

Nicht nur neu, sondern gut und seinen Preis wert: Welche Instrumente und Regelungen braucht das deutsche Gesundheitssystem?

**Prof. Dr. med. Reinhard Busse
Lehrstuhl für „Management im Gesundheitswesen“
an der Technischen Universität Berlin**

**- Kurzfassung -
anlässlich des Expertenforums des MDS
„Implantate und Co: Wer gibt Antworten bei Risiken und
Nebenwirkungen?“
am 3. Mai 2007 in Berlin**

Leistungen der gesetzlichen Krankenversicherung müssen gemäß § 12 SGB V „ausreichend, zweckmäßig und wirtschaftlich“ sein. Laut § 135 müssen ihr Nutzen, ihre medizinische Notwendigkeit sowie ihre Wirtschaftlichkeit nachgewiesen sein, damit sie im Rahmen der vertragsärztlichen Versorgung zu Lasten der Krankenkassen erbracht werden dürfen. Bezüglich der stationären Versorgung ist der Gesetzgeber in § 137c nicht ganz so streng; hier darf prinzipiell jede Leistung erbracht werden und nur wenn eine Überprüfung ergibt, dass sie für eine Versorgung nach § 12 nicht erforderlich ist, darf sie fortan nicht mehr erbracht werden. Diese gesetzlichen Vorgaben bilden den Hintergrund einer Entwicklung, die unser Gesundheitssystem in der Zukunft vermutlich so prägen wird wie wenige andere: die Bewertung von Technologien und Verfahren, kurz „Health Technology Assessment“ (HTA) genannt.

Schon diese kurze Einführung zeigt, dass HTA zwar seit 10 Jahren Schritt für Schritt in das deutsche Gesundheitssystem eingeführt wurde, dass aber der derzeitige Stand keineswegs als das Optimum angesehen werden kann. Dieser Beitrag soll diskutieren, was sich in Deutschland ändern sollte bzw. müsste? Dies lässt sich m.E. auf einen Nenner bringen, der – auch wegen seiner scheinbar paradoxen Forderungen – näher zu erläutern sein wird: Wir müssen gleichzeitig rigider und flexibler werden, aber auch innovativer!

Rigidität: Die Einsicht in die Notwendigkeit der Evaluation bzw. die Akzeptanz der Evaluationskultur muss bei den Entscheidungsträgern noch gefestigt werden. Denn trotz der relativ klaren Aussagen im SGB V gelangen viele Leistungen immer noch unevaluiert in den Leistungskatalog

bzw. die tatsächliche Versorgung. Zum Beispiel hätte der neue Einheitliche Bewertungsmaßstab 2000plus (EBM 2000plus) eigentlich „nur“ eine neue Gebührenliste für bereits vorhandene Leistungen sein sollen. Ein Blick in den EBM 2000plus zeigt aber, dass nicht nur alte Leistungen entfielen, sondern auch neue ungeprüft aufgenommen wurden. Dies betrifft insbesondere den (sehr) langen Anhang mit ambulanten Operationen.

Noch offensichtlicher geschieht die ungeprüfte Leistungserbringung im stationären Sektor. Die Einführung der DRGs hat nicht – wie ggf. zu erwarten oder, je nach Sichtweise, zu befürchten gewesen wäre – zu einer Quasi-Positivliste geführt; im Grunde werden durch das DRG-System nur die häufigen Diagnose-/Therapie-Paare gruppiert, wodurch die seltenen aber nicht unmöglich geworden sind. § 17 Krankenhausfinanzierungsgesetz (KHG) stellt dazu fest, dass man sich über zusätzliche Leistungen auf lokaler Ebene einigen soll. Und genau dies geschieht – und nicht gerade selten. In der Klinik, in der Professor A, Autor des Operationsverfahrens A, arbeitet, wird diese Leistung vergütet und in der Nachbarklinik das Verfahren B des dort tätigen Privatdozenten B.

Ein beliebter Weg, um eine Leistung über die lokale Ebene hinaus zu etablieren, besteht darin, im OPS-Katalog des DIMDI (Deutsches Institut für Medizinische Dokumentation und Information), in dem Prozeduren katalogisiert sind, eine Ziffer für diese Leistung zu erhalten. Mit einer solchen Ziffer können dann alle Krankenhäuser diese Leistung dokumentieren. Was aber alle Krankenhäuser dokumentieren können, ist tatsächlich vorhanden und lässt sich bei der nächsten Runde der DRG-Berechnung berücksichtigen. Die Leistung aber, die hinter dieser OPS-Ziffer steht, ist bei deren Vergabe nicht evaluiert worden.

Ein ebenfalls beliebtes Instrument zur Leistungssetzung ist die Weiterbildungsordnung. Wenn darin steht, dass bestimmte Maßnahmen für die Facharztqualifikation nachgewiesen werden müssen, müssen diese Maßnahmen in den Krankenhäusern auch stattfinden. Daher sollten auch die Dinge, die in eine Weiterbildungsordnung neu aufgenommen werden, rigider als bisher einem HTA unterzogen werden.

Aus alledem ist zu schließen, dass die Evaluationsrichtlinien vereinheitlicht und dann strenger angewendet werden müssen. Die seit Oktober 2005 gültige neue Verfahrensordnung des Gemeinsamen Bundesausschusses (G-BA) ist diesbezüglich noch nicht weitgehend genug, da sie zwar den Bewertungsprozess angleicht, bezüglich des stationären Sektors jedoch den Verbotsvorbehalt aufrecht erhält.

Flexibilität: Mehr Flexibilität ist vor allem bei der Beurteilung der Evidenz geboten: Die Evaluation der Wirksamkeit kann und muss nicht immer nur auf der Grundlage von randomisierten kontrollierten Studien (RCTs) erfolgen. Entscheidungsbedarf besteht auch dort, wo keine RCTs vorliegen. Routinedaten oder eine gut durchgeführte Kohortenstudie können zur Beantwortung bestimmter Fragestellungen durchaus „besser“ als eine schlecht durchgeführte RCT sein. Diese Aussage darf aber keinesfalls so missverstanden werden, dass wir – für Medizinprodukte – *keine* RCTs brauchen.

Der Wunsch nach mehr Flexibilität betrifft noch einen weiteren Punkt, zu dem die Diskussion im G-BA bereits angelaufen ist. Es geht darum, dass neben reinen Ja- und Nein-Entscheidungen zum Leistungskatalog auch Zwischenstufen möglich sein sollten. Wenn eine neue Technologie zur Entscheidung ansteht, liegen nicht immer (bzw. zumeist nicht) die Daten vor, die alle Unsicherheiten beseitigen. Auf sichere Ergebnisse zu warten, ist aber oft nicht möglich. Eine neue Technologie könnte auf dieser Basis nur „blind“ zugelassen oder abgelehnt werden. Alternativ sollte es möglich sein, die Technologie z. B. nur für bestimmte Indikationen oder zunächst nur an Zentren mit bestimmten Voraussetzungen oder nur an namentlich bezeichneten Zentren zeitlich begrenzt zuzulassen – mit dem verbindlichen Auftrag, die Ergebnisse zu evaluieren. Oder die Freigabe beschränkt sich auf die Teilnehmer an einer z. B. vom G-BA genehmigten prospektiven, multizentrischen Evaluationsstudie. Der G-BA ist – insbesondere bei seinen Akupunktur-Entscheidungen – bereits von reinen Ja-/Nein-Entscheidungen abgewichen. Klar definierte Zwischenschritte jenseits vom Modellvorhaben (vgl. erster Akupunktur-Beschluss), ersten Einschränkungen der Indikation (z. B. zweiter Akupunktur-Beschluss, aber auch zur Kernspintomographie der Mamma) und Aussetzen der Beschlussfassung müssen aber erst noch definiert werden.

Innovationen: Die Bewertung neuer Technologien sollte idealiter in drei Stufen erfolgen:

- In der ersten Stufe wird die Wirksamkeit (efficacy) bewertet, die sich aus klinischen Studien ergibt. Bisher erfolgt die Bewertung in Deutschland ganz überwiegend auf dieser Stufe, wobei – zumindest bei Arzneimitteln – sich hier eine Ungereimtheit ergibt, da oftmals die gleichen Studien zur Zulassung (durch das Bundesinstitut für Arzneimittel und Medizinprodukte – BfArM) und zur GKV-Erstattungsfähigkeit (Bewertung durch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen – IQWiG, Entscheidung durch den G-BA) genutzt werden. Bei Medizinprodukten könnte man hier etwas anders

argumentieren, da die Wirksamkeit in der EU – anders als in den USA – derzeit (noch) nicht Voraussetzung ist.

- In der zweiten Stufe wird die Effektivität unter Alltagsbedingungen (community effectiveness) – auch oft der „Nutzen“ (benefit) genannt – bewertet. Hier geht es im Kern darum, ob die Technologie auch jenseits der klinischen Studien mit ausgewählten Leistungserbringern (die motiviert und überdurchschnittlich geschult sind) und Patienten (die zumeist nicht multimorbide, sehr alt etc. sind) „wirkt“. Zur Messung des Nutzens werden auch ein längerer Zeithorizont als in klinischen Studien und patientennähere Outcomes erwartet.
- In der dritten Stufe erfolgt die Abschätzung der Summe der Auswirkungen (impact). Dazu gehören neben der Wirksamkeit vor allem: Die Sicherheit (für Patienten, aber auch für Anwender), ethische und soziale Implikationen, organisatorische Bedingungen (z. B.: Wird es zu einer Verschiebung von einem Sektor in einen anderen kommen? Sind überhaupt genügend Leistungserbringer vorhanden?), Kosten (für den einzelnen Patienten und für das Gesamtsystem) sowie die – zumindest im Ausland auf dieser Stufe im Mittelpunkt des Interesses stehende – Kosten-Effektivität.

Methodisch ergibt sich hierbei eine Reihe von Herausforderungen. So wird bezüglich der für die Beurteilung der Effektivität notwendigen Vergleichsgruppen oft darüber gestritten, ob der Vergleich mit einer Placebogruppe ausreicht oder ob ein Head-to-Head-Vergleich mit einem tatsächlich genutzten Präparat bzw. Verfahren erforderlich ist. Auch dabei müssen wir in den Stufen denken. Für den Wirksamkeitsnachweis bzw. die Zulassung reicht meist eine Placebogruppe. Je näher wir aber an Entscheidungen kommen, wo es um Erstattungsfähigkeit geht, desto weniger reichen Studien gegen Placebo und desto eher sind Head-to-Head-Studien gegen die derzeitige „Standardtherapie“ bzw. die bisher kosten-effektivste Therapie notwendig. Damit ergeben sich aber auch neue Schwierigkeiten: Wer ist z. B. für die Durchführung solcher Studien verantwortlich? Die Firma, die zuletzt mit ihrem Präparat bzw. ihrem Produkt an den Markt will? Beide zu gleichen Teilen? Woher sollen diese aber wissen, was zukünftig – also dann, wenn die Studie fertig ist – der wichtigste Kompetitor sein wird? Müsste hier nicht doch die öffentliche Hand die Finanzierung übernehmen, wenn es z. B. um die Evaluation von Penicillin im Vergleich zu neuen, deutlich teureren Antibiotika geht? Wie gehen wir mit dem Problem um, dass wir bei der Beurteilung der Effektivität unter Alltagsbedingungen die Limitationen von RCTs – also die Konzentration auf bestimmte Patientenpopulationen, d.h. zumeist nicht die ganz Alten und nicht die Multimorbiden – vermeiden wollen und müssen (sonst wären es ja keine Alltagsbedingungen)?

Trotz dieser Fragen und Schwierigkeiten ist aber heute bereits klar: Bei Arzneimitteln werden wir uns in Zukunft nicht mehr mit einem reinen Plazebovergleich begnügen können. Bei Medizinprodukten muss die Evaluationskultur überhaupt erst Einzug halten – bei Herstellern, Anwendern (insbesondere im Krankenhaus) und Entscheidungsträgern (im G-BA). Solange wir nicht genügend Studien mit direkten Vergleichen haben (und diese nicht so weit wie möglich verzerrungsfrei sind), können den Entscheidungsträgern auf der zweiten und dritten Stufe der Bewertung auch Modellierungen helfen, wie sie von vielen Gesundheitsökonomen favorisiert werden. Für diese gilt aber: Wir wissen noch gar nicht, welche Modellierungen sich später in der Realität als zutreffend heraus stellen. Die Ergebnisse von Modellierungen dürfen daher nicht als sakrosankt betrachtet werden, sondern müssen sich dem Wirklichkeitstest unterziehen. Es müssen also so schnell wie möglich tatsächliche Daten geliefert werden und verglichen werden. Dafür muss dringend auch eine Nutzung von Routinedaten, wie sie im Rahmen des § 21 Krankenhausentgeltgesetz (KHEntgG) von den Krankenhäusern erhoben und über die Datensammelstellen an das Institut für das Entgeltsystem im Krankenhaus (InEK) weiter gegeben werden, aber auch der separat von der Bundesgeschäftsstelle Qualitätssicherung (BQS) erhobenen Daten, erforderlich.

Das deutsche Gesundheitssystem ist heute sicherlich evidenzbasierter als vor 10 Jahren. Es gibt jedoch keinen Grund, sich bequem zurückzulehnen. Die konsequente, aber zielgerichtete Anwendung von HTA – insbesondere für Medizinprodukte – bleibt eine der Herausforderungen, um unser System im Sinne der §§ 12 und 135ff zu erhalten und zu optimieren.