

Prof. Dr. Bernhard Ruger
Institut fur Statistik
Universitat Munchen

Buro:
Kunigundenstrae 2
80802 Munchen

Munchen, 13.03.2008

Anmerkungen zur

IQWiG-Methodik fur die Bewertung von Verhaltnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung (Version 1.0 vom 24. Januar 2008)

Die Kosten-Nutzen-Analyse des IQWiG besteht im Wesentlichen aus den folgenden sieben Schritten:

- 1) Eine (moglichst scharfe) Diagnosestellung; die gesamte Analyse bezieht sich auf die jeweilige spezifische Diagnose.
- 2) Erfassung der vorhandenen (ublichen) Therapien (A, B, C usw.), mit denen eine bestimmte (neue oder problematische) Therapie X verglichen werden soll.
- 3) Festlegung einer eindimensionalen kardinalskalierte Nutzen- oder Werteskala fur diese Krankheit (unabhangig von den in Frage kommenden Therapien).
- 4) Aufstellung der Gesamtkosten (in erster Linie die direkten erstattungsfahigen Kosten) fur jede der in Betracht gezogenen Therapien.
- 5) Bestimmung der (auf die Therapien A, B, C usw. bezogenen) Effizienzgrenze im Kosten-Nutzen-Koordinatensystem.
- 6) Beurteilung der Therapie X nach ihrer Lage im Kosten-Nutzen-Diagramm: Liegt X oberhalb oder auf der Effizienzgrenze, so wird X prinzipiell zugelassen, anderenfalls nicht.
- 7) Bei Zulassung von X: Aus der Effizienzgrenze bestimmte Richtlinien zur Festlegung von Fest- oder Hochstbetragen fur die Erstattung der Kosten von X.

Zentrale Bedeutung in der IQWiG-Kosten-Nutzen-Analyse hat die Bestimmung einer Effizienzgrenze. Dieses streng entscheidungstheoretische Instrument stellt die *Kurve der „admissible decision rules“* dar und wird durch diejenigen

„decision rules“ (Handlungsalternativen, hier Therapien) gebildet, die nicht dominiert werden. (Eine Therapie A wird durch eine Therapie B dominiert, wenn B *besser und billiger* ist als A). Dieses im ökonomischen Bereich (mit den dort herrschenden streng quantitativen Bewertungsmöglichkeiten) sehr bewährte Konzept wird in aktuellen Veröffentlichungen immer stärker auch für den Bereich der Gesundheitsökonomie einzusetzen gefordert. Das IQWiG-Konzept liegt in diesem Sinne im Trend der Entwicklung. Gleichwohl entstehen aber hier aus einer Anwendung der IQWiG-Methodik zur Bewertung von Therapien die folgenden Probleme.

Erstes Problem:

Der **Nutzen** einer Therapie oder der **Wert**, der dem Nutzen beigemessen wird, muß nach IQWiG auf einer **eindimensionalen Kardinalskala** gemessen werden (die auch modellabhängig festgelegt werden darf). Wie problematisch diese Voraussetzung gerade für psychotherapeutische Behandlungen ist, zeigt das folgende Zitat aus der IQWiG-Methodik (Seiten 15/16): *„Wenn Intervention A doppelt so viel wie Intervention B kostet, dann wird doppelt so viel an alternativem Wert eines Nutzens geopfert, sollte sich jemand eher für den Erwerb von A statt von B entscheiden. Damit sich diese Entscheidung gegenüber dem Leistungsempfänger rechtfertigen lässt, sollte A mindestens zweimal so viel Nutzen stiften wie B. Um in die Lage versetzt zu werden, zu beurteilen, ob dies der Fall ist, ist es erforderlich, dass die Messgröße der Wertbestimmung des Nutzens kardinalskaliert ist, so dass die Behauptung, dass A nicht nur mehr Wert eines Nutzens als B erbringt (d.h. ein ordinal skaliertes Maß), sondern um x-mal mehr wert ist, aussagekräftig wird. Auch wenn x selten mehr als eine approximative Zahl sein wird (d.h. eine grobe Bewertung), wird dadurch beträchtlich mehr Information generiert als durch eine bloße ordinale Bewertung. Der Vergleich des erzielten Wertes des Nutzens mit den dadurch entstandenen Kosten stellt den wesentlichen Inhalt der vorgeschlagenen Methodik dar.“* Danach müsste beispielsweise für den Vergleich einer tiefenpsychologischen Behandlung, die, sagen wir einmal, doppelt so teuer ist wie eine verhaltenstherapeutische, feststellbar sein, ob erstere einen mindestens doppelt so hohen Wert hervorbringt als letztere! Vielleicht kann hier wie in vielen anderen Vergleichsfällen gerade noch eine Ordinalskala, sicherlich aber keine Kardinalskala zu Grunde gelegt werden. Man kann eben nicht so ohne weiteres, wie im IQWiG-Text unterstellt wird, aus einer ordinalen Nutzenskala durch eine zahlenmäßige Umbewertung des Nutzens zu einer kardinalen Werteskala gelangen. Dazu müssen vielmehr ganz bestimmte Eigenschaften (wie sie z.B. in den bekannten Neumann-Morgenstern-Axiomen enthalten sind) vorausgesetzt werden können. In allen anderen Fällen stellt die im obigen Text erwähnte „Generierung von mehr Information“ eine Selbsttäuschung dar.

Dieses Problem (nämlich die im IQWiG-Konzept vorauszusetzende kardinale Nutzenskala) entschärft sich nur in dem folgenden sehr speziellen Fall einer Dominanz: Es werden lediglich *zwei* Therapien X und A verglichen wobei entweder *X kostengünstiger und nützlicher als A ist (oder umgekehrt A kostengünstiger und nützlicher als X)*. In diesem (allerdings sehr trivialen) Fall kann auf den Einsatz einer Effizienzgrenze und der dazu erforderlichen kardinalen Nutzenfunktion verzichtet werden, es genügt eine lediglich ordinal skalierte Nutzenfunktion, um eine Entscheidung zwischen X und A zu treffen.

Zweites Problem:

Eine gewisse Abschwächung erfährt die strenge Annahme einer kardinalen Werteskala lediglich an einer Stelle und dort eher beiläufig: Man darf statt des Nutzens selbst auch die **Wahrscheinlichkeit eines Response** (Überschreitung eines Schwellenwertes) einführen und diese auf der Nutzenachse als Ersatz für eine kardinale Werteskala abtragen. Dann aber ergeben sich sofort andere offene Fragen, die den Ansatz einer Responsewahrscheinlichkeit stark relativieren: Wie soll ein oder „der“ Schwellenwert festgelegt werden? Sollen ein oder mehrere Schwellenwerte festgelegt werden? Genügt die einmalige Überschreitung eines Schwellenwertes? Wie sollen Rückfälle in die Bewertung einbezogen werden? Wie kann die Dauerhaftigkeit eines Response eingebracht werden? Gegenüber diesen (und anderen Fragen) müsste die „Überschreitungswahrscheinlichkeit“ als Werteskala robust sein! Außerdem stellt die Wahrscheinlichkeit eines Response (im Sinne einer kardinalen Werteskala) selbst ein Problem dar: Rechtfertigt eine doppelt so hohe Erfolgswahrscheinlichkeit auch wirklich doppelt so hohe Kosten? Könnte stattdessen (wie in der Literatur vorgeschlagen wird) nicht der Logarithmus oder eine andere konkave Funktion (die dem „Gesetz“ des fallenden Grenznutzens entspricht), gewählt werden? Diese Fragen werden in dem IQWiG-Methoden-Papier nicht behandelt. Dieser Umstand zeigt auch, wie wenig Bedeutung dem Ansatz einer Responsewahrscheinlichkeit seitens des IQWiG beigemessen wird. Grundlage des IQWiG-Konzeptes ist eben die Annahme einer „harten“ eindimensionalen kardinalen Werteskala.

Drittes Problem:

Selbst wenn sich der Nutzen oder Wert einer Therapie kardinal messen lässt, kann es sich bei seiner Bestimmung nur um den **Mittelwert des Nutzens** (in einer Population) handeln. Hier sollte als zweite Zielgröße wenigstens noch die **Streuung** und damit das **Risiko** eingeführt werden. Geschieht dies, so müsste aus Mittelwert und Streuung eine „künstliche“ Größe als „Nutzen“ eingeführt und auf der Nutzenachse abgetragen werden. Diese werden zwar unter dem bekannten Namen „Effektgrößen“ in nahezu allen klinischen Studien gemessen.

Sie dienen dort aber nur einer vergleichenden Feststellung der Wirksamkeit und nicht des „absoluten“ Nutzens einer Therapie und taugen daher nicht für eine Kosten-Nutzen-Analyse. Als Ausweg bleibt eben nur der von IQWiG vorgeschlagene Weg, die betreffenden Streuungen erst gar nicht zu berücksichtigen. Ich halte das für problematisch. Noch größere Probleme treten auf, wenn der Nutzen oder Wert einer Therapie nicht auf einer einzigen, sondern realistischerweise auf mehreren Skalen gemessen wird; eine damit gegebene **mehrdimensionale Nutzenskala** ist im IQWiG-Konzeptes nicht vorgesehen.

Viertes Problem:

Das IQWiG Konzept geht davon aus, dass sich der Nutzen oder Wert einer Therapie auf einer „allgemeingültigen“ (kardinalen) Skala messen lässt, die zwar von der behandelten Krankheit abhängen darf, nicht aber von der eingesetzten Therapie. Ist dieses Postulat wirklich realistisch? Hängt nicht gerade in dem großen Bereich psychischer oder psychiatrischer Erkrankungen die Nutzenbetrachtung auch von der eingesetzten Therapie ab? Kann man z.B. einen Psychiater dazu „zwingen“, die OPD-Skala zu verwenden oder einen Psychoanalytiker, sich (allein) auf die BDI-Skala zu verlassen? Ich meine das keineswegs „schulenspezifisch“, sondern entscheidungstheoretisch: **Wenn für die verschiedenen Handlungsalternativen unterschiedliche Nutzenskalen vorausgesetzt werden müssen, gerät man in ein Dilemma.**

Fünftes Problem:

Nach dem IQWiG Konzept bezieht sich der Nutzen einer Therapie und die betreffende Skala stets nur auf eine bestimmte Krankheit (Diagnose). **Die gesamt Kosten-Nutzen-Analyse ist streng diagnosespezifisch.** Diese Spezifizierung ist problematisch: Wie kann eine Therapie beurteilt werden, die bei verschiedenen Krankheiten (zwar mit unterschiedlichen Nutzen, aber doch sehr gut) wirkt? Und umgekehrt: Wie wird eine Therapie beurteilt, die bei einer bestimmten Krankheit einen sehr großen (maximalen) Nutzen unter allen anderen Therapien hat, aber nur auf diese eine Diagnose „zugeschnitten“ und anderenfalls untauglich ist? (Gemeint sind hier nicht „Nebenwirkungen“; diese werden im IQWiG System entweder in der Nutzenbewertung oder unter den entstehenden Kosten erfasst.) Das Dilemma zwischen **reduktionistischem oder holistischem** Therapieansatz wird im IQWiG-Konzept klar zu Gunsten des reduktionistischen Ansatzes entschieden.

Sechstes Problem:

Die Kosten-Nutzen-Analyse des IQWiG-Konzeptes bezieht sich (fast) nicht auf das gesamte Leben eines „Patienten“: Auf der Nutzenseite wird eine gewonnene Lebenszufriedenheit, insbesondere eine erhöhte **Resilienz** nicht in Rechnung gestellt, auf der Kostenseite bleiben die **kumulativen Gesamtkosten** (unter Einbeziehung mehrerer Krankheiten) unberücksichtigt. Damit hängt auch zusammen, dass im IQWiG-Konzept eine **Salutogenese** („was macht einen Menschen fähig, gesund zu bleiben?“) als Ansatz gar nicht erst einbezogen werden kann. Ähnlich wie im fünften Problem schon betont wird ein streng diagnose- oder störungsspezifischer Ansatz verfolgt und der Wert einer Therapie auf die spezielle Störung und nicht auf die Lebenszeit (oder auch nur Versicherungszeit) eines Patienten insgesamt bezogen.

Zusammenfassung:

Das IQWiG-Konzept zur Kosten-Nutzen-Analyse von Therapien überträgt ein bekanntes entscheidungstheoretisches, und vor allem im betriebswirtschaftlichen Bereich bewährtes Entscheidungskonzept auf die vergleichende Beurteilung medizinischer Behandlungen ohne kritisch zu hinterfragen, ob die dazu erforderlichen strengen Voraussetzungen erfüllt sind oder erfüllt werden können. Aus diesem Versäumnis ergeben sich die oben angeführten Probleme. Diese betreffen grundsätzlich den gesamten Bereich medizinischer oder psychotherapeutischer Behandlungen, allerdings nicht in gleichem Ausmaß für verschiedene Teilbereiche: Im Vergleich mit medikamentösen oder rein somatischen Therapien können für psychotherapeutische Behandlungen, vor allem für tiefenpsychologisch fundierte oder psychoanalytische Therapien, die erforderlichen Voraussetzungen viel schwerer, oft nur unter relativ willkürlichen Annahmen oder sogar gar nicht erfüllt werden. Dadurch ergeben sich deutliche Benachteiligungen (Willkürlichkeiten, Unsymmetrien bis hin zu subjektiven Einstufungen) von vergleichenden Bewertungen sowohl von Psychotherapien untereinander als auch (und hier erst recht) zwischen Psychotherapien und rein medikamentösen oder somatischen Behandlungen und deren oft „härter“ evaluierbaren Therapien. Diese Benachteiligungen (bis hin zu Verzerrungen) sind durch die oben angeführten Probleme klar und deutlich belegt. Es ist zu befürchten, dass bei einer Verwirklichung des IQWiG-Konzeptes zur Kosten-Nutzen-Analyse Therapien mit einer reduktionistischen, störungsspezifischen und deshalb (scheinbar) harten (sprich kardinalen) Erfolgsskala dominieren. **Streng genommen ergibt sich die Folgerung, dass die IQWiG-Methodik in der vorliegenden Form nicht auf psychotherapeutische Behandlungen und deren Vergleich mit anderen Behandlungsarten anwendbar ist.**