

Technischer Anhang Unsicherheit

Version 1.0

09.10.2008

Kontakt:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Dillenburger Straße 27

D-51105 Köln

Tel.: +49 (0)221 / 35685-0

Fax: +49 (0)221 / 35685-1

E-mail: knb-methoden@iqwig.de

Inhaltsverzeichnis

Inhaltsverzeichnis.....	i
1 Einleitung	1
2 Eine kurze Übersicht über die Literatur zur Effizienzgrenze und die empirischen Entwicklungen.....	4
3 Das Thema der Effizienzmessung und ihr Grad an Unsicherheit	7
3.1 Umgang mit Erwartungstreue von Schätzungen.....	7
3.2 Konstruktion von Konfidenzintervallen.....	9
4 Effizienzgrenzen mit kleinen Datensätzen basierend auf Sekundärdaten.....	11
4.1 Konstruktion der Effizienzgrenze	11
4.2 Umgang mit Unsicherheit	14
Literaturverzeichnis.....	22

1 Einleitung

Wie in dem Hauptdokument aufgeführt, wurde das Konzept der Effizienzgrenze „für die Bewertung von Gesundheitstechnologien durch das IQWiG [...] ausgewählt, da es die Anforderungen der deutschen Gesetzgebung erfüllt, gleichzeitig aber auch mit den grundlegenden internationalen wissenschaftlichen Standards konsistent ist. Insbesondere ermöglicht die Methode die Berücksichtigung einer rationalen Mittelverwendung innerhalb einer Indikation unter Berücksichtigung der vom IQWiG durchgeführten Nutzenbewertung und Vermeidung von Diskriminierungen“ [1].

In dem Kontext einer Effizienzgrenze wird jede Evaluation eines neuen Produktes unter der Annahme durchgeführt, dass alle zur Verfügung stehenden Technologien effizient sind. Ohne diese Annahme wäre es möglich, entweder den Mix der Technologien durch den Entscheidungsträger anzupassen bevor eine neue Technologie berücksichtigt wird, oder aufseiten der Hersteller die neue Technologie mit einer Alternative zu vergleichen, die sich nicht auf der Effizienzgrenze befindet. Folglich könnte die neue Technologie zur Erstattung akzeptiert werden ohne tatsächlich die Gesamteffizienz zu steigern.

Die Methode der Effizienzgrenze ist ein international akzeptiertes und empfohlenes Evaluationsinstrument. Sie informiert Entscheidungsträger über das Ausmaß der Effizienz einer bestimmten Technologie unter der Voraussetzung, dass das Entscheidungsmodell „unverzerrte Schätzer der zu erwartenden Kosten und Effekte sowie der Entscheidungsunsicherheit auf zeitgemäße Weise und im Rahmen von beschränkten Ressourcen bietet, welche von den Entscheidungsträgern bestimmt werden, die das Modell in Auftrag geben“ [2] (Anmerkung: deutsche Übersetzung des englischen Zitats). Erwartungstreue der Schätzer und der Grad ihrer Genauigkeit sind dabei die wichtigsten Aspekte, die zu berücksichtigen und zu prüfen sind, wenn eine Effizienzgrenze für Entscheidungsträger informativ sein soll. Leider wird unter dem Begriff „Unsicherheit“ recht häufig die Unsicherheit, die sich auf die Schätzung des wahren Wertes eines Parameters bezieht, mit der Unsicherheit in Bezug auf Konfidenzintervalle im Rahmen der Schätzung (welche statt der Reliabilität den Grad der Präzision der Schätzung angeben sollten) verwechselt. Tatsächlich können „erwartungstreue“ Schätzer mit breiten Konfidenzintervallen vorliegen, die daher nicht für die Entscheidungsfindung nützlich sind und „verzerrte“

Schätzer mit engen Konfidenzintervallen, die aus anderen Gründen ebenfalls nicht für die Entscheidungsfindung nützlich sind.

Entscheidungsträger sollten nicht nur unverzerrte Informationen aus den Punktschätzern der Effizienzgrade, sondern auch über deren Konfidenzintervalle erhalten. Konfidenzintervalle sind von Interesse, weil sie Auskunft über die Präzision des Schätzers geben. Wenn daher berichtet wird, „dass eine Firma zu 80 % effizient ist, würden wir gerne zwischen Fällen, bei denen unsere Unsicherheit gering ist (zum Beispiel, wenn wir relativ sicher sein können, dass der wahre Effizienzlevel zwischen 78 % und 82 % liegt) und Fällen mit großer Unsicherheit (zum Beispiel, wenn wir nur relativ sicher sein können, dass dieser zwischen 65 % und 95 % liegt) differenzieren können“ [3, S. 91] (Anmerkung: deutsche Übersetzung des englischen Zitats).

Es ist außerdem sehr wichtig, sich darüber bewusst zu sein, dass die Bewertung der Reliabilität von Punktschätzern, zum Beispiel das Verzerrungspotenzial bei der Schätzung einer Effizienzgrenze, den Grad der Unsicherheit in einer Weise reduziert, die sich von den Informationen aus einem Konfidenzintervall über einen Punktschätzer unterscheidet (also wie ein Konfidenzintervall um die Schätzungen zu bilden ist). In den folgenden Kapiteln liegt das Hauptinteresse in diesem technischen Anhang in dem ersten Thema, nämlich wie Informationen über den Grad an Verzerrungen bei den Schätzungen gewonnen werden; das zweite Thema ist in unserem Kontext weniger relevant. Vielmehr ist der Grad der Genauigkeit im Falle eines erwartungstreuen Schätzers sowie eines Schätzers mit effizienten und konsistenten Schätzungen einfach eine Frage der Stichprobengröße!

Wie im Folgenden gezeigt wird, gibt es verschiedene Wege, Effizienzgrenzen zu berechnen oder zu schätzen. Generell funktionieren all diese Methoden unter der Annahme, dass Forscher über eine große Anzahl an Beobachtungen verfügen, um die Modellparameter zu schätzen. Leider ist dies nicht immer der Fall, insbesondere bei der Aufgabe, die das IQWiG innehat. Wann immer die Zahl der verfügbaren Beobachtungen beschränkt ist (wir können uns Beispiele mit weniger als 10 Beobachtungen vorstellen), ist es schwierig, Formeln aus der Inferenzstatistik zur Gewinnung von Konfidenzintervallen einzusetzen. Aus diesem Grund wird im vorliegenden Anhang zunächst eine Übersicht über die Standardmethoden zur Schätzung oder Berechnung von Effizienzgrenzen mit großen Datensätzen erstellt.

Anschließend werden alternative Ansätze zur „Konstruktion“ von Effizienzgrenzen und zur Generierung von Unsicherheitsmaßen basierend auf kleinen Datensätzen untersucht.

Der folgende Teil des technischen Anhangs ist in drei Kapitel aufgeteilt. In Kapitel 2 wird eine kurze Übersicht über die Literatur zur Effizienzgrenze und zu relevanten empirischen Entwicklungen geboten. Kapitel 3 diskutiert umfassend das Thema „Messung“ und die Rolle der Unsicherheit bei Effizienzgrenzen bei großen Datensätzen. Demgegenüber konzentriert sich das Kapitel 4 auf die Ableitung der Effizienzgrenze und die Messung von Unsicherheit im Zusammenhang mit kleinen Datensätzen.

2 Eine kurze Übersicht über die Literatur zur Effizienzgrenze und die empirischen Entwicklungen

Für ein besseres Verständnis der Behandlung des Themas „Unsicherheit“ in unserem Kontext ist es sinnvoll, eine kurze Übersicht über die Literatur zu theoretischen und empirischen Aspekten der Effizienzgrenze zu geben.

Die Messung von Daten zur ökonomischen Effizienz geht zurück auf eine Originalarbeit von Farrell [4], der als erster Wissenschaftler den engen Zusammenhang zwischen Effizienz und der Anwendung von Grenzfunktionen erkannte und der eine Methode einführte, welche die Gesamteffizienz einer Produktionseinheit in technische und allokativen Komponenten zerlegt.

Seitdem ist die Literatur hierzu stark angewachsen. Ausgehend von dem initialen Konzept der Isoquante der Effizienz, entwickelt von Farrell [4], entstanden verschiedene Alternativen, die technologische Ausstattung eines Produzenten zu spezifizieren, z. B. Produktions-, Kosten-, Ertrags- oder Gewinnfunktionen. Abstandsfunktionen wurden ebenfalls als Maße für die technische und allokativen Effizienz entwickelt. Eine weitere wichtige Entwicklung war die Umwandlung von dem einfachen One-input-one-output-Modell, das von Farrell aufgestellt wurde, zu seiner Multi-input-multi-output-Generalisierung, deren mathematische Beschreibung von Charnes, Cooper und Rhodes [5] dargelegt wurde.

Während das theoretische Konzept der Definition der Funktion einer Effizienzgrenze zur Messung der aktuellen Leistungsfähigkeit von Produktionseinheiten seit der anfänglichen Arbeit von Farrell nahezu unverändert geblieben ist, haben sich auf der anderen Seite die „empirischen Forschungsaktivitäten“ in verschiedene Richtungen weiterentwickelt. Dies ging mit der Entwicklung unterschiedlicher Techniken zur Berechnung oder Schätzung von Effizienzgrenzen einher.

In einer umfassenden Übersichtsarbeit über die Literatur zur Effizienzgrenze bietet Murillo-Zamorano [6] eine sehr klare und vollständige Taxonomie der verschiedenen Arten zur Klassifikation dieser Techniken. Laut Autor bezieht sich eine erste wichtige Klassifikation auf den Unterschied zwischen parametrischen und nicht-parametrischen Methoden,

das heißt zwischen Techniken, bei denen die Form der Funktion der Effizienzgrenze a priori festgelegt ist, und solchen, bei denen die Form der Funktion nicht vorgegeben ist,

sondern auf empirischem Wege über Beobachtungen in Stichproben ermittelt wird. Der nicht-parametrische Ansatz hat sich traditionell der Data Envelopment Analysis (DEA) angeglichen, ein auf beobachtete Daten bezogenes mathematisches (Programmierungs-) Modell, welches sowohl einen Weg für die Konstruktion von Produktionsfunktionen als auch für die Kalkulation von Effizienz-Scores bezüglich dieser konstruierten Funktionen liefert. In Bezug auf parametrische Ansätze kann eine Unterteilung in deterministische und stochastische Modelle vorgenommen werden. Die erstgenannten werden auch „full frontier“-Modelle genannt. Sie umfassen alle Beobachtungen, indem sie den Abstand zwischen der beobachteten Produktion und der maximalen Produktion, definiert durch die Grenze und die verfügbare Technologie, als technische Ineffizienz identifizieren. Die deterministische Spezifizierung besteht daher in der Annahme, dass alle Abweichungen von der Effizienzgrenze auf den Agierenden zurückzuführen sind. Allerdings gibt es weitere Umstände außerhalb des Einflusses des Agierenden, welche ebenfalls eine suboptimale Leistung der Einheiten verursachen können. Ein wettbewerbsregelndes Umfeld, Wetter, Glück, sozio-ökonomische und demografische Faktoren, Unsicherheit usw. sollten nicht wirklich als Einheiten der technischen Effizienz betrachtet werden. Der deterministische Ansatz tut dies jedoch. Darüber hinaus wird aus Sicht deterministischer Techniken jedes Spezifikationsproblem als ineffizient betrachtet. Im Gegensatz dazu modellieren stochastische Verfahren mit Grenzansätzen sowohl Spezifikationsfehler als auch unkontrollierbare Faktoren unabhängig von der technischen Effizienzkomponente, indem ein doppelseitiger Zufallsfehler bei der Spezifikation des Grenzmodells einbezogen wird. Eine weitere Klassifikation von Grenzmodellen kann entsprechend den verwendeten Instrumenten zur Berechnung dieser vorgenommen werden, und zwar die Unterscheidung zwischen mathematischer Programmierung und ökonometrischen Ansätzen. Deterministische Grenzfunktionen können entweder durch die Anwendung mathematischer Programmierungen oder mittels ökonometrischer Techniken berechnet werden. Die stochastische Spezifikation kann nur mittels ökonometrischer Techniken geschätzt werden. In dem Großteil der Literatur, die sich mit der Messung ökonomischer Effizienz befasst, basieren die Analysen entweder auf einer der oben genannten parametrischen oder auf nicht-parametrischen Methoden. Die Wahl der Schätzmethode ist ein umstrittenes Thema, wobei einige Wissenschaftlicher den parametrischen (z. B. Berger, 1993) und andere den nicht-parametrischen (z. B. Seiford und Thrall, 1990) Ansatz präferieren. Der wesentliche Nachteil nicht-parametrischer Ansätze ist ihre deterministische Eigenschaft. Die Data Envelopment Analysis beispielsweise unterscheidet nicht zwischen technischer Ineffizienz und statistischen Störeinflüssen.

Demgegenüber benötigen parametrische Grenzfunktionen eine Definition der spezifischen Funktionsform für die Technologie und des Ineffizienz-Fehlerterms. Der Bedarf einer Funktionsform verursacht sowohl Spezifikations- als auch Schätzprobleme [6, S. 35, 36] (Anmerkung: deutsche Übersetzung des englischen Zitats).

Wie in den nächsten Kapiteln dargelegt wird, hängt der Umgang mit Unsicherheit bei der Schätzung von Effizienzgrenzen erheblich von der jeweiligen zur Anwendung kommenden Technik ab.

3 Das Thema der Effizienzmessung und ihr Grad an Unsicherheit

Wie oben gesehen, ist seit dem wegweisenden Artikel von Farrell [4] ein großes Angebot an Literatur entstanden, welche das Thema der Berechnung oder der Schätzung des Effizienzgrades und seines Präzisionsgrades behandelt.¹ Unser Ziel ist es, die zur Verfügung stehenden Techniken zum Umgang mit Erwartungstreue und Präzision von Schätzungen innerhalb der eingeschränkten Auswahl an Modellen von Effizienzgrenzen zu präsentieren, welche das IQWiG als Instrument für den Vergleich alternativer medizinischer Interventionen akzeptiert. Dem in dem Hauptdokument dargestellten Ansatz zur Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung folgend impliziert dies, dass nur Single-input-single-output-Grenzen und deren Techniken zur Schätzung herangezogen werden.

Aus empirischer Sicht benötigt die Berechnung von Effizienzgrenzen mittels mathematischer Programmierung oder ökonometrischer Techniken die Verfügbarkeit von Informationen zum Input und Output. Je größer die Menge an Informationen ist, desto besser wird das Ergebnis der Analyse sein.² Bei der Anwendung der ökonometrischen Methode müssen die standardstatistischen Eigenschaften eines optimalen Schätzers beachtet werden: a) Erwartungstreue, b) Effizienz und c) Konsistenz.

3.1 Umgang mit Erwartungstreue von Schätzungen

Bei der Berechnung oder Schätzung einer Effizienzgrenze können verschiedene Störelemente in den Prozess eingreifen; die resultierenden Effizienzgrade werden dadurch beeinträchtigt. Diese Störelemente können den Wert der Schätzungen beeinflussen und zu „verzerrten“ Schätzungen führen. Bei einem Modelling-Ansatz können die Störelemente durch 1) die Wahl

¹ Es ist nicht das Ziel dieses Beitrags eine Übersicht über alle entsprechenden Studien zu bieten. Unter den verschiedenen Arbeiten kann der interessierte Leser z. B. auf die von Murillo-Zamorano [6] oder Kim und Schmidt [3] und die Literatur, die dort angegeben ist, verwiesen werden.

² Den ökonometrischen Ansatz betreffend haben standard-ökonometrische Techniken wie Ordinary Least Squares (OLS) den Nachteil, dass „ein durchschnittliches Verhältnis“ statt „des optimalen Verhältnisses“ geschätzt wird. Die Anwendung eines Grenzfunktionsansatzes ist insofern attraktiver, als diese Parameter näher an dem theoretischen Konzept des optimalen Verhältnisses zwischen Input und Output bei einem Produktionsprozess sind.

des Modells, 2) fehlende Variablen und 3) Messfehler sowohl bei den abhängigen als auch bei den erklärenden Variablen entstehen.

Das Problem der Wahl des Modells bezieht sich auf die bestimmte Funktionsform, die der Modellierer als formale Approximation des tatsächlichen Verhältnisses zwischen abhängigen und unabhängigen Variablen festlegen muss. In dem speziellen Kontext der Effizienzgrenzen wurden enorme Fortschritte erzielt mit der Verwendung flexibler Funktionsformen, welche lineare Approximationen zweiter Ordnung der unbekanntenen wahren Funktionen sind. Begleiterscheinung solcher Funktionsformen ist die größere Zahl an Parametern, die zu schätzen sind. Allerdings können, solange die Zahl der Beobachtungen groß genug ist, nur Berechnungsprobleme deren Schätzung einschränken.

Das Problem fehlender Variablen bezieht sich auf die Unmöglichkeit für den Wissenschaftler, alle relevanten Informationen zu beachten, die zur Bestimmung des wahren Verhältnisses zwischen Input und Output nützlich sind. Was immer man beobachten kann, sollte als erklärende Variable in das Modell einfließen. Dies ist insbesondere dann richtig, wenn Daten auf Patientenebene behandelt werden, die aufgrund persönlicher Charakteristika eine hohe Heterogenität hinsichtlich der Ansprechrate auf eine Behandlung widerspiegeln. Was immer bezüglich Heterogenität durch den Wissenschaftler beobachtbar ist, sollte in das Modell einbezogen werden, um Abweichungen zu erklären.

Zum Beispiel kann das Risiko an Krebs zu erkranken von der Familienanamnese abhängen. Grundsätzlich kann dies beobachtet werden und anschließende Entscheidungen, wie z. B. die Entscheidung zur Durchführung eines Screenings, darauf zurückgeführt werden. Dies steht im Gegensatz zu Variationen in der Rate einer Krankheitsprogression, welche zu der Zeit, zu der die Entscheidung zum Screenen getroffen wird, nicht beobachtbar ist. Daher könnte man nicht die Entscheidung treffen, nur solche Patienten zu screenen, bei denen die Krebserkrankung schneller voranschreitet. Bei der Schätzung der Kosteneffektivität einer Intervention für eine heterogene Population kann man dies von den beobachteten Charakteristika abhängig machen und die gesamte Gruppe in homogene Subgruppen, in denen die Patienten identische beobachtbare Charakteristika besitzen, aufteilen. Das Modell kann dann getrennt für jede homogene Gruppe berechnet werden, um Schätzungen der Kosteneffektivität für den jeweiligen Satz an beobachteten Charakteristika zu generieren. Annahmeentscheidungen können dann für jede dieser sich gegenseitig ausschließenden

und identifizierbaren Patientengruppen getroffen werden [2, S. 246]. (Anmerkung: deutsche Übersetzung des englischen Zitats).

Zuletzt bezieht sich das Problem der Messfehler auf die Fähigkeit des Wissenschaftlers alle Variablen richtig zu messen. In vielen Fällen kann die Messung einiger Erscheinungen sehr schwierig sein. Abhängig von dem Ausmaß der angewandten Approximation können Verzerrungen entstehen. Wann immer ein Fehler im Bereich der Variablen auftaucht, welcher nicht durch Messungen verringert werden kann, können Wissenschaftler diesen durch die Verwendung empirischer Verteilungen beschreiben.

Zum Beispiel wird der Grad, zu dem sich ein individuelles Cervixkarzinom weiterentwickelt, zwischen Patienten variieren. Wir können eine Verteilung des Grades der Karzinomprogression durch Berechnung der Patientenzahl mit unterschiedlicher Progressionsrate beschreiben. Dennoch würden weitere Untersuchungen die Variation in der Progressionsrate nicht reduzieren. Ein anderes Beispiel ist, wenn man sich bei einem Ereignis, das mit einer bestimmten Wahrscheinlichkeit eintritt, wie z. B. Tod, den Eintritt dieses Ereignisses als bestimmt durch ein Lotteriespiel vorstellt. So können wir mit Sicherheit wissen, dass die Wahrscheinlichkeit des Todes z. B. 5 % beträgt, wissen aber nicht, welche 5 % der Personen sterben werden [2, S. 245] (Anmerkung: deutsche Übersetzung des englischen Zitats).

3.2 Konstruktion von Konfidenzintervallen

Unter der Annahme, dass unsere Schätzungen erwartungstreu sind, ist die Konstruktion eines Konfidenzintervalls eine eher einfache Aufgabe. Murillo-Zamorano [6] bietet eine recht vollständige und aktuelle Übersicht der Methoden zur Berechnung von Konfidenzintervallen sowohl mit mathematischen Programmierungen als auch mit ökonometrischen Ansätzen. Im ersten Fall zeigten neueste Entwicklungen, dass es möglich ist, ein statistisches Modell zu definieren, welches die Bestimmung statistischer Eigenschaften der nicht-parametrischen Grenzschatzer ermöglicht, obwohl nicht-parametrische Techniken deterministischer Natur sind und dies Wissenschaftler dazu veranlasst hat, diese als nicht-statistische Methoden zu beschreiben. Diese neuen Methoden werden in Kapitel 2.3 des Artikels von Murillo-Zamorano [6] dargelegt.

In Bezug auf den ökonometrischen Ansatz sollten wir zwischen deterministischen und stochastischen Grenzansätzen unterscheiden. Der deterministische ökonometrische Ansatz

nutzt das gleiche theoretische Gerüst, das für mathematische Programmierungsansätze entwickelt wurde, erlaubt aber eher das Schätzen als das „Berechnen“ der Parameter der Grenzfunktionen. Wichtiger ist, dass statistische Rückschlüsse basierend auf solchen Schätzungen möglich und leicht abzuleiten sind. Verschiedene Techniken wie die Modified Ordinary Least Squares (z. B. [7]), die Corrected Ordinary Least Squares (z. B. [8]) und die Maximum Likelihood Estimation (z. B. [9]) wurden in der ökonometrischen Literatur entwickelt, um diese vollständig deterministischen Modelle, die Grenzüberschneidungen folgen, zu schätzen. Auf jeden Fall bieten weder programmierte Modelle noch deterministisch ökonometrische Ansätze exakte Messungen des Input-Output-Verhältnisses, da beide außerstande sind, zufällige Einwirkungen einzukalkulieren.

Aigner, Lovell und Schmidt (1977), Meeusen und van den Broeck (1977) und Battese und Corra (1977) entwickelten gleichzeitig ein Stochastic Frontier Model (SFM), das neben der Aufnahme des Begriffs der Effizienz in der Analyse (wie bei den deterministischen Ansätzen) ebenso die Effekte durch exogene Einflüsse außerhalb der Kontrolle der analysierten Einheiten einbezieht. Darüber hinaus erfasst dieser Modelltyp auch Fehler bei den Beobachtungen und Messungen der Outputs [6, S. 48] (Anmerkung: deutsche Übersetzung des englischen Zitats).

Kapitel 3.1 in dem Artikel von Murillo-Zamorano [6] diskutiert die wesentlichen Ergebnisse in Hinblick auf Rückschlüsse bei der Behandlung von Querschnittsdaten.

4 Effizienzgrenzen mit kleinen Datensätzen basierend auf Sekundärdaten

4.1 Konstruktion der Effizienzgrenze

Wenn es notwendig ist, mit wenigen Beobachtungen zu arbeiten, ist es schwierig, die Konstruktion einer Effizienzgrenze auf mathematischen Programmierungen und ökonometrischen Techniken aufzubauen. In all diesen Fällen kann die Effizienzgrenze empirisch durch Verwendung des Konzeptes des inkrementellen Kosten-Effektivitäts-Verhältnisses (Incremental Cost Effectiveness Ratio, ICER) konstruiert werden. Briggs [10] und Briggs et al. [11] folgend kann man eine Effizienzgrenze durch einfache Invertierung der Achsen einer Kosten-Effektivitäts-Ebene erhalten. Eine Kosten-Effektivitäts-Ebene kann wie in Abbildung 4-1 dargestellt werden. Sie zeigt einfach die Differenz (Behandlung minus Kontrolle) in der Effektivität (ΔE) pro Patient in Abhängigkeit der Differenz in den Kosten (ΔC) pro Patient. Mit der Übernahme der Regel, dass die horizontale Achse die Effektivitätsdifferenz widerspiegelt, entspricht die Steigung der Linie, die jeglichen Punkt auf der Ebene mit dem Ursprung verbindet, dem inkrementellen Kosten-Effektivitäts-Verhältnis (ICER), $\Delta C/\Delta E$, der relevanten statistischen Kenngröße in Kosten-Effektivitäts-Studien. Verschiedene Interventionen können in der gleichen Ebene abgetragen werden, jede von ihnen mit ihrem eigenen ICER. Ein Beispiel für einen Multi-Interventions-Graphen ist in Abbildung 4-2 dargestellt.

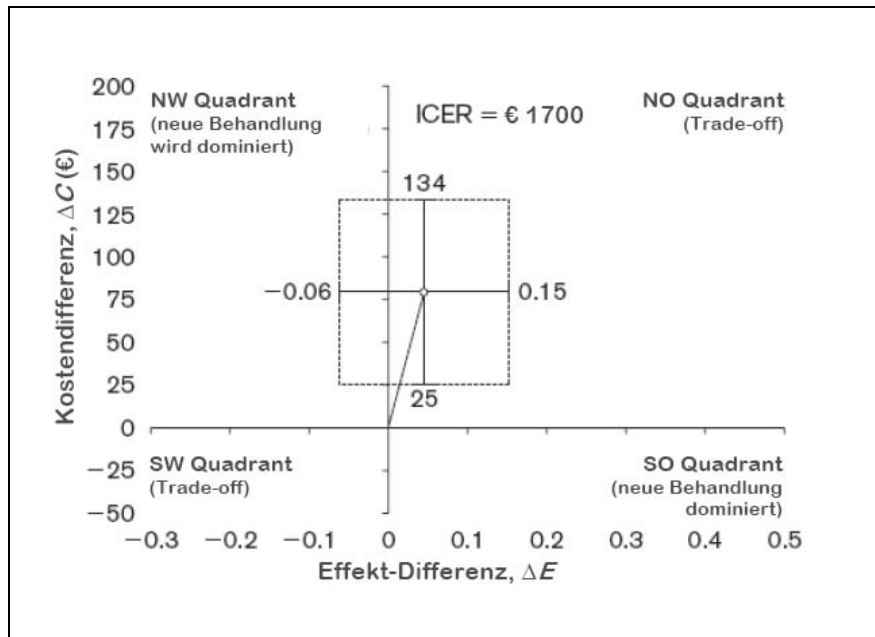


Abbildung 4-1 Kosten-Effektivitäts-Ebene mit der Ergebnislage für das Beispiel des frühen Endoskopieversuchs. ICER = inkrementelles Kosten-Effektivitäts-Verhältnis. NW = Nord-West; NO = Nord-Ost; SW = Süd-West; SO = Süd-Ost (in Anlehnung an [10]).

In Abbildung 4-1 wird ein entsprechendes inkrementelles Kosten-Effektivitäts-Verhältnis in Höhe von 1.700 € erreicht, indem man berücksichtigt, dass eine frühe Endoskopie bei dyspeptischen Patienten zusätzlich 80 € pro Patient kostet und dazu führt, dass zusätzlich 4,7 % der Patienten für 12 Monate frei von dyspeptischen Symptomen sind. Eine Effizienzgrenze erhält man aus der Kosten-Effektivitäts-Ebene durch einfaches Invertieren der Achsen. Abbildung 4-2 gibt ein klares Beispiel, wie eine Effizienzgrenze mit nur sechs Beobachtungen zu konstruieren ist!

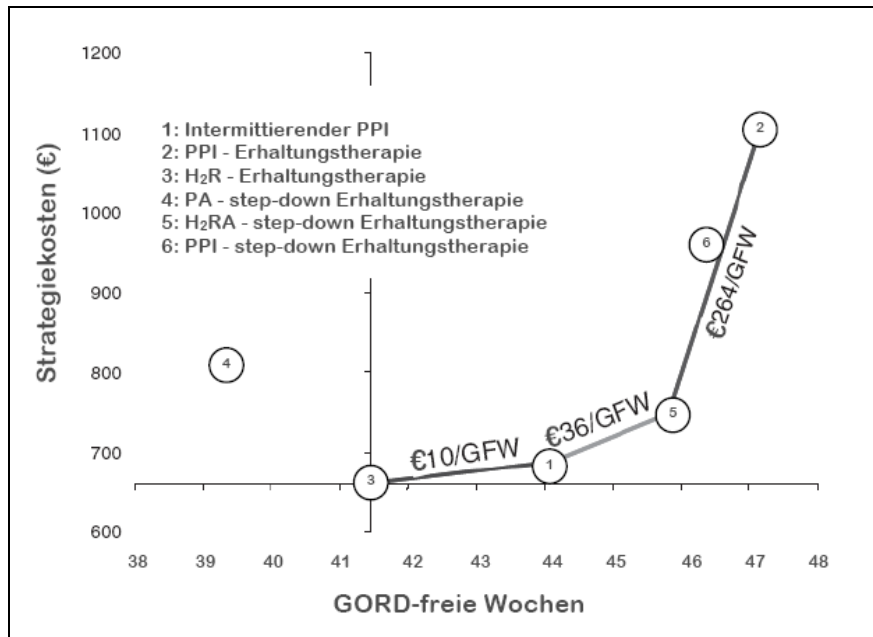


Abbildung 4-2 Kosten-Effektivitäts-Ergebnisse der Basisanalyse in der Kosten-Effektivitäts-Ebene als „Effizienzgrenze“ dargestellt. PPI = Protonenpumpeninhibitor; H₂RA = H₂-Rezeptor Antagonist; PA = prokinetisches Mittel; GORD = Refluxkrankheit; GFW = GORD-freie Wochen (in Anlehnung an [11]).

Ähnliche Beispiele könnten unter Verwendung von Datensätzen aus anderen Bereichen erstellt werden. Abbildung 3 zeigt ein Beispiel von Delea [12], basierend auf Lipidsenkern.

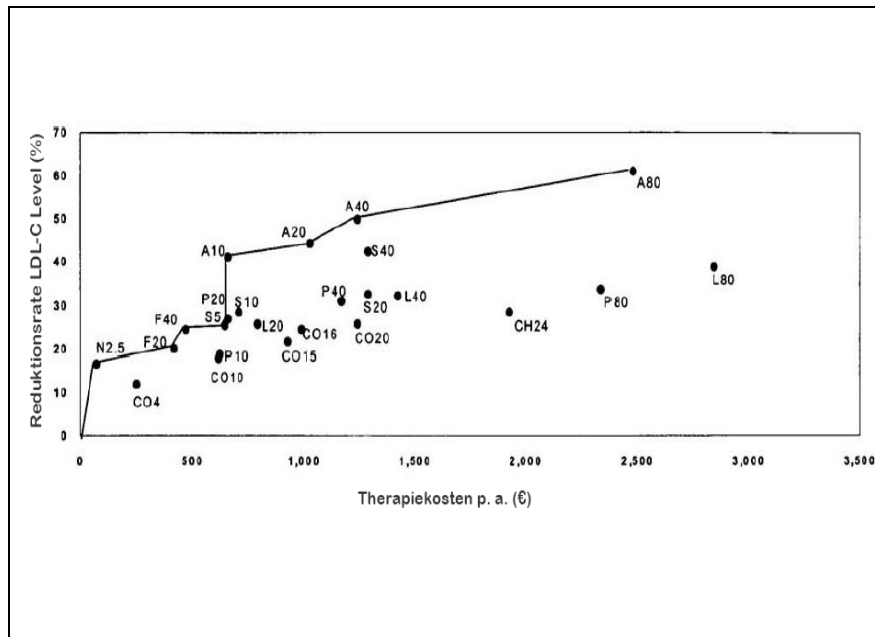


Abbildung 4-3 Absolute Effizienzgrenze, gegeben durch die Stufenform, für lipidsenkende Therapien: Effektivität wird durch Anteil der Reduktion im LDL-C-Level gemessen. LDL-C = LDL-Cholesterin; N = Niacin (2,5 mg); CO = Colestipol (4, 10, 15, 16, 20 mg); CH = Cholestyramin (24 mg); F = Fluvastatin (20, 40 mg); S = Simvastatin (5, 10, 20, 40 mg); P = Pravastatin (10, 20, 40, 80 mg); A = Atorvastatin (10, 20, 40, 80 mg); L = Lovastatin (20, 40, 80 mg) (in Anlehnung an [12]).

4.2 Umgang mit Unsicherheit³

Die oben genannten Beispiele stellen für jede Intervention die Werte der Kosten und Nutzen (oder inkrementellen Kosten- und Nutzenwerte) dar. Die Effizienzgrenze wird basierend auf diesen Werten konstruiert. Informationen, die aus diesen Daten ableitbar sind, sind für Entscheidungsträger wertvoll, solange diese Informationen reliabel (oder erwartungstreu) sind. Die Unsicherheit bei der Reliabilität dieser Schätzungen unterstützt die Analyse der Rolle der Unsicherheit in ökonomischen Evaluationen. Briggs [10] gibt eine Übersicht über diese statistischen Ansätze zum Umgang mit Unsicherheit, sowohl für primäre Studien, bei

³ Für weitergehende Details bezüglich der in diesem Kapitel präsentierten Ergebnisse kann der interessierte Leser auf [10,11,13] verwiesen werden.

denen Informationen zu Kosten und Effekten auf Patientenebene verfügbar sind, als auch für sekundäre Studien, welche typischerweise entscheidungsanalytische Methoden zur Informationssynthese aus der publizierten Literatur verwenden.

Mit der zunehmenden Verwendung klinischer Studien als Vehikel für ökonomische Evaluationen wächst auch das Interesse am Einsatz statistischer Methoden zum Umgang mit Unsicherheit von Daten auf Patientenebene, sowohl in Bezug auf Kosten als auch auf Effekte. Allerdings ist sogar für Studien basierend auf sekundären Daten die Verwendung statistischer Methoden zur Spezifizierung der Verteilungen der Parameter in sogenannten probabilistischen Analysen zunehmend üblich und wird aktuell als Methode von einer Reihe Good-Practice-Leitlinien und regulativen Institutionen empfohlen [10, S. 551] (Anmerkung: deutsche Übersetzung des englischen Zitats).

Unter den am häufigsten verwendeten statistischen Methoden zum Umgang mit Unsicherheit schlagen Briggs [10] und Briggs et al. [11] die Verwendung des „Akzeptanzkurven-Ansatzes“ vor, da dieser sehr viel mehr Informationen zur Unsicherheit liefert als Konfidenzintervalle. Für ein besseres Verständnis des Konzeptes müssen zunächst einige Grundsätze in Bezug auf die Fieller'schen Konfidenzintervallgrenzen und auf Bootstrap-Konfidenzintervalle besprochen werden. Dabei folgen wir Briggs [10].

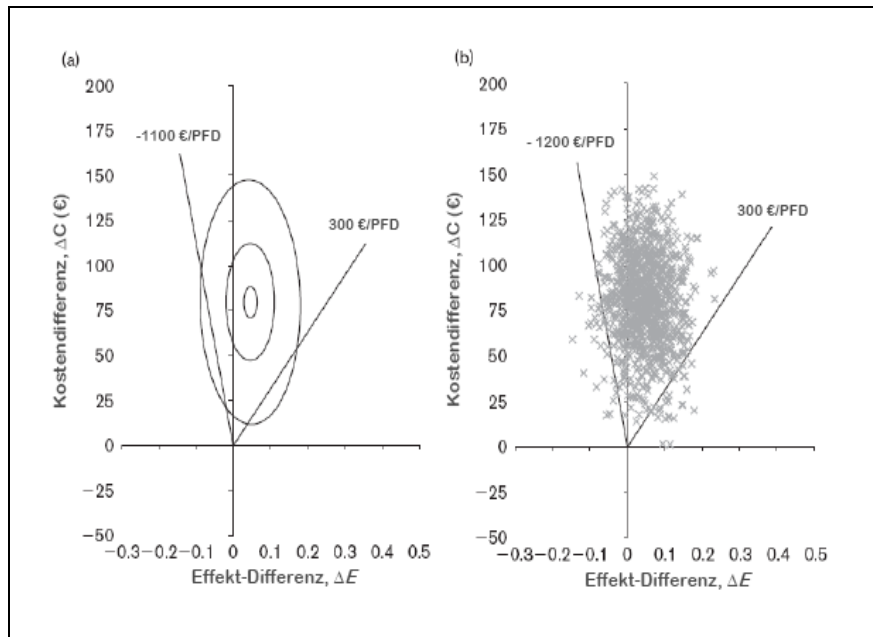


Abbildung 4-4 Konfidenzintervalle für Kosten-Effektivitäts-Verhältnisse für das Beispiel der frühen Endoskopie. (a) Parametrisches Fieller'sches Theorem, und (b) nicht-parametrisches Bootstrapping. PFD = dyspepsiefreie Patienten (in Anlehnung an [10]).

Fieller'sche Konfidenzintervallgrenzen

Fieller's Ansatz basiert auf der Annahme, dass die Differenzen in Kosten und Effekten eher einer gemeinsamen Normalverteilung folgen, als dies das Kosten-Effektivitäts-Verhältnis selber tut. Abbildung 4-4(a) zeigt die Annahme einer gemeinsamen Normalverteilung in der Kosten-Effektivitäts-Ebene für die in Abbildung 4-1 aufgeführten geschätzten ICER: drei eingezeichnete Ellipsen mit gleicher Dichte, die 5 %, 50 % und 95 % der integrierten gemeinsamen Dichte abdecken. Ebenso eingezeichnet sind die geschätzten Konfidenzintervalle unter Verwendung von Fieller's Theorem (300 € bis -1100 €), dargestellt durch die Steigung der durch den Ursprung verlaufenden Geraden in der Ebene. Man beachte, dass der „Keil“, der durch die Konfidenzintervallgrenzen bestimmt wird, in die 95 % Ellipse fällt; dies ist darauf zurückzuführen, dass der Fieller'sche Ansatz automatisch dafür adjustiert, dass 95 % der integrierten gemeinsamen Dichte in die Keilfläche fallen.

Bootstrap-Konfidenzintervalle

Dies ist eine Re-Sampling Methode, welche eine empirische Stichprobenverteilung für die interessierenden statistischen Kenngrößen abschätzt statt auf parametrische Annahmen zu vertrauen. Mit einer großen Anzahl gezogener Stichproben wird ein Vektor der Bootstrap-Wiederholungen generiert, welcher die empirische Schätzung der statistischen Stichprobenverteilung darstellt. Tausend Bootstrap-Effektdifferenzen und Kostendifferenzen für das in Abbildung 4-1 aufgeführte geschätzte ICER sind in der Kosten-Effektivitäts-Ebene in Abbildung 4-4(b) eingezeichnet. Konfidenzintervallgrenzen erhält man, indem die 2,5 und 97,5 Perzentile der Bootstrap-Wiederholungen, welche in steigender Reihenfolge der Kosten-Effektivitäts-Verhältnisse angeordnet sind, ausgewählt werden; dies stellt sicher, dass 95 % der geschätzten gemeinsamen Dichte in die Keilfläche der Kosten-Effektivitäts-Ebene, die durch die Konfidenzintervallgrenzen gebildet wird, fallen. Wie klar ersichtlich aus Abbildung 4-4(b) sind die Bootstrap-Schätzungen der gemeinsamen Dichte und die Bootstrap-Konfidenzintervallgrenzen sehr ähnlich zu denen, die nach dem Fieller'schen Theorem entstehen.

Bezüglich der Abbildungen 4-2 und 4-3 sollten diese Verfahren für jeden Punkt auf dem Graphen wiederholt werden. Dies führt zu einem Graphen wie in Abbildung 4-5 dargestellt. (Die Bedeutung dieser Abbildung wird später ausführlicher diskutiert.)

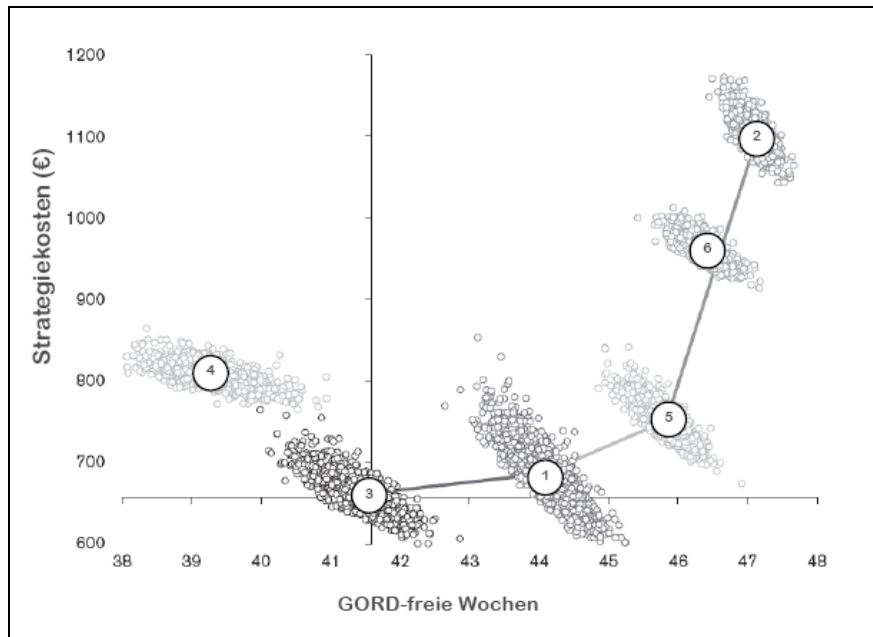


Abbildung 4-5 Zehntausend Monte-Carlo-Simulationen der sechs Behandlungsstrategien in dem Beispiel der Refluxkrankheit (GORD) eingetragen in der Kosten-Effektivitäts-Ebene. 1 = intermittierender Protonenpumpeninhibitor (PPI); 2 = PPI – Erhaltungstherapie; 3 = H₂ Rezeptor-Antagonist (H₂RA) - Erhaltungstherapie; 4 = prokinetisches Mittel - step-down Erhaltungstherapie; 5 = H₂RA - step-down Erhaltungstherapie; 6 = PPI step-down Erhaltungstherapie (in Anlehnung an [11]).

Die Akzeptanzkurve

Fieller'sche Konfidenzintervallgrenzen und Bootstrap-Konfidenzintervalle funktionieren ausgezeichnet, bis ICERs des gleichen Merkmals verglichen werden müssen, die auf beiden Seiten der Effizienzgrenze liegen. In diesem Fall gelingt es diesen Ansätzen nicht, ICERs richtig zu vergleichen. Wie von Briggs [10] vorgeschlagen, kann eine Lösung für dieses Problem gefunden werden. In der Tat könnten wir hinsichtlich der Bootstrap-Wiederholungen in der Kosten-Effektivitäts-Ebene in Abbildung 4-4(b) Unsicherheit zusammenfassen, indem geprüft wird, wie viele der Bootstrap-Wiederholungen unterhalb und rechts der Linie mit einer entsprechenden Steigung eines vorab festgelegten Höchstverhältnisses (R_C) fallen, was Hinweise für die Kosteneffektivität der Intervention liefert.

Alternativ, falls wir glücklich mit der Annahme der gemeinsamen Normalverteilung von Kosten und Effekten sind, können wir den Anteil der parametrischen gemeinsamen Dichte, der im kosteneffektiven Bereich der Kosten-Effektivitäts-Ebene liegt, betrachten. Natürlich ist der angemessene Wert von R_C selber unbekannt. Allerdings kann er variiert werden, um zu zeigen, wie die Evidenz für die Kosteneffektivität einer Intervention mit R_C variiert. Die resultierende Kurve für das Beispiel der frühen Endoskopie in Abbildung 4-1, basierend auf der Annahme der gemeinsamen Normalverteilung wie in Abbildung 4-4(a) gezeigt, ist in Abbildung 4-6 dargestellt und wird als „Kosten-Effektivitäts-Akzeptanzkurve“ bezeichnet, da sie direkt die Evidenz zugunsten der Intervention, die kosteneffektiv für alle möglichen Werte der Entscheidungsrichtlinien ist, zusammenfasst. [10, S. 555-556]. (Anmerkung: deutsche Übersetzung des englischen Zitats).

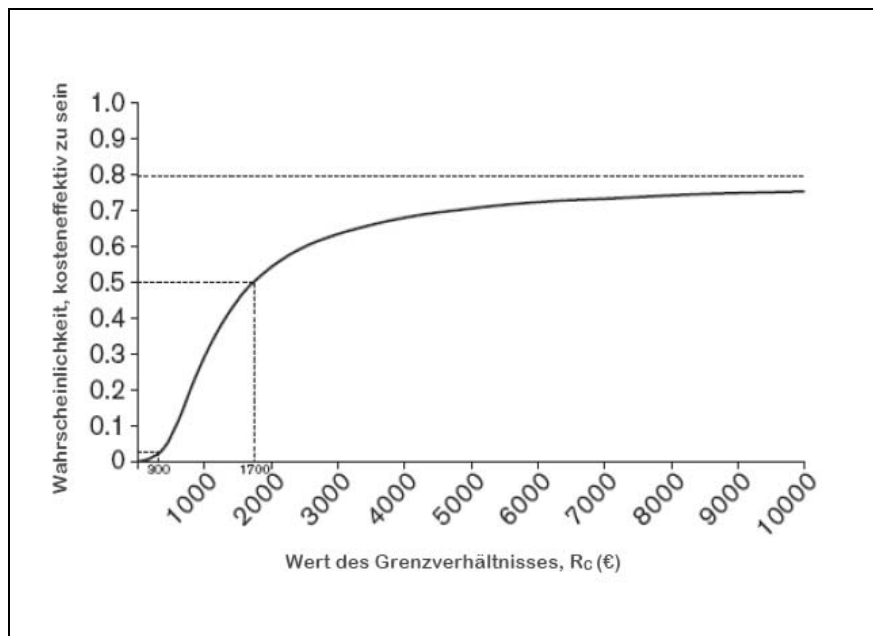


Abbildung 4-6 Kosten-Effektivitäts-Akzeptanzkurve für das Beispiel der frühen Endoskopie (in Anlehnung an [11]).

Für jede der in Abbildung 4-5 aufgezeigten Interventionen könnten mehrere dieser Kurven generiert werden. Dies führt zu einer Kurve wie die in Abbildung 4-7. Wie erwartet ist Strategie 4 nicht in Abbildung 4-7 dargestellt, was darauf hinweist, dass sie niemals im Wettbewerb um Kosteneffektivität steht (sie liegt unterhalb der Effizienzgrenze). Strategie F ist abgebildet, obwohl sie niemals mehr als 13 % der Simulationen erreicht, die auf eine Kosteneffektivität hinweisen, sogar bei dem günstigsten Grenzverhältnis (etwa 260 € pro GORD-symptomfreien Tag). In der Praxis werden Entscheidungsträger höchstwahrscheinlich

zwischen den Strategien 3, 1, 5 und 2 wählen wollen, und ihre Wahl wird höchstwahrscheinlich durch die Strategie, die derzeit als Standardpraxis angesehen wird sowie das Grenzverhältnis und die Risikoeinstellung gesteuert.

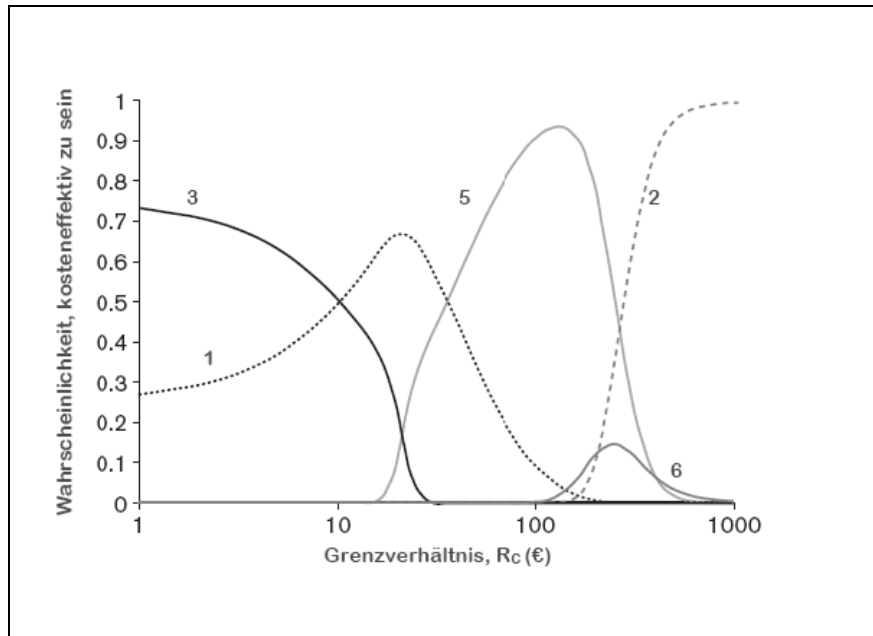


Abbildung 4-7 Multiple Akzeptanzkurven für die sechs Behandlungsstrategien in dem Beispiel der Refluxkrankheit (GORD), die den Anteil der Häufigkeit, mit der jede Strategie als kosteneffektiv angesehen wurde, als eine Funktion des Grenzverhältnisses darstellen; eine Log-Skala wird eingesetzt zur besseren Darstellung der kleinen Werte (siehe Legende zu Abb. 4-5 für kurze Beschreibung der Strategien) (in Anlehnung an [11]).

Die Information, die in der Analyse für Entscheidungsträger noch fehlt ist das Grenzverhältnis oder der Schwellenwert für die Zahlungsbereitschaft. Diese Schwellenwerte sind exogen bezogen auf die Analyse der Unsicherheit und sollten vorab bestimmt werden. Unter diesen Informationen stellt sich Abbildung 4-7 wie Abbildung 4-8 dar. In diesem Fall impliziert ein Grenzverhältnis von 10, dass Strategie „3“ ausgewählt wird, während bei dem Fall eines Grenzverhältnisses von über 264 € Strategie „2“ ausgesucht werden könnte.

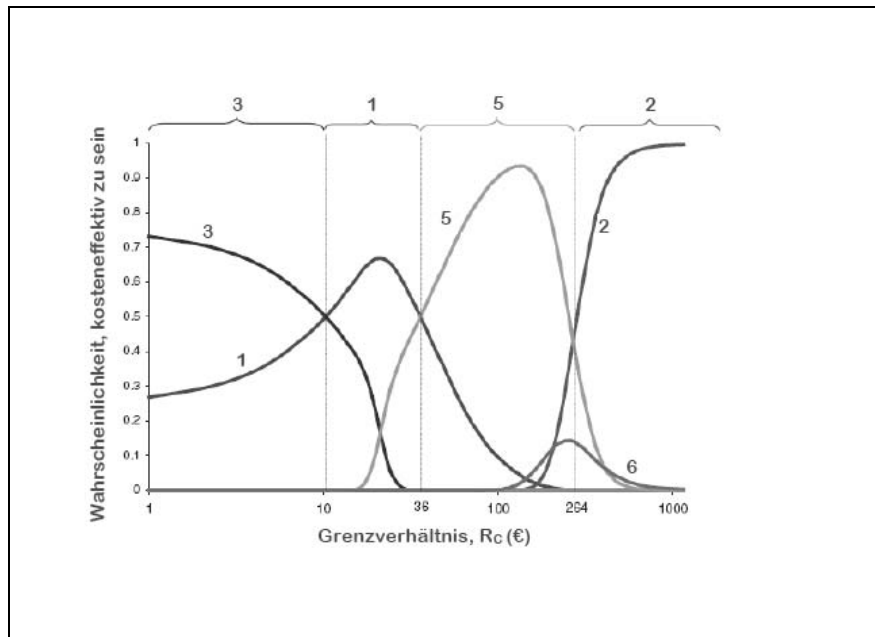


Abbildung 4-8 Multiple Akzeptanzkurven für die sechs Behandlungsstrategien in dem Beispiel der Refluxkrankheit (GORD) mit Schwellenwerten für die Zahlungsbereitschaft (siehe Legende zu Abb. 4-5 für kurze Beschreibung der Strategien) (in Anlehnung an [11]).

Literaturverzeichnis

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung [Online]. 24.01.2008 [Zugriff am 09.09.2008]. URL: http://www.iqwig.de/download/08-01-24-Methoden_Kosten-Nutzen-Bewertung_Version_1_0.pdf.
2. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic analysis and computationally expensive models: necessary and required? *Value Health* 2006; 9(4): 244-252.
3. Kim Y, Schmidt P. A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. *Journal of Productivity Analysis* 2000; 14(2): 91-118.
4. Farrell MJ. The measurement of productive efficiency. *J R Stat Soc (Ser A)* 1957; 120: 253-281.
5. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *Eur J Oper Res* 1978; 2(6): 429-444.
6. Murillo-Zamorano LR. Economic efficiency and frontier techniques. *J Econ Surv* 2004; 18(1): 33-77.
7. Richmond J. Estimating the efficiency of production. *International economic review* 1974; 15(2): 515-551.
8. Gabrielsen A. On estimating efficient production functions: working paper A-35. Bergen: Chr. Michelsen Institute; 1975.
9. Greene WH. Maximum likelihood estimation of econometric frontier functions. *J Econom* 1980; 13(1): 27-56.
10. Briggs AH. Statistical approaches to handling uncertainty in health economic evaluation. *Eur J Gastroenterol Hepatol* 2004; 16(6): 551-561.
11. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making* 2002; 22(4): 290-308.
12. Delea TE. Using of surrogate measures in cost-effectiveness analyses of lipid-lowering therapies. *Value Health* 1998; 1(3): 175-180.
13. Van Hout BA, Maiwenn JA, Gordon GS, Rutten FH. Costs, effects and c/e-ratios alongside a clinical trial. *Health Econ* 1994; 3(5): 309-319.