

Gäfgen, Gérard

**Working Paper**

Effektivität und Nettonutzen medizinischer Innovationen: Eine  
Besinnung auf die Grundlagen der Bewertung und eine Anwendung  
auf diagnostische Verfahren

Diskussionsbeiträge - Serie I, No. 235

**Provided in Cooperation with:**

Department of Economics, University of Konstanz

*Suggested Citation:* Gäfgen, Gérard (1987) : Effektivität und Nettonutzen medizinischer Innovationen: Eine Besinnung auf die Grundlagen der Bewertung und eine Anwendung auf diagnostische Verfahren, Diskussionsbeiträge - Serie I, No. 235, Universität Konstanz, Fakultät für Wirtschaftswissenschaften und Statistik, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/68936>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

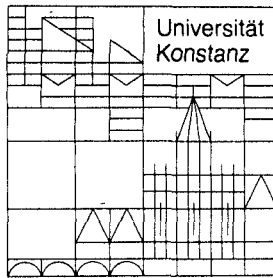
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**Fakultät für  
Wirtschaftswissenschaften  
und Statistik**

---

Gérard Gäfgen

**Effektivität und Nettonutzen  
medizinischer Innovation:**

Eine Besinnung  
auf die Grundlagen der Bewertung  
und eine Anwendung  
auf diagnostische Verfahren

Diskussionsbeiträge

---

Effektivität und Nettonutzen medizinischer Innovationen:

Eine Besinnung auf die Grundlagen der Bewertung  
und eine Anwendung auf diagnostische Verfahren.

Gérard Gäfgen

Serie I - Nr. 235

Dezember 1987

Ag 562/88  
Weltwirtschaft  
Kiel  
12/87

## 1. Vorbemerkungen zu den Einsatzbedingungen der Evaluierungsverfahren.

Nutzen-Kosten-Analysen im weitesten Sinne sind in den letzten Jahren vielfach zur Beurteilung von gesundheitspolitischen Programmen sowie von neuen medizinischen Verfahren oder Hilfsmitteln herangezogen worden. Dies geschah in Form von vollständigen Effizienzbeurteilungen mittels Untersuchungen des Verhältnisses ausgewählter nicht monetärer Effekte zu monetären Kosten, und schließlich monetär ausgedrückter Cost-Benefit-Relationen sowie Punktwertmethoden der Nutzwertanalyse. Dabei kam es zunehmend zu einer Verfahrensroutine, welche sich der Grundprobleme eines solchen Vorgehens nicht mehr recht bewußt ist, andererseits aber auch zu Verfeinerungen der Verfahren und neuen Anregungen. Beides legt eine Neubesinnung auf einige Grundsatzfragen nahe, die sich heute besser klären lassen als zu Beginn der Anwendung der Bewertungsverfahren. Das soll hier im Hinblick auf die Beurteilung von medizinischen Neuerungen - also einen spezifischen und besonders wichtigen Anwendungsbereich - geschehen, wobei kein Anspruch auf eine vollständige Klärung aller Methodenfragen erhoben werden kann.

Der Anwendungsbereich medizinischer Neuerung umfaßt neue Verfahren oder Hilfsmittel in Therapie und Diagnostik, welche den Anspruch erheben, echte Verbesserungen im Vergleich zur bisherigen medizinischen Praxis zu bieten. Im therapeutischen Bereich heißt das - vorläufig definiert -

- (1) bessere medizinische Ergebnisse (größere Effektivität), soweit nicht eine dadurch hervorgerufene Verteuerung im Vergleich stärker ins Gewicht fällt;
- (2) Verbilligung angewandter Verfahren durch: Einsparungen von "inputs"; Ersetzung durch neuartige preisgünstigere inputs oder durch nunmehr billiger herzustellende und zu beziehende

schon bekannte inputs, wiederum soweit dies nicht mit einer stärker zu gewichtenden Verschlechterung der medizinischen Ergebnisse verbunden ist:

Eine Verbesserung im diagnostischen Bereich liegt entsprechend vor,

- (1) wenn die Qualität des diagnostischen Urteils (Präzision und Sicherheit) erhöht wird, ohne daß die damit eventuell verbundene Kostensteigerung stärker zu veranschlagen wäre. Wiederum in Analogie zur therapeutischen Neuerung kann zweitens von einer diagnostischen Verbesserung gesprochen werden,
- (2) wenn das diagnostische Vorgehen durch Einsatz neuer Hilfsmittel oder Materialien verbilligt wird, ohne daß die Qualität des diagnostischen Urteils wesentlich verschlechtert wird; hierher zählt auch die Einsparung von nicht so leicht erfaßbaren Zeitkosten durch schnellere Verfügbarkeit der diagnostischen Befunde.

Die Feststellung, ob vorgeschlagene oder realisierte Neuerungen auch Verbesserungen im soeben definierten Sinne darstellen, ist Sinn jeder Evaluierung. Sie dient der Entscheidung darüber, ob Neuerungen erst entwickelt, schon entwickelte angewandt, und bereits angewandte beibehalten werden sollen. Demnach finden Kosten-Nutzen-Analysen im Zuge des Gesamtprozesses der Entstehung und Verwendung medizinisch-technischer Innovationen zu ganz verschiedenen Zeitpunkten statt:

"ex ante" zur Steuerung von Forschungs- und Entwicklungsaktivitäten oder zur Planung des Einsatzes einer Neuerung; "ex post" zur Kontrolle des Erfolges ihres Einsatzes. Ein Dilemma jeglicher Evaluierung besteht darin, daß sie gerade zu jedem Zeitpunkt besonders dringlich ist, zu dem die Daten über Kosten und Wirkungen noch am unsichersten sind; und daß man umgekehrt über die sichersten Daten erst verfügt, wenn die Neuerung schon in voller Breite praktiziert wird. Die ex ante bestehende Unsicherheit der einer Bewertung zugrunde liegenden Daten oder Annahmen erstreckt sich bei der Planung einer erst zu entwickelnden Neuerung zusätzlich noch auf: den vermuteten Forschungsaufwand, die vermutete Zeit der Entwicklung bis zur Anwendungsreife und die Wahrscheinlichkeit, überhaupt ein brauchbares Produkt oder Verfahren entwickeln zu können. Die Setzung von Forschungsprioritäten kann sich also nicht nur am erhofften gesellschaftlichen Nettonutzen eines Forschungsergebnisses orientieren, zumal die

Wahrscheinlichkeit brauchbarer Forschungsergebnisse nicht durch einen erhöhten Forschungsaufwand pro Periode beliebig gesteigert werden kann. Die gesellschaftliche Bedeutung eines Medikaments gegen eine verbreitete Krankheit, welche hohe gesellschaftliche Kosten verursacht, mag z.B. noch so groß sein; sind die wissenschaftlichen Grundlagen für die Entwicklung eines solchen Medikaments nur mühsam und zeitraubend zu verbessern, so mag trotz dessen sozialer Bedeutung ein hoher in kurzer Frist einzusetzender Forschungsaufwand nicht sinnvoll sein<sup>1)</sup>. Dieses Beispiel eines verbreiteten Trugschlusses der Forschungspolitik diene als Hinweis auf den frühesten und damit unsichersten Anwendungszeitpunkt einer Kosten-Nutzen-Analyse. Im folgenden wird unter Ausklammerung der speziellen Bewertungsfragen, die nur bei Forschungsprojekten auftreten, von einer bereits anwendungsreifen oder gar von einer bereits partiell erprobten Neuerung ausgegangen, um nur allgemeinere Fragen j e g l i c h e r Bewertung von Neuerungen anzusprechen.

Die Bewertungsprobleme stellen sich aber nicht nur anders zu jedem Zeitpunkt des Innovationsprozesses, sie unterscheiden sich auch nach dem Adressaten, dessen Entscheidungen durch das Bewertungsergebnis erleichtert werden sollen: Was vom Standpunkt eines Betriebes, z.B. eines Krankenhauses, die Wirtschaftlichkeit seiner Aktivität erhöht, kann für einen Fiskus, der etwa dem Betrieb Aufträge erteilt, z.B. für eine Krankenkasse, zu finanziellen Einbußen führen; und was für einen Fiskus eine finanzielle Einsparung bedeutet, kann gesamtgesellschaftlich eine Einbuße an Effizienz darstellen. Wir beschränken uns im folgenden auf Urteile vom sogenannten gesamtgesellschaftlichen Standpunkt, für welche bekanntlich Vorschläge im Rahmen der gesamtwirtschaftlich orientierten Kosten-Nutzen-Analyse entwickelt worden sind (vgl. unten Abschn. 4.) Damit stoßen wir auf die Schwierigkeit, daß es - empirisch gesehen - keinen Adressaten gibt, der wirklich im Interesse der Gesamtgesellschaft zu handeln bestrebt ist. Ist also eine Neuerung gesamtgesellschaftlich positiv zu bewerten und soll diese Bewertung zum Leitfaden auch praktischer Entscheidungen werden, so muß immer zusätzlich überlegt werden, wie der potentielle Anwender der Neuerung dazu gebracht werden kann, sie auch tatsächlich in bestimmtem Umfang einzuführen. Die Einführung einer neuartigen ärztlichen Leistung mag die Gesamtausgaben der Krankenkassen oder des staatlichen Gesundheits-

dienstes erhöhen; sie mag unter bisherigen Bedingungen anstelle besser honorierter Leistungen treten und so das Einkommen des einzelnen Arztes schmälern. Ist sie nach den genannten Bewertungsgrundlagen positiv zu beurteilen, so müssen eben die Anreizstrukturen einschließlich gesetzlicher Zwänge so geändert werden, daß die Neuerung in der Praxis Anwendung findet. Wegen dieses Adressatenproblems ist demnach eine bewertende Analyse einer Neuerung immer zu ergänzen durch das, was wir uns angewöhnt haben, als politische Ökonomie der Anwendung eines Gestaltungsvorschlages zu bezeichnen. Es läßt sich methodisch allerdings nicht regeln, wie weit man dabei gehen sollte, ob man z.B. nicht nur die Notwendigkeit einer bestimmten Änderung der ärztlichen Gebührenordnung betont, sondern auch einen vielleicht zu erwartenden Widerstand von Instanzen oder Verbänden mit einbezieht.

## 2. Die Nutzung erfahrungswissenschaftlicher Erkenntnisse.

Im Rahmen der bislang skizzierten Einsatzbedingungen besteht die Evaluierung in der Bewertung der bei Anwendung einer Neuerung entstehenden Kosten auf der einen und der Erträge auf der anderen Seite. Bevor jedoch überhaupt Bewertungen vorgenommen werden können, muß bekannt sein, die Aufwendung welcher Kosten welche Erträge nach sich zieht, also welcher "output" an gesundheitlichen Erträgen erzielt werden kann, wenn bestimmte "inputs" getätigt werden. Input-output-Beziehungen sind aber Darstellungen *t e c h n i s c h e r* Möglichkeiten und damit entweder natur- oder sozialwissenschaftliche empirisch bewährte Aussagen über die maximal erzielbaren Ertragsfolgen eines bestimmten Aufwands oder über die zur Erzielung bestimmter Erträge mindestens erforderlichen Aufwendungen (Mindestkosten). Die Rede ist hier zunächst nur von heterogenen physischen Einheiten, z.B. dem input an geopferter Freizeit oder an verbrauchten Materialien, dem output an gesundheitlichen Ergebnissen, wie größere Schmerzfreiheit und mehr Tage der Arbeitsfähigkeit pro Jahr. Die Frage ist, wie man hier zu verlässlichen medizintechnischen und/oder sozio-technischen Aussagen über die Bestimmungsgründe der Kostenseite und die gesundheitlichen Effekte (zusammengefaßt "Effektivität") eines neuen medizinischen Vorgehens gelangt.

Die Beziehung zwischen den verschiedenen Kostenkomponenten einerseits und den verschiedenen Dimensionen von Effektivität andererseits hängt von den Besonderheiten des behandelten Falles ab; z.B. ist der lebensverlängernde Effekt einer Therapie vom Lebensalter des Patienten abhängig. Für ein Gesamturteil sind daher auch Erkenntnisse über die Verbreitung der Fälle nach Schwere der Erkrankung, Alter des Patienten usw. in der Gesamtbevölkerung erforderlich, also Inzidenz- bzw. Prävalenzdaten. Bei sehr breit anwendbaren Neuerungen beziehen sich diese nicht auf spezifische Krankheiten, sondern auf ganze medizinische Gebiete oder gar auf behandlungsbedürftige Krankheitsfälle überhaupt. Zwar lassen sich letztlich vielleicht Kosten und Erträge auf den durchschnittlichen Einzelfall beziehen, durchschnittliche input-output-Relationen ergeben sich aber streng genommen erst aus Mittelwerten der potentiellen Anwendungsfälle oder einer repräsentativen Stichprobe von Fällen. Auf jeden Fall erlaubt aber erst die Erfassung der Kosten und Erträge aller Anwendungsfälle zusammengenommen ein vollständiges evaluatives Urteil. Nur hilfsweise mag man vereinfachend auf eine geschätzte Durchschnittstechnik, auf Durchschnittskosten und Durchschnittserträge abstellen, wenn die Struktur des Anwendungsfeldes einer Neuerung noch wenig bekannt ist; beispielsweise, wenn man eine hohe Dunkelziffer von nicht erfaßten behandelbaren Fällen vermutet.

Die ideale medizin-technische Information würde aus einem Therapieversuch an einer Zufallsstichprobe unter Vergleich mit einer ebensolchen Kontrollgruppe hervorgehen. Es ist bekannt, wie selten sich solche Idealbedingungen realisieren lassen und daß es vielfach sogar ethisch nicht zu vertreten ist, den Nutzen einer zu erprobenden Neuerung einer Kontrollgruppe von Patienten vorzuenthalten. Unter Idealbedingungen würde man auch Informationen darüber erhalten, wie häufig und in welchem Ausmaß bestimmte Wirkungen einer Behandlung bei bestimmten Falltypen auftreten und ob bestimmte Effekte auch ohne Behandlung eingetreten wären. Eine therapeutische Neuerung wird zumeist weder bei allen Fällen zu einem vollen oder doch einem gleichen Heilungserfolg führen, noch wird dieser Erfolg mit absoluter Sicherheit, sondern nur mit Wahrscheinlichkeit auftreten; und schließlich gibt es in manchen Anwendungsfeldern "Spontanheilungen", die ohne Behandlung zustande kommen. Im letzteren Falle ist der Gesamt-



nutzen, welcher der neuen Behandlungsmethode zuzuschreiben ist, nicht identisch mit dem Nutzen aus allen "geheilten" Fällen. Geringere Erfolgsgrade bei bestimmten Prozentsätzen von Fällen müssen auch letztlich als geringerer Nutzen veranschlagt werden, und die Häufigkeit von Mißerfolgen muß ebenfalls Berücksichtigung finden. In der Regel kann man also nicht von der Vereinfachung ausgehen, daß die Neuerung auf sämtliche Fälle angewandt wird und bei allen den gleichen vollen Effekt erzielt. In manchen Anwendungsbereichen stellt die Anwendung der Neuerung bei jedem Krankheitsfall einen langwierigen Prozeß von z.B. mehrjähriger Dauer dar, während dessen der Patient verschiedene Stadien mit unterschiedlichem Gesundheitszustand durchläuft. Damit erhöht sich die Dimensionalität der erzielten Effekte noch um die Zeitdimension. Treten auch hierbei erzielte Zustandsänderungen nur mit gewisser Wahrscheinlichkeit auf, so stellt sich der Heilungsvorgang als stochastischer Prozeß dar.

Die sozialen Vorgänge bei der Anwendung einer Neuerung sind mit diesen medizin-technischen Erwägungen noch nicht berücksichtigt. Selbst wenn das potentielle Anwendungsfeld ohne Dunkelziffer erfaßt ist, so hängt die effektive Anwendung doch vom Verhalten der Anwender ab. Dies sind außer den behandelnden Ärzten auch die Patienten - vor allem dann, wenn eine Therapie deren aktive und womöglich sorgfältige Mitwirkung verlangt. Solange eine Neuerung nicht von allen Ärzten aufgegriffen und in allen ihren Indikationsgebieten angewandt wird und solange nicht mit vollständiger "compliance" aller Patienten gerechnet werden kann, wird eine Nutzen-Kosten-Bewertung, die dies nicht beachtet, nur einen potentiellen Kosten- und Nutzenstrom und nicht den faktisch zu erwartenden erfassen. Dieses Anwendungspotential wird, selbst wenn es schließlich ausgeschöpft wird, bei keiner Innovation von Anfang an voll erschlossen. Typisch ist vielmehr ein Diffusionsprozeß der allmählichen Verbreitung von Informationen und von Anwendungsfähigkeiten sowie steigender Anwendungswilligkeit. Das bedeutet einen im Zeitablauf variierenden Strom von Kosten und Nutzen, z.B. zunächst in den Anfangsperioden hohen Fallkosten der Anwendung durch geringen Grad der Ausnutzung eines neuen Geräts. Rechnet man hier nur mit den schließlich erreichten Endzuständen voller Nutzung, so unterschlägt man die durch schlechte anfängliche Nutzung entstehenden Einführungskosten. Geht man aber von zusätzlichen Maßnahmen

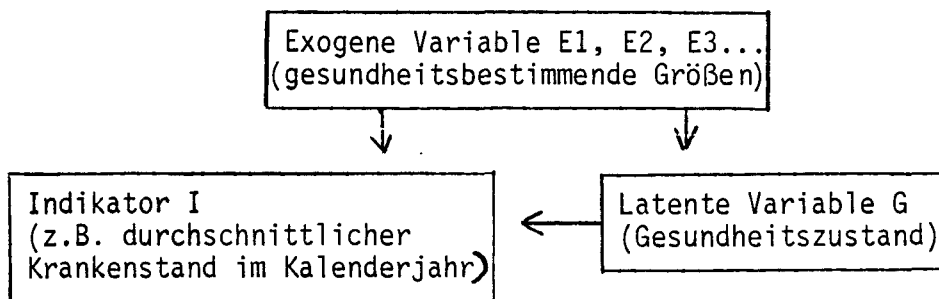
aus, welche die Einführung beschleunigen sollen, so ist zu beachten, daß auch diese wiederum Kosten verursachen. Zu bewerten ist also nicht die Anwendung einer Neuerung, sondern ein ganzes Programm zu ihrer Einführung, ihrer Verbreitung und schließlich dauerhaften Anwendung.

Es geht aber um mehr als um diese Einführungsphasen, nämlich schlechthin um die Nachfrage nach der neuen Substanz, oder dem neuen Gerät. Diese hängt nämlich nicht nur von der medizinischen Indikation ab, sondern auch von der qualitativen Ausgestaltung des Produkts, von seinem Preis, von Einkommen und Versicherungsgrad der Patienten, der versicherungsrechtlichen Erstattungsfähigkeit und der Honorierung der zu erbringenden komplementären ärztlichen Leistungen. Fälschlicherweise unterstellen die Kosten-Nutzen-Analysen hier zumeist, es würde alles so ablaufen, als ob die Neuerung einfach in den Leistungskatalog eines die gesamte Bevölkerung umfassenden staatlichen Gesundheitsdienstes aufgenommen würde. Typischerweise wurde die Cost-Benefit-Analysis gerade für derartige Entscheidungen einer staatlichen Instanz entwickelt, bedarf aber für Gesundheitssysteme, die nicht vom Typ des britischen NHS sind, der genannten Ergänzungen. (Im übrigen haben sich auch in Großbritannien oft fiskalische Erwägungen der Behörden gegenüber der gesamtgesellschaftlichen Perspektive der Cost-Benefit-Analyse durchgesetzt). In jedem Fall sind politische und soziale Einstellungen gegenüber Krankheiten und Behandlungsmethoden und ihre Finanzierbarkeit von Bedeutung, z.B. existiert oft eine hohe Bereitschaft, für die Bekämpfung seltener, aber erschreckender und in der Öffentlichkeit bekannter Krankheiten Opfer zu bringen, auch wenn eine Nutzen-Kosten-Analyse dies nicht rechtfertigen sollte.

### 3. Weiterführende Urteile über die Effektivität einer Neuerung.

Gehen wir vorläufig davon aus, daß die Kostenseite leicht zu definieren, zu erfassen und in einem einheitlichen Ausdruck zusammenzufassen sei. Ein evaluierendes Urteil ist dann nur möglich, wenn auch die verschiedenen positiv zu bewertenden Wirkungen, die eine neue Therapie zeitigt, in einem einheitlichen Maß der medizinischen Effektivität zusammenzufassen sind. Der begrifflichen Intention nach geht es um ein Maß für die

erzielte Verbesserung der "Volksgesundheit", also für die durchschnittlich im Einzelfall erzielte gesundheitliche Verbesserung und die Zahl der gebesserten Fälle. Mit den Methoden der empirischen Wissenschaft läßt sich eine solche gedachte Maßgröße nur definieren als eine sogenannte "latente" Variable, deren Niveau von den Werten abhängt, welche exogene sie determinierende Variablen annehmen, und die ihrerseits Wirkungen auf andere Variable ausübt, die somit ihre Veränderung "anzeigen" ("Indikatorvariable"):



Bei einer solchen Formulierung lassen sich unter bestimmten einschränkenden Annahmen die Parameter der zwischen den genannten Größen E1, E2, E3..., G und I bestehenden als linear unterstellten Beziehungen ökonometrisch schätzen; G wird dann zu einer bis auf eine lineare Transformation bestimmten Maßgröße, die sich auf einer kardinalen Skala abbilden läßt und einen empirisch begründeten Gesundheitsindex darstellt<sup>2)</sup>. Das System linearer Gleichungen beruht dabei auf durch Schätzung präzisierten Hypothesen über die Beziehungen zwischen den beteiligten Größen (Strukturmodell).

Verfügt man nicht über gut begründbare Hypothesen, die sich in einem Strukturmodell abbilden lassen, so bleibt nur übrig, von der empirischen Methodik zu intuitiven Urteilen und begrifflichen Konventionen überzugehen. Man mißt dann die Effektivität

- durch einen als besonders geeignet angesehenen einzelnen Indikator des Gesundheitszustandes, oder
- durch eine gewichtete Kombination von Indikatoren ("Index"), wobei zusätzlich zur Auswahl der Größen auch ihre Gewichtung einem intuitiven Expertenurteil entstammt.

Da eine Indikatorvariable in der Regel nicht nur vom Gesundheitszustand, sondern auch von Größen beeinflusst wird, und da diese nur teilweise mit "Gesundheit" korrelieren, läßt sich eine vollkommen valide Messung von Gesundheit durch einen Indikator nicht erreichen. Die Höhe des Indikators "krank verbrachte Tage im Jahr pro Versicherter" hängt beispielsweise u.a. auch davon ab, in welchem Ausmaß Lohnfortzahlung im Krankheitsfall gewährt wird, wie weit die Beschäftigung bei häufiger Abwesenheit vom Arbeitsplatz als gefährdet angesehen wird u.ä.m..

Kombiniert man mehrere Indikatoren zu einem Index  $G' = G'(I_1, I_2, I_3, \dots)^3$ , im einfachsten Falle eine Linearkombination  $G' = \sum_i a_i I_i$ ,  $i = 1, 2, 3, \dots$ , so hängt dessen Eignung nicht nur von den Eigenschaften der einzelnen Indikatoren  $I_i$  ab, sondern zusätzlich noch von den ihnen beigegebenen Gewichten, im linearen Beispiel also den Parametern  $a_i$ ,  $\sum_i a_i = 1$ . Solche Gewichte lassen sich aus dem Durchschnitt von Expertenvorschlägen bilden oder indirekt aus der relativen Bewertung von Kombinationen aus den Werten verschiedener Indikatoren durch Experten ableiten. Einen Mittelweg stellt ein Teilindex dar, der eine in sich komplexe Dimension von Gesundheit als gewichtete Summe ihrer Teilaspekte abbildet, z.B. die Summe qualitätsgewichteter Lebensjahre.

Diese Urteile zur Festlegung der Verfahren zur Effektivitätsmessung beruhen zum Teil auf objektivierbaren fachwissenschaftlichen Grundlagen, auch wenn die objektive Überprüfung mit Schwierigkeiten verbunden ist. Ein Beispiel ist die Zusammenstellung der spezifischen Behinderungen, die mit einer bestimmten Krankheit verbunden sind, und die Erhebung von deren Bedeutung für das Wohlergehen des Patienten. Sobald jedoch konventionelle Festsetzungen der Bedeutung von Variablen stattfinden, wird zugleich die Grenze zwischen der objektivierbaren Messung der Wirkung einer neuen medizinischen Maßnahme und deren gesundheitspolitischer Bewertung überschritten. Damit stellen Effektivitätsmessungen in mancher Hinsicht bereits einen Übergang von der Nutzung der erfahrungswissenschaftlichen Informationen über die input-output-Zusammenhänge zur eigentlichen Bewertung der Aufwands- und Ertragskomponenten dar. Vermeidbar sind diese Probleme nur in solchen Fällen, wo alle Effekte in Richtung einer Verbesserung der Gesundheit weisen.

#### 4. Die Begründung und Ableitung von Bewertungskriterien.

Am wenigsten Bewertungsfragen werfen wir auf, wenn wir uns darauf beschränken, die durchschnittliche Effektivität des Einsatzes einer aufgewandten Kosteneinheit bei Anwendung der Neuerung zu ermitteln: Kosteneffektivität = Gesamteffektivität: Kostensumme. Dennoch enthält die Zusammenfassung aller einzelnen inputs zu einem einheitlichen Kostenausdruck in Geldeinheiten konstanter Kaufkraft in nuce schon alle Probleme der Nutzenbewertung, da Kosten entgangenen Nutzen in der nächstwichtigen Verwendung darstellen (Opportunitätskosten) und da umgekehrt manche Nutzenelemente aus vermiedenen Kosten bestehen. Benennen wir beispielhaft einige Kostenkategorien: direkte Behandlungskosten; Einkommensausfälle der Patienten während der Behandlung; körperliche und psychische Unannehmlichkeiten der Behandlung; zusätzliche Kosten von Krankheiten, die durch Lebensverlängerung infolge Heilung auftreten; Einkommenseinbußen für hochspezialisierte Anbieter, deren Leistungen durch die eingeführte Neuerung oder durch deren Erfolge überflüssig werden. Selbst die direkten Behandlungskosten sind eigentlich nur dann durch Bewertung der eingesetzten Ressourcen (Faktoren) mit deren Marktpreisen ermittelbar, wenn die Marktpreise die Opportunitätskosten des Faktoreinsatzes richtig widerspiegeln. Es gibt Fälle, wo dies nicht der Fall ist, z.B. die Faktorentlohnung deutlich von dem Preis abweicht, zu dem Leistungsbereitschaft des Faktors besteht. Glücklicherweise ändert sich durch Anwendung einer medizinischen Neuerung die Nachfrage und damit der Faktorpreis zumeist nur unwesentlich, so daß kein zusätzlicher Unterschied zwischen zu zahlender Entlohnung und der zur Herstellung von Leistungsbereitschaft nötigen Mindestentlohnung (die sich an den Opportunitätskosten der Faktorbereitstellung orientiert) entsteht. Wäre dies der Fall, so entstünde nämlich eine zusätzliche Faktorrente, die analog der noch zu benennenden Konsumentenrente auf der Nutzenseite zu behandeln wäre.

Sind dagegen physische Verbrauchsmengen von Faktorleistungen und Hilfsmitteln leicht in Geld zu bewerten und vernachlässigt man geldlich schwer feststellbare Kosten, wie etwa Kompensationen für die Unannehmlichkeiten einer Behandlung für den Patienten, so ergibt das Verhältnis von Effektivität  $E$  zu Gesamtkosten  $K$  ein Bewertungsmaß, das einen Vergleich mit existierenden oder anderen neuen medizinischen Verfahren ermöglicht: Bei gleicher

oder größerer Effektivität ist das Verfahren mit den geringeren Kosten vorzuziehen, bei gleichen oder geringeren Kosten das Verfahren mit der höheren Effektivität. Ist nur das *V e r h ä l t n i s*  $E:K$  günstiger, sind jedoch Effektivität und Kosten beide größer oder beide kleiner als bei anderen Lösungen für das gleiche medizinische Problem, so erlaubt der Maßstab keine so eindeutigen Urteile über die relative Vorzugswürdigkeit der Einführung einer Neuerung. Er ermöglicht es allerdings, bei einem feststehenden Budget für sämtliche Gesundheitsleistungen in einem Behandlungsbereich alle diejenigen Projekte zu realisieren, die zusammengenommen die Kosteneffektivität des Gesamtbudgets maximieren - eine nur selten praktikable Anwendung.

Ein *a b s o l u t e s* Urteil darüber, ob eine Neuerung einzuführen ist oder nicht, setzt voraus, daß die Ertragsseite (Nutzen  $N$ ) in den gleichen Einheiten bewertet wird wie die Aufwandsseite (Kosten  $K$ ), also in der Regel in Geldeinheiten: Nutzen-Kosten-Analyse (NKA) im engeren Sinne. Es läßt sich dann nämlich feststellen, ob ein positiver Nettonutzen als Differenz von Nutzenstrom und Kostenstrom  $N - K$  existiert. (Auf die Vergleichbarmachung von zu unterschiedlichen Zeitpunkten anfallenden Nutzen- und Kostengrößen durch Diskontierung auf den gleichen Zeitpunkt und die dabei zu unterstellende Konvention über die gesellschaftliche Zeitpräferenz sei hier nur hingewiesen). Es spielt dann auch keine Rolle, ob z.B. Einsparungen an Krankheitskosten durch vermiedene Krankheiten als Minderkosten bei  $K$  abgezogen oder bei  $N$  hinzugefügt werden. Bei Benutzung einer Verhältnisgröße, also  $N:K$ , wäre das nämlich von Belang, so daß eine solche Formulierung als Bewertungsmaßstab besser nicht benutzt werden sollte. (Damit wird einer gelegentlich geäußerten Kritik an der Uneindeutigkeit der NKA Rechnung getragen).

Die einzige aus sozialetischen Grundpostulaten ableitbare Meßanweisung für monetär formulierten Nutzen ist die Ermittlung der Summe, welche ein Individuum für die Anwendung der Neuerung zu zahlen bereit ist, oder - bei negativem Nutzen - der Summe, die ihm gezahlt werden muß, um es auf gleichem Nutzenniveau zu halten wie bisher. Nach Einführung der Neuerung besteht dann ein Zustand, bei dem diejenigen, die bessergestellt sind als bisher, diejenigen, die schlechtergestellt sind, durch Kompensationszahlungen entschädigen könnten, ohne schlechtergestellt zu sein als bisher: Die Neuerung ist dann eine potentielle paretianische Verbesserung des Wohlergehens der

Bevölkerung<sup>4)</sup>. Die Zahlungsbereitschaft enthält einen Geldpreis für alle Vorteile, die ein Individuum aus der Anwendung einer medizinischen Neuerung ziehen könnte: Als Kranker könnte es sich besser fühlen oder finanziell geringer belastet sein, als Erwerbstätiger wäre es arbeitsfähiger, als Angehöriger und Freund sähe es Mitmenschen weniger leiden, als krankheitsbedrohter Mensch sieht es sich einem geringeren Risiko von Vorkommen, Dauer oder Schwere einer Erkrankung gegenüber. Die Zahlungsbereitschaft übersteigt auch die tatsächlich zu zahlenden Preise für Gesundheitsleistungen, da nur der sich am Markt bildende Preis zu zahlen ist; die entstehende Differenz ist die Konsumentenrente.

Diese an sich allgemein bekannte Grundlage wird hier deshalb betont, weil in praktischen Anweisungen zur NKA gelegentlich die Ermittlung der Zahlungsbereitschaft als nur e i n e Methode der Nutzenfeststellung unter mehreren anderen erwähnt wird. Alle praktizierten Nutzenerfassungsmethoden sind aber nur Näherungs- oder Ersatzverfahren für die in idealer Weise nicht durchführbare Ermittlung der Zahlungsbereitschaft. Es gibt Befragungsmethoden und theoretisch konzipierte Steuerzahlungen als Entgelt für eine Entscheidungsteilnahme, welche die Subjekte veranlassen sollen, annähernd ihre wahre Zahlungsbereitschaft mitzuteilen. Man kann versuchen, Nachfrage und Angebot und deren Preisabhängigkeit am Markt zu ermitteln, um die Höhe von Konsumentenrenten (und gegebenenfalls Faktorrenten) festzustellen; jedoch auch hier gibt es gerade im Gesundheitswesen Verzerrungen; z.B. wird die Zahlungsbereitschaft der Nachfrager zu hoch geschätzt, wenn diese versichert sind und daher den zu zahlenden Preis ersetzt bekommen<sup>5)</sup>. Für den Nutzen aus verminderten Mortalitäts- oder Morbiditätsrisiken ist ein Anhaltspunkt die Zusatzentlohnung, die in Berufen geboten wird, die mit solchen Risiken belastet sind. Es geht hingegen nicht an, lediglich andere geldwerte Vorteile wiederhergestellter Gesundheit einfach mit der Wahrscheinlichkeit des Heilungserfolges zu gewichten, da auch die negative Einstellung der Menschen zum Risiko in ihre Nutzenvorstellungen eingeht.

Eine wichtige Komponente monetären Nutzens ist die Verbesserung der Erwerbsfähigkeit gemessen an besseren Einkommensmöglichkeiten. Diskontiert auf die Gegenwart wird sie als Verbesserung des Humankapitals bezeichnet. Unter der Annahme, daß das Erwerbseinkommen den Beitrag eines Individuums

zum Sozialprodukt mißt, stellt sich als Ergebnis einer medizinischen Maßnahme eine Erhöhung des Sozialprodukts ein. Daher wird manchmal die Erwünschtheit einer Humankapitalerhöhung auch aus dem Ziel eines möglichst hohen Sozialprodukts abgeleitet<sup>6)</sup>; die Maximierung des Sozialprodukts läßt sich jedoch auf keinerlei akzeptiertes sozialetisches Axiom zurückführen. Deshalb ist die Methode der Humankapitalermittlung nur ein Verfahren zur hilfsweisen Ermittlung e i n e r wichtigen Komponente des Nutzens als stellvertretend für die übrigen Komponenten und daher eventuell als eine niedrigstmögliche Schätzung der Zahlungsbereitschaft.

Bei anderer Auffassung würde ja ein gerettetes Lebensjahr dann geringer bewertet, wenn ein Betroffener einer Schicht mit niedrigerem Einkommen angehört oder gar überhaupt kein Einkommen aus Produktionsleistung bezieht. Ohnehin würde selbst eine vollständige Erfassung der Zahlungsbereitschaft diejenigen individuellen Nutzen geringer gewichten, die nicht einfach aus einer geringeren Beachtung von Gesundheitsschäden resultieren, sondern aus einer niedrigeren Zahlungs f ä h i g k e i t des Individuums. Dadurch kommt der Bekämpfung von Krankheiten, die vorwiegend Schichten mit niedrigerem Einkommen betreffen, auch ein geringeres Gewicht zu. Man kann aber nicht unterstellen, die interpersonelle Einkommensverteilung - selbst n a c h Steuern und Transfers als Ausgleich - wäre ethisch voll gerechtfertigt. Es ist daher erforderlich, neben dem monetären gesamtwirtschaftlichen Nettutzen einer medizinischen Neuerung als eine weitere Komponente, die in der Evaluierung berücksichtigt werden sollte, die induzierten Veränderungen in der Verteilung des Volkseinkommens und im Gesundheitszustand verschiedener Schichten hinzuzufügen. Eine andere Möglichkeit bestünde darin, die Zahlungsbereitschaftsurteile verschiedener Schichten mit korrigierenden Gewichten zu versehen<sup>7)</sup>. Das würde aber sowohl ein verteilungspolitisches Urteil wie auch eine Manipulierung der individuellen Präferenzen durch den Analytiker implizieren - und damit mehr Willkür in der Methodik. Um ein extremes Beispiel zu erwähnen: Man könnte ein gleiches Gewicht aller individuellen Urteile über Zahlungsbereitschaft anstreben, womit Gleichheit in Gesundheitsversorgung und Einkommenserzielung als gewünschtes Ideal unterstellt wäre.

Neben der Ermittlung des monetären Nutzens durch die NKA und die Beschreibung der Verteilungsfolgen können noch weitere Dimensionen für die Evaluierung



eine Rolle spielen, die nicht in den individuellen Präferenzen für medizinische Maßnahmen schon zum Ausdruck kommen oder, praktisch bedeutsamer, welche durch die unvollständigen Hilfsmethoden der monetären Nutzenermittlung z.B. die Humankapitalmethode nicht erfaßt wurden, obwohl die Bürger sie allgemein als wichtig ansehen. Dann müssen die einschlägigen in der NKA nicht erfaßten Wirkungen zusätzlich beschrieben werden, z.B. der "intrinsische" Wert eines neuen Verfahrens für die medizinische Wissenschaft. Je mehr solche "intangibles" aber als Beschreibungsdimensionen in das Ergebnis der Evaluierung eingehen, umso weniger eindeutig fällt das Ergebnis aus: Der monetäre Nutzen dient ohnehin nur der Berücksichtigung des gesamtwirtschaftlichen Aspektes einer Neuerung, er wird aber nun sozusagen abgewertet zu einem von vielen Merkmalen, die für die Beurteilung relevant sein können. Auch ergibt sich nun wieder das Bedürfnis, die vielen Dimensionen zu einem systematischen Gesamturteil zusammenzufassen. Das kann logischerweise wiederum nur dadurch geschehen, daß man die unterschiedlichen Attribute als Teilindikatoren auffaßt, denen bestimmte politische Bedeutungsgewichte zuzumessen sind, um zu einem Index des übergeordneten gesellschaftlichen Gesamtnutzens oder -wertes zu gelangen. Als lineare Form eines solchen gesellschaftlichen Nutzenindex ist die mit konstanten Gewichten arbeitende N u t z w e r t a n a l y s e entwickelt worden. Auf die relative Willkür und die Politikanmaßung, die mit derartigen Festsetzungen verbunden sind, wurde bereits bei der Erwähnung des Gesundheitsindex als Effektivitätsmaß hingewiesen. Es sei hier nur zusätzlich vermerkt, daß die Folgen der Festlegung konstanter Gewichte von Bewertungsattributen bei Anwendung auf die Evaluierung von Neuerungen auch für denjenigen kaum übersehbar sind, der die Festsetzungen vornimmt. Beispielsweise kann bei einer linearen gesellschaftlichen Nutzenfunktion die Bewertung bereits bei kleinsten Änderungen der Eigenschaften alternativer Projekte in ihr Gegenteil umschlagen; es kann so leicht zur Befürwortung von Projekten mit sehr einseitigen Vorzügen und vielen kleinen Nachteilen kommen. Die Analyseergebnisse sind hierdurch besonders leicht manipulierbar.

#### 5. Besonderheiten bei der Bewertung diagnostischer Neuerungen.

Die bisherigen Überlegungen zur Feststellung besserer Effektivität und erhöhten Nettonutzens, welchen eine medizinische Innovation mit sich bringt, bezogen sich implizit auf therapeutische Neuerungen. Das hat seinen guten Grund; denn diagnostische Verbesserungen bestehen nur darin, daß sie die Informationsgrundlagen für die Anwendung von Therapien verbessern. Ihre Bewertung setzt also die der möglichen Therapien schon voraus. Knüpfen wir an die eingangs (Abschn. 1.) gegebene vorläufige Definition einer diagnostischen Verbesserung an, so haben wir Kosteneinsparungen und Erhöhungen der Qualität des diagnostischen Urteils zu berücksichtigen.

Zunächst mag also ein neues Diagnoseverfahren mit `g e r i n g e r e n` `K o s t e n` verbunden sein. Soweit es die direkten Kosten des Ressourceneinsatzes betrifft, kann die Einführung eines neuen Verfahrens kostspieligere ältere Verfahren ersetzen. Hierbei ist auch daran zu denken, daß es sonst oft nötige weitere diagnostische Schritte überflüssig machen kann; man muß also stets vergleichen mit der gesamten bisherigen diagnostischen Prozedur. Dadurch werden auch Kostenelemente beachtet wie etwa die Hospitalisierung zu diagnostischen Zwecken, die vielleicht abgekürzt oder vermieden wird, und die damit für den Patienten verbundenen weiteren Kosten wie etwa Einkommenseinbußen. Eine erste Schwierigkeit, die hier auftaucht, besteht darin, daß der Anwendungsbereich des neuen Verfahrens nur mit richtigem gezieltem Einsatz der Untersuchungsmethoden definiert ist und daß er ein Gebiet abdeckt, welches vielleicht einer ganzen Reihe bislang je nach Verdachtsdiagnose verschiedener bisheriger Prozeduren entspricht. Dabei kann es Teilbereiche geben, in denen der Nettonutzen des neuen Verfahrens geringer als der älterer Vorgehensweisen ist; da dies erst nach Evaluierung festgestellt werden kann, muß der vorläufig unterstellte Einsatzbereich vielleicht schrittweise korrigiert werden. Zu den mit dem diagnostischen Vorgehen verbundenen Kosten zählen auch die für den Patienten entstehenden Unannehmlichkeiten und Gefahren; ein neues Verfahren kann dem Patienten nicht nur Zeit ersparen, sondern auch Schmerzen und Angst (auch beim Warten auf das Ergebnis) verringern und das Risiko von Folgewirkungen diagnostischer Eingriffe, z.B. invasiver Verfahren, herabsetzen oder vermindern. All diese Fragen sind z.B. aufgetaucht

bei der Evaluierung der Einführung der computer-tomographischen Verfahren. Diese bieten zugleich ein gutes Beispiel für die Abhängigkeit der fallbezogenen Diagnosekosten (einschließlich der Zugangskosten für den Patienten) von der erwähnten Breite des Anwendungsfeldes und der Zahl sowie der Verteilung der Großgeräte - was beides zusammen deren Ausnutzungsgrad bestimmt<sup>8)</sup>.

Die Nutzenseite bezieht sich im wesentlichen zunächst auf eine verbesserte Präzision der Diagnose: Eine bessere Spezifität oder Sensitivität des diagnostischen Verfahrens vermindert die Zahl der falsch positiven oder falsch negativen Diagnosen und erhöht damit die Wahrscheinlichkeit, daß auch die geeignetste Therapie zur Anwendung gelangt bzw. mindert die Wahrscheinlichkeit, daß falsch therapiert wird: Der erwartete therapeutische Nutzen steigt, die erwarteten therapeutischen Kosten (aus Fehlbehandlung) sinken. Die Steigerung des Nettonutzens durch Einführung des neuen Diagnoseverfahrens läßt sich nur mit Hilfe von statistischen Erwartungswerten des Nutzens<sup>9)</sup> ausdrücken und erfordert die Analyse der zweistufigen Entscheidungen aus diagnostischem Urteil und Entscheidung für eine Therapie. Eine genaue Evaluierung ist daher nur möglich mit Hilfe der statistischen (bayesianischen) Entscheidungstheorie<sup>10)</sup>. Dies läßt sich zeigen an dem als Anhang beigefügten vereinfachten Beispiel, das bestimmte Eigenschaften eines diagnostischen Tests und der verfügbaren Therapietechniken (Effektivitätsermittlung, vgl. auch oben Abschn. 2.) unterstellt. Diese Daten werden oft schwer feststellbar sein, da das Beispiel eines einzelnen Tests nicht berücksichtigt, daß dieser Test neben vielen anderen Erhebungen und Erwägungen in das diagnostische Vorgehen eingebettet ist, so daß seine Wirkungen auf die Veränderung des erwarteten Therapienutzens kaum isolierbar sind. Unser Beispiel ist daher auch nur als Erläuterung eines Bewertungsprinzips zu verstehen. Schon ein Komplex aus mehreren systematisierbaren diagnostischen Schritten (Stufendiagnose) führt rasch zu einer starken Komplikation der Bewertungsschritte. Leicht zu handhaben ist hier nur der Fall der sequentiellen Wiederholung des gleichen Tests<sup>11)</sup> und der Vergleich des dadurch steigenden prognostischen Werts des Vorgehens mit den ebenfalls steigenden Diagnosekosten. (Zum Begriff des prognostischen Wertes vgl. das angefügte Beispiel im Anhang). Die Betrachtung des erwarteten therapeutischen Nutzens bedarf zudem der Ergänzung durch die

unmittelbaren Wirkungen verbesserter Diagnose auf den Patienten, der geringerer Ungewißheit ausgesetzt ist, für den also vor allem negative Testergebnisse glaubwürdiger werden und eher eine beruhigende Wirkung ausüben. Die skizzierten wahrscheinlichkeitstheoretischen Erwägungen werden in praxi oft nur approximiert werden können durch Erfassung der besseren Anpassung der Therapien an die Besonderheiten des Falles: Die genauere Indikationsstellung zwischen mehreren verfügbaren Therapien vermindert die Häufigkeiten, mit denen weniger geeignete oder unnötige Maßnahmen zur Anwendung kommen. Sind solche Häufigkeiten aus der nachträglichen Analyse bisheriger Erfahrungen vorhanden, so kann der Fachmann Vermutungen über ihre Änderung durch präzisere Diagnosemethoden anstellen und diese mit Hilfe der laufenden Erfahrungen (kontrollierende Evaluierung) überprüfen. Man schließt so direkt von der neuen Vorgehensweise auf die Erwartungswerte für therapeutische Effekte (und damit letztlich für therapeutischen Nutzen), praktisch also indem man praktizierende Ärzte zur Bedeutung der neuen Diagnosetechnik befragt.

Ein Vorgehen der zuletzt genannten Art wird vor allem da angebracht sein, wo das neue Verfahren dazu führt, daß die Testergebnisse *s c h n e l l e r* vorliegen. *R a s c h e r e V e r f ü g b a r k e i t* kann in Form neuer Verfahren auftreten, welche überhaupt erst ermöglichen, eine Krankheit frühzeitig zu erkennen (Früherkennungs- im Unterschied zu Schnell-diagnose-Verfahren); der Vergleich bezieht sich dann auf eine bisherige Diagnosepraxis, welche in ihrer ersten Phase sozusagen im Warten auf erste Manifestationen der noch gar nicht vermuteten Erkrankung bestand.

Bei Krankheiten, deren frühzeitige Behandlung zu wesentlich besseren Ergebnissen führt, ist der therapeutische Nutzen einer Frühdiagnose beträchtlich; der Nettonutzen hängt aber insgesamt davon ab, mit welcher Häufigkeit solche Fälle entdeckt werden, welche Gesamtheit von Patienten damit untersucht werden und wie teuer der einzelne Test ist. Diese Bewertungsfragen sind im Zusammenhang mit der Bewertung von screening-Programmen vielfach diskutiert worden<sup>12)</sup> und brauchen daher hier nur angedeutet zu werden.

Wird hingegen bei schon bestehendem Krankheitsverdacht das Diagnoseverfahren beschleunigt (z.B. "Sofortdiagnostik"), so treten andere Vorteile auf: Es entfallen die bei langwierigen diagnostischen Abklärungen vorsorglich

auf Verdacht eingeleiteten Therapien, soweit sie bei Vorliegen der endgültigen Diagnose aufgegeben werden. Damit entfallen auch die damit für den Patienten verbundenen Unannehmlichkeiten ebenso wie die Periode der Ungewißheit bis zum Erhalt der Testergebnisse. Voraussetzung ist, daß die Präzision der Diagnose zumindest die gleiche ist wie bei den langwierigeren alten Verfahren. Die vermiedenen Kosten (Nutzeneinbußen) vorübergehender Falschtherapien in einem bestimmten Prozentsatz der Fälle werden umso mehr ins Gewicht fallen, je mehr in diesen Fällen auch der endgültige therapeutische Erfolg schlechter ausgefallen wäre, als er bei richtiger Soforttherapie hätte sein können. Im übrigen sind die hier benutzten Urteile "falsch" und "richtig" vereinfachende Annahmen, da sie implizieren, daß die Diagnoseprozedur keine statistischen Fehler 1. oder 2. Art (Falschpositive oder -negative) enthält.

Diese Fiktion wird hier nicht nur aus praktischen Gründen fehlender Daten sinnvoll sein, sondern auch deshalb, weil hier Verfahren verglichen werden, die sich nur in der Schnelligkeit der Durchführung, nicht im Präzisionsgrad, unterscheiden.

Der Zeitbedarf für die Durchführung diagnostischer Tests wird verringert, wenn auf die Kostenvorteile sequentieller Schritte in der Diagnose (Stufendiagnostik) verzichtet wird, indem man auch solche Parameter sogleich erhebt, die eventuell unnötig sind, weil sie nur bei bestimmten Werten zuerst erhobener Parameter relevante differentialdiagnostische Informationen darstellen. Die damit erreichten Vorteile entsprechen der Art nach denen eines schneller verfügbaren einzelnen Testergebnisses. Dieses Vorgehen mittels zahlreicher simultaner Tests kann gegenüber der Stufendiagnostik an Vorteilen einbüßen, wenn das Ergebnis des Einzeltests rascher verfügbar wird.

Zu beachten ist allerdings, daß bei solchen Vergleichen auch die **e c h t e n** direkten und Zeitkosten - neben dem erwähnten therapeutischen Nutzen - verglichen werden. Ist etwa die bisherige Praxis der Diagnose nicht an diesen Kosten und Nutzen orientiert, so kann man auch einen bewertenden Vergleich der herrschenden Praxis mit einem neuen Verfahren mit rascher verfügbaren Testergebnissen nicht vornehmen, da in der aktuellen Praxis der potentiell beste Nettonutzen infolge von Fehlanreizen (z.B. aus der Gebührenordnung für Laborleistungen) gar nicht realisiert ist (vgl. hierzu oben Abschn. 1.).

Es zeigt sich, daß bei sachgerechter Interpretation die Evaluation im diagnostischen Bereich zwar besondere Anwendungsbedingungen aufweist, aber die allgemeinen Methodenfragen, welche durch die NKA aufgeworfen werden, stets die gleichen sind. In diesen Ausführungen sollte nur auf einige davon hingewiesen werden, welche, obwohl längst bekannt, in der Praxis der Evaluation medizinischer Neuerungen allzu leicht in Vergessenheit geraten.

ANHANG

Ermittlung des Nutzens eines diagnostischen Verfahrens  
(vereinfachtes Beispiel verbesserter Präzision der Diagnose)

I. Daten-Grundlagen

A. Eigenschaften des diagnostischen Tests

Test- ergeb- nis \ Krankheits- zustand	gutartig $k_1$	bösartig $k_2$
eher gutartig $e_1$	Sensitivität $w(e_1/k_1)$	$w(e_1/k_2)$
eher böartig $e_2$	$w(e_2/k_1)$	Spezifizität $w(e_2/k_2)$
relative Häufigkeit des Vorkommens von $k_1$ und $k_2$	$h(k_1)$	$h(k_2)$

Mit den Wahrscheinlichkeiten des Vorliegens von  $k_1$  oder  $k_2$  ohne Vornahme des Tests:  $h(k_i)$ ,  $i = 1,2$ , und den bedingten Wahrscheinlichkeiten der Testergebnisse  $e_j$  bei Vorliegen von  $k_i$ :  $w(e_j/k_i)$ ,  $j = 1,2$ , sind auch definiert; Der prognostische Wert des Verfahrens:  $[w(e_1/k_1)] : [1-w(e_2/k_2)]$  sowie die bedingten Wahrscheinlichkeiten des Vorliegens von  $k_i$  beim jeweiligen Testergebnis  $e_j$ :  $w(k_i/e_j)$  und die Wahrscheinlichkeiten, mit denen das Auftreten des jeweiligen Testergebnisses  $e_j$  überhaupt erwartet werden kann:  $w(e_j)$ .

- B. Eigenschaften des therapeutischen Verfahrens  
 (Effektivität  $E_{mi}$ ,  $m = 1, 2$  optimaler Therapien bei Vorliegen von  $k_1$  und  $k_2$ )

Therapien \ Krankheitszustände	$k_1$	$k_2$
beste Therapie für $k_1$ : $t_1$	sehr gut $E_{11}$	sehr schlecht $E_{12}$
beste Therapie für $k_2$ : $t_2$	schlecht $E_{21}$	gut $E_{22}$

## II. Evaluierungsschritte

- A. Ermittlung des Nutzens  $N_{mi}$  der Ergebnisse  $E_{mi}$ :  
 Feststellung von  $N(E_{11})$ ,  $N(E_{12})$ ,  $N(E_{21})$ ,  $N(E_{22})$ .
- B. Aus den Nutzengrößen  $N_{mi}$  und den Wahrscheinlichkeiten  $w(N_{mi}/e)$  läßt sich die Therapie mit dem günstigsten Erwartungswert des Nutzens  $EN$  bei Vorliegen eines bestimmten Testergebnisses  $e_j$  bestimmen; z.B. für  $e_1$ :  

$$\text{Max} \{ (EN(t_1/e_1), EN(t_2/e_1)) \}$$
- C. Jedem Testergebnis  $e_j$  ist so ein bester Erwartungswert  $E_0 N(t_i)/e_j$  zugeordnet (therapeutische Entscheidungsregel für die Auswertung von Testergebnissen). Da die Wahrscheinlichkeiten  $w(e_j)$ , mit denen ein bestimmtes  $e_j$  überhaupt erwartet werden kann, bekannt sind, läßt sich so der Vornahme des Tests ein bestimmter Erwartungswert des Nutzens aus den bestmöglichen Therapien zumessen, nämlich:  

$$\sum_j w(e_j)(E_0 N/e_j).$$
- D. Dieser erwartete Nutzen des neuen Verfahrens ist dessen Kosten gegenüberzustellen und das Nettoergebnis mit dem eines Vorgehens ohne diesen Testeinsatz, also nur mit bisherigen Diagnoseverfahren, zu vergleichen.



## ZUSAMMENFASSUNG

Medikamentöse bzw. therapeutische Verbesserungen, definiert als Nettoergebnis aus Veränderungen der Effektivität und der Kosten der Therapie, sowie diagnostische Verbesserungen, ihrerseits definiert als ein solches Nettoergebnis aus kostengünstigerer, präziserer oder rascher verfügbarer Diagnose, werfen verschiedene Bewertungsprobleme auf, je nachdem, zu welchem Zeitpunkt im langwierigen Prozeß des medizin-technischen Fortschritts sie eingeführt werden und an welche Entscheidungsträger sich die bewertende Analyse richten soll.

Voraussetzung jeder Evaluation ist die Nutzung der erfahrungswissenschaftlichen Erkenntnisse der Medizin und der Sozialwissenschaften, da nur so die technischen Beziehungen zwischen Aufwendungen und Erträgen im diagnostisch-therapeutischen Prozeß der Behandlung einer Krankheit erfaßt werden können.

Nur auf dieser Grundlage ist ein Urteil über die Effektivität eines bestimmten medizinischen Vorgehens und damit auch über die Anwendung einer medizinischen Neuerung möglich. Sobald die Wirkungen (Effekte) eines solchen Vorgehens zu einem Gesamtmaß der Effektivität zusammengefaßt werden sollen, müssen jedoch zusätzliche wertende Urteile zur Aufstellung eines Ergebnisindex oder eines Ergebnisindikators herangezogen werden.

Sollen Aufwands- und Ertragsseite voll vergleichbar gemacht werden (Nutzen-Kosten-Analyse NKA im engeren Sinne), so wird damit zwar ein absolutes Urteil über die Vorteilhaftigkeit einer vorgeschlagenen Neuerung ermöglicht; jedoch müssen alle Kostenkomponenten im Sinne von Opportunitätskosten voll beachtet werden. Streng genommen sind Kostengrößen nur mit den gleichen Bewertungsprämissen ermittelbar, wie sie bei Nutzengrößen erforderlich werden. Letztlich beruhen alle bewertenden Urteile "vom gesamtgesellschaftlichen Standpunkt" auf dem Grundpostulat der Vorzugswürdigkeit einer potentiellen paretianischen Verbesserung der Wohlfahrt aller Individuen zusammengenommen. Dabei vernachlässigte weitere gesellschaftspolitische Aspekte, insbesondere Wirkungen neuer Maßnahmen auf

die Verteilung von Einkommen und Gesundheitszuständen in der Bevölkerung, können nur als zusätzliche Bewertungsdimensionen mitbeschrieben werden. Ihre Zusammenfassung in sogenannten Nutzwertanalysen ist nur relativ willkürlich möglich und mit Manipulationsgefahren verbunden.

Die Bewertung diagnostischer Neuerungen erfordert keine anderen Methoden als die in der NKA entwickelten. Die Erfassung von Kosten oder deren Einsparung im diagnostischen Bereich ist nur etwas subtiler zu interpretieren. Ferner ist der Nettonutzen einer besseren Präzision der Diagnose eigentlich nur mit Hilfe der statistischen Entscheidungslehre (ihrerseits ein Zweig der Nutzentheorie) als Erwartungsgröße erfaßbar; praktisch wird man sich aber nur auf fachmännische Urteile über veränderte Häufigkeiten von Therapien abstützen. Die Vorteile von Früherkennungsverfahren und Schnelldiagnosen stellen keine zusätzlichen Bewertungsprobleme dar, wirken aber vor allem auf bestimmte Kostenelemente ein, z.B. auf die mit vorübergehenden Fehltherapien verbundenen Kosten. Insgesamt treten immer nur neue Varianten der gleichen grundsätzlichen Evaluierungsfragen auf.

Anmerkungen (Fußnoten)

- 1) Zu den zusätzlichen Bewertungsdimensionen, die sich gegenüber bekannten Neuerungen bei Forschungsvorhaben ergeben, vgl. etwa GÄFGEN (1984), S. 177 ff.
- 2) Zur Ermittlung einer solchen Maßgröße als latente Variable vgl. die Benutzung des sogen. MIMIC-Index bei LEU, DOPPMANN, KELLER und DEUTSCHMANN (1986).
- 3) Eine Übersicht über die Methoden solcher Indexbildung findet sich bei CHEN und BUSH (1979).
- 4) Diese Grundlagen werden besonders deutlich herausgearbeitet bei MISHAN (1975), S. 382 ff.
- 5) Art und Umfang dieser Verzerrung der Nutzenschätzung weist nach: P. ZWEIFEL (1984).
- 6) Diese Fehlinterpretation findet sich z.B. bei BAPST (1986), S. 15.
- 7) Zu diesen Verteilungsgesichtspunkten und ihrer Berücksichtigung siehe etwa DRUMMOND (1980), S. 113 ff.
- 8) Zu diesem Beispiel und den genannten Anwendungs- und Bewertungsfragen siehe WAGNER (1983).
- 9) Zur Bewertung nur wahrscheinlicher und sequentieller Nutzengrößen siehe etwa den Beitrag von FISHER (1979).
- 10) Eine allgemeine Einführung in diese medizinischen Methoden findet sich bei PATRICK (1979).
- 11) Die Kosten- und Effektivitätsänderungen, die durch Testwiederholung bewirkt werden, sind dargestellt bei MOONEY, RUSSELL und WEIR (1980), S. 92 ff.
- 12) Vgl. den Überblick bei DRUMMOND (1980), S. 87 ff.

## Literaturverzeichnis

- BAPST, L. (1986), Die mehrdimensionale Kosten-Nutzen-Analyse als Evaluationsinstrument im Gesundheitswesen, in: HORISBERGER und VAN EIMEREN (1986), S. 1-50.
- CHEN, Milton M. und BUSH, James W. (1979), Health Status Measures, Policy, and Biomedical Research, in: MUSHKIN und DUNLOP (1979), S. 15-42.
- CULYER, A. J. und HORISBERGER, B. (Hrsg.) (1983), Economic and Medical Evaluation of Health Care Technologies, Springer-Verlag, Berlin u. a.
- DRUMMOND, M. F. (1980), Principles of Economic Appraisal in Health Care, Oxford University Press, Oxford u.a.
- FISHER, Gregory W. (1979), Willingness to Pay for Probabilistic Improvements in Functional Health Status: A Psychological Perspective, in: MUSHKIN und DUNLOP (Hrsg.) (1979), S. 167-202.
- GÄFGEN, G. (1984), Zur Beurteilung medizinischer Fortschritte und Forschungen: Notwendigkeit, Ansatzpunkte und Probleme normativer Aussagen, in: MÜNNICH, F. E. und OETTLE, K. (Hrsg.), Ökonomie des technischen Fortschritts in der Medizin, Beiträge zur Gesundheitsökonomie, Band 6, Bleicher Verlag, Gerlingen, S. 146-188.
- HORISBERGER, B. und VAN EIMEREN, W. (Hrsg.) (1986), Die Kosten-Nutzen-Analyse. Methodik und Anwendung am Beispiel von Medikamenten, Springer Verlag Berlin u.a.
- LEU, R. E., DOPPMANN, R., KELLER, T. und DEUTSCHMANN, R. (1986), Die quantitative Erfassung von Gesundheitszustand und Lebensqualität, illustriert am Beispiel von Psoriasis-kranken, in: HORISBERGER und VAN EIMEREN (Hrsg.) (1986), S. 153-248.
- MISHAN, E. J. (1975), Cost-Benefit-Analysis, 2. Auflage, George Allen & Unwin Ltd., London.
- MOONEY, Gavon H., RUSSELL, Elizabeth M., und WEIR, Roy D. (1980), Choices for Health Care, The Macmillan Press Ltd., London und Basingstoke.
- MUSHKIN, Selma J. und DUNLOP, David W. (Hrsg.) (1979), Health: What Is It Worth? Measures of Health Benefit, Pergamon Press, New York u.a.

PATRICK, Edward A. (1979), Decision Analysis in Medicine: Methods and Applications, CRC Press, Boca Raton (Florida).

WAGNER, J. L. (1983), Macroeconomic Evaluation of Computerized Tomographic Scanning, in: CULYER und HORISBERGER (Hrsg.) (1983), S. 321-334.

ZWEIFEL, P. (1984), Technological Change in Health Care: Why are Opinions so Divided ?, Managerial and Decision Economics, Vol. 5, No. 3, S. 177-182.