AEP). URL: <http://www.BFV-online.de/BFV/AEP>
10. Santos-Eggimann B, Sidler M, Schopfer D, Blanc T (1997) Comparing results of concurrent and retrospective designs in a hospital utilization review. *Int J Qual Health Care* 2: 115
11. Schneeweiss S, Sangha O, Siebert H, Hübner M, et al (2000) Übereinstimmung einer zeitgleichen Bewertung von Fehlbelegung mit deren retrospektiver Beurteilung anhand von Patientenakten. *Gesundheitswesen* 62: 207
12. Schneeweiss S, Sangha O, Siebert H, Hübner M, et al (2000) Erfassung der Notwendigkeit von stationären Behandlungen in der Inneren Medizin. Reproduzierbarkeit eines für Deutschland entwickelten Verfahrens. *Dtsch Med Wochenschr* 125: 894
13. Schneeweiss S, Sangha O, Siebert H, Hübner M, et al (2000) Reproduzierbarkeit eines deutschen Instruments zur Erfassung der Notwendigkeit von stationären Behandlungen in der Chirurgie. *Soz Präventivmed* 45: 258
14. Streiner DL, Norman GR (1994) *Health measurement scales*. Oxford University Press, Oxford

S. Schneeweiss MD, ScD  
Department of Medicine  
Brigham and Women's Hospital  
and Harvard Medical School  
221 Longwood Ave (BLI341)  
Boston, MA 02115  
USA  
E-Mail: [schneeweiss@post.harvard.edu](mailto:schneeweiss@post.harvard.edu)