

Beispiele methodischer Probleme bei der Fehlbelegungsprüfung im Krankenhaus und Ansätze zu deren Behebung

O. Sangha, M. Wildner

Bayerischer Forschungsverbund Public Health – Öffentliche Gesundheit, Ludwig-Maximilians-Universität München

Zusammenfassung: Spätestens seit der Veröffentlichung der Ergebnisse des Modellvorhabens zur Prüfung der Notwendigkeit der Krankenhausbehandlung nach § 275 a SGB V durch den Medizinischen Dienst der Spitzenverbände der Krankenkassen (MDS) im Frühjahr 1997 ist die Fehlbelegungsdiskussion zu einem Dauerthema geworden. Bundesweit werden seither vermehrt Fehlbelegungsprüfungen durch den Medizinischen Dienst der Krankenkassen (MDK) durchgeführt.

In der sachgerechten Begutachtung von Krankenhausfehlbelegungen ist der Ausgangspunkt eine sorgfältige, umfassende und methodisch gültige Erhebung der tatsächlichen Situation. Inwieweit die in Deutschland eingesetzten Verfahren dem derzeitigen internationalen wissenschaftlichen Stand der Epidemiologie, Biometrie bzw. der Versorgungsforschung entsprechen, war zunächst offen. Nach eingehender wissenschaftlicher Prüfung einzelner typischer Prüfverfahren mußte jedoch festgestellt werden, daß die derzeit verwendeten Prüfverfahren methodisch unzureichend sind und demzufolge keine gültigen Aussagen über das Vorliegen von ungerechtfertigten Krankenhausaufnahmen bzw. Krankenhausbehandlungstagen zulassen. Diese Feststellung erfolgt ungeachtet der Tatsache, daß es in Krankenhäusern möglicherweise doch Fehlbelegung gibt; das verwendete Prüfverfahren erlaubt jedoch keine Aussage hierzu. Ziel dieser Arbeit ist die Vorstellung häufiger Fehler bezogen auf grundsätzliche Anforderungen einer gültigen Fehlbelegungsprüfung anhand von ausgewählten Beispielen. Vorschläge zur Verbesserung der Untersuchungsmethoden und der Darstellung der Ergebnisse werden auf dem Boden der kritischen Untersuchungen gemacht.

Examples and Proposed Solutions for Methodological Problems in Hospital Appropriateness Assessment: Inappropriate hospital care has become a recurring theme since the publication of the results of the model project on the appropriateness of hospital care (§ 275 a SGB V) by the medical service of the National Associations of Health Insurers (MDS) in Spring 1997. Nationwide, an increased number of appropriateness assessments have since been conducted by the medical service of the health insurers (MDK).

The basis for an objective assessment of the appropriateness of hospital care is a diligent, comprehensive, and methodologically valid examination of the actual situation. At first, it was unknown to what extent current methods in Germany conformed to international scientific standards in Epidemio-

logy, Biometry, and Health Services Research. Following a thorough scientific examination of several typical methodologies, it became clear that current methods are insufficient, and hence do not allow valid conclusions regarding inappropriate hospital admissions or hospital days. This statement does not deny that inappropriate hospital use is very likely, but our current assessment tools do not permit further conclusions to be drawn. The purpose of this article is to identify typical mistakes in the fundamental aspects of a valid assessment based on selected examples. Proposals are made regarding methodology and the presentation of results that follow from this critical examination.

Key words: Public Health – Health Services Research – Appropriateness of Hospital Care

Spätestens seit der Veröffentlichung der Ergebnisse des Modellvorhabens zur Prüfung der Notwendigkeit der Krankenhausbehandlung nach § 275 a SGB V durch den Medizinischen Dienst der Spitzenverbände der Krankenkassen (MDS) im Frühjahr 1997 ist die Fehlbelegungsdiskussion zu einem Dauerthema avanciert. Basierend auf den Ergebnissen dieses Vorhabens definiert das gemeinsame Aktionsprogramm des MDS unter dem Titel „Einsparpotential durch Abbau von Fehlbelegung im Krankenhaus“ als Ziel künftiger Prüfverfahren eine Kürzung der Aufwendungen um ca. 20%.

Im vergangenen Jahr ist es daher bundesweit zu vermehrten Fehlbelegungsprüfungen durch den Medizinischen Dienst der Krankenkassen (MDK) gekommen. Die Prüfungen basieren dabei überwiegend auf § 17a KHG, wonach die Krankenkassen durch gezielte Einschaltung des MDK darauf hinwirken, daß Fehlbelegungen vermieden und bestehende Fehlbelegungen zügig abgebaut werden. Paragraph 17a KHG stellt dabei auf die Notwendigkeit stationärer Krankenhausbehandlungen ab, wobei vor- und nachstationäre sowie teilstationäre Behandlungsformen einer stationären Behandlung zuzuordnen sind. Ungeachtet unzähliger Prüfverfahren, die auf § 17a KHG beruhen, besteht weiterhin Rechtsunsicherheit, welchen Umfang eine „gezielte“ Prüfung beinhalten darf. Insbesondere wird kontrovers diskutiert, inwieweit eine flächendeckende oder einer Ausforschung gleichstehende Prüfung, z.B. einer gesamten Abteilung oder eines gesamten Krankenhauses, überhaupt zulässig ist.

Während systematische Fehlbelegungsprüfungen den Leistungsanbietern in Abhängigkeit der Güte des verwendeten

Prüfverfahrens durchaus wichtige Informationen im Sinne qualitätssichernder Maßnahmen liefern können, werden sie zunehmend als Bedrohung verstanden, da die entsprechenden Prüfergebnisse häufig in laufende oder bevorstehende Budgetverhandlungen integriert werden.

In der sachgerechten Begutachtung von Krankenhausfehlbelegungen ist der Ausgangspunkt, wie allgemein in Begutachtungsfragen, eine sorgfältige, umfassende und methodisch gültige Erhebung der tatsächlichen Situation. Dies kann mit der „Befunderhebung“ in der Individualmedizin verglichen werden. Im Unterschied zur typischen Begutachtung in der Medizin ist jedoch nach den Erfahrungen stattgefundener Fehlbelegungsprüfungen Gegenstand bzw. Subjekt der Prüfung nicht ein einzelner Patient, sondern eine Gesamtheit von Patienten, beispielsweise alle innerhalb eines Jahres stationär behandelten Patienten. Diese zugrundeliegende Gesamtheit wird dabei aus praktischen Erwägungen anhand einer *Stichprobe* untersucht.

Inwieweit die in Deutschland eingesetzten Verfahren dem derzeitigen internationalen wissenschaftlichen Stand der Epidemiologie, Biometrie bzw. der Versorgungsforschung entsprechen, ist zunächst offen. Nach eingehender wissenschaftlicher Prüfung einzelner typischer Prüfverfahren muß jedoch festgestellt werden, daß die derzeit verwendeten Prüfverfahren methodisch unzureichend sind und demzufolge keine gültigen Aussagen über das Vorliegen von ungerechtfertigten Krankenhausaufnahmen bzw. Krankenhausbehandlungstagen zulassen. Diese Feststellung erfolgt ungeachtet der Tatsache, daß es in Krankenhäusern möglicherweise doch Fehlbelegung gibt; das verwendete Prüfverfahren erlaubt jedoch keine Aussage hierzu.

Wesentliche methodische Kritikpunkte an dem geübten Vorgehen betreffen das *Verfahren der Stichprobenziehung*, die *Stichprobengröße* und die *Gültigkeit (Validität) und Zuverlässigkeit (Reliabilität) des Prüfverfahrens*.

Ziel dieser Arbeit ist die Vorstellung häufiger Fehler bezogen auf grundsätzliche Anforderungen einer gültigen Fehlbelegungsprüfung anhand von ausgewählten Beispielen. Vorschläge zur Verbesserung der Untersuchungsmethoden und der Darstellung der Ergebnisse werden auf dem Boden der kritischen Untersuchungen gemacht.

Methodische Probleme der Fehlbelegungsprüfung

Es wird keine empirische Aussage für die Bundesrepublik versucht, sondern es sollen anhand von aus tatsächlichen Kasuistiken abgeleiteten Beispielen methodische Schwächen dargestellt und der Anstoß zu einer zielführenden Diskussion gegeben werden. Die gegebenen Beispiele sind allesamt realen Gutachten entnommen und im Einzelfall belegt. Im Interesse der Vertraulichkeit wollen wir jedoch keine näheren Angaben machen, was auch für die intendierte Aussage bzw. den Anstoß einer Diskussion nicht notwendig ist.

Die zugrundeliegenden Sachverhalte betreffen das Modellvorhaben zur Prüfung der Notwendigkeit der Krankenhausbehandlung nach § 275a SGB V durch den Medizinischen Dienst der Spitzenverbände der Krankenkassen vom Frühjahr 1997, zu welchem seitens des Bayerischen Forschungsverbundes

Public Health – Öffentliche Gesundheit eine kritische gutachtliche Stellungnahme erfolgt ist, die detaillierte gutachtliche Untersuchung der Fehlbelegungsprüfung von drei weiteren einzelnen Krankenhäusern mit zehn geprüften Fachabteilungen, telefonische Kontakte zu einer Reihe von Krankenhäusern im Zusammenhang mit einer Fehlbelegungsprüfung, die Empfehlungsvereinbarungen zwischen einer Landeskrankengesellschaft und dem Medizinischen Dienst der Krankenkassen zur Durchführung von Leistungsuntersuchungen, sowie die Kooperation mit dem Medizinischen Dienst der Krankenkassen zweier Bundesländer zur gemeinsamen Entwicklung eines reliablen und validen Prüfinstrumentes für Fehlbelegungsprüfungen, welches in mehreren Pilotprojekten zum Einsatz gekommen ist.

Methodische Probleme der Fehlbelegungsprüfung

1. *Die Stichprobenziehung orientiert sich nicht an der Fragestellung, inwieweit Krankenhausaufnahmen oder aber Krankenhausbehandlungstage überprüft werden sollen. Im Falle der Beantwortung beider Fragestellungen müssen mindestens zwei unabhängige Stichproben gezogen werden.*

Prüfungsgrundlage ist eine repräsentative Stichprobe der zu prüfenden Grundgesamtheit, z.B. sämtlicher vollstationärer Patienten eines Prüfzeitraumes. Wichtig ist hierbei eine Unterscheidung zwischen zwei Prüfgegenständen: zum einen dem der gerechtfertigten *Krankenhausaufnahme*, zum anderen dem des gerechtfertigten stationären *Behandlungstages*. Ist die Prüfung der Krankenhausaufnahme das Ziel, muß sich die Stichprobe auf die stationären Patienten des Prüfzeitraumes beziehen, unabhängig von ihrer jeweiligen Krankenhausverweildauer. Ist die Prüfung des Krankenhausbehandlungstages das Ziel, darf sich die Stichprobe nicht mehr auf die Patienten beziehen, welche ja unterschiedlich viele, statistisch miteinander „verbundene“ Behandlungstage beitragen, sondern muß sich auf die Gesamtheit aller Behandlungstage eines Prüfzeitraumes beziehen.

Beispiel 1: In einem Krankenhaus A, welches bezüglich der Fehlbelegung eines Kalenderjahres untersucht werden soll, werden die Krankenakten aller Patienten, welche sich an drei Stichprobentagen in der Klinik befanden, einer zurückschauenden (*retrospektiven*) Prüfung unterzogen. Auf dem Boden der Prüfergebnisse dieser Akten werden etwa 25% Fehlbelegung bei den Krankenhausaufnahmen und 10% Fehlbelegung bei den stationären Behandlungstagen konstatiert.

Kritisch ist hier anzumerken, daß die Stichprobengewinnung in mehrfacher Hinsicht ungeeignet ist, die angestrebte Grundgesamtheit zu repräsentieren. Zum einen muß bei einer Beschränkung auf drei Tage eines Jahres von Wochentagseffekten und Jahreszeiteffekten ausgegangen werden, die den Case-Mix beeinflussen. Des weiteren ist bei einem Querschnitt der zu einem bestimmten Tag stationären Patienten eine Überrepräsentation von Patienten mit langer Krankenhausverweildauer zu erwarten. Diese haben erfahrungsgemäß eine höhere Wahrscheinlichkeit, fehlbelegte Tage aufzuweisen. Gleiches gilt für eine Selektion bestimmter Diagnosen mit höherer Wahrscheinlichkeit der Fehlbelegung. So überrascht nicht, daß die durchschnittliche Liegezeit der geprüften Patienten annähernd doppelt so hoch war wie in der abzubildenden Grundgesamtheit. Zudem waren die Fehlbe-

gungsquoten für Krankenhausaufnahmen und für Krankenhausbehandlungstage aus einer einzigen Stichprobe berechnet worden. Dies ist ein statistisch unzulässiges Verfahren, da auf Ebene der Behandlungstage die statistische Forderung der Unabhängigkeit der Untersuchungseinheiten nicht mehr gegeben ist: Ein Patient, der als „Langlieger“ einen Fehlbelegungstag aufweist, wird mit großer Wahrscheinlichkeit auch weitere im Verlauf einer stationären Behandlung aufweisen.

2. Der Festlegung der Stichprobengröße geht keine statistische Berechnung voraus, in der u. a. die zu erwartende Fehlbelegungsquote und ein akzeptabler Fehlerbereich berücksichtigt werden.

Die Festlegung der Stichprobengröße, durch welche eine Beurteilung der Grundgesamtheit erfolgen soll, unterliegt statistischen Gesetzen. Festlegungen wie „maximal 10% der Krankenhausaufnahmen des untersuchten 12-Monats-Zeitraumes“, wie sie in Vereinbarungen gefunden werden können, sind statistisch unsinnig. Einflußgrößen auf die Stichprobengröße sind die vermutete Fehlbelegungsquote, ein akzeptierter Fehler durch Zufallseinflüsse (Typ-I- und Typ-II-Fehler) und die angestrebte Genauigkeit der Aussage. So beträgt für eine Genauigkeit von 2 Prozentpunkten die benötigte Stichprobengröße 456 Patienten bei einer angenommenen Fehlbelegung von 5%, aber bereits 864 Patienten bei einer angenommenen Fehlbelegung von 10% (Irrtumswahrscheinlichkeit des Typ-I-Fehlers jeweils 5%). Diese notwendigen Fallzahlen ändern sich, wenn der Typ-I-Fehler nur 1% betragen soll – die Fallzahlen betragen dann 190 bzw. 149/ Patienten, und natürlich auch, wenn die angestrebte Genauigkeit anders festgelegt wird.

Beispiel 2: In einem Krankenhaus B, welches bezüglich der Fehlbelegung eines 12-Monatszeitraumes untersucht werden soll, werden die Krankenakten von 87 Patienten einer Abteilung eines Kalendermonats untersucht, 9 Krankenhausaufnahmen werden dabei als unnötig eingestuft. Die Fehlbelegungsquote wird mit 10% angegeben.

Kritisch anzumerken ist, daß die Untersuchung der Patienten eines Kalendermonats ebensowenig wie die Untersuchung nach Stichtagen geeignet ist, die angestrebte Grundgesamtheit repräsentativ abzubilden, es gilt dabei das oben Gesagte. Auch hier fanden sich erhebliche Abweichungen zwischen Stichprobe und Grundgesamtheit, z.B. bei der Krankenhausverweildauer. Im speziellen ist aber zu bemängeln, daß als Mindestanforderung weder eine Fallzahlberechnung vor Festlegung der Stichprobengröße noch ein Vertrauensbereich (Konfidenzintervall) nach erfolgter Prüfung um die ermittelte Quote gelegt wurde, welcher den Einfluß des Zufalls abschätzt. So wird üblicherweise ein 95%-Vertrauensbereich angegeben, welcher den wahren Wert der Grundgesamtheit mit 95% Wahrscheinlichkeit enthält. Im vorliegenden Fall hätte dieser Bereich die Werte 5%–19% umfaßt und die erhebliche Unsicherheit der genannten Fehlbelegungsquote allein aufgrund von Zufallseinflüssen deutlich gemacht. Dies setzt eine gültige (valide) Prüfmethodik voraus, eine Annahme, die ebenfalls kritisch hinterfragt werden muß.

3. Durch eine rückwirkende Beurteilung von Krankenunterlagen können nicht alle Aspekte, die in dem konkreten multifaktoriellen Prozeß zur Entscheidung zu einer stationären Aufnahme oder der Legitimation eines bestimmten Behandlungstages führen, berücksichtigt werden. In der Regel kommt es dabei zu einer Überschätzung einer etwaigen Fehlbelegung.

Die Gültigkeit (Validität) von Prüfverfahren ist eine Aussage, welcher eine rigorose metrische bzw. testtheoretische Evaluation vorausgehen muß. Die Gültigkeit eines Verfahrens ohne eine solche Evaluation anzunehmen ist nicht akzeptabel. An der Gültigkeit bzw. Zuverlässigkeit eines Prüfverfahrens bestehen bis zum Nachweis des Gegenteils erhebliche Zweifel.

Beispiel 3: In einem Krankenhaus C wird nach standardisierten Prüfprotokollen vorgegangen, die jedoch nicht einer metrischen Testung nach gängigen Kriterien unterzogen worden sind. Unvermeidliche Diskrepanzen bei der Beurteilung mit den behandelnden Krankenhausärzten versucht man durch nachträgliche Konsensgespräche zu lösen, die z.T. zu überhitzten Reaktionen auf beiden Seiten führen. Im Verlauf der Diskussion wird seitens der Prüfarzte u.a. mit den Empfehlungen aus Standardlehrbüchern argumentiert, ohne daß dadurch die Diskrepanzen aufgelöst werden können.

Das geschilderte Beispiel macht das methodische Ungenügen der derzeitigen Prüfverfahren deutlich. Wie jedes Testverfahren muß auch ein Erhebungsinstrumentarium zur Beurteilung der Notwendigkeit von medizinischen Leistungen zwei wesentliche Gütekriterien nachweisen: Gültigkeit (Validität) und Zuverlässigkeit oder Reproduzierbarkeit (Reliabilität)

Gültigkeit beschreibt dabei die Fähigkeit eines Verfahrens, das zu messen, was es vorgibt zu messen, in diesem Fall die Notwendigkeit einer stationären Behandlung. Erschwert wird die Definition einer notwendigen Krankenhausbehandlung durch das Fehlen allgemein gültiger Kriterien („goldener Standard“). Um diesen Mangel zu kompensieren, sollte ein Instrument hinsichtlich der Gültigkeit zumindest die folgenden Eigenschaften besitzen (Payne 1987):

Augenscheinvalidität: Dafür beurteilt ein Expertengremium das Ausmaß, in welchem das Instrument oder Verfahren überhaupt in der Lage ist, nicht notwendige Krankenhausbehandlungen zu messen. Kritisch ist anzumerken, daß dieses Validitätskriterium nicht in einem strengen Sinn empirisch faßbar ist.

Inhaltsvalidität: Dieser Begriff beschreibt das Ausmaß, in welchem das Instrument alle relevanten Bereiche berücksichtigt, die für die Definition von nicht notwendigen Krankenhausbehandlungen erforderlich sind.

Konvergenzvalidität: Hierfür ist das Ausmaß zu prüfen, in welchem sich die auf dem Instrument oder Verfahren basierten Entscheidungen, z.B. durch ein Expertengremium, rekonstruieren lassen.

Bei vielen Diagnosen und Vorgehensweisen sind sich auch namhafte Experten uncinig, was richtig bzw. angemessen und falsch bzw. unangemessen ist. Diese Problematik wird auch in

der aktuellen Diskussion über Leitfäden in der Medizin verdeutlicht. Zwischenzeitlich gibt es kaum eine medizinische Fachgesellschaft, die nicht Leitlinien für die Diagnostik und Therapie definierter Krankheitsbilder entwickelt. Die Bundesärztekammer zählt inzwischen weit mehr als 500 solcher Leitlinien und Richtlinien, die sich jedoch – obwohl meistens von einer medizinischen Fachgesellschaft sanktioniert – in ihrer Qualität erheblich unterscheiden.

Lehrbuchgestützte Kriterien sind nach Ansicht vieler Experten heutzutage völlig unzureichend. So verweist Ian Chalmers, einer der weltweit führenden Autoritäten auf dem Gebiet der „Evidence-based Medicine“, 1995 in einer gutachterlichen Stellungnahme vor dem House of Lords in England (Oberhaus) darauf hin, daß viele Patienten ihr Leben verlieren, weil sich Mediziner auf inhaltlich überholte Lehrbücher beziehen. Er belegt diese Behauptungen u.a. damit, daß in der aktuellen Ausgabe eines der führenden medizinischen Lehrbücher, dem Oxford Textbook of Medicine, Medikamente, die eine Verstopfung der Herzkranzgefäße aufzulösen vermögen (sog. Thromobolytika), als „wissenschaftlich ungeprüft und von zweifelhafter Sicherheit“ bezeichnet würden. Mediziner, die sich auf dieses Lehrbuch beziehen, würden ihren Patienten eine durchaus wirksame und unter Berücksichtigung entsprechender Vorkehrungen sichere Therapie vorenthalten, die in vielen Fällen lebenserhaltend ist. Der medizinische Fortschritt ist so schnell, daß die meisten Lehrbücher nach Veröffentlichung z.T. schon wieder überholte Informationen enthalten. Weatherall et al. vertreten daher in einem Kommentar (The Lancet) den Standpunkt, daß Lehrbücher bestenfalls als Orientierungshilfe zum Verständnis von Krankheiten dienen; wenn es aber darum geht, die aktuelle Diagnostik und Therapie zu verstehen, müßten unbedingt medizinische Fachzeitschriften und andere Hilfen und Informationsquellen beigezogen werden (Weatherhall 1995).

An der Gültigkeit der Aussagen auf dem Boden der üblichen Begutachtung müssen daher zunächst erhebliche Zweifel geäußert werden. Erst wenn ein Panel ausgewiesener Experten die zugrunde gelegten Kriterien bzw. Resultate der Prüfung autorisiert, könnten diese als Grundlage für die Festlegung von Fehlbelegungen verwendet werden.

Die *Zuverlässigkeit* eines Instrumentes oder Verfahrens wird primär durch dessen Reproduzierbarkeit definiert. Ein Instrument oder Verfahren ist zuverlässig, wenn es, mehr als einmal unter vergleichbaren Bedingungen bei denselben Patienten eingesetzt, zu den selben Ergebnissen führt. Im Falle der Beurteilung von nicht notwendigen Krankenhausbehandlungstagen muß also nachgewiesen werden, daß ein Prüfer, der dieselbe Krankenakte an unterschiedlichen Zeitpunkten bewertet, zu demselben Prüfergebnis kommt (Intra-Beurteiler Reliabilität) bzw. daß zwei unterschiedliche Prüfer für eine Krankenakte unabhängig zum gleichen Prüfergebnis kommen (Inter-Beurteiler Reliabilität).

Die Diskrepanz in der Einschätzung der Fehlbelegungstage zwischen den Krankenhausärzten und den Prüferärzten läßt auch bzgl. der Zuverlässigkeit begründete Zweifel entstehen. Hätten beispielsweise Fachärzte mit guter Kenntnis des betroffenen Krankenhauses sowie der lokalen Versorgung die Beurteilung durchgeführt, so ist anzunehmen, daß unter Verwendung desselben Erfassungsinstrumentariums ein anderes

Ergebnis erreicht worden wäre, was die Reliabilität des verwendeten Instrumentariums in Frage stellt.

Gewöhnlich werden mindestens drei Verfahren zur Bestimmung der Zuverlässigkeit eines Instrumentes durchgeführt und dokumentiert:

Allgemeine Übereinstimmung: Der Anteil von Patienten in Prozent, bei denen zwei oder mehr Prüfer unabhängig zu derselben Entscheidung bezüglich des Substitutionspotentials gekommen sind.

Spezielle Übereinstimmung: Die Übereinstimmungsrate bei den Patienten, bei denen eine nicht notwendige Krankenhausaufnahme durch wenigstens einen Prüfer festgestellt wurde.

Kappa-Statistik: Diese Analyse bestimmt den Übereinstimmungsgrad und macht eine Aussage dazu, welcher Anteil dieser Übereinstimmung über eine zufällige Übereinstimmung hinausgeht (Cohen 1960).

Die eingesehenen Gutachten selbst machten keinerlei Aussagen zur Validität und Reliabilität des angewandten Prüfverfahrens bzw. Erfassungsinstrumentariums. Obschon das Nichterwähnen dieser elementaren Eigenschaften das gesamte Prüfverfahren in Frage stellt, kann zunächst nicht abschließend festgestellt werden, inwieweit der MDK über diese Informationen verfügt. Ist jedoch keine Validierung vorgenommen worden, ist das Prüfverfahren prinzipiell als gegenstandslos zu betrachten, da die Untersuchungsergebnisse im Extremfall rein zufällig oder Folge einer systematischen Verzerrung sein könnten. Dies ist nicht damit gleichzusetzen, daß es keine Fehlbelegung geben könnte. Es verdeutlicht aber, daß aufgrund eines mangelhaften Prüfinstrumentariums und bzw. oder einer in Methodik und Umfang unzureichenden Stichprobenziehung nicht auszuschließen ist, daß das Konfidenzintervall um die festgestellte Fehlbelegungsquote als Maß für die Unsicherheit einer Aussage über den wahren Wert im unteren Bereich bis auf null Tage reicht.

Grundsätzlich läßt sich eine Prüfung auf Notwendigkeit von Krankenhausaufnahmen bzw. die Notwendigkeit einzelner stationärer Behandlungstage prospektiv (*ex ante*) oder retrospektiv (*ex post*) durchführen. Während die retrospektive, d.h. die auf einer rückwirkenden Beurteilung anhand von Krankenakten basierte Untersuchung, Vorteile hinsichtlich der logistischen Effizienz des Prüfverfahrens aufweist, führt dieses Verfahren im Vergleich zu prospektiven (*ex ante*) Untersuchungen zu fälschlicherweise erhöhten Fehlbelegungsquoten. Ursache dieses Fehlers ist, daß chronologisch erst die nach dem jeweiligen zu prüfenden Tag vorliegende Information mit zur Beurteilung herangezogen wird, also ein wirklichkeitsfremdes Informationsszenario geschaffen wird.

Bei prospektiven Designs würde ein Gutachter den Entscheidungsprozeß eines aufnehmenden Krankenhausarztes synchron beobachten und bewerten. Zwischenzeitlich konnte in verschiedenen internationalen Studien gezeigt werden, daß Krankenakten, welche prinzipiell nicht für den Zweck einer Fehlbelegungsprüfung konzipiert sind, für eine solche nur bedingt geeignet sind. Sie dienen neben einer Dokumentation der medizinischen und verwaltungsmäßigen Vorgänge im

Krankenhaus vorrangig der haftungsrechtlichen Absicherung der behandelnden Ärzte, des Pflegepersonals und des Krankenhausträgers. In den Krankenakten sind nicht zwingend alle notwendigen Informationen zur „*ex post*“-Beurteilung der Notwendigkeit von einzelnen Behandlungstagen enthalten. Untersuchungen aus der Schweiz haben dokumentiert, daß retrospektive Beurteilungen das Ausmaß an nicht notwendigen Krankenhausbehandlungen eher überschätzen (Santos-Eggimann et al. 1997). Weitere wissenschaftliche Untersuchungen konnten belegen, daß Krankenakten bei der retrospektiven Beurteilung mit einem Standardverfahren ein höheres Ausmaß an nicht notwendigen Krankenhausbehandlungen annehmen lassen (Ramos-Cuadra et al. 1995). Hierin begründet sich eine prinzipiell eingeschränkte Eignung von Krankenakten für eine systematische Fehlbelegungsprüfung.

Vorschläge zur Verbesserung der Untersuchungsmethoden und der Darstellung der Ergebnisse

Die Vorschläge zur Verbesserung der derzeit ungenügenden Situation der Fehlbelegungsprüfung lassen sich im wesentlichen aus dem oben Gesagten ableiten.

1. Zum einen betreffen sie das Verfahren einer repräsentativen Stichprobenziehung. Diese sollte nur mit statistischer Expertise vorgenommen werden, da gezeigt werden konnte, daß selbst eine einfache Zufallsstichprobe nach Aufnahmen und Behandlungstagen unzureichend ist. Zu fordern ist eine geschichtete Zufallsstichprobe, welche wichtige Einflußgrößen explizit berücksichtigt, sowie eine vorhergehende explizite Fallzahlberechnung.
2. Des weiteren sollte die Präsentation der Untersuchung bzw. der Ergebnisse detailliert auf die Methodik, d.h. die Eignung des Studiendesigns, die metrischen Eigenschaften des Untersuchungsinstrumentariums, die Verfahren zur Stichprobenziehung und die statistische Auswertung eingehen. Liegt ein solches validiertes Prüfinstrument für den deutschen Sprachraum nicht vor, ist seine Entwicklung unverzichtbar. Auch muß ausgeführt werden, wie die Prüfärzte während der Auswertung der Krankenakten mit Zweifelsfällen umgegangen sind.
3. Das Gutachten sollte sich auch explizit zu den Limitationen des Prüfverfahrens und das daraus resultierende Fehlinterpretationspotential äußern. Dieser Mangel wird insbesondere dann relevant, wenn Krankenkassenvertreter solche Gutachten im Rahmen von Budgetverhandlungen als Verhandlungsgrundlage verwenden. Als Minimalanforderung sollte ein 95%-Konfidenzintervall um die festgestellten Fehlbelegungsquoten gelegt werden. Durch die Angabe von Vertrauensbereichen bzw. einer Fehlerrechnung wird es erst möglich, das Ausmaß etwaiger Fehler zu quantifizieren, sozusagen im Sinne eines idealen („*best case*“) bzw. denkbar schlechten („*worst case*“) Szenarios. Ansonsten wird durch die Angabe von Fehlbelegungsquoten bis auf die zweite Stelle nach dem Komma genau in den Gutachten eine Präzision suggeriert, die ein solches Verfahren nie zu leisten vermag.

Literatur

- 1 Chalmers I. In: House of Lords Select Committee of Science and Technology. Sub-Committee I: Medical Research and the NHS Refers. London: HM Stationary Office 1995.
- 2 Cohen JA. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; 20: 37.
- 3 Payne SMC. Identifying and managing inappropriate hospital utilization: A policy synthesis. *Health Services Research* 1987; 22: 710-769.
- 4 Ramos-Cuadra A, Marión-Buen J, García-Martín M et al. (1995). The effect of completeness of medical records on the determination of appropriateness of hospital days. *Int J Quality of Care* 7: 267-275.
- 5 Santos-Eggimann B, Sidler M, Schöpfer D, Blanc T. Comparing results of concurrent and retrospective designs in a hospital utilization review. *International Journal for Quality in Health Care* 1997; 9: 115-120.
- 6 Weatherall DJ, Ledingham JGG, Warrell DA. On dinosaurs and medical textbooks. *Lancet* 1995; 346: 4-5.

Manfred Wildner

Bayerischer Forschungsverbund Public Health – Öffentliche Gesundheit
Ludwig-Maximilians-Universität München
Tegernseer Landstraße 243
D-81549 München

E-mail: bfv@ibe.med.uni-muenchen.de

BUCHBESPRECHUNG

- 1 Ludwig, Malte: **Angiologie in Klinik und Praxis**. 1998, 347 S., Georg Thieme Verlag Stuttgart, Gebunden, DM 198,00/öS 1445,00/sFr 178,00, ISBN 3 13 110191 1-694

Dieses Werk entstand aus der Zusammenarbeit eines Angiologen, eines Gefäßchirurgen und eines Radiologen. Diese Zusammenstellung von Autoren verschiedener Fachgebiete spiegelt die Notwendigkeit einer Zusammenarbeit von Gebietsärzten bei der Behandlung von Gefäßkrankheiten. Ziel des vorliegenden Lehrbuches ist es, praxisnah grundlegende Weiterbildungsinhalte des Schwerpunktfaches Angiologie zu vermitteln.

Arterienerkrankungen und Venenerkrankungen sind die Themen des ersten und zweiten Kapitels. Danach werden Lungenembolie und Lymphgefäßerkrankungen behandelt. Weitere Kapitel widmen sich den speziellen Gefäßerkrankungen, angeborenen Gefäßmalformationen, Gefäßverletzungen und nicht invasiven apparativen angiologischen Untersuchungsverfahren.

Die epidemiologische und volkswirtschaftliche Bedeutung von Gefäßkrankheiten, insbesondere von peripheren Gefäßkrankheiten, wird unterschätzt. Aus diesem Grund ist es erfreulich, wenn dem noch jungen Gebiet der Angiologie eine wachsende Bedeutung zugestanden wird. Das Lehrbuch von Ludwig erweitert das Literaturangebot für diesen Bereich der Inneren Medizin auf erfreuliche Weise.

C. Michel, Kiel