

国際医療福祉大学審査学位論文（博士）

大学院薬学研究科博士課程

# 薬剤経済評価の不確実性とその対処法

平成28年度

医療・生命薬学専攻

学籍番号：13R3001 氏名：荒西 利彦

研究指導教員：池田 俊也 教授

## 和文抄録

薬剤経済評価には様々な原因により不確実性が含まれるが、評価の結論は償還可否などの結論に影響を与える可能性があり、重要な課題である。本研究では、現状の把握と解決策の提案のため、その原因、対応方法の整理を行った後、不確実性の評価の現状調査を行った。更に、既存の評価法における課題を整理し、その解決策として新たな方法を提案した。この結果、今後国内における評価で不確実性の評価がより詳細に行われるために、確率的感度分析に関するガイドライン等の整備が望まれることが結論付けられた。また、不確実性の評価指標であるExpected Value of Perfect Information (EVPI) を改変したEVPI<sub>r</sub>を提案し、その性質を調べ、追加情報取得のための研究投資の評価に用いる場合に論理的に正しい結論を導くことを示した。

## キーワード

薬剤経済評価, 感度分析, 情報の価値

## **Abstract**

In pharmacoeconomics evaluations, many factors cause uncertainty on the analysis. The results may affect decision about reimbursement hence it is a key issue. In this study, to figure out the current situation and to suggest solutions for that, we summarized the cause of uncertainty and the methods to address them. After that, we conducted a literature review to reveal how uncertainty is evaluated on the literatures. In addition, we pointed out the issue with the current methods of evaluation and suggested new measure of uncertainty. In result, it is recommended to prepare probabilistic sensitivity analysis guideline to help more appropriate evaluation of uncertainty in a pharmacoeconomic study. At last, we suggested a modified expected value of perfect information (EVPI) which we call  $EVPI_f$  and compared its characteristics with original EVPI. The comparison found that  $EVPI_f$  leads right conclusion when it is used on the evaluation of the effectiveness of investment on a study to get more clinical evidence.

## **Key words**

Pharmacoeconomics, Sensitivity Analysis, Value of information

## 目次

1	序論.....	1
1.1	背景.....	1
1.2	本研究の研究課題.....	2
1.3	本研究の目的.....	3
1.4	薬剤経済評価の方法.....	3
1.5	費用効用分析 (Cost Utility Analysis; CUA) の方法.....	6
1.6	薬剤経済評価の手順.....	8
1.7	薬剤経済評価における不確実性.....	9
1.8	不確実性の評価と対応.....	10
1.8.1	Methodological uncertainty の評価、対応.....	10
1.8.2	Structural uncertainty の評価、対応.....	13
1.8.3	Parameter uncertainty の評価、対応.....	32
1.9	まとめ.....	41
2	研究 1 確率的感度分析の使用状況.....	42
2.1	緒言.....	42
2.2	PSA の手法と結果の解釈.....	44
2.2.1	PSA.....	44
2.2.2	PSA から得られる結果とその解釈.....	47
2.3	PSA に関する各国ガイドライン.....	50
2.4	PSA の利用状況.....	52
2.4.1	本邦における邦文で報告された PSA の使用状況.....	52
2.4.2	本邦における英文で報告された PSA の使用状況.....	55
2.5	結果のまとめ.....	56
3	研究 2 頻度論流 EVPI の提案.....	58
3.1	緒言.....	58
3.1.1	EVPI(Expected Value of Perfect Information).....	58
3.1.2	EVPI の解釈.....	60
3.1.3	EVPI の例: cost-utility analysis between Linezolid and Vancomycin in the empiric treatment of nosocomial pneumonia.....	63
3.2	方法.....	66
3.2.1	頻度論に基づく EVPI の定義.....	66
3.2.2	EVPI と $EVPI_f$ の比較.....	67
3.3	比較結果.....	68
3.4	$EVPI_f$ の例.....	69
3.5	考察.....	70
3.6	結果のまとめ.....	73
4	考察.....	74
5	本研究の限界.....	75

6	結論.....	76
7	謝辞.....	77
8	倫理上の配慮・利益相反.....	78
9	略称一覧.....	79
10	参考文献.....	81

# 1 序論

## 1.1 背景

Health Technology Assessment (医療技術評価; HTA) とは、EU 諸国の HTA 機関のネットワークである EUnetHTA によると、「HTA とは、医療技術の利用に関する医学的・社会的・経済的・倫理的な問題についての情報を、システムティックに、透明性をもって、偏見なく、着実にまとめていく学際的なプロセスである。その目的は、患者中心の安全で効率的な医療政策を作るために情報を提供し、最良の価値を達成しようとするものである。」とされている<sup>1,2)</sup>。HTA の枠組みの中において、医療技術を経済的な側面から評価することを医療経済評価と呼ぶ。本研究においては医療経済評価のうち薬剤を評価するものに着目し、薬剤経済評価と呼ぶ（ただし、研究 1 においては薬剤経済評価を包含する概念として医療経済評価という呼称を必要に応じ使用する）。

薬剤経済評価において評価される課題は、多くの場合「ある治療が、別の治療と比べて費用対効果に優れているか」というリサーチクエスチョンを評価する相対的なものである。評価対象の治療について適切な比較対照を設定し、治療間で有効性、安全性に代表される利得と、コストの差を分析することで、評価対象の治療が要求する追加的なコストが、追加的に得られる利得に対し受け入れ可能なものであるかを判断する。

薬剤経済評価は、医薬品の償還の可否あるいは償還価格の決定の評価材料の一つとして採用される動きが近年各国で広がっている。薬剤経済評価が政策に用いられている最も有名なイギリスの例を挙げると、医薬品の価格は利益率の上限は設定されるものの、各製薬会社の自由価格が認められているが、National Health Service (NHS) が指定した品目について National Institute for Health and Care Excellence (NICE) が薬剤経済評価などの根拠をもとに NHS の支払いの下での使用を推奨するか否かをガイドラインにまとめる。製薬会社は償還を推奨されなかった品目について値引き等の交渉の余地は残されているが、交渉が合意に至らなかった場合はイギリス国内における償還は実質的には行われなくなると考えてよい。イギリスの医

療はすべて税金を財源とし、行政側から予算の管理が可能であるためこのような制度が実現しているが、費用対効果が優れていないことを理由に新薬へのアクセスが阻害されることに関して患者団体などからの反発は強い。このため、例えば抗がん剤に対しては Cancer Drugs Fund (CDF)<sup>3)</sup>を医療費の予算とは別に設定し、患者アクセスを担保している。一方で日本においては、薬事承認を獲得した医薬品・医療機器は原則として一定の公定価格のもと保険償還されるというシステムのもと、薬剤経済評価の観点からの保険償還価格の設定及び保険償還の可否は行われてこなかったが、中央社会保険医療協議会において2012年より費用対効果評価の活用が検討され、2016年から試行的導入が行われており、2018年度の薬価改定時に活用される予定である。

## 1.2 本研究の研究課題

医学研究において、不確実性を完全に取り除くことはできない。Cochrane library<sup>4)</sup>のように多くの臨床試験のメタ・アナリシスを行う理由は、臨床試験の結果は不確実性を含むため、臨床試験1つのみでは誤った結論に至る可能性があるからである。このようなメタ・アナリシスの結果であっても、誤った結論に至る可能性を完全には排除できないが、統計学的にある程度以上の確信が得られた段階で「現段階での最善のエビデンス」としてその結果を受け入れ、治療ガイドライン等に組み込まれる。これはメタ・アナリシスによって得られた結論が誤っている可能性が十分に小さければ、誤った結論を採択したために生じる損失の期待値よりも、メタ・アナリシスの結論が正しい場合に得られる利得の期待値が上回っているため正当化される。新薬の開発において有効性の検証段階である第III相臨床試験においても、両側5%の第1種の過誤（実際には新薬と対照薬の間に差は無いが、誤って差があると結論づけてしまう誤り）を許容するのも同様の理由による。

薬剤経済評価は様々な臨床試験、データベースを統合するものであるため、これらに含まれる不確実性の影響を当然受ける。さらに、薬剤経済評価では疾患の予後をモデル化して評

価を行うことが多いが、患者ごとに千差万別である予後を、シンプルな（例えば癌の費用対効果評価においてはしばしば【無再発】、【再発】、【死亡】の3つの疾患状態のいずれかにすべての患者が分類されると設定する）モデルで表現する過程において失われる情報も多く、薬剤経済評価自体における方法論や前提条件が異なれば結果も異なる。よって一般の医学研究に比べ不確実性の影響を大きく受けるといえる。前述のとおり、薬剤経済評価の結論によって新薬を使用しないという意思決定につながる場合があり、このことを勘案すると薬剤経済評価の結論の誤りは臨床試験における第2種の過誤（新薬は真に有効であるが、仮説検定の結果その有効性を検出できない過り）に相当する大きな影響を与えうるといえるため、薬剤経済評価における不確実性の考察は重要なポイントである。そこで本研究では、薬剤経済評価における不確実性について整理し、解決策を提案することを課題とする。

### 1.3 本研究の目的

本研究では、薬剤経済評価における不確実性について整理した後、不確実性に対する解決策を提案するため以下の研究テーマを設定した。

1. 不確実性の評価が日本における研究にてどのように行われているかを調査する。
2. 既存の評価法における課題を整理しその解決策として新たな方法を提案する。

### 1.4 薬剤経済評価の方法

薬剤経済評価では、評価対象の治療（以下、新治療とする）を費用と効果の両面から対照治療と比較することで費用対効果に優れているか否かを検討する。費用と効果のいずれにおいても、「新治療を選択」と「対照治療を選択」の意思決定がもたらす結果を統合し、評価する。以下にその具体的な説明を行う<sup>5)</sup>。

「費用」とは、治療そのものの費用を含めることは当然であるが、副作用が発生した場合の治療費や、診療費、検査費、通院にかかる交通費および疾患に伴う労働損失なども含める



ことがあり、費用として計上する範囲は分析の視点（perspective）に応じて変化することになる。例えば公的医療費の支払い者の立場に立てば患者が疾患のため働けない損失（労働損失）は計上されないが、社会の立場、すなわち社会全体の利益を考えるならば労働損失は社会的損失として費用に計上される。

「効果」とは、治療により獲得される医療上の利得である。これは評価の方法や疾患により異なる尺度が使用される。臨床検査値の改善、生存期間の延長、無再発生存期間の延長などが考えられるが、質調整生存年（Quality-adjusted Life Years; QALYs）が疾患を問わず比較できる点、有効性だけでなく安全性面からの評価を含んでいる点から、有用な指標として活用されている<sup>6)</sup>。

「費用」と「効果」の取り扱いにより以下に示す費用最小化分析、費用効果分析、費用効果分析、費用便益分析の4つの薬剤経済評価の方法が提案されており、評価の目的、評価対象治療の性質等により使い分けが行われる。

- 費用最小化分析（Cost Minimization Analysis; CMA）

新治療と対照治療の要求する費用を比較し、より費用が少ないものを優れているとする方法。新治療と対照治療で得られる効果が同等であると認められる場合には簡便な分析方法として有効な方法である。

- 費用効果分析（Cost Effectiveness Analysis; CEA）

効果として臨床検査値や生存年などの臨床指標を採用し、効果1単位を追加的に獲得するために必要な追加費用を算出する。これを Incremental Cost-Effectiveness Ratio（増分費用効果比; ICER）と呼ぶ。ICERは以下の式で算出される。

$$ICER = \frac{C_{new} - C_{com}}{E_{new} - E_{com}}$$

$C_{new}$ 、 $C_{com}$ ：新治療、比較対照となる治療の費用

$E_{new}$ 、 $E_{com}$ ：新治療、比較対照となる治療によって得られる効果

費用効果分析は、効果 1 単位当たりの費用を算出するため、例えば糖尿病治療において HbA1c を効果とした評価と、高血圧治療において拡張期、収縮期血圧を効果とした評価の間での比較ができないという課題はあるが、臨床において通常使用されている効果指標をそのまま使用するため解釈が容易であり、また効果のデータの入手も比較的容易であるという利点がある。

- 費用効用分析 (Cost Utility Analysis; CUA)

CEA のうち、効果として QALYs を用いるものを費用効用分析と呼び識別する。QALYs は、以下の式で表わされる通り各疾患状態における Quality Weight (QW) とその疾患状態での生存年数の積を合計したものである。なお、QW とは、完全な健康状態を 1、死亡を 0 として各健康状態に QOL に基づく重みづけをする指標である。

$$QALY = \sum (\text{生存年}) \times (QOL)$$

QW は疾患の種類を問わずに定義ができる指標であり、QW に基づいた QALYs も疾患を問わず比較可能であるという優れた性質を持っている。このため CUA の結果は前述のように CEA では難しかった糖尿病治療と高血圧治療の間での疾患横断的な比較にも活用することができる。QW は EQ-5D<sup>7)</sup>や HUI<sup>8)</sup>、SF-6D<sup>9)</sup>などの QOL 尺度から算出されるが、これらは通常診療で用いられることが少ない評価指標であることから効果データとして QW を獲得することが CUA における課題である。

- 費用便益分析 (Cost Benefit Analysis; CBA)

ある利得を獲得するために支払ってもよいとする上限の費用を Willingness To Pay (支払い意思額; WTP) と呼ぶ。CEA や CUA において効果 1 単位あたりの金銭的価値を定めることができれば、効果を金銭に換算し、費用との差をとることで、対照治療を新治療に切り替えることによってどれだけの金銭的利益が生まれるのかを明らかにすることができる。

## 1.5 費用効用分析 (Cost Utility Analysis; CUA) の方法

前節に列挙した薬剤経済評価の方法のうち、多くの国において標準的な手法とされている CUA の方法を述べる。

前節にて触れたとおり、CUA では新治療と対照治療の間で費用と効果 (QALYs) を比較するものであるが、これを図示すると図 1-1 の通りとなる。平面の原点に対照治療を置き、X 軸で効果 (QALYs) を、Y 軸で費用を示す。2次元の平面において X 軸と Y 軸で区切られた区間を、X、Y ともに正である第 1 象限から反時計回りに第 1 象限、第 2 象限、第 3 象限、第 4 象限と呼ぶが、ここでは表記を簡単にするため (効果 X, 費用 Y) の正負をもとに (効果+, 費用+) から (効果-, 費用+)、(効果-, 費用-)、(効果+, 費用-) と表示する。このような図を費用効果平面と呼ぶ。

新治療が (効果+, 費用-) の区間に位置する場合、新治療の導入によって対照治療よりも高い効果が得られるうえ、費用も節約されることになる。このようなケースにおいては薬剤経済的観点からは新治療への切り替えが望ましいことに議論の余地はなく、これを **dominant** であるという。一方で、新治療が対照治療よりも費用が増加するにもかかわらず、効果は低い (効果-, 費用+) の場合、薬剤経済的観点からは新治療に切り替えないことが最善の選択であり、これを **dominated** であるという。また、新治療のほうが効果が低く、費用も低い (効果-, 費用-) であるケースも可能性としてはあり得るが、費用節減のため現行治療 (対照治療) を効果の劣る治療 (新治療) に切り替えるかという倫理的議論については本稿では触れない。

**Dominant** や **dominated** の場合は、新治療と対照治療のいずれかを選択することで高い効果をより低い費用で達成することができるため意思決定は容易であるが、多くの新治療は (効果+, 費用+)、すなわち対照治療よりも効果が高いが費用も高くなるであろう。このような場合に「対照治療を新治療に切り替えることにより、効果を 1QALY 獲得するためにいくらの費用を追加で支払うことになるのか」という観点から評価を行うのが CEA の考え方で

ある。図 1-1 に示す新治療 A と新治療 B を見ると、新治療 A は対照治療と比較して効果は大きく優れるが追加費用は少なくてすむのに対し、新治療 B は効果の増加が少ないにもかかわらず追加で多くの費用がかかることがわかる。これを定量的に表すため、対照治療と新治療を結ぶ直線の傾きを示したものが ICER である。1QALY を獲得するのに支払ってもよいとする金額を支払い意思額として定義すると、支払い意思額に対応する半直線を原点から引くことができるが、費用効果平面上のうち（効果+，費用+）の領域でこの半直線（片方は点で止まり、もう片方はどこまでも伸びている線）よりも上にあるか下にあるかが、新治療が費用対効果に優れているか否かに対応する。

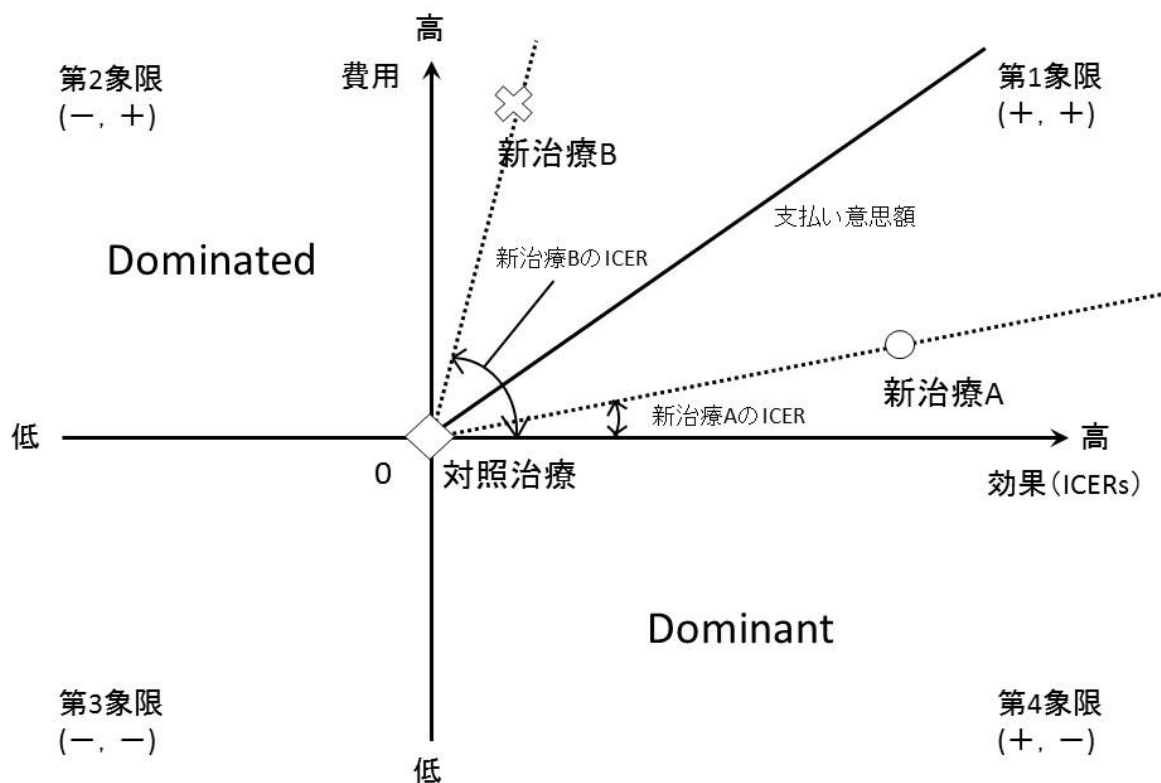


図 1-1 費用効果平面の例

## 1.6 薬剤経済評価の手順

薬剤経済評価に伴う不確実性を議論するため、薬剤経済評価の手順を CUA の例を挙げ簡潔にまとめる。本稿で例示される薬剤経済評価の手法として断りのない場合は CUA を用いるが、他の手法においても一般性を失わない。

### 1. リサーチクエスチョンの設定

評価を行う患者集団、新治療と対照治療、アウトカムを決定し、臨床的に明らかにしたい課題を明確にする。あわせて評価を行う期間（時間地平）や、未来に発生する費用や効果を現在価値に換算するための割引率も決定する必要がある。

### 2. 疾患モデルの作成

時間地平は治療間の差異を評価するために「十分長い分析期間」を設定することが本邦の薬剤経済評価ガイドラインにおいても推奨されているが<sup>10, 11</sup>、「十分長い分析期間」とは多くの場合数十年にわたり、患者の生涯を通した評価とする場合も珍しくはない。一方で臨床試験や臨床研究の試験期間は治療の有効性を判断するのに十分な期間として設定されるため、多くの場合薬剤経済評価において要求される「十分長い分析期間」より短い。よって臨床試験や臨床研究から得られたデータを薬剤経済評価で用いるためには、仮に患者を長期間追跡した場合にどのような予後をたどるのかを予測する必要がある。また、リサーチクエスチョンで設定した患者集団と完全に一致した臨床試験や臨床研究が先行研究として必ずしも存在するとは限らないため、患者集団を絞った（あるいはひろげた）場合の結果を予測する必要がある。このような問題を解決するため、疾患のたどる予後を一定の仮定のもとでモデル化することが一般的に行われる<sup>12</sup>。モデル化の手法としてしばしば使用されるものとして、ある治療を受けた場合の予後を枝分かれで示し、最終的な疾患状態に至るまでの確率とそれまでにかかる費用、効果をモデル化する決定樹モデルと、「健康」、「再発」、「死亡」などの疾患状態間を一定の期間（cycle length）ごとに遷移するように予後をモデル化するマルコフモデルがある<sup>12</sup>。

### 3. 疾患モデルに効果、費用のデータ（パラメータ）を挿入する

疾患モデルで作成したいずれのモデルにおいても、いくつかの疾患状態を定義して、治療によってそれらの疾患状態間を遷移する確率を変動させることで治療の効果を表現する。このための確率は、臨床試験や臨床研究から推定する必要がある。また、それぞれの疾患状態における QW や、かかる費用も推定する。

### 4. ICER の算出

疾患モデルに効果、費用のデータを挿入すると、興味のある患者集団において時間地平の期間にわたった評価を行った場合に、新治療と対照治療で得られた効果と、要した費用が算出できる。これらの治療間差の比をとることで、ICER を算出する。

## 1.7 薬剤経済評価における不確実性

薬剤経済評価における不確実性は以下の 3 種に分類される<sup>13)</sup>。

- **Methodological uncertainty**

割引率の設定や、含める費用の範囲など、薬剤経済評価をどのように行うかの手順は一意ではなく、不確実性の原因となる。

- **Structural uncertainty**

疾患モデルに含まれる疾患状態の定義は必ずしも単一にはならない。より細かな疾患状態を定義すれば疾患モデルの上では現実の予後を正しく示すことができるように見えるが、パラメータの獲得が困難となったり、またパラメータが精度よく推定できなくなる。またそもそも設定した疾患状態が現実的ではない可能性もある。

- **Parameter uncertainty**

疾患モデルに挿入するパラメータは臨床試験、臨床研究から獲得されることが通常であるが、臨床試験、臨床研究はいくら良く計画され、正しく実行されたとしても結果には必ず不確実性を含む。特に薬剤経済評価は多くの臨床試験、臨床研究の結果を集約して

結果を得るプロセスであることから、それぞれの臨床試験、臨床研究に含まれる不確実性が最終的な結果、結論に与える影響は大きい。

## 1.8 不確実性の評価と対応

薬剤経済評価における不確実性はその発生原因から 3 種に大別できることを前節で示した。それぞれの不確実性について、その特性を述べ、評価と対応の方法について述べる。

### 1.8.1 Methodological uncertainty の評価、対応

Methodological uncertainty とは、薬剤経済評価の方法に伴う不確実性であり、以下のようなものがあげられる。

- 評価の視点
- 比較対照の選定
- 評価指標
- 費用、効果の割引率
- 時間地平
- 費用の通貨換算レート

Methodological uncertainty は 3 種の不確実性のうちもっとも科学的な取り扱いが困難な不確実性であるといえる。割引率を例に挙げると、将来の価値を現在価値に直したときにどれだけの割合を割り引いて考える必要があるかは、評価を行う国や経済の状況により変わることから、科学的に正しい値を得ることは根本的に無理がある。

割引率の場合であれば国債の利率などを参考にするなど、一定の理由づけをもって何らかの値を定義することは可能である。薬剤経済評価を行った結果の比較可能性を担保するという目的を達成するためには評価の条件を揃えることである程度の解決が可能であることから、妥当と考えられる方法に基づく値をガイドラインに設定することにより methodological

uncertainty の現実的な対処とすることができる。日本における研究者用ガイドラインにおいては年率 2%、英国 NICE のガイドラインでは年率 3.5% が割引率の標準として推奨されている<sup>6, 10)</sup>。

ガイドラインに従った methodological uncertainty の例として、英国 NICE の評価のため行われた、F508del 変異ホモ接合の嚢胞性線維症患者における lumacaftor と ivacaftor 併用療法の薬剤経済評価を例示する<sup>14)</sup>。NICE に指定された薬剤は、まずその薬剤を製造販売するメーカー（この例では Vertex Pharmaceuticals 社）から費用対効果評価の結果が提出され、その結果が NICE から委託を受けたグループ（Evidence Review Group; ERG）により批判的吟味される。メーカーは原則として NICE ガイドラインに従う方法による分析（reference case と呼んでいる）を行うこととされており、必要に応じ reference case 以外の方法による評価を補助的に提出することができる。表 1-1 に、ERG によって企業が提出した分析が NICE の reference case と合致していることの確認結果を示す。薬剤経済評価においてもっとも適した方法は状況により変わる可能性はあるが、まずはどのような状況であっても reference case の結果を示すことで、methodological uncertainty に対処をしている。一方で、状況に応じ標準から逸脱した方法を用いた分析結果を示すことが適切と考えられる場合は、その結果を参考資料として総合的な判断が下されることになる。



表 1-1 ERG による reference case とメーカーの提出した分析の突合(参考文献 14 より抜粋)

5.2.1 NICE reference case checklist (TABLE ONLY)

Attribute	Reference case and TA Methods guidance	Does the <i>de novo</i> economic evaluation match the reference case
Comparator(s)	Therapies routinely used in the NHS, including technologies regarded as current best practice for the two populations	The comparator in the decision problem, which is SoC, matches the comparator in the NICE scope. The intervention examined here is intended as an adjunct to SoC, so the comparison is between LUM-IVA with SoC versus SoC alone.
Patient group	As per NICE final scope	People with CF aged 12 years and older who are homozygous for the <i>F508del</i> mutation
Perspective costs	NHS & Personal Social Services	Yes.
Perspective benefits	All health effects on individuals	Yes.
Form of economic evaluation	Cost-effectiveness analysis	Cost-utility analysis.
Time horizon	Sufficient to capture differences in costs and outcomes	Yes (lifetime horizon).
Synthesis of evidence on outcomes	Systematic review	Primary clinical evidence from a meta-analysis of relevant trials
Outcome measure	Quality adjusted life years	Yes.
Health states for QALY	Described using a standardised and validated instrument	Yes. Health states are evaluated using EQ-5D data collected during the clinical effectiveness studies, as well as values from the literature.
Benefit valuation	Time-trade off or standard gamble	The standard UK EQ-5D tariff is used, which is based upon time-trade off.
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	Yes.
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes.
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes.
Probabilistic modelling	Probabilistic modelling	Yes.
Sensitivity analysis		A range of sensitivity and scenario analyses are presented.

ガイドラインにより統一した手法で評価を行うことは一見して薬剤経済評価間の完全な比較可能性を保つように見えるが、手法の統一のみで公平な判断ができるというわけではない。割引率の例を用いると、割引率が高く設定された場合、一時的に費用が多くかかるが将来的には費用が節約されるような治療に対しては、将来に節約される費用の現在価値が比較的少なく評価されることになることから不利に働く。つまり手法の選択の時点で既に治療の特性による有利、不利が発生するのである。このようなことを勘案するため、methodological uncertainty の解決には学問領域を超えた議論が必要であるといえる。

### 1.8.2 Structural uncertainty の評価、対応

手持ちのデータから、長い時間地平や異なる患者集団における疾患の予後を得るため疾患モデルを作成することがしばしば行われる<sup>15)</sup>。疾患モデルは疾患の予後をいくつかの疾患状態により表現し、疾患状態の間を遷移することで予後を表現するが、疾患モデルが実際の予後を正しく表現できていないことによる不確実性を structural uncertainty と呼ぶ。簡単な例を図 1-2 に示す。モデル 1 では疾患を「健康」「入院」「死亡」の 3 つの状態を推移するとしてモデル化している一方、モデル 2 では「入院」が心血管イベントを原因とするか否かで 2 つの状態に分けられており、「健康」「入院（心血管）」「入院（その他）」「死亡」の 4 つの状態を推移するモデルとなっている。心血管イベントを原因とする入院と、その他の理由による入院の間で効果（例えば QW）や治療にかかる費用が大きく変わるため同じ状態であるとはみなしがたい場合、モデル 2 のほうが実際の疾患状態を正しく反映したモデルであると言えるが、一方で図中では矢印であらわされている推定すべきパラメータの数が 4 つから 7 つと 3 つ増加していることがわかる（同じ疾患状態に戻る矢印は除く）。先行研究などの限られた情報からパラメータを推定する場合、モデルを精緻に作りこめば作りこむほどより個別のパラメータの取得は難しくなる。また、モデルの複雑さを増すことにより手持ちのデータを説明する能力は増加するが、未来に生じる現象を予測する能力が増加するとは限ら

ない。疾患状態を増やし、より複雑なモデルとするべきであるか否かは、疾患状態を増やすことにより実際の予後を反映できるメリットと不確実性の増加によるデメリットのいずれが大きいかにより判断する必要がある。Structural uncertainty に対して、いくつかの評価、対処の方法が提案されている。

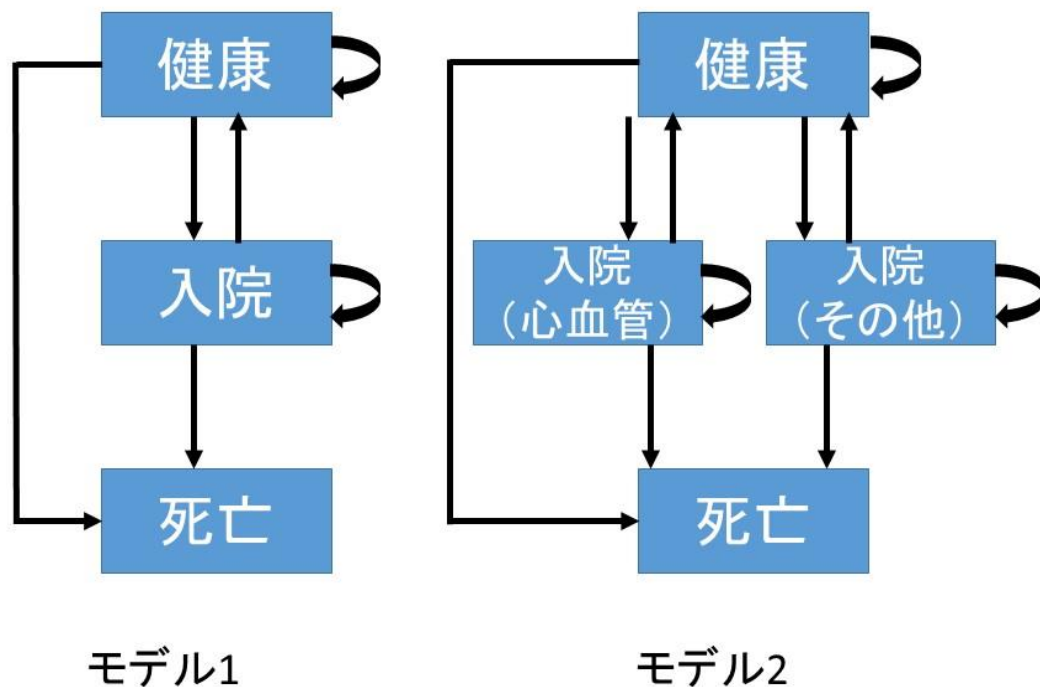


図 1-2 マルコフモデルの例

以後の議論のため、マルコフモデルを Strong らの報告<sup>16)</sup>を参考に定式化する。

- 全ての対象者はいずれかの疾患状態のうちいずれか1つに含まれる
- Cycle length ごとに、一定の確率で今いる疾患状態から別の疾患状態に遷移する
- 疾患状態の遷移確率は、今いる疾患状態のみに依存する

という3つの特徴をもつマルコフモデルでは、治療: $d = (1, 2, \dots, D)$ のもとで時間: $t = (0, 1, \dots, T)$ に疾患状態: $n = (1, 2, \dots, N)$ にいる患者の割合: $\boldsymbol{\pi}_{dt} = (\pi_{d1t}, \pi_{d2t}, \dots, \pi_{dNt})'$ は以下のような式で表すことができる。

$$\boldsymbol{\pi}'_{dt} = \boldsymbol{\pi}'_{d0} \mathbf{M}_d^t$$

ただし、 $\mathbf{M}_d$ は治療 $d$ において1回の cycle で疾患を遷移する確率を示す行列であり、

$$\mathbf{M}_d = \begin{pmatrix} p_{d11} & \cdots & p_{d1N} \\ \vdots & \ddots & \vdots \\ p_{dN1} & \cdots & p_{dNN} \end{pmatrix}$$

で示される。このモデルのもと、さらに CUA を定式化する。治療 $d$ を受けており、疾患状態  $n$  で1回の cycle を過ごした場合に支払う費用と、効果をそれぞれ  $\mathbf{C}_d = (c_{d1}, c_{d2}, \dots, c_{dN})'$ 、

$\mathbf{E}_d = (e_{d1}, e_{d2}, \dots, e_{dN})'$ と示すと、ICER の推定値は

$$\widehat{ICER}_d = \frac{\sum_{t=0}^T \boldsymbol{\pi}'_{d0} \mathbf{M}_d^t \mathbf{C}_{dt}}{\sum_{t=0}^T \boldsymbol{\pi}'_{d0} \mathbf{M}_d^t \mathbf{E}_{dt}}$$

と示すことができる。一方で効果 1 単位あたりに支払ってもよいとする費用  $\omega$  を設定すると、CBA は以下のように示されることがわかる。

$$\widehat{NB}_d = \omega \sum_{t=0}^T \boldsymbol{\pi}'_{d0} \mathbf{M}_d^t \mathbf{E}_{dt} - \sum_{t=0}^T \boldsymbol{\pi}'_{d0} \mathbf{M}_d^t \mathbf{C}_{dt}$$

### 1.8.2.1 シナリオ分析

薬剤経済評価の主たる解析として設定するモデル (base case) を 1 つ選択し、他に考えるいくつかのシナリオによる解析結果と base case による解析結果に大きな乖離や矛盾がないことを確認することで、base case の結論の妥当性と、もし base case が適切ではなく、他のシナリオが正しかった場合も結論には大差がないことを担保する方法をシナリオ分析と呼ぶ。この手法は NICE ガイドライン<sup>6)</sup>等でも推奨されているものである。この方法の優れている点はそのシンプルさにあり、base case と代替のシナリオが臨床的観点からどのように異なっているかが明確にできるところにあるといえる。図 1-2 に示すマルコフモデルの例をみると、すべての入院を同一とみなすモデル 1 を base case とし、心血管イベントを原因とする入院とそれ以外の入院を分けたモデル 2 の解析結果と比較したとき、大きな乖離がみられなければ

入院を2つの状態に分けることが解析結果に大きな影響を与えなかったことおよび、仮にモデル2のほうが適切であったとしても結論には影響しないことが確認される。

### 1.8.2.2 Model selection

前節(1.8.2.1)で説明したシナリオ分析では base case 以外のシナリオはあくまで base case による結論を裏打ちするために準備されるもので、シナリオ分析で大きな問題点が発見されない限り原則として base case の結果を採用する。つまりシナリオ分析においては base case がモデルの構造の観点から妥当であるかについての検証がなされることはない。よって base case の妥当性は臨床的観点からの慎重なレビューにより担保される必要がある。一方で model selection (モデル選択) は、いくつかのシナリオに基づくモデルを、「将来得られるデータをどれだけ説明しているか」、「手持ちのデータにどれだけ適合しているか」という観点から評価し、手持ちのデータをもとに最適なモデルを選択する方法である。

Model selection の基準として赤池情報量 (Akaike's Information Criterion; AIC)<sup>17)</sup>とベイズ情報量 (Bayesian Information Criterion; BIC)<sup>18)</sup>が知られている。いずれもモデルの対数尤度  $f(x|\hat{\theta})$  にモデルが含むパラメータの数  $p$  によるペナルティ (パラメータ数が少ないほうが AIC、BIC が小さくなる補正) を付けたものであり、それぞれ以下の通り示され、いずれも数値が小さいモデルがよりよいと解釈される。

$$\text{AIC} = -2\log\{f(x|\hat{\theta})\} + 2p$$

$$\text{BIC} = -2\log\{f(x|\hat{\theta})\} + p\log(n)$$

BIC の第2項に含まれる  $n$  はサンプル数を示す。AIC にはサンプル数はペナルティに寄与しないため、同一の母集団からのサンプルであっても、サンプル数が増えるとより複雑なモデルが選ばれる。一方で BIC はサンプル数が増加するとペナルティが増加する。複雑なモデルにおいてはサンプル数が増えることで第1項が小さくなるが、第2項のペナルティが増加することでこれが調整され、BIC に基づくモデルの評価はサンプル数が増加しても一貫性が

保たれる。AIC と BIC の使い分けはモデル選択の目的により、未来のデータの予測力が高いモデルを選択したい場合は手持ちのデータの大きさに応じ相応に複雑なモデルが選択される AIC を使用すればよい。一方でモデルの候補の中に 1 つ正しいモデルが必ず含まれているという仮定のもと、そのモデルを特定したい場合、サンプル数により正しいモデルが変わるわけではないので、サンプル数によらず一貫したモデルが選択される BIC を用いる<sup>15)</sup>。

### 1.8.2.2.1 個票データを用いたモデル選択

多くの場合、薬剤経済評価においては未来のデータを予測し、費用と効果を評価することが目的であることから、AIC を用いた model selection について詳細を述べる。AIC は前述のとおりパラメータ数によりペナルティを付けた対数尤度である。尤度を求めるためには、通常であればモデルに患者個人ごとの臨床データ（個票データ）を代入する。例えばある疾患状態からの死亡率を個票データから求める際に、生存関数に指数分布を仮定したモデルとワイブル分布を仮定したモデルのいずれが良いかを選択したいとする。

個票データから生存時間 $(t_1, t_2, \dots, t_n)$ が得られたとき、指数分布、ワイブル分布の尤度はそれぞれ

$$f(t|\lambda) = \prod_{t=1}^n \lambda \exp(-\lambda t)$$

$$f(t|\lambda, p) = \prod_{t=1}^n \lambda p (\lambda t)^{p-1} \exp\{-(\lambda t)^p\}$$

である。

よって最大尤度は

$$f(t|\hat{\lambda}) = \prod_{t=1}^n \hat{\lambda} \exp(-\hat{\lambda} t)$$

$$f(t|\hat{\lambda}, \hat{p}) = \prod_{t=1}^n \hat{\lambda} \hat{p} (\hat{\lambda} t)^{\hat{p}-1} \exp\{-(\hat{\lambda} t)^{\hat{p}}\}$$

と得られる。

パラメータの数が指数分布では 1 つ ( $\lambda$ )、ワイブル分布では 2 つ ( $\lambda, p$ ) であるから、AIC はそれぞれ

$$AIC_{Exponential} = -2\log\{f(t|\hat{\lambda})\} + 2 = -2\log\{\prod_{t=1}^n \hat{\lambda} \exp(-\hat{\lambda}t)\} + 2$$

$$AIC_{Weibul} = -2\log\{f(t|\hat{\lambda}, \hat{p})\} + 4 = -2\log\{\prod_{t=1}^n \hat{\lambda} \hat{p}(\hat{\lambda}t)^{\hat{p}-1} \exp\{-\hat{\lambda}t\}^{\hat{p}}\} + 4$$

と得られる。これらの AIC の値を比較し、小さいほうのモデルを採用する。

### 1.8.2.2.2 論文データを用いたモデル選択

前節のとおり、AIC や BIC は、個票データをもとに算出されるため、個票データがない場合これらを算出することはできない。薬剤経済評価においては先行研究の報告など、論文データを用いた疾患モデルの構築が行われることが多いが、この場合、先行研究の著者などを通して個票データを入手することができればよいが、そうでない場合は論文に報告された要約統計（ハザード比など）を用いてモデルを構築することになる。よって作成したモデルに対して AIC や BIC を算出することはできない。しかしながら、モデル選択においては候補となるモデルの間で AIC や BIC の値がわからずとも、大小が判断できればモデルを選択するのに足る（これらの値が小さいモデルを選択する）。本節では論文から取得可能な推定量と標準誤差を用いて AIC のモデル間における差を求める方法を述べる。モデル  $M_1$  に含まれるパラメータのうちの 1 つ ( $\theta_r$ ) を取り除いたモデル  $M_2$  を考え、モデル  $M_1$  とモデル  $M_2$  の間で AIC を比較する。取り除かれたパラメータ  $\theta_r$  はモデル  $M_1$  では先行研究の論文から得られた最尤推定量と標準誤差を用いているとする。モデル  $M_2$  ではこの  $\theta_r$  は効果が 0 であり、例えば群間差であれば 0、ハザード比であれば 1 と見なしていることに相当する。

モデル  $M_k$  ( $k = 1, 2$ ) の AIC を  $AIC_k$  として、

$$AIC_1 = -2\log\{f(x|\widehat{\theta}_r, \widehat{\theta}_{i \neq r})\} + 2p_1$$

$$AIC_2 = -2\log\{f(x|\widehat{\theta}_{i \neq r})\} + 2p_2$$

とできるから、2つのモデルの AIC の差  $\Delta AIC_{21}$  は以下のように変形され、

$$\begin{aligned} \Delta AIC_{21} &= AIC_2 - AIC_1 = \{-2\log\{f(x|\widehat{\theta}_{i \neq r})\} + 2p_2\} - \{-2\log\{f(x|\widehat{\theta}_r, \widehat{\theta}_{i \neq r})\} + 2p_1\} \\ &= 2\log \frac{f(x|\widehat{\theta}_r, \widehat{\theta}_{i \neq r})}{f(x|\widehat{\theta}_{i \neq r})} + (2p_2 - 2p_1) \end{aligned}$$

モデルM<sub>1</sub>とモデルM<sub>2</sub>の対数尤度比とパラメータ数の差で説明されることがわかる。

ここで、モデルM<sub>1</sub>とモデルM<sub>2</sub>のWald検定のp値 (**p**) が得られれば、Wald検定の性質より  $\chi_1^2(1 - \mathbf{p})$  は漸近的に対数尤度比の2倍に一致するので、

$$\Delta AIC_{21} = \chi_1^2(1 - \mathbf{p}) + (2p_2 - 2p_1)$$

とできる。Wald検定のp値は  $\widehat{\theta}_r$  とその標準誤差  $S_r$  を用いて、

$$\mathbf{p} = \frac{\widehat{\theta}_r}{S_r}$$

と求められる。すなわち、2つのモデルを比較したハザード比とその信頼区間が報告されていれば、点推定値が  $\widehat{\theta}_r$  であり、信頼区間幅から  $S_r$  が導かれるので、上記式に代入すると、モデル間のAICの差が算出できることになる。

実際の研究での活用例を示す。EVER1試験<sup>19)</sup>は腹部大動脈瘤に対するステントグラフト内挿術と開腹手術による人工血管置換術の比較試験であり、この試験をもとにした薬剤経済評価がEpsteinら<sup>20)</sup>が報告している。Jacksonら<sup>15)</sup>は、Epsteinらが評価している8つのシナリオ分析のうち、base caseと結果が同様であった5つのシナリオを除く3つについてAICおよびBICを算出している。図1-3に示す通り、腹部大動脈瘤に対する手術 (Initial aneurysm repair) 後、イベントのない状態 (No event) から、

- 非致死性の冠動脈イベント (Non-fatal CVD event) およびイベント後、回復した状態 (No event (after CVD))
- 非致死性の腹部大動脈瘤再発イベント (Non-fatal AAA event) およびイベント後、回復した状態 (No event (after AAA))

を経て、冠動脈イベントによる死亡 (Death (CVD))、腹部大動脈瘤再発イベントによる死亡 (Death (AAA))、その他理由による死亡 (Death (other)) のいずれかに至るまでをモデル化している。



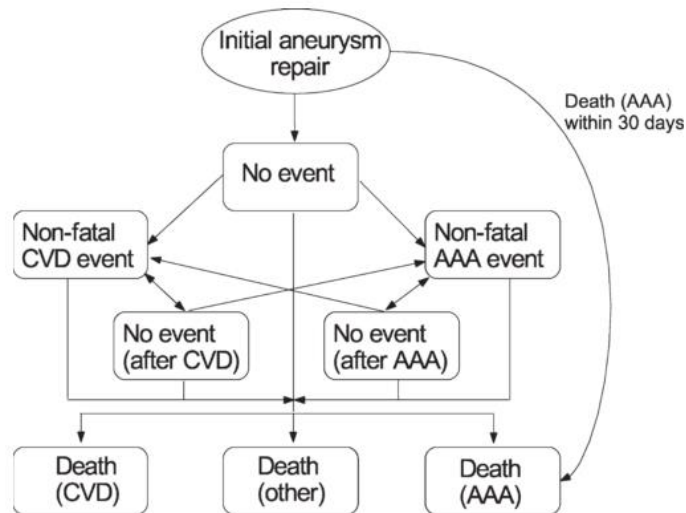


Fig. 1. Representation of the Markov decision model for aneurysm repair: states and permitted transitions between them

図 1-3 腹部大動脈瘤に対する手術後の予後をモデル化するマルコフモデル

(Jackson ら<sup>15)</sup> より引用)

Epstein らによるシナリオ分析の結果のうち、Jackson らによる評価が行われたシナリオは base case に加え以下の 3 つのシナリオである。

- (a) Base case において、EVER1 試験の対象患者は一般の集団に対して冠動脈イベントによる死亡リスクが高く、ハザード比 2.00 (95%信頼区間 0.83 - 4.83) であると設定しているが、この設定は誤りで、EVER1 の対象患者と一般の集団での冠動脈イベントによる死亡のハザード比は 1 であるというシナリオ。
- (b) Base case において、ステントグラフ内挿術は開腹手術よりも冠動脈イベントによる死亡リスクが高く、ハザード比 3.06 (95%信頼区間 1.12 - 8.36) であると設定しているが、この設定は誤りで、ステントグラフ内挿術と開腹手術での冠動脈イベントによる死亡のハザード比は 1 であるというシナリオ。
- (c) Base case において、ステントグラフ内挿術は開腹手術よりも腹部大動脈瘤再発イベントによる死亡リスクが高く、ハザード比 5.84 (95%信頼区間 0.70 - 48.50) であると設定しているが、この設定は誤りで、ステントグラフ内挿術と開腹手術での腹部大動脈瘤再発イベントによる死亡のハザード比は 1 であるというシナリオ。

Base case ではステントグラフ内挿術は開腹手術よりも費用は高く、得られる QALY が少ない dominated の結果となったが、これらのシナリオの結果ステントグラフ内挿術は費用はより高いが得られる QALY も高くなり、それぞれ ICER は 239,000、42,000 および 48,000 ポンド/QALY と得られた。Jackson らはこれらシナリオと base case との間の AIC と BIC の差を算出した。結果を表 1-2 に示す。

表 1-2 Base case に対する各シナリオの AIC、BIC の差 (Jackson ら<sup>15)</sup> より引用、日本語に改変)

パラメータ	ハザード比 (95%信頼区間)	両側p値	症例数 n	情報量基準 の差		共変量を含む モデルの事後確率	
				AIC	BIC	AIC	BIC
(a)一般集団とEVAR1試験の 対象患者の冠動脈イベント による死亡リスクの差	2.00 (0.83, 4.83)	0.123	1139	0.37	-4.6	0.546	0.090
(b)冠動脈イベントによる死亡 に対する治療効果	3.06 (1.12, 8.36)	0.029	1016	2.79	-2.1	0.801	0.256
(c)腹部大動脈瘤再発による 死亡に対する治療効果	5.84 (0.70, 48.50)	0.102	1016	1.83	-3.1	0.714	0.175

いずれのシナリオも base case よりも AIC が増加していることから、AIC の基準によると base case がこれらシナリオよりもよいモデルであることがわかる。

この方法が薬剤の評価においても有用であることを示すため、統計解析ソフトウェアである R<sup>21)</sup>の survival パッケージ<sup>22)</sup>にサンプルデータ (ovarian) として提供されている卵巣がんデータ<sup>23)</sup>を例に、Wald 検定の p 値と AIC を実際に算出し、Wald 検定の p 値のみからでもモデル間の AIC の差が求められることを示す。サンプルデータ ovarian は、ステージ IIIB および IV の進行卵巣がん患者に対し、cyclophosphamide 単剤と、cyclophosphamide に adriamycin を併用した場合の生存を比較した試験のデータである。サンプルデータには生存時間、打ち切り/死亡の別、年齢、残存病変の有無、治療群、ECOG PS (Eastern Cooperative Oncology Group Performance Status)<sup>24)</sup>が格納されているが、ここでは生存時間を治療群と ECOG PS により説明するモデルと、治療群のみにより説明するモデルの 2 つを作成、前者のモデルにおける ECOG PS の Wald 検定の p 値を用いて AIC の差を実際に算出する (表 1-3)。

表 1-3 データ ovarian による Cox 回帰の例

	HR	95%信頼区間	p 値
治療、ECOG PS によるモデル			
治療	0.5609	0.1772; 1.775	0.325
ECOG PS	1.4474	0.4582; 4.573	0.529
治療のみによるモデル			
治療	0.5508	0.1743; 1.740	0.310

まず、Wald 検定の  $p$  値から AIC の差を求める。 $p = 0.529$ 、 $p_1 = 2$ （治療、ECOG PS）、 $p_2 = 1$ （治療）であるので、治療、ECOG PS によるモデルの AIC ( $AIC_1$ ) と治療のみによるモデルの AIC ( $AIC_2$ ) の差  $\Delta AIC_{21}$  は、

$$\begin{aligned}\Delta AIC_{21} &= \chi_1^2(1 - p) + (2p_2 - 2p_1) \\ &= \chi_1^2(1 - 0.529) + (2 - 4) \\ &= -1.60\end{aligned}$$

となった。このケースでは生データを利用可能であるので、それぞれの AIC を算出したところ、治療と ECOG PS を用いたモデルの AIC ( $AIC_1$ ) は 72.52、治療のみによるモデルの AIC ( $AIC_2$ ) は 70.92 であったので、差を取ると  $\Delta AIC_{21} = 70.92 - 72.52 = -1.60$  となり、Wald 検定の  $p$  値から求めた  $\Delta AIC_{21}$  と、生データから算出した  $\Delta AIC_{21}$  が一致していることが確認された。このことから、生データを入手できない限り算出できない AIC ( $AIC_1$  と  $AIC_2$ ) について、生データが無い場合においても Wald 検定の  $p$  値が文献から得られれば、AIC の差である  $\Delta AIC_{21}$  を算出することは可能であることが確認された。

### 1.8.2.3 Model averaging

いくつかの候補シナリオをもとに作成したモデルの中から、AIC や BIC によって最も良いモデルを特定し、その結果を採用するのが model selection であったが、疾患の予後という複雑な事象をモデル化している以上、最も良いモデルであっても単一のモデルにより完全に疾患の予後を説明することは難しい。Model averaging とは、作成した複数のモデルについて、データをもとにその妥当性を定量的に評価し、その妥当性を重みとしてモデルから得られる

出力（CUA の場合は ICER）の加重平均を求める方法である。Buckland ら<sup>25)</sup>や Burnham ら<sup>26)</sup>の報告において、AIC から

$$A_k = \exp(-0.5AIC_k)$$

を求め、それぞれのモデルの  $A_k$  をもとに

$$p(M_k|x) = \frac{A_k}{\sum_k A_k}$$

を加重平均の重みとして用いている。

$A_k$  は各モデルの AIC から算出されるが、以下の通り各モデル間の差  $\Delta AIC$  から算出可能であることから、前節の方法により論文データにより  $\Delta AIC$  のみが得られる場合でも応用が可能である。

$$\begin{aligned} \frac{A_k}{\sum_i^n A_i} &= \frac{\exp(-0.5AIC_k)}{\exp(-0.5AIC_k) + \sum_{j \neq k} \exp(-0.5(AIC_j + \Delta AIC_{jk}))} \\ &= \frac{\exp(-0.5AIC_k)}{\exp(-0.5AIC_k) + \exp(-0.5AIC_k) \sum_{j \neq k} \exp(-0.5\Delta AIC_{jk})} \\ &= \frac{1}{1 + \sum_{j \neq k} \exp(-0.5\Delta AIC_{jk})} \end{aligned}$$

データから得られるこのような重みによって候補モデルの加重平均をとることにより、より妥当なモデルの結果に大きな重みを、妥当性の低い結果に小さな重みを与え、平均的により良いモデリングを実現することができる。候補モデルの中に正しいモデル（そのモデルが単独で真の値を導くことができる）が含まれる場合を **M-closed** と呼び<sup>27)</sup>、データが無限に得られた場合に **model averaging** の手法が正しいモデルに全ての重みを与え、その結果真の値を導くのは、この場合に限られる。一方で候補モデルのすべてが現実と乖離している場合（**M-open** の場合）においてはデータが無限に得られたとしても真の値を導くことはできない。前述のとおり単一のモデルにより複雑な疾患の予後を完全に説明することは難しく、ほとんどの薬剤経済評価は **M-open** の場合であるといえる。よって **model averaging** の利点は、臨床的

観点からいずれも妥当であると判断される複数のモデルから、いずれのモデルが正しいかを完全に特定することなく平均的に良い値を獲得することが出来る、という点に限定される。

#### 1.8.2.4 Discrepancy approach

1.8.2.3 で述べた model averaging は、実際のアウトカムがモデルによりどれだけうまく説明されるかを評価した結果によりモデルの重みづけを行う data driven な手法である。しかしながら、薬剤経済評価においてしばしばモデルを作成する目的は測定されていないアウトカム（例えば2年間のアウトカムしか得られていない状況における10年後のアウトカム）を推定することであり<sup>26,28)</sup>、実際のアウトカムのデータが得られている状況は薬剤経済評価においてはむしろ稀である。一般的な臨床試験やアウトカム研究によるアウトカムの収集は倫理的課題や実現可能性の問題からある程度の期間に限定されるのに対し、薬剤経済評価においてガイドラインが推奨する評価期間は相当に長い期間、時に生涯にわたる評価が求められるため、真のアウトカムのデータのない中モデルの評価を行わなければならない状況がしばしば発生する。Strong らは、真のアウトカムデータのない状況において適応可能な方法として、疾患の予後やプロセスに関する知識をもとにする law driven な方法である discrepancy approach による解決を提案している<sup>29)</sup>。

データ（のベクトル） $X$ によってアウトカム $Z$ を説明するモデルを作成した時、モデルのアウトプット $Y = f(X)$ とアウトカム $Z$ との間の discrepancy（乖離）は structural error  $\delta_z$ によって以下の式で記述される。

$$Z = f(X) + \delta_z$$

この $\delta_z$ を調べるのが discrepancy approach である。

Strong ら<sup>29)</sup>の報告では discrepancy approach の実例として NICE による運動介入の費用対効果評価を用いた例が示されている<sup>30)</sup>。図 1-4 に費用対効果評価で用いる決定樹を示す。□は decision node を、○は chance node を示す。最初の運動介入の有無の decision（図左）により、その後運動を行う確率および運動が継続される確率が変動する（図中央）。運動の有無、継

続の有無は、その後の CHD、脳卒中、糖尿病の発症確率変動する（図右）というモデルである。

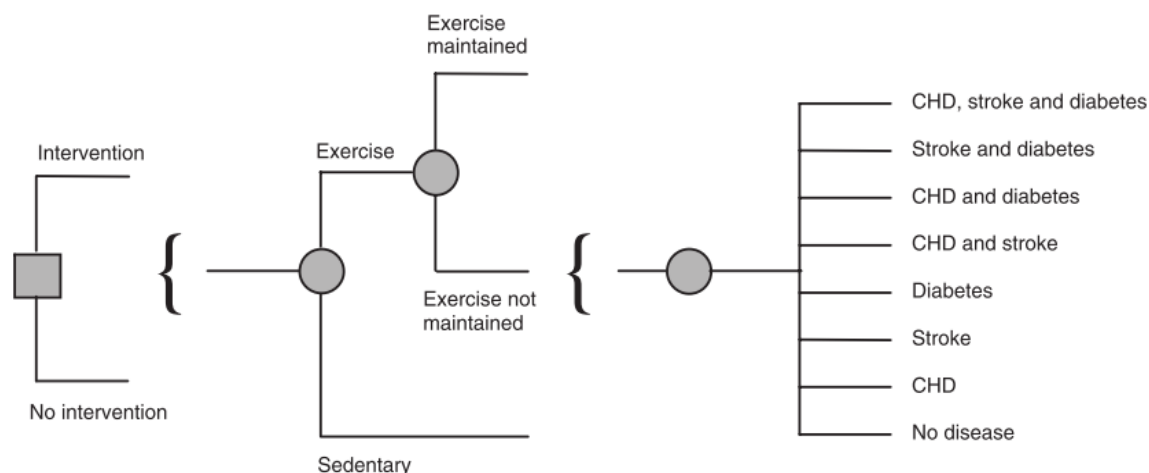


図 1-4 運動介入の費用対効果評価の決定樹

このモデルが導く最終的なアウトカムは金銭に換算した利得

$$Z = \lambda(E_2 - E_1) - (C_2 - C_1) = \lambda\Delta E - \Delta C$$

であるとする。ただし、 $E_d$ は運動介入 $d = \{1: 介入なし, 2: 介入あり\}$ により得られる平均的な効果（QALY）、 $C_d$ は平均的な費用、 $\lambda$ は WTP である。

このモデルに対応するデータは、例えば運動介入を受けた人が運動を始める割合、運動の有無ごとの疾患発症率、疾患ごとの医療費などが考えられるが、これらについて十分な数のデータを集め、モデルにより $f(X)$ を得たとして、最終的なアウトカム $Z$ 、すなわち運動介入の有無でどれだけの利得が発生するか（あるいは失われるか）との乖離を評価することで、モデルの改良が可能である。しかしながら、前述のとおり $Z$ は未測定であることがほとんどであるため、 $Z$ と $f(X)$ の差をとることはできない。更に、金銭に換算した利得は、もともとのデータから複雑なプロセスを経た結果として現れるため、データと見比べた際に臨床的観点からもっともらしい値になっているかを判断することさえも容易ではない。このような課題を解決するため、Strongらはモデルをいくつかの subfunction に分割し、分割された subfunction の与える中間変数ごとに乖離を評価する方法（Strongらはこの方法を internal

discrepancy approach と呼んで最終的なアウトカムZからの乖離を評価する external discrepancy approach と区別している) を提案している。

モデル $f(X)$ の subfunction への分割は以下のような例を用いて説明される。

$$Y = f(X_1, \dots, x_7) = \left\{ (X_1 X_2 + X_3 X_4) \left( \frac{1}{1 + X_5} \right)^{-X_6} \right\} - X_7$$

は、例えば $Y_1 = f_1(X_1, \dots, X_4) = X_1 X_2 + X_3 X_4$ 、 $Y_2 = f_2(X_5, X_6) = \left\{ \frac{1}{1 + X_5} \right\}^{-X_6}$  および $Y_3 = f_3(Y_1, Y_2, X_7) = Y_1 Y_2 - X_7$ の通り $f_1, f_2, f_3$ の3つの関数の組み合わせに分けることが可能である。これにより7つの入力ベクトル $X = (X_1, \dots, X_7)$ から3つの中間パラメータ $Y_1, Y_2, Y_3$ が得られた。Subfunction をそれぞれ「subfunction に真の値を入れた場合に、subfunction の出力もまた真の値であるか」という観点から評価し、もしモデルの制約により真の値を入力しても出力が真の値から乖離する場合は、discrepancy  $\delta_i$ を導入する。Discrepancy の導入は $Y_i = f_i(\cdot) + \delta_z$ のように加算の形をとっても、 $\log(Y_i) = \log\{f_i(\cdot)\} + \log(\delta_i)$ のような乗算の形をとってもよい。この結果、モデルは以下のように書き直される。

$$Y = f^*(X_1, \dots, x_7) = \left\{ (X_1 X_2 + X_3 X_4 + \delta_1) \left( \left( \frac{1}{1 + X_5} \right)^{-X_6} + \delta_2 \right) \right\} - X_7 + \delta_3$$

乖離が起きると見なされ、discrepancy の導入が必要である subfunction を特定した後に、それぞれの $\delta_i$ がどの程度の幅で変動するかを $\sqrt{\text{Var}(\delta_i)} = v_i |E(Y_i)|$ のように中間パラメータ $Y_i$ に対する比の形で決定し、また $\delta_i$ が従う分布を設定する。

$\delta_i$ について設定した分布に基づき、 $\delta_i$ を1つずつ固定し、それ以外のすべての discrepancy  $\delta_{-i}$ を変動させるシミュレーションを行うことで、 $\delta_i$ を与えたもとでのZの分散の低下割合  $\text{Var}_{\delta_i}\{E(Z|\delta_i)\}/\text{Var}(Z)$ を得る ( $\delta_i$ が $\delta_{-i}$ と独立であれば固定する $\delta_i$ はその取りうる範囲の中であればどの値をとっても結果は変わらないが、独立でない場合は $p(X, \delta_{-i}|\delta_i)$ の分布からのサンプリングを行う必要があるが、この分布は未知である。この問題へ対応するアルゴリズムは Strong らの報告の Appendix A を参照されたい<sup>29)</sup> )。Strong らはこれを'main effect index'と呼んでいるが、main effect index を比較することで、 $\delta_i$ のうちいずれがZのばらつきに寄与す

るか、すなわちどの中間パラメータ $Y_i$ を導く subfunction  $f_i$ がモデルの中でもっとも改善の余地があるポイントであるかを明らかにすることができる。また、同様に $\delta$ 所与のもとでの $Z$ の分散と、データ $X$ 所与のもとでの $Z$ の分散の比 $\text{Var}_\delta\{E_X(Z|\delta)\}/\text{Var}_X\{E_\delta(Z|X)\}$ を'structural parameter uncertainty ratio'と呼び、これを評価することで、薬剤経済評価全体としての不確実性がデータによるものであるのか、モデルによるものであるかを判断することができる。データによる不確実性に対して、モデルに起因するものの割合が少ないことが分かれば、そのモデルは現在手持ちのデータに比して十分な精度を持っていると結論づけることができる。

Strong らの例において discrepancy approach を用いたモデルの改善により薬剤経済評価のばらつきが軽減された様子を図 1-5 に示す。運動介入に関する図 1-4 の決定樹モデルに基づき、入力データの不確実性のみについて感度分析を行った結果が実線で示されている（感度分析については 1.8.3 節にて述べる）。このモデルを subfunction に分解し、それぞれの中間変数につき discrepancy  $\delta$ を導入した結果、入力データとモデルの両方の不確実性により破線の如くばらつきは増加している。この後すべての $\delta_i$ について main effect index を調べた結果を表 1-4 に示す。中間変数 $p_{dj}$ に相当する $\delta_{p_{dj}}$ がアウトカムにおいて比較的大きな影響を持つことが分かる。中間変数 $p_{dj}$ は運動の有無 ( $d$ ) 別の、特定の疾患 ( $j$ ) を発症する確率であり、これはすなわち作成したモデルにおいては、疾患発症確率にあたる箇所のモデリングが実際の機序から大きく乖離していることを示唆する。よって、疾患発症確率に関するモデルの見直しを行い、運動の有無だけでなく運動継続年数によっても疾患発症確率が決まるようにモデルを改変した結果が、図 1-5 において一点鎖線 (・-・-) で示されている。実線で示される最初のモデルよりも、不確実性の減少が見られていることがわかる。本手法はモデルの構造を、疾患に基づく知見を元に修正し structural uncertainty を改善する有効な方法であるが、薬剤経済評価においては使用例が見られない。今後薬剤経済評価への応用が期待される。



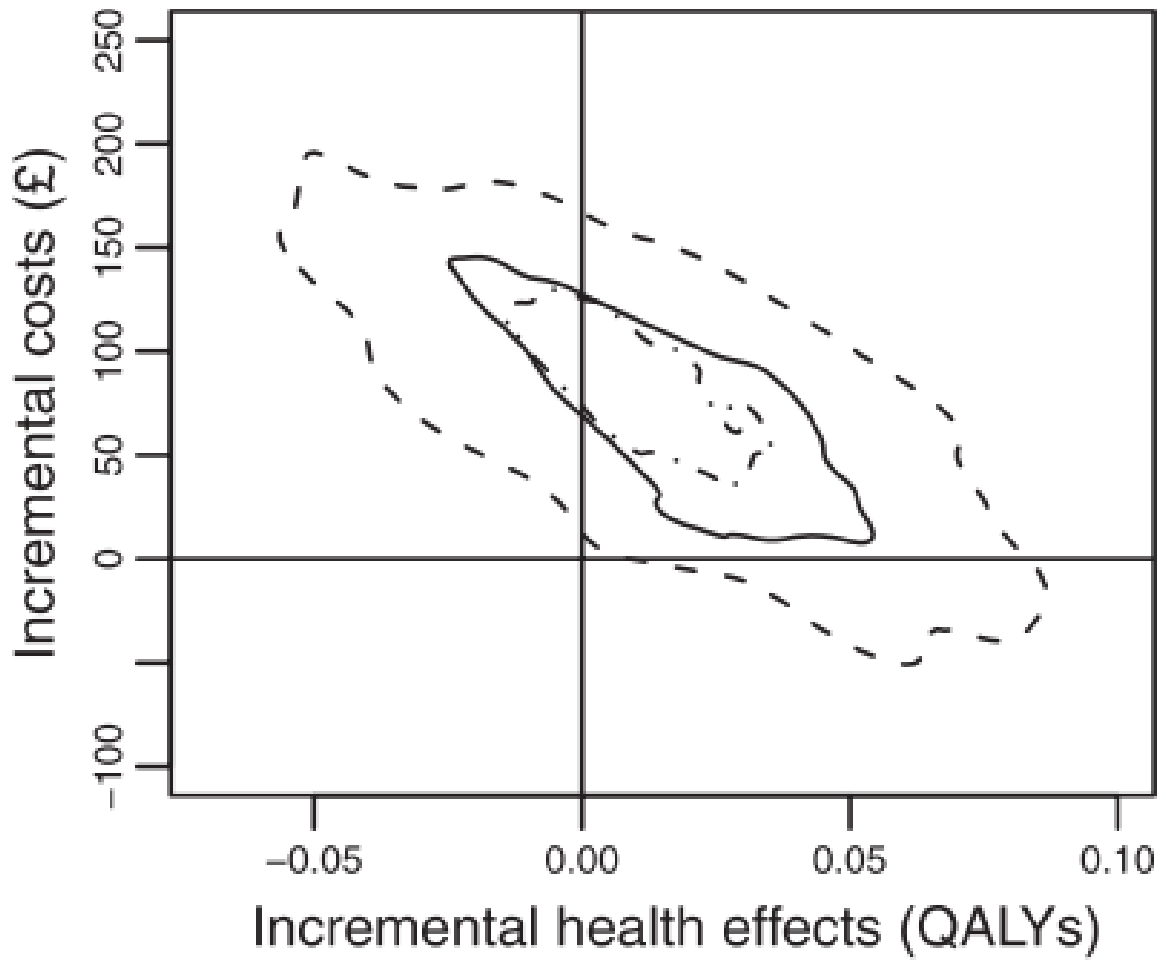


図 1-5 Discrepancy approach によるモデル改善の結果 (参考文献<sup>29)</sup> より抜粋)

表 1-4  $\delta$ ごとの main effect index

<i>Discrepancy</i>	<i>Main effect</i>	<i>Discrepancy</i>	<i>Main effect</i>	<i>Discrepancy</i>	<i>Main effect</i>
$\delta_{r_1}^{(ex)}$	0.002	$\delta_{p1,1}$	<i>0.266</i>	$\delta_{c_{chd}}^*$	0.002
$\delta_{r_2}^{(ex)}$	0.002	$\delta_{p1,2}$	<i>0.128</i>	$\delta_{c_{str}}^*$	0.002
$\delta_{r_3}^{(ex)}$	0.003	$\delta_{p1,3}$	<i>0.076</i>	$\delta_{c_{dm}}^*$	0.001
$\delta_{r_4}^{(ex)}$	0.002	$\delta_{p1,4}$	0.002	$\delta_{d_{chd}}$	0.002
$\delta_{r_5}^{(ex)}$	0.003	$\delta_{p1,5}$	<i>0.054</i>	$\delta_{d_{str}}$	0.002
$\delta_{r_6}^{(ex)}$	0.002	$\delta_{p1,6}$	0.025	$\delta_{d_{dm}}$	0.002
$\delta_{r_7}^{(ex)}$	0.003	$\delta_{p1,7}$	0.014	$\delta_{q_5}$	0.002
$\delta_{r_8}^{(ex)}$	0.004	$\delta_{p1,8}$	0.010	$\delta_{q_6}$	0.002
$\delta_{r_1}^{(sed)}$	0.002	$\delta_{p2,1}$	<i>0.257</i>	$\delta_{q_7}$	0.002
$\delta_{r_2}^{(sed)}$	0.002	$\delta_{p2,2}$	<i>0.124</i>	$\delta_{q_8}$	0.002
$\delta_{r_3}^{(sed)}$	0.002	$\delta_{p2,3}$	<i>0.076</i>	$\delta_{c_5}$	0.002
$\delta_{r_4}^{(sed)}$	0.002	$\delta_{p2,4}$	0.002	$\delta_{c_6}$	0.002
$\delta_{r_5}^{(sed)}$	0.002	$\delta_{p2,5}$	0.049	$\delta_{c_7}$	0.002
$\delta_{r_6}^{(sed)}$	0.002	$\delta_{p2,6}$	0.025	$\delta_{c_8}$	0.002
$\delta_{r_7}^{(sed)}$	0.002	$\delta_{p2,7}$	0.013	$\delta_{\Delta q}$	0.003
$\delta_{r_8}^{(sed)}$	0.002	$\delta_{p2,8}$	0.008	$\delta_{\Delta c}$	0.001

†Values greater than 5% are shown in italics.

### 1.8.2.5 Expected Value of Perfect Information (EVPI)

前節においては、モデル全体をいくつかの **subfunction** に分解し、それらにより導かれる中間変数が真の値からどれだけ乖離するかを評価し、その乖離が全体の不確実性のうちで占める割合 **main effect index** を比較することでモデルの中で改善を要する箇所を特定、改善することにより **structural uncertainty** を解消する方法を紹介した。中間変数ごとの真の値からの乖離を評価する尺度は、**main effect index** に限られるものではなく、**Expected Value of Perfect Information (EVPI)** により評価を行う方法も Strong らの報告において提案されている<sup>16)</sup>。EVPIは **structural uncertainty** の評価に限らず後述する **parameter uncertainty** の評価においても

利用が可能な、薬剤経済評価の不確実性を定量するための有用な尺度であるため、本節にその詳細を示す。

EVPIの基となる Value of Information (VoI) は、ある情報についてそれを追加的に得られたとき、その情報をもたらす意思決定上の価値のことであり<sup>31)</sup>、統計的意思決定論において体系立てられた後<sup>32, 33)</sup>、工学や環境リスク分析へ応用されている<sup>34, 35, 36)</sup>。この中で EVPI は、あるパラメータについて完全な情報が得られた場合に得られる価値の期待値である。薬剤経済評価は、多くのパラメータに関する情報を集約し、いずれの意思決定（新治療を採用するか、対照治療の使用を継続するか）が最適であるかを判断するが、情報には不確実性が含まれることから、その判断が必ずしも正しいとはいえない。新治療の採用を支持するデータはあくまで偶然に得られたもので、多くのデータが収集された後においては新治療採用の判断が誤りであったということが事後的にわかるようなことは、情報の不確実性の大きさによりその頻度は異なるものの不可避なリスクであるといえる。この枠組みの中において、あるパラメータに関する不確実性が完全に取り除かれた場合、正しい判断に至る確率が高まることを通して価値が生じるといえる。この価値の期待値を EVPI と呼ぶ。

EVPIを説明するための例として、手持ちの現金を株式か債券に投資する際に、一定期間後に得られる利益が景気動向（好況/不景気）に左右されるようなケースを考える。景気別に株式と債券の投資から得られる利益は事前に予測可能であり、株式は好況の場合に大きな利益を生むが不景気の場合に大きな損失を被ること、一方で債券は好況の場合に得られる利益は株式に劣るが、不景気の場合でも利益は得られるとし、また好況になる確率は60%であることもわかっているとす。手持ちの情報をまとめると表 1-5 の通りであるが、この情報から株式と債券のいずれに投資すればよいかを利益の期待値のみから判断すると、株式に投資した場合の利益の期待値は  $10,000 \times 0.6 + (-5,000) \times 0.4 = +4,000$  円となり、一方で債券に投資した場合は  $3,000 \times 0.6 + 1,000 \times 0.4 = +2,200$  円となり、利益の期待値が高い株式に投資する、という結論に至るであろう。そしてこの投資から得られる利益の期待値は+4,000円である。

表 1-5 投資判断に利用可能な情報

		景気	
		好況 60%	不景気 40%
投資先	株式	+¥10,000	-¥5,000
	債券	+¥3,000	+¥1,000

ここで仮に、将来の景気を間違いなく予測できるとして、この予測はいくらの価値を持つだろうか、というのが EVPI の考え方である。すなわち、景気動向について「好況になる確率が 60%、不景気になる確率が 40% である」という情報はあるものの、実際にどちらになるかはわからないが、これらのうちどちらが実現するかがはっきりする予測はこの投資の利益を期待値でいくら増加させるだろうか、ということである。

この予測の結果、「好況」と出れば株式に投資し、不景気と出れば債券に投資すれば最大の利益が得られるため、この投資の利益の期待値は、 $10,000 \times 0.6 + 1,000 \times 0.4 = +6,400$  円であることが分かる。景気に関する予測を得ることで、不景気になるとの予測が出た場合に株式ではなく債券に投資する、という判断に切り替えることで、+4,000 円の利益よりも 2,400 円多い +6,400 円の利益が期待されることになる。よってこの予測には 2,400 円の価値があるといえるが、これが（景気動向に関する情報の）EVPI である。EVPI は

- 情報の不確実性の大きさ

ほぼ確実に好況になるようなケースにおいて、景気動向に関する情報の EVPI は低い

- 情報が意思決定に与える影響度

不景気になったとしても株式から得られる利益が債権と大差がない場合、景気動向に関する情報の EVPI は低い

に依存する。薬剤経済評価において多くのパラメータについて EVPI を算出することが出来るが、これらを算出することにより意思決定に対するパラメータの影響を判断することが出

来る。EVPIが大きなパラメータは、手持ちのデータでは不確実性が高いか、あるいは意思決定に大きな与えるパラメータであるということである。

薬剤経済評価においてEVPIはいくつかの用途があるが、その一つが、1.8.2.4節で論じたモデルの中間変数の評価の尺度としての活用である。EVPIの大きい中間変数を導く

subfunctionはより大きなstructural uncertaintyを含んでいるため、改善の余地が大きい。

Strongらの2014年の報告<sup>16)</sup>では、Strongらの2012年の報告<sup>29)</sup>で単純な決定樹モデルに適用していたdiscrepancy approachをより複雑なマルコフモデルに拡張するに伴い、discrepancy term  $\delta$  の評価にmain effect indexに代わりEVPIを活用している。ある中間変数に関する情報が完全に得られた場合に追加的に得られる効用がEVPIであるから、これは当該箇所のsubfunctionの改善により得られる効用の最大値であるともいえる。よってstructural uncertaintyの議論においてEVPIはExpected Value of Model Improvement (EVMI)とも呼ぶことが可能である。

### 1.8.3 Parameter uncertainty の評価、対応

Parameter uncertaintyとは、薬剤経済評価で推定されるパラメータの不確実性であり、その発生の理論は臨床試験の結果の不確実性と同様である。臨床試験においては、よくデザインされた試験デザインと適切な症例数設計により不確実性をコントロールすることは可能であるが、既存の論文データなどをもとにパラメータを推定することの多い薬剤経済評価では、しばしばパラメータの推定が精度よく行えない局面が発生する。しかしながら、薬剤経済評価の真の目的は、適切な意思決定を行うことにあり、パラメータを精度よく推定することはその手段に過ぎない。パラメータの不確実性を定量、評価し、その大きさにより適切な判断を行うことが重要となる。本章では、parameter uncertaintyを定量する方法について述べる。

薬剤経済評価の信頼性を評価するという本来の目的とは異なるが、parameter uncertaintyを評価することにより、薬剤経済評価の結果に大きな影響を与えるパラメータを抽出すること

も可能である。例えば新薬の臨床開発の途中、第 II 相臨床試験データが得られた時点で薬剤経済評価を行い、ICER に大きな影響を与えるパラメータを特定することができれば、そのパラメータを重点的に取得するような第 III 相臨床試験を設定することでより良い薬剤経済評価に繋がるであろうし、開発戦略や製品プロファイルの改善材料にも用いることができる。

#### 1.8.3.1 確率によらない感度分析 (deterministic sensitivity analysis)

1次元感度分析は、各パラメータを一定の範囲で変動させた時に、ICER がどのように変化するかを評価するものである。1次元感度分析により、パラメータの不確実性が ICER にどれだけ大きな影響を与えるかを評価することができる。いずれのパラメータを臨床的に妥当と判断される範囲で変動させたとしても、ICER が意思決定上許容可能な範囲に収まるのであれば、その薬剤経済評価の結果と、そこから判断される意思決定は相当の信頼性を持って受け入れられるといえる。

1次元感度分析においてはパラメータ間の相互作用については考慮されないが、2つのパラメータの組み合わせを同時に変化させることにより相互作用による想定外の不確実性がないかを評価することができ、これを2次元感度分析と呼ぶ。3つ以上のパラメータを同時に変化させることにより、3次元以上の感度分析を行うことも理論上は可能ではあるが、紙上に図示することが難しく、パラメータの組み合わせ数が膨大になることから現実的ではない。よって平面上に ICER の等高線図で示すことが可能な2次元感度分析まで行われるのが通常である。これらのように、パラメータの変動幅を決定して行う感度分析を *deterministic sensitivity analysis* と呼ぶ。

#### 1.8.3.2 確率的感度分析 (probabilistic sensitivity analysis)

確率的感度分析とは、シミュレーション実験による不確実性の評価である。ある仮定の下で乱数に基づきパラメータを発生させ、それらの仮定の値の組み合わせから ICER を算出す

る。これを何度も繰り返すことで、パラメータが持つ不確実性が最終的に算出される ICER の不確実性にどのように反映されるかを評価することができる。薬剤経済評価において費用と効果はしばしば多くのパラメータによって説明されることから、これらの不確実性を同時に評価することのできる確率的感度分析は薬剤経済評価の質評価において重要な役割を果たす。また、Claxton, 2008<sup>37)</sup>において指摘されているように、マルコフモデルを含む多くのモデルにおいてパラメータと ICER の関係は線形ではないことから、例えば 1 次元感度分析で適切なパラメータの範囲を設定して ICER の取りうる範囲を得たととしても、その範囲の中で ICER が対称に分布しているのか、あるいは大きな値や小さな値に局在しているのかを知ることにはできない。確率的感度分析を行うことにより ICER の分布を得ることができるため、より適当な意思決定につなげることができる。

以後、確率的感度分析の最も一般的な方法であるモンテカルロシミュレーションによる確率的感度分析の例を示す。臨床試験や臨床研究の結果、比較する治療間の効果と費用の差について表 1-6 の通り得られたとする。この例では新治療は既存治療と比べて効果においては 0.7QALY 分すぐれているが、費用も 300 万円多くかかり、それぞれの標準誤差は 0.2QALY 及び 100 万円である。更に効果と費用の相関係数は 0.5 であるとして、効果が高い結果が得られたとしたらそれにかかる費用も相応に高いという傾向がみられると考える。

表 1-6 モンテカルロシミュレーションの設定例

新治療－既存治療	点推定値	標準誤差	相関係数
効果(QALY)	0.7	0.2	0.5
費用(円)	3,000,000	1,000,000	-

効果、費用が表 1-6 のパラメータを持つ二変量正規分布に従うと仮定し、シミュレーションにより 1,000 個のデータを発生させた結果を図 1-6 に示す。この設定のもとでは、当然効果と費用の群間差はそれぞれ 0.7、3,000,000 を中心に分布し、右上がりの中程度の効果と費

用の間に中程度の相関がみられること、効果の群間差が負の値をとったり、費用の群間差が負の値をとること、つまり新治療が既存治療より効果が劣っていたり、費用が安かったりする結果が得られる確率がほとんどないことを見て取ることができる。

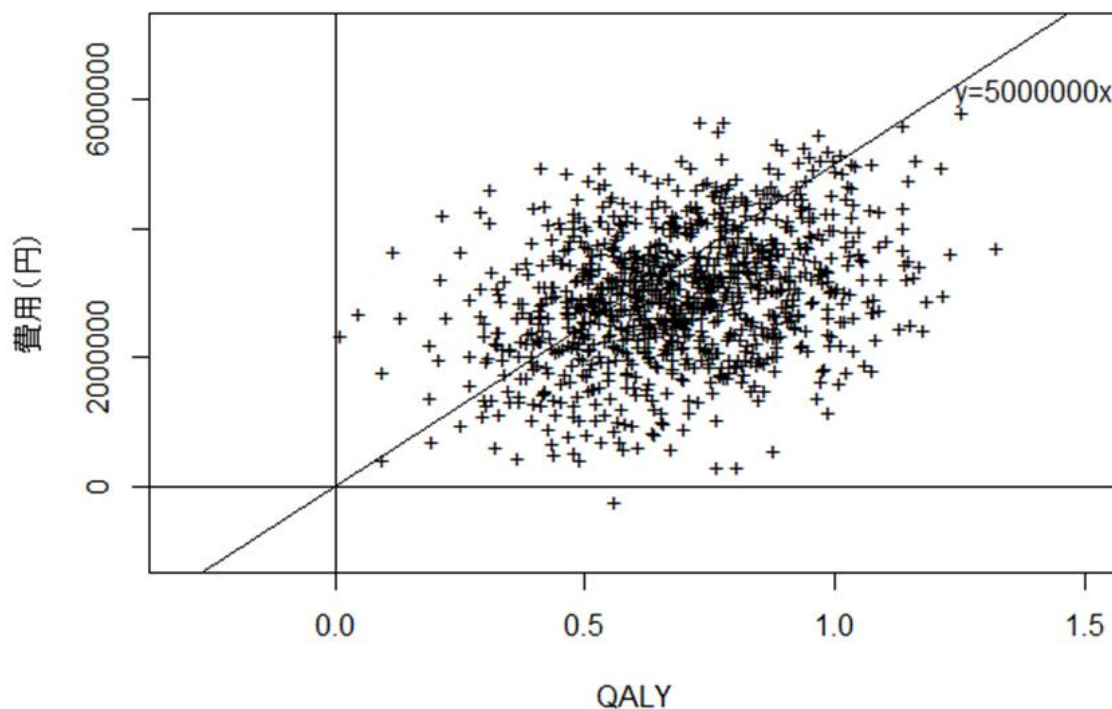


図 1-6 モンテカルロシミュレーションの結果を費用効果平面にプロットした例

確率的感度分析の結果の示し方は、図 1-6 のように費用効果平面上のプロット（しばしばある一定の支払い意思額を示す直線が参照として示される）や、各象限に現れたプロットの割合を示すことが一般的である。一方でより意思決定に近い表現として、ある支払い意思額のもとでの新治療の受け入れ可能性がどれだけ高いかを示す、Cost-effectiveness acceptability curve (CEAC)を示すことも有効である。図 1-7 に図 1-6 の例をもとにした CEAC を示す。X 軸に支払い意思額を、Y 軸に ICER が支払い意思額を下回ったプロットの割合を「新治療の受け入れ確率」として示している。CEAC を見ることで、支払い意思額がある値であった場合に新治療の受け入れ確率がどれだけ高いかを評価し、意思決定につなげることができる。



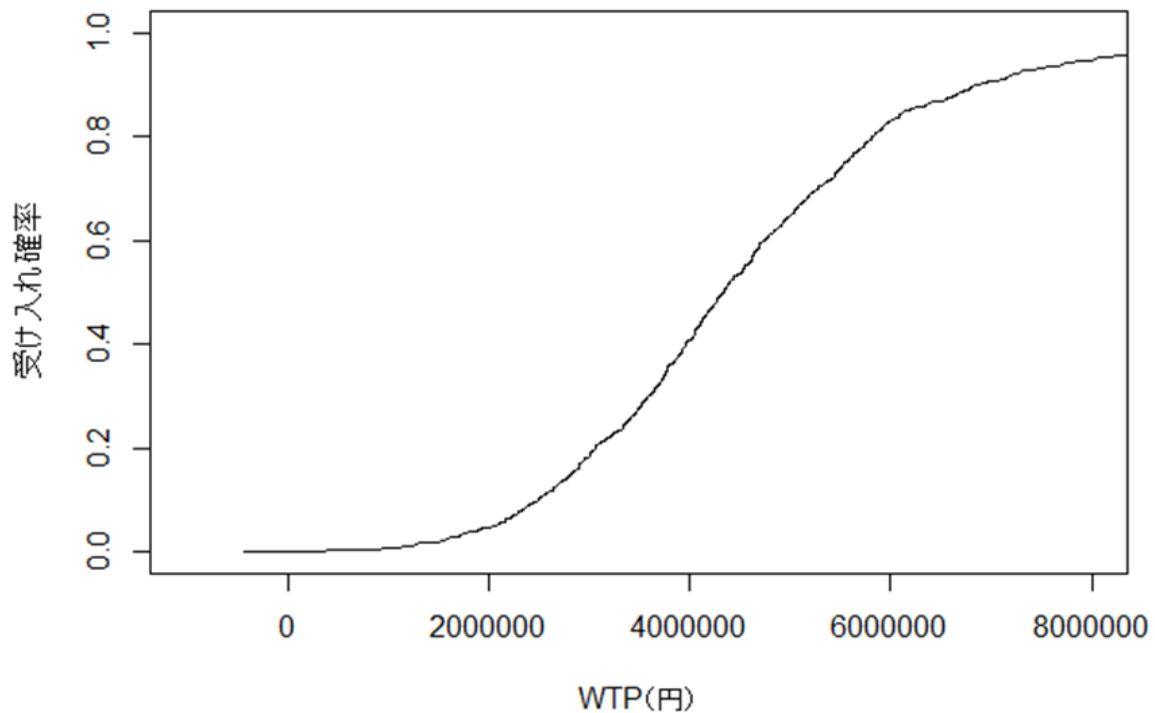


図 1-7 図 1-6 の例に対応する CEAC

### 1.8.3.3 ICER の信頼区間の構築

ICER は費用と効果の増分の比に基づく統計量であることから、一般的には信頼区間を構成することは容易ではない。Obenchein ら<sup>38)</sup>は、費用、効果それぞれの信頼区間が費用効果平面に描く矩形の角を用いて信頼区間を定義する **box approach** や比の信頼区間を与える **Fieller's theorem** を用いた方法、**bootstrap** 法を用いたシミュレーションの結果を基に、費用効果平面上で 95% のシミュレーション結果を含む楔形の範囲を得る方法が紹介されており、**bootstrap** 法に基づく方法のみが常に妥当な結果を与えるとしている。一方で、モンテカルロシミュレーションにより得られた費用と効用の同時分布を用いて、ICER の信頼区間を構築することが行なわれている。すなわち、図 1-6 のように得られた PSA の結果をもとに、原点を通り且つ分布を 2.5% と 97.5% に切り分ける 2 つの直線の組の傾きを ICER の 95% 信頼区間

とするものである。95%信頼区間を表す2本の直線を費用効果平面上に示したものを図1-8に示す。この信頼区間の解釈は、1000回のモンテカルロシミュレーションの結果のうち、ICERの高いものから25番目と975番目の点を抽出したものと理解され、この例であれば傾きの小さい直線、大きい直線の傾きがそれぞれ95%信頼区間の下限と上限であるといえる。

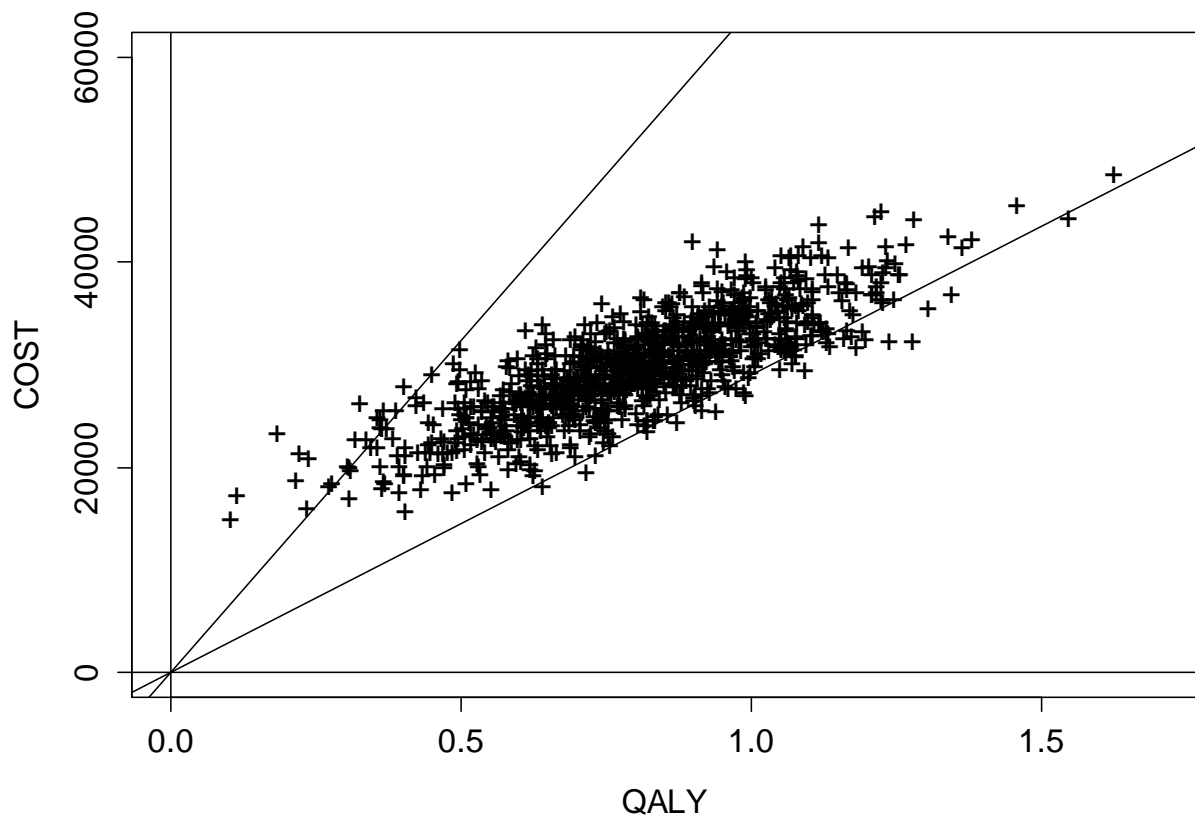


図1-8 信頼区間を示す2本の直線

#### 1.8.3.4 ICERの信頼区間に伴う課題

モンテカルロシミュレーションの結果第1象限のみに結果がプロットされた単純な例を示したが、実際の薬剤経済評価においてはこのようなケースはまれであり、第2-4象限にもプロットが打たれることが多いであろう。この場合、ICERの信頼区間の構築にはいくつかの課題が発生する。

## 1. 信頼区間が離れること

図 1-9 は、対照薬と比べた新薬の効用に際立った差が無いケースである（費用、効果について平均がそれぞれ 20000、0.0、標準偏差がそれぞれ 5000、0.2、相関係数 0.9）。前節の図 1-8 のような場合であれば、原点を通りかつ分布を 2.5% と 97.5% に切り分ける 2 つの直線の組は、1000 個のプロットから ICER をそれぞれ算出し、2.5% 点と 97.5% 点を求めることによりそれらの傾きを得ることができた。しかしながら、このようなケースではその方法によって信頼区間を求めることはできない。x→0 の時と x→-0 の時では ICER の符号が逆転するため、Y 軸を挟む 2 点の ICER の符号が逆転してしまい、近傍の点であるにも関わらず ICER が大きく離れた値をとることが原因である。

このような場合であっても、図 1-9 に示すように ICER の信頼限界を示す直線の組を求めることは可能であるが、この場合の ICER は、 $-\infty < \text{信頼上限}$ 、 $\text{信頼下限} < \infty$  のように、離れた値をとる<sup>39)</sup>。

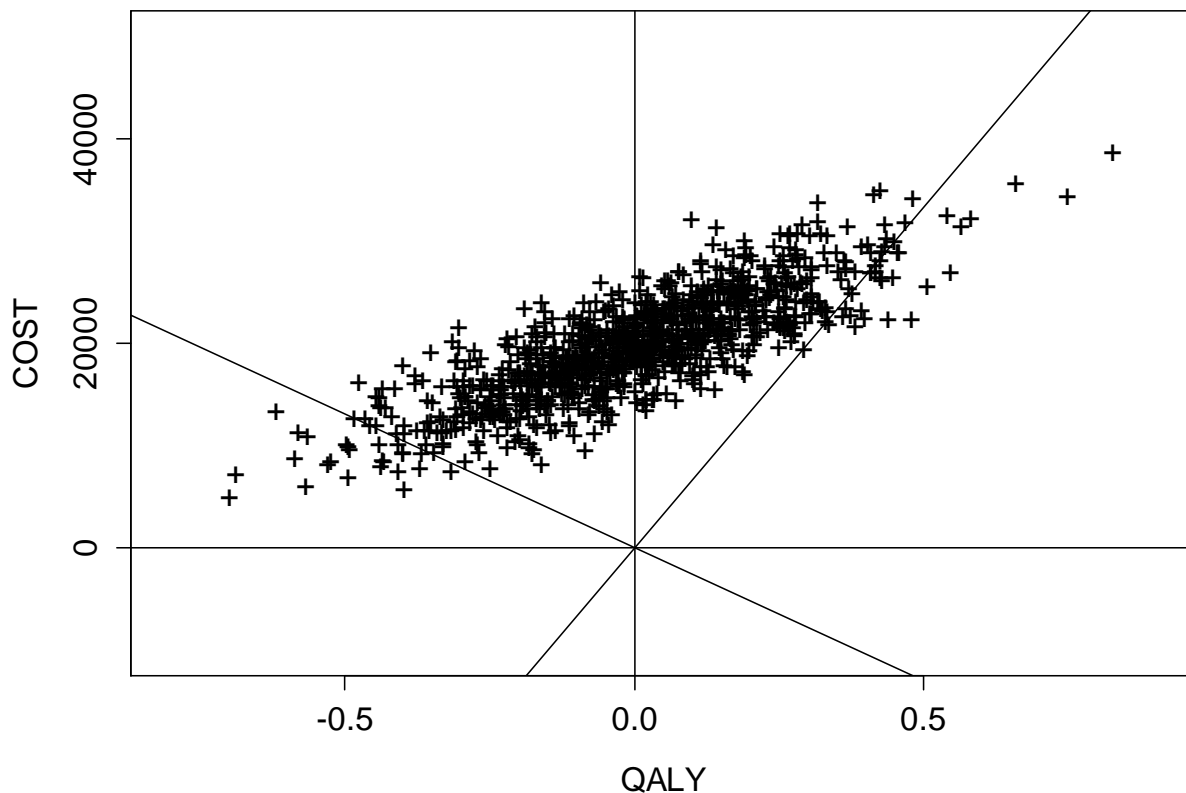


図 1-9 信頼区間が離れる例

2. 信頼区間にバイアスが入っていること

図 1-10 の例では、ICER の信頼区間という観点では、第 1 象限において 2 本の直線が挟む区間に 95% のプロットが入っているべきであると考えられるが、このケースでは第 3 象限のプロットが、95% 信頼区間内として扱われているため、実際には第 1 象限で直線に挟まれるプロットは 95% に満たない。

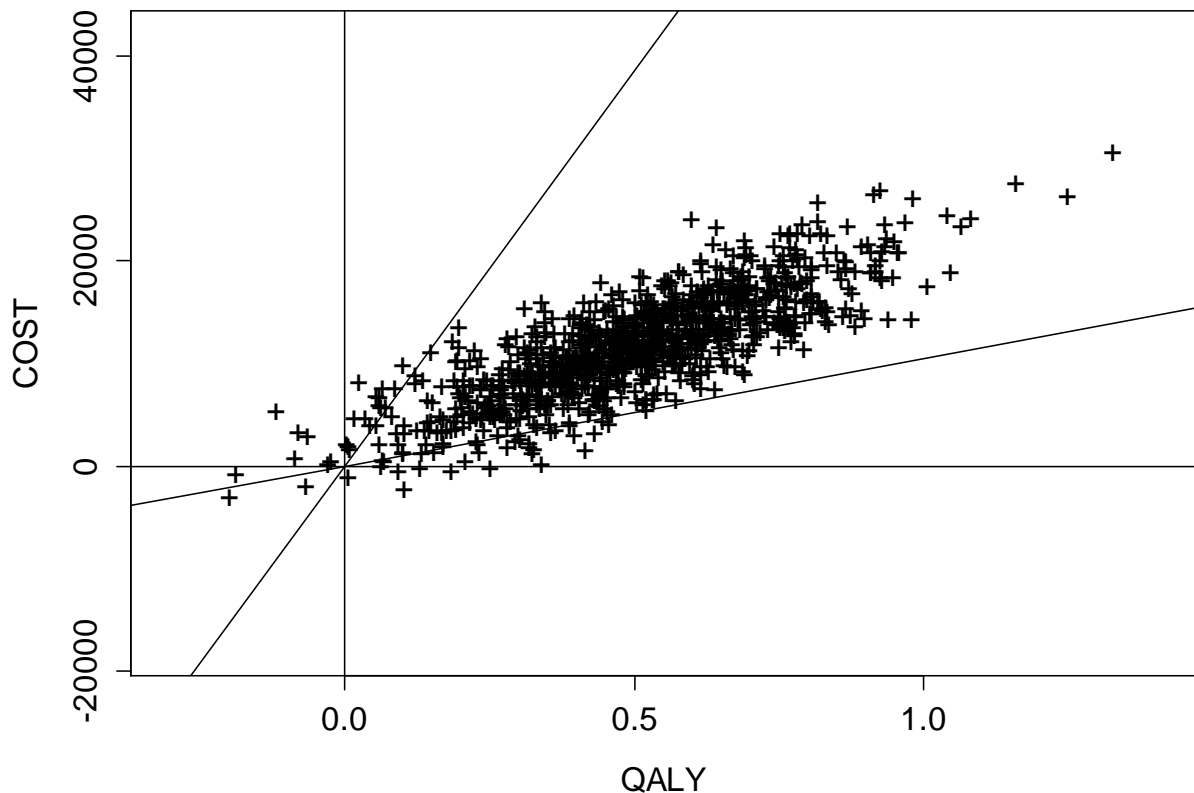


図 1-10 信頼区間にバイアスが入っているとみなせるケース

3. 象限が異なるプロットは異なる意思決定に至るにもかかわらず ICER の信頼区間としては同様に扱われること

図 1-9 では、第 1 象限のプロットでは新治療を、第 3 象限のプロットでは対照治療を採用することになるため、意思決定上反対の結果であるといえるが、ICER の値や、信頼区間だけを見るとこれらは混在してしまい、信頼区間と意思決定の結果が合致しないことがある。また、従来治療と比べて新治療が優れている場合の ICER の値と、新治療が劣っているが費用節減的である場合の ICER の値が互換可能なのかについても議論が必要であろう。

4. 原点を通りかつ分布を 2.5% と 97.5% に切り分ける 2 つの直線の組が定義できないことがあること

図 1-11 のようなケースでは（費用、効果について平均がそれぞれ 0、0.0、標準偏差がそれぞれ 5000、0.2、相関係数 0.9）、原点を通る直線は分布を 2.5%と 97.5%に切り分けることはできない。この場合、95%信頼区間を定義することはできない。

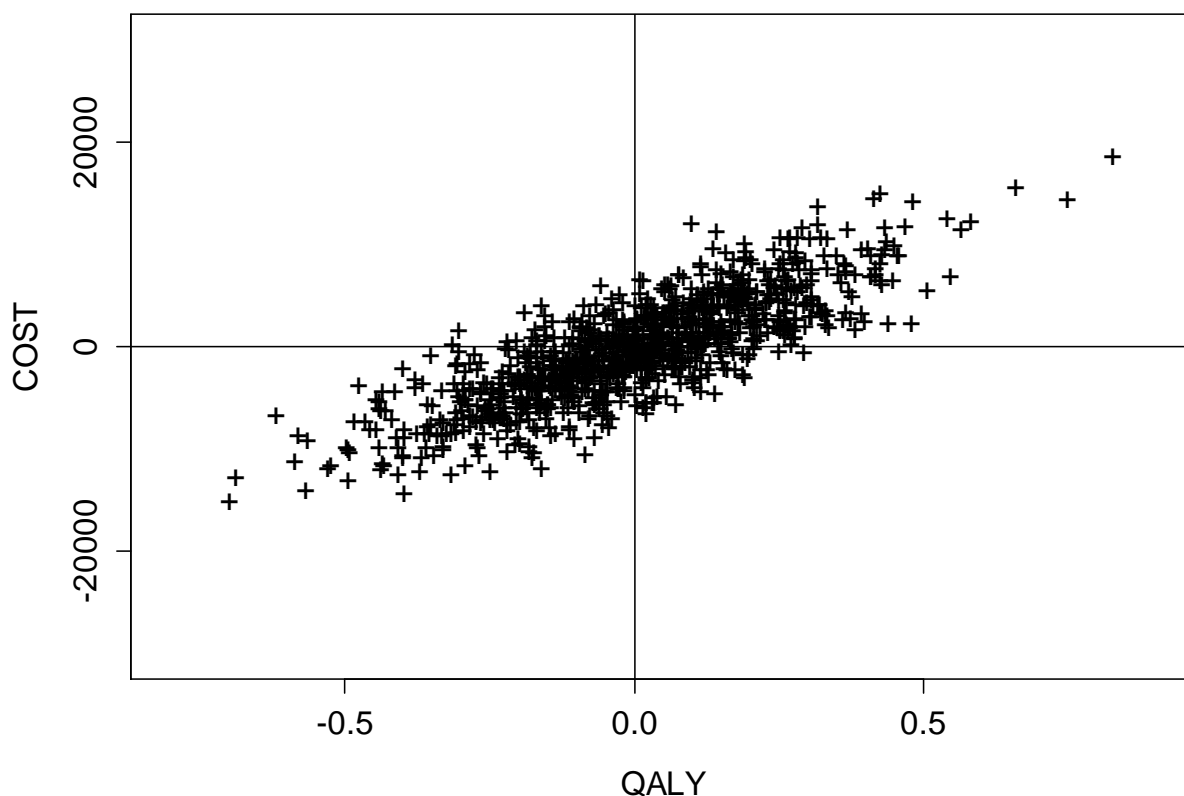


図 1-11 信頼区間が定義できない場合

## 1.9 まとめ

本章においては、不確実性の発生要因（methodological uncertainty, structural uncertainty 及び parameter uncertainty）ごとにその評価と対応方法をまとめた。また、ICER の信頼区間の構築法について、現在使用されている方法には特定の状況においてバイアスが含まれたり、信頼区間が定義されないという課題があることを示した。薬剤経済評価には様々な要因による不確実性が含まれることを考えると、これらの評価については複数の方法を用い総合的に評価、判断することが求められる。

## 2 研究 1 確率的感度分析の使用状況

### 2.1 緒言

薬剤経済評価では効果、安全性、費用など、治療にかかわる多くのパラメータについての推定が行われ、最終的にそれらのパラメータを以下のような式であらわされる効果と費用の2つの比である増分費用効果比(ICER; Incremental Cost Effectiveness Ratio)のような単一の値に集約し、意思決定を行う。

$$ICER = \frac{C_{new} - C_{com}}{E_{new} - E_{com}}$$

$C_{new}$  : 新治療の費用、 $C_{com}$  : 比較対照となる治療の費用、

$E_{new}$  : 新治療によって得られる効果、

$E_{com}$  ; 比較対照となる治療によって得られる効果

効果としてしばしば使用される指標として、QALY (Quality Adjusted Life Year) があげられる。これは質で調整した生存年を示し、

$$QALY = \sum (\text{生存年}) \times (QOL)$$

の式で示される。本稿ではこの後効果の指標として QALY を用いる。

評価に用いられるパラメータが確定した数値で得られることはまれであり、多くの場合は臨床試験の結果から推定するか、文献データをもとにメタ・アナリシスを行って推定することから、それぞれのパラメータには不確実性が含まれる。医学研究においては、不確実性の大きさは信頼区間で示されるが、薬剤経済評価で用いられる増分費用効果比や増分費用効用比は、効果と費用に関する多くのパラメータが複雑に影響するモデルを元に算出されるため、一般に信頼区間を示すことは困難である。このため、多くの場合 ICER は点推定値のみが示され、評価の対象となってきた。一方で薬剤経済評価の結果が、パラメータの推定の不確実性によってどの程度の影響を受けるかは研究の質を反映し、研究の結論がどの程度の信頼性

を持つかを示すことから、感度分析を行い、薬剤経済評価の頑健性を定量的に分析することが重要となる。

感度分析は確率によらない感度分析 (Deterministic Sensitivity Analysis) と確率的感度分析 (PSA; Probabilistic Sensitivity Analysis) に大別される。確率によらない感度分析では、効果や費用、割引率など薬剤経済評価に用いるパラメータを十分に広い範囲で変化させ、その上限や下限において評価の結論が変わりうるか、あるいはどの程度極端な値をとった場合に結論が変わりうるのかを評価する方法である。この方法は簡便であるので広く利用されているが、どの程度極端な値であれば結論が変わっても許容されるのかを定量的に判断することが難しいこと、また多くのパラメータを同時に評価することが難しいことから、費用と効果の両方を同時に評価に用いる必要のある薬剤経済評価においては十分にその不確実性を評価できない場合がある。

複数のパラメータの不確実性を同時に評価するため、それらのパラメータが従う分布に基づき評価を行うことを確率的感度分析と呼ぶ<sup>40)</sup>。複数のパラメータの同時分布を考える場合、同時分布の確率密度関数は複雑になり、解析的に評価することが難しくなることから、分析にはモンテカルロシミュレーションや MCMC (マルコフ連鎖モンテカルロ) 法などのコンピュータシミュレーションが多く利用される<sup>41)</sup>。シミュレーションにより得られた多くの仮想データを増分費用効果平面にプロット、分析することにより、ICER の信頼区間を算出することが可能である。ある利得、この場合では IQALY を獲得するために支払ってもよいとする上限の費用を Willingness To Pay (支払い意思額; WTP) と呼ぶが、シミュレーションを行うことにより、ある WTP に対して評価対象の治療が受け入れられる確率がどの程度であるかをプロットに示すことができる。これを受容曲線 (Acceptability Curve) と呼び、一般的に許容されるとされる WTP を設定したもとの当該治療の受容確率と、十分な確率で当該治療が受容される WTP の範囲の両面から分析、評価することができる。

このように、PSA は複数のパラメータに起因する薬剤経済評価の不確実性を同時に評価可



能であるという望ましい性質を持っていること、解釈しやすい分析を行うことができることから、昨今では各国の薬剤経済評価ガイドラインで使用が推奨されている。よって国内においても PSA の使用は推進されることが望ましいと考えられるが、現段階においてその利用状況は不明である。本研究では、国内における薬剤経済評価のレビューを行い、PSA の利用状況を明らかにする。

2.2 章では PSA の方法、結果の解釈について説明を行い、2.3 章で各国ガイドラインにおける PSA の扱いについてまとめる。2.4 章では PSA の本邦での利用状況を、邦文にて論文が出版された研究のレビューにより示す。

## 2.2 PSA の手法と結果の解釈

### 2.2.1 PSA

本章では、新治療と既存治療を比較するケースを想定し、PSA の例を挙げ、方法を順を追って説明する。臨床試験や臨床研究のデータにおける不確実性は、患者ごとに治療の効き目やかかる費用のばらつきが存在することや、もともとの患者の背景因子にばらつきがあることに起因する。この不確実性は、試験に組み込む例数を適切に確保することや、試験や研究のデザインの設計により軽減することが可能だが、不確実性を完全になくすことは事実上不可能である。薬剤経済評価にはこのように多種多様な原因による不確実性が含まれることから、不確実性の要因となるパラメータについて分布の仮定をおき、シミュレーション実験を行うことでその挙動を評価し、薬剤経済評価の結論が不確実性によってどの程度影響を受けるかを確認することが薬剤経済評価の結論の頑健性の評価につながる。

#### 2.2.1.1 モンテカルロシミュレーション

はじめに、モンテカルロシミュレーションにより、効果と費用が従う同時分布を設定し、

同時分布からランダムに発生した標本をもとに ICER のばらつきを評価する方法を見てみる。臨床試験や臨床研究の結果、比較する治療間の効果と費用の差について以下の表 2-1 の通り得られたとする。この例では新治療は既存治療と比べて効果においては 0.7QALY 分すぐれているが、費用も 300 万円多くかかり、それぞれの標準誤差は 0.2QALY 及び 100 万円である。更に効果と費用の相関係数は 0.5 であるとして、効果が高い結果が得られたとしたらそれにかかる費用も相応に高いという傾向がみられると考える。

表 2-1 Example: Effectiveness and cost used for simulation

新治療－既存治療	点推定値	標準誤差	相関係数
効果(QALY)	0.7	0.2	0.5
費用(円)	3,000,000	1,000,000	-

効果、費用が上記のパラメータを持つ二変量正規分布に従うと仮定し、シミュレーションにより 1,000 個のデータを発生させた結果を図 2-1 に示す。この設定のもとでは、当然効果と費用の群間差はそれぞれ 0.7、3,000,000 を中心に分布し、右上がりの中程度の効果と費用の間に中程度の相関がみられること、効果の群間差が負の値をとったり、費用の群間差が負の値をとること、つまり新治療が既存治療より効果が劣っていたり、費用が安かったりする結果が得られる確率がほとんどないことを見て取ることができる。

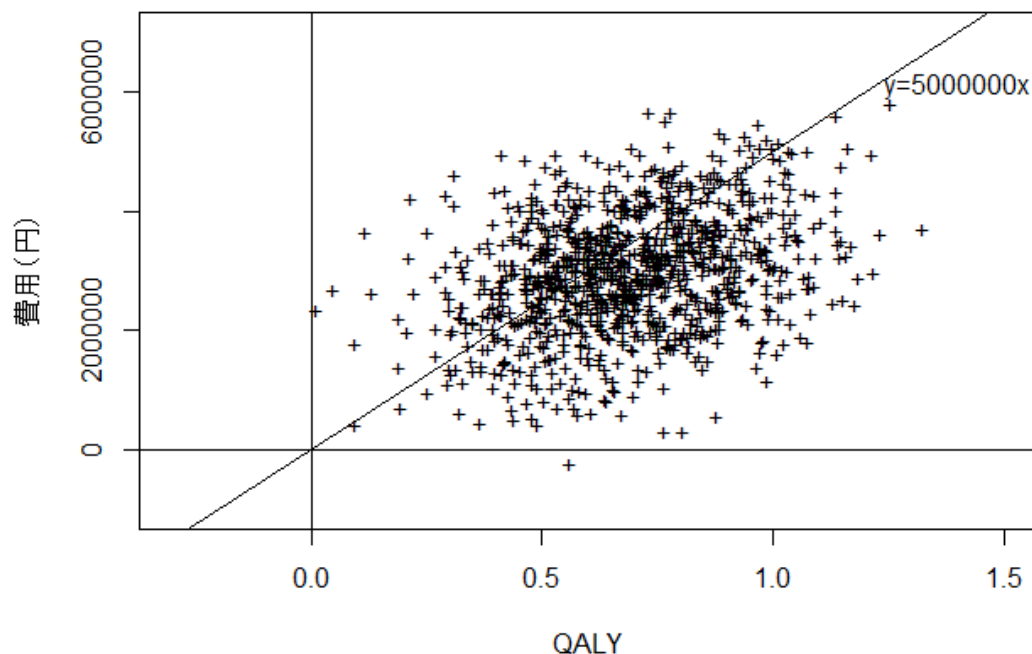


図 2-1 Scatter plot of effectiveness and cost based on Monte Carlo simulation

文献などから効果、費用に関するデータを推定した場合も、それぞれのパラメータの従う分布を仮定することで上記と同様のシミュレーションを行い、PSA を行うことが可能である。たとえば論文のメタ・アナリシスなどから効果とその標準誤差の推定を行ったり、官庁統計などを元に生命予後を推定したり、あるいは診療報酬や薬価の積み上げで費用を推定することになるが、それらの推定それぞれに含まれる不確実性を、コンピュータシミュレーションによって最終的には ICER という単一の指標のばらつきに落とし込むことが可能なのである。

この例では議論を簡単にするため効果と費用が二変量正規分布に従うと仮定したが、必ずしもこの仮定が成立するとは限らず、データの性質から別の分布に従うと考えるほうが適切なケースもあると考えられる。例えば、費用データについては平均値を中心に均等に分布している正規分布よりも、頻度は高くないが高額な費用も発生しうると考えるほうがむしろ自然な場合が多いと考えられるが、この場合は対数正規分布のような分布を検討することも必

要となる<sup>42)</sup>。

二変量正規分布については R や SAS などの一般的な統計ソフトを利用して乱数を発生させることが容易であるが、通常は二変量の同時分布の乱数を簡単に発生させることは容易ではない。このため、乱数発生のためのアルゴリズムとしてマルコフ連鎖モンテカルロ法 (MCMC 法) がしばしば利用される<sup>40,41)</sup>。R パッケージ”BCEA”を用いることで MCMC 法の実行が可能である<sup>41,43)</sup>。

### 2.2.1.2 ブートストラップ法

前節で説明したモンテカルロシミュレーションでは、効果と費用が二変量正規分布に従うと仮定し、推定された点推定値と標準誤差から同時分布を構成したが、それぞれの患者における健康状態の推移、費用、予後などを取得可能であれば、ブートストラップ法<sup>44)</sup>を用いることでそれぞれのパラメータの分布を未知のまま ICER の不確実性を評価することも可能である。ブートストラップ法の具体的なプロセスは以下の通りである。分布の仮定が不要なため、パラメータの分布について特段の知見が無い場合には有効な方法である。

1. 臨床試験から取得した新治療、既存治療それぞれの群の個別データから、それぞれの例数と同数のサンプルを復元抽出によって得る。
2. 得られたサンプルを用いて評価を行い、効果と費用の算出を行う。
3. 1～2 を反復して (たとえば 10,000 回) 行い、図 2-1 と同様のプロットを作成する。

## 2.2.2 PSA から得られる結果とその解釈

2.2.1 章では、シミュレーションの結果図 2-1 のような効果と費用の同時分布が得られた。ベイズ統計学的解釈によると、この分布は、手持ちのデータから判断して、効果と費用がど

の値をとるかについてどの程度の確信が得られているかを示す。一方、頻度論的解釈を行なうと、もし薬剤経済評価を反復して行った場合に結果がどのように分布するかを示しているものであると解釈できる。

これを利用して、ICERがある一定のWTPを下回る確率を評価できる。たとえば、WTPを500万円としたとき、WTPはグラフ上では図2-1に示したように、原点を通り傾きが5,000,000である直線であらわされる。この直線よりも下にあるプロットはすべてICERがWTPを下回っている、つまり新薬の増分費用が1QALYあたりのWTPよりも低い金額を達成していることを示す。図2-1の例では1,000回のシミュレーションのうち646回がWTPを下回っていることから、費用対効果の観点から新治療が受け入れられる確率は64.6%であるといえる。

このように、あるWTPのもとで費用対効果の観点から新治療が受け入れられる確率が得られることから、WTPごとの新治療の受け入れ確率をグラフに表示することができ、これを費用対効果受容曲線と呼ぶ。日本においては500万円や600-700万円が費用対効果を判断する上での閾値とされることが多い<sup>45,46)</sup>が、もしWTPをより小さな値に設定したとしても新治療が受け入れられる確率が高いことがわかれば、結論はWTPの値によって変化しないと結論付けることが可能となる。図2-2のような場合は、WTPが600万円以上あれば十分に受け入れ可能性は高いが、400万円では受け入れ可能性は50%を切る、ということが読み取れることになる。

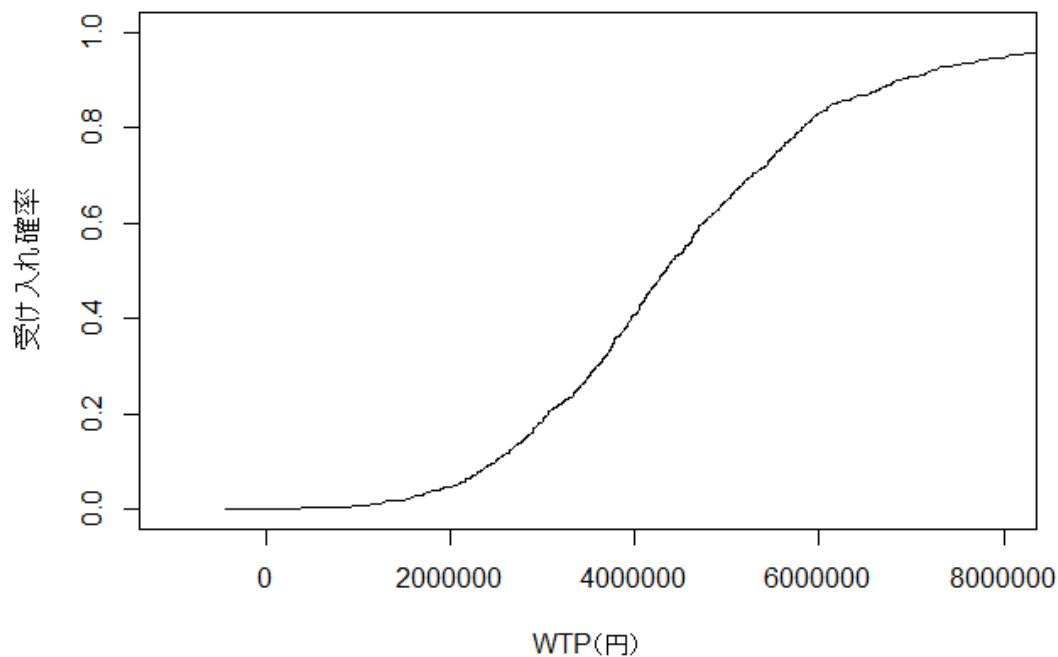


図 2-2 Cost effectiveness acceptability curve

同様の考え方により ICER の 95%信頼区間を推定することが可能である。効果と費用のプロットにおいて、原点を通る直線は ICER が同一である効果と費用の組を示す。原点を通る直線のうち、ある傾きの直線の下に 2.5%の点が含まれているとき、手持ちの情報から判断して ICER がこの直線が示す ICER よりも低い値である確率は 2.5%に過ぎないと解釈できることから、95%信頼下限であるとみなせる。同様の考え方により、直線の下に 97.5%の点を含む傾きから、95%信頼上限を推定することが可能である。

図 2-1 に 95%信頼区間を示す直線を加えたものを図 2-3 に示す。95%信頼区間の下限と上限は、1000 個のプロットからそれぞれ 25 個のプロットを切り取る点を通る 2 本の直線の傾きであり、ICER はそれぞれ 1,601,727 円/QALY 及び 9,350,484 円/QALY であった。

この方法により「信頼下限<ICER<信頼上限」の形の 95%信頼区間が得られるのは、シミュレーションにより得られたデータのうち 95%以上が QALY>0 である場合、すなわち Y 軸

より右側に 95%のデータがある場合に限られる。この条件を満たさない場合は、原点を通る 2本の線分で分布の 95%を切り取る対を探し、95%信頼区間は「 $-\infty < ICER < \text{信頼上限}$  または  $\text{信頼下限} < ICER < \infty$ 」の範囲となる<sup>39)</sup>。ただし、図 1-11 に例示したように原点を通るいかなる直線も分布を 95%と 5%に切り分けることができない場合、95%信頼区間は定義されない。

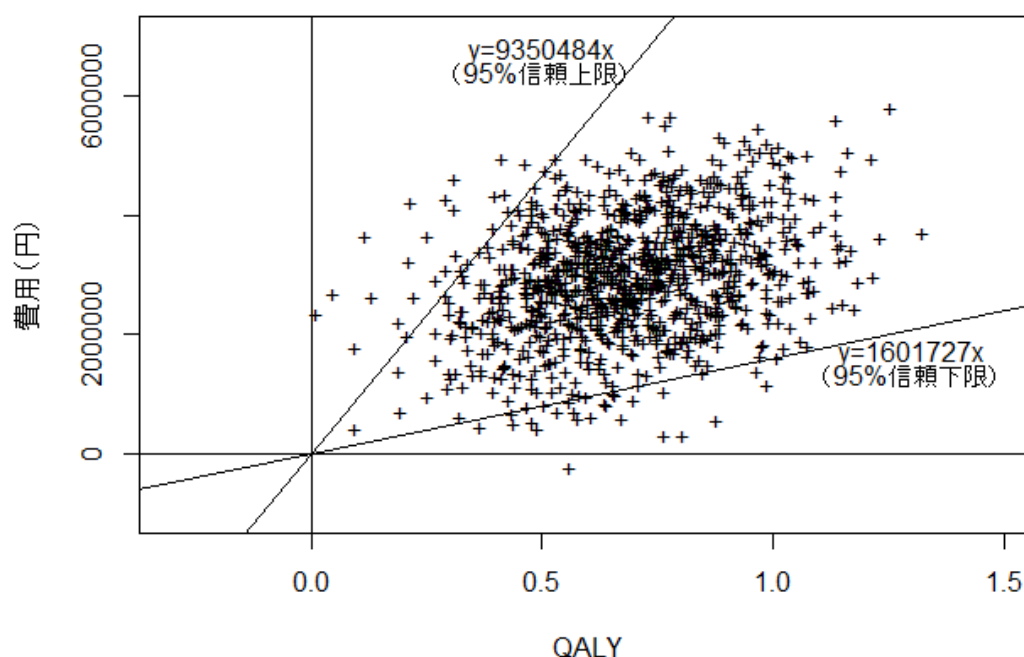


図 2-3 Scatter plot of effectiveness and cost with lines describe 95% confidence limits

### 2.3 PSAに関する各国ガイドライン

薬剤経済評価を包含する医療経済評価の方法については、多くの国で独自にガイドラインが発行されており、その中で感度分析の方法についても言及されている。オーストラリア、フランス、イギリス、米国および日本における医療経済評価のガイドラインのレビューを行い、PSAの使用に関する各国の状況をまとめた。ガイドラインによっては複数回の改版を経ているものがあるが、最新版をレビューの対象とした。

各国のガイドライン発行者に関する情報をまとめる。オーストラリアの Pharmaceutical Benefits Advisory Committee (PBAC)は、医薬品給付制度における医薬品リストへの掲載を推奨する。推奨がない場合医薬品は給付されない。一方でフランスは Haute Autorité de Santé (HAS)が医療上の有用性 (SMR; Service Médical Rendu/Service) と既存の薬剤または治療法と比較した医療上の貢献度 (ASMR; Amélioration du Service Médical Rendu) の2つの軸をもとに最終評価を行い、全国疾病保険金庫連合 (UNCAM; Union Nationale des Caisses d'Assurance Maladie) が SMR をもとに公的保険による償還率を、契約交渉委員会 (CEPS; Comité Économique des Produits de Santé) が企業と交渉する際に ASMR を参考として価格の決定を行う。イギリスの National Institute for Health and Care Excellence (NICE)は National Health Service (NHS)での償還可否についての推奨を行う<sup>47)</sup>。米国においてはマネジメント保険会社が新薬を保険収載するかどうかを判断する際に薬剤経済学研究データを活用する動きが広がっており、マネジメント薬局協会(Academy of Management Care Pharmacy; AMCP)は製薬企業が薬剤経済学研究データを提出するための指針を発表した<sup>48)</sup>。

オーストラリアで 2008 年に発行されている”Guidelines for preparing submissions to the pharmaceutical benefits advisory committee (Version 4.3)”では<sup>49)</sup>、「PSA を行った場合にはパラメータが従うと想定した分布を示すこと」としており、これらから 2008 年時点ではオーストラリアでは、医療経済評価を行う際の感度分析について推奨はするものの、その方法（確率によらない感度分析を行うか PSA を行うか）ははっきりと定めることはなく、任意とされていたと解釈される。

一方、フランスで 2011 年に発行されている”Choices in methods for economic evaluation”（英語版は 2012 年に発行）においては<sup>50)</sup>、確率によらない感度分析には、パラメータ間の相互作用の評価に限界があることを理由に、モンテカルロシミュレーションを用いた PSA を推奨している。米国 AMCP のガイドライン<sup>51)</sup>では、確率的感度分析を推奨しており、費用効果平面へのプロットと費用効果受容曲線での報告を推奨しており、複雑なモデルの場合は特に



推奨されるとしている。また、イギリスのガイドラインは2013年に改版され、“Guide to the methods of technology appraisal 2013”が発行されたが<sup>6)</sup>、ここでははっきりとPSAを推奨する旨が記載されており、その根拠として非線形のモデルの場合にPSAが平均の費用、効用値のもっともよい推定値を与える方法であるからであるとしている。このことから、2011年以降に発行されている最新のフランス、イギリスのガイドラインではいずれもPSAを積極的に推奨していることがわかる。

日本においては2013年に医療経済評価研究における分析手法に関するガイドラインが発行されたが<sup>52)</sup>、「可能であれば確率的感度分析もあわせておこなうこと」とされており、その推奨度合いは★★☆（星二つ：本ガイドラインではそのように取り扱うことを推奨するが、状況によっては他の選択肢もとりのもの）である。

## 2.4 PSAの利用状況

PSAの利用について調査を行うため、医療経済評価における研究におけるPSAの使用状況を調査した。

### 2.4.1 本邦における邦文で報告されたPSAの使用状況

本邦におけるPSAの使用状況を調査するため、医中誌Webにて以下の検索により邦文の文献検索を行った。

検索式：((質調整生存年/TH or QALY/AL) or (質調整生存年/TH or 質調整生存年/AL)) and (PT=原著論文)

2013年7月13日時点で、上記の条件を満たす72件の論文が確認され、それらについてレビューを行った。23件は医療経済評価を行った研究ではなかったことから、レビューから除外された。

医療経済評価を行っていた 49 件の論文について、感度分析の利用状況から、「1.PSA を使用している」、「2.感度分析を行っているが、PSA ではない」、および「3.PSA も感度分析も行っていない」に分類した。結果を図 2-4 に示す。

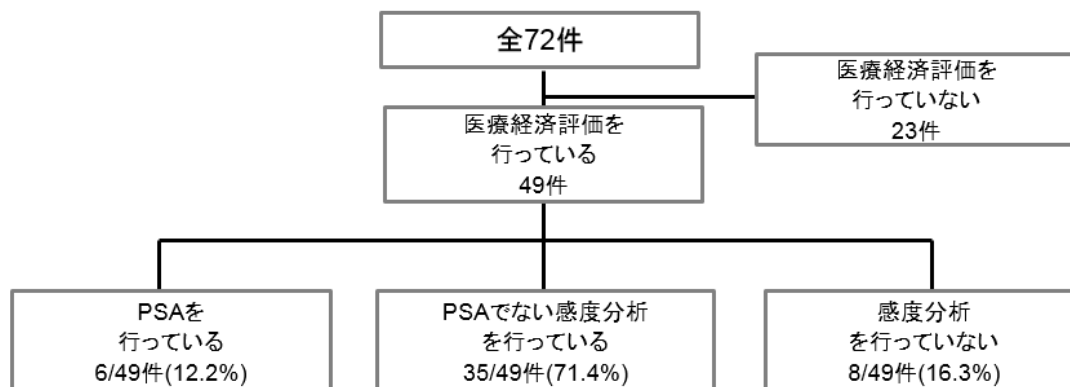


図 2-4 The usage of probabilistic sensitivity analysis in Japan reported in Japanese

医療経済評価を行っていた 49 件のうち、PSA による感度分析が行われているものは 6 件 (12.2%)、確率によらない感度分析を行っているものは 35 件 (71.4%)、感度分析を行っていないものは 8 件 (16.3%) であった。PSA の結果の報告の内容は、6 件のうち 3 件が PSA によって ICER の信頼区間を算出し、報告しており、費用対効果受容曲線およびモンテカルロシミュレーションの結果が散布図で示された研究がそれぞれ 1 件、残りの 1 件は WTP の閾値を下回る確率が示されていたのみであった。

PSA に限定するとその使用割合は 12.2%であり、高いとはいえない。一方で確率によらない感度分析を行っている研究は全体の 70%を超えており、方法を PSA に限定しなければ、感度分析の使用割合は高いといえる。

PSA を行っていた論文の一覧を表 2-2 に示す。

表 2-2 Listing of studies which used probabilistic sensitivity analysis in sensitivity analysis

	疾患領域	出版年	PSA の結果の示し方	PSA の方法の記載	パラメータ等の記載
Okubo Ichiro ら <sup>53)</sup>	がん	2005	ICER の信頼区間	あり	あり
Amemiya Shiori ら <sup>54)</sup>	循環器	2009	ICER の信頼区間	あり	あり
Tamayama Kazuhiro ら <sup>55)</sup>	花粉症	2010	ICER の信頼区間	あり	あり
池田 俊也ら <sup>56)</sup>	疼痛	2011	費用対効果受容曲線	あり	なし
井上 裕智 <sup>57)</sup>	ワクチン	2011	費用対効果受容曲線、モンテカルロシミュレーションにより得られたプロット	あり	あり
駒場 大峰 <sup>58)</sup>	副甲状腺機能亢進	2012	WTP の受容確率	なし	なし

6 報の扱う疾患領域はすべて異なっていた。PSA の結果の示し方としては、3 報が ICER の信頼区間として報告しており、2 報が費用対効果受容曲線（うち 1 方はモンテカルロシミュレーションのプロットも提示していた）、1 報が特定の WTP（500 万円）が受容される確率を報告していた。PSA の方法は、5 報が報告しており、いずれもモンテカルロシミュレーションで、ブートストラップ法による PSA を行っていたと報告したものは無かった。パラメータの分布の仮定等、PSA の再現を可能とする情報については、4 報が記載していた。以上

のことから、PSA を行っている論文においてはその方法論、再現性を担保するためのパラメータの叙述は高い割合で行われていることがわかった。

PSA の使用状況の経時的変化を評価するため、PSA を行った論文の出版年度を確認したが、2005 年、2009 年、2010 年および 2012 年にそれぞれ 1 報、2011 年に 2 報であり、経時的な増減について定量的に評価できるほどの数ではなかった。

#### 2.4.2 本邦における英文で報告された PSA の使用状況

日本で行われ、英文で報告された研究について調査するため、Pubmed にて以下の検索により 136 件の論文を確認し、レビューをおこなった。

(qaly AND japan) AND ("1900/1/1"[Date - Publication] : "2013/7/13"[Date - Publication])

38 件は医療経済評価を行った研究ではなかったこと、17 件は日本語で執筆された論文であることから前節との重複を避けるため、3 件は日本における研究ではなかったこと、1 件は医療経済評価の総説であったことにより、レビューから除外された。

以上を除外した、医療経済評価を行っていた 77 件の論文について、邦文と同様に分類した結果を図 2-5 に示す。

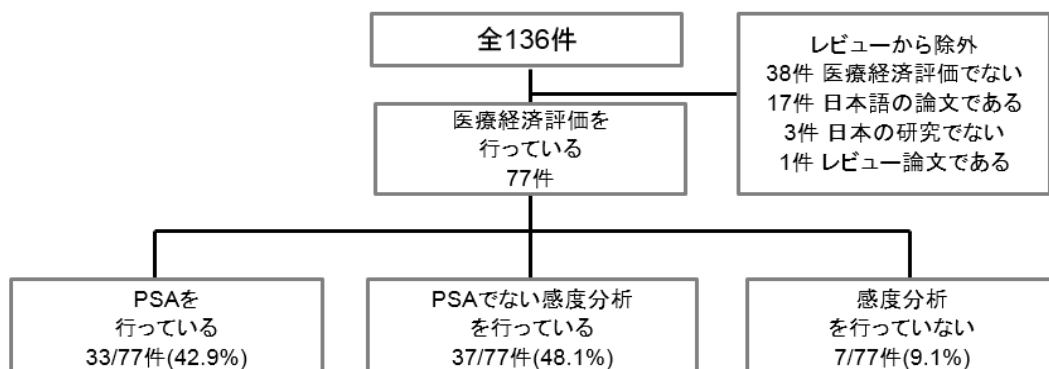


図 2-5 The usage of probabilistic sensitivity analysis in Japan reported in English.

医療経済評価を行っていた 77 件のうち、PSA による感度分析が行われているものは 33 件 (42.9%)、確率によらない感度分析を行っているものは 37 件 (48.1%)、感度分析を行っていないものは 7 件 (9.1%) であった。PSA の結果の報告の内容は、33 件のうち 24 件が図を示しており、その内訳は 12 件が費用効果平面へのプロットと費用対効果曲線の両方を示していたのに対し、費用効果平面へのプロットのみを示していたものは 3 件、費用対効果受容曲線のみを示していたものは 9 件であった。一方で 9 件は結果の図示はないが、6 件が特定の WTP が受容される確率、2 件がシミュレーション結果の最大値と最小値、1 件が QALY 及び費用の 95%信頼区間を示していた。

英文で報告された研究は邦文で報告されたものと比べ PSA の使用割合が高く、また結果の表示も図示により明確に示されることが多いことが分かった。しかしながら、確率によらない感度分析の使用割合は邦文により報告された研究と同様に PSA の使用割合を上回っていた。

## 2.5 結果のまとめ

本研究では、PSA の方法論について述べたのち、各国の PSA に対する考え方をまとめ、また邦文にて出版された医療経済評価の研究での PSA または感度分析の利用状況をまとめた。

PSA はコンピュータシミュレーションによって乱数に基づく仮想のデータを多数発生させ、その分布から薬剤経済評価の信頼性に関する知見を得るものである。各種ソフトウェアの乱数発生関数は充実してきており、シミュレーションの方法の選択さえ適切に行うことができれば実行することは難しくない。PSA を用いることにより、複数のパラメータの不確実性を ICER の信頼区間という単一の指標に落とし込んで評価することができるので、研究から得られた結果の頑健性をわかりやすく判断可能であるという望ましい性質をもつ。一方で本稿にて例示したモンテカルロシミュレーションのケースは効果と費用の分布が二変量正規分

布に従い、またその共分散が予測可能であるという強い仮定をおいたが、複数のパラメータが従う同時分布の予測はそもそも困難なことが多く、これはモンテカルロシミュレーションによる PSA の限界であるといえる。また、ブートストラップ法による PSA も、その適用は臨床試験などにより患者個別データを入手しているケースに限られる。

各国のガイドラインのレビューの結果、感度分析の必要性に関して以前より指摘されていたことが読み取れるが、手法を PSA に限定するものではなかった。2011 年以降に発行されたガイドラインでは PSA を推奨する旨が明記されており、今後の薬剤経済評価においては PSA を用いた感度分析を行い、結果の不確実性が結論に与える影響を評価することが望まれる。

本邦で行われた医療経済評価研究のレビューを行った結果、邦文で報告されたものは医療経済評価を行った 49 件のうちで PSA を行った研究は 6 件（12.2%）にとどまることから、PSA が国内で広く使われているとは言いがたい。一方で PSA でない感度分析を行った研究は 35 件（71.4%）あることから、感度分析自体は多くの研究で用いられている。英文による報告においても、PSA の使用割合は高かったものの、PSA でない感度分析を行った研究の使用割合が上回っていることは同様であった。よって PSA が感度分析のよりよい方法論として受け入れられるようになれば利用は促進されると考えられる。このためには PSA を行うためのガイドラインなどの作成が望ましい。

### 3 研究2 頻度論流 EVPI の提案

1.8.2.5 節において、EVPI を疾患モデルの比較の指標として用いる例を示したが、EVPI の用途はそれに限らず、特定の不確実性を解消することが、医療システム全体としてどの程度の価値につながるかの期待値としても活用可能である。既存の薬剤経済評価の結果に基づいてある薬剤が選択されている状況において、例えばより例数の大きな臨床試験を行うことで有効性についてより精度の高い情報を得たり、薬剤疫学調査により有害事象発生率を得たりするなどして、薬剤経済評価の精度を向上させると、既存の結果による意思決定が誤りであることがわかり薬剤の選択が変わる可能性がある。このようによりよい意思決定を行うことができることを通して寄与する追加情報の価値を EVPI により評価することができる。

モンテカルロシミュレーションに基づく EVPI はベイズ統計の考え方に基づくと解釈できるが、EVPI から得られる結論は論理的に最適な意思決定につながらない場合があると考えられる。本章では、この問題を指摘し、解決策として頻度論に基づく EVPI を提案する。

#### 3.1 緒言

##### 3.1.1 EVPI(Expected Value of Perfect Information)

薬剤経済評価において、不確実性の評価は必要不可欠であり、不確実性を評価するための様々な方法が提案され、活用されている。薬剤経済評価の不確実性は、

- Methodological uncertainty  
方法論の不整合による薬剤経済評価の不確実性
- Model uncertainty  
決定樹やマルコフモデルなどの構造による不確実性
- Parameter uncertainty  
費用や効果などの推定に伴う不確実性

の3つの不確実性に分類される<sup>59)</sup>。

本研究ではパラメータの不確実性を評価する方法の一つとして用いられる、Value of Information Analysis<sup>60)</sup>に基づく EVPI (Expected Value of Perfect Information)<sup>61)</sup>に着目する。

薬剤経済評価の結果をもとに意思決定 $j$ をとった場合に結果的に得られる効用 (Net Benefit)  $NB(j)$ は、有効性 $E_j$ と支払った費用 $C_j$ 、有効性を1単位獲得するために支払ってもよいとする費用 $\lambda$  (支払意思額) をもとに、 $NB(j) = \lambda E_j - C_j$ で得られる。ここで薬剤経済評価に用いたパラメータ $\theta$ について不確実性が存在するとし、 $\theta$ が不確実であるもとで効用を最大にする意思決定をとると、その期待効用は $\max_j E_\theta NB(j, \theta)$ で示される。一方で仮に $\theta$ がばらつきなく推定されるとして (これを完全な情報を持つと呼ぶ)、 $\theta$ の完全な情報のもとで最適の意思決定をとるものとする、その期待効用は $E_\theta \max_j NB(j, \theta)$ である。

EVPIは $\theta$ が不確実性を持つ状態で意思決定を行った時の効用の期待値と、完全な情報のもとで最適の意思決定を選択した場合の効用の期待値の差

$$EVPI = E_\theta \max_j NB(j, \theta) - \max_j E_\theta NB(j, \theta)$$

で示される<sup>62)</sup>。

薬剤経済評価における不確実性は、しばしば確率的感度分析と呼ばれる、乱数に基づきパラメータを変動させるシミュレーション実験の結果の挙動を見ることにより評価される<sup>6, 41)</sup>。確率的感度分析の結果の表示は、費用効果平面へのシミュレーション結果のプロット、新しい治療が費用対効果に優れているとされる確率を支払意思額に対しプロットする Cost-Effectiveness Acceptability Curve (CEAC) などがあるが、EVPIもまた不確実性の評価の指標として用いられる。費用効果平面のプロットやCEACが薬剤経済評価全体の不確実性を評価するものであるのに対し、EVPIは薬剤経済評価の要素の1つであるパラメータ $\theta$ の不確実性を評価するという点で用途が異なる。薬剤経済評価は費用データや効果、効用値など多くのパラメータを統合して行われるが、EVPIはそれらのパラメータのうち特定のパラメータ $\theta$ に関して完全な情報が得られた場合に、その薬剤経済評価から新しい治療を採用するか否かをより正しく決断することによってどの程度の追加的効用が見込まれるかを予測するものであ



る。実際の適用例としては、費用効果分析で用いたモデルを構成するパラメータのうち、完全な情報が得られると追加的効用が大きいものを発見し<sup>63)</sup>、追加的効用が大きいと特定されたパラメータを重点的に検証することでモデルの改善を効率よく行うことができる。また、より巨視的な活用としては、研究の優先付けに用いることもできる。EVPIの大きな情報は薬剤経済上より価値の高いものであると見なすことができるから、いくつかの研究計画に優先順位をつけるため、EVPIの大きいものに優先して投資するというのは理にかなった戦略であるといえる<sup>62)</sup>。あるいは、EVPIと研究対象の疾患の患者数の積を取ると、その研究により得られる価値の最大値と考えることができるため、研究にかかる費用と比較して検討することも有効である。

### 3.1.2 EVPIの解釈

EVPIは通常確率的感度分析の枠組みの中で、モンテカルロシミュレーションの結果から算出される。EVPIの理解のため試行回数5回のみの仮想のシミュレーションの結果を示す(表3-1)。

表 3-1 シミュレーション結果の例

Trial	Additional Effectiveness (QALY)	Additional Cost (USD)	Net Benefit (USD)	$\max_j NB(j, \theta)$
1	0.2	5,000	5,000	5,000
2	0.1	10,000	-5,000	0
3	0.3	10,000	5,000	5,000
4	0.2	15,000	-5,000	0
5	0.4	18,000	2,000	2,000

この例においては、パラメータ $\theta$ を乱数に基づき5回発生させ、パラメータ $\theta$ 所与の下で対照治療を新治療に切り替えることにより追加的に得られる効果(Additional Effectiveness (QALY))と、追加的に支払う費用(Additional Cost (USD))を求めた。ここでは支払意思額を\$50,000としている。各試行における効用が正の値をとる場合、新治療への切り替えは費

用対効果に優れるため意思決定は「新治療に切り替える」、となる。一方で負の値であった場合最適な意思決定は「新治療に切り替えない」である。例えば、1回目の試行において新治療に切り替えると\$5,000の効用が得られるため、 $\max_j NB(j, \theta) = \$5,000$ とできる。一方で2回目の試行では新治療に切り替えるとその効用は\$-5,000であり、損をするため、新治療に切り替えずにこの損失を回避することで、損得は現状と変わらず $\max_j NB(j, \theta) = \$0$ である、とできる。このように、パラメータ $\theta$ 所与のもとそれぞれの試行において最適な意思決定を行うことにより、

$$E_{\theta} \max_j NB(j, \theta) = \$2,400$$

を獲得することができる。

一方、パラメータ $\theta$ が不明の場合に所得可能な情報は、新治療を採用すること ( $j = 1$ ) により

$$E_{\theta} NB(1, \theta) = \$400$$

となり、採用しない場合 ( $j = 0$ )

$$E_{\theta} NB(0, \theta) = \$ - 400$$

に比べて期待値的には利得が大きくなることのみである。よってこの場合は新治療を選択することが最善の意思決定であり、

$$\max_j E_{\theta} NB(j, \theta) = \$400$$

となる。結果、この例では、EVPIは

$$EVPI = E_{\theta} \max_j NB(j, \theta) - \max_j E_{\theta} NB(j, \theta) = \$2,000$$

となり、よってパラメータ $\theta$ の情報は\$2,000の価値を持つといえる。この価値は、試行2, 4のように新治療を選択することによりNBが負の値をとる、すなわち新治療に切り替えないほうが薬剤経済の観点から最適の意思決定である場合に、そちらを選択することができることによる。この価値はパラメータ $\theta$ の情報なしには得られないものである。一方で試行1, 3, 5の場合に得られるNBはパラメータ $\theta$ に関する情報の有無にかかわらず得られる効用であるので、パラメータ $\theta$ のEVPIには寄与しない。パラメータ $\theta$ の変動が新治療の採否に大きな影

響を与える場合、また誤って最適でない意思決定を採択した場合に獲得できなかった NB の機会損失が大きい場合に EVPI が大きくなる。

表 3-1 の例をもとに EVPI を図示すると図 3-1 のように示される。

プロットの脇に示す番号は試行を示す。各試行のうち、パラメータ  $\theta$  の情報がない場合に最適でない意思決定を行う試行 2, 4 から、支払意思額 \$50,000 を示す直線までの鉛直方向の距離の (試行 1, 3, 5 を含むすべての試行についての) 期待値が EVPI に相当する。よって EVPI は正しい意思決定が「新治療を採用する」であった場合 ( $\lambda\hat{\mu}_e - \hat{\mu}_c > 0$ ) と「新治療を採用しない」であった場合 ( $\lambda\hat{\mu}_e - \hat{\mu}_c < 0$ ) 別に、以下のようにも示すことができる。

$$EVPI = \begin{cases} \frac{\sum_{i=1}^n \#(\lambda e_i < c_i)(c_i - \lambda e_i)}{n} & (\text{if } \lambda\hat{\mu}_e - \hat{\mu}_c > 0) \\ \frac{\sum_{i=1}^n \#(\lambda e_i > c_i)(\lambda e_i - c_i)}{n} & (\text{if } \lambda\hat{\mu}_e - \hat{\mu}_c < 0) \end{cases}$$

ただし、 $e$  は効果、 $c$  は費用を示し、それらの平均値の推定値を  $\hat{\mu}_e$ 、 $\hat{\mu}_c$  とする。また、 $\#(\cdot)$  は括弧内の式が真のとき 1 を、偽の時 0 を返す演算子、 $n$  はシミュレーションの試行回数を示すとする。

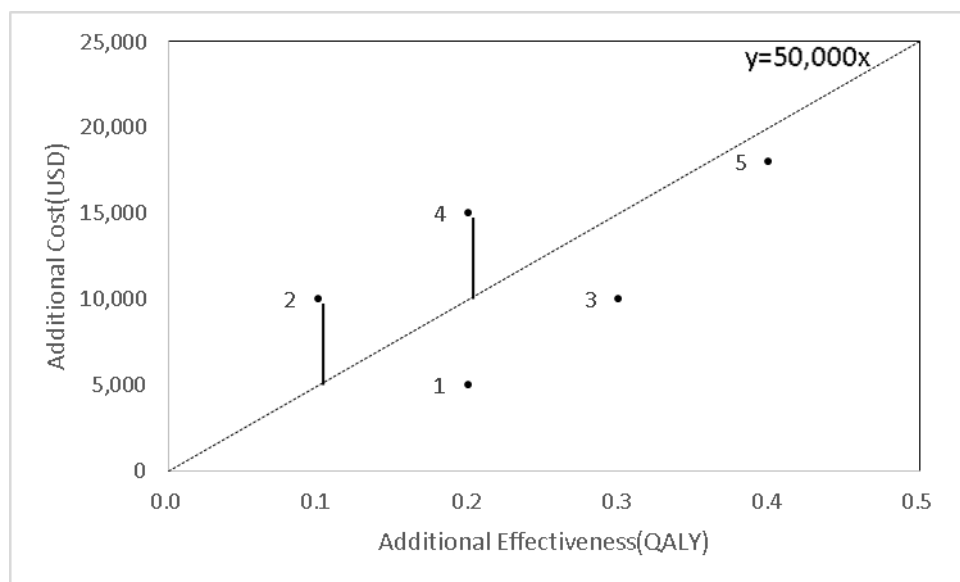


図 3-1 図による EVPI の説明

### 3.1.3 EVPI の例: cost-utility analysis between Linezolid and Vancomycin in the empiric treatment of nosocomial pneumonia

院内感染肺炎への治療として、起炎菌の検査結果を待たずに広域抗菌薬を投与する治療を経験的治療と呼ぶ。Collins ら<sup>64)</sup>は経験的治療における vancomycin に対する linezolid の費用対効果の評価を行なった。分析は決定樹モデルにより行なわれ、院内感染肺炎発症後 vancomycin または linezolid を投与した場合の質調整生存年と費用が比較された。この分析において、費用、肺炎及び副作用治療費、決定樹の移行確率、効用値、治療期間、生存年について確率的感度分析が行なわれた。これらのうち決定樹の移行確率につき Collins らが確率的感度分析に用いた分布により EVPI の算出を行い、性質の比較を行なった。Collins らが移行確率が従うとした分布に従ったシミュレーションデータを 100,000 個発生させ、WTP の値ごとの EVPI を算出、図示した。結果を図 3-2 に示す。なお、この例における WTP の範囲は一般的に 1QALY あたりの ICER のいわゆる閾値とされる値と異なるが、本研究の目的は WTP と EVPI の関係性を論じることにより、WTP の受け入れ可能な上限値である閾値によらず一般化可能な議論であることから、この例により説明を行う。



図 3-2 Collins らの例をもとにした EVPI

EVPIの性質に着目すると、WTPが\$0から\$10,129まで上昇するにつれ、EVPIは上昇している。\$10,129がこの確率的感度分析の結果得られたICERの点推定値であり、WTPがこの領域を取る場合の判断は、linezolid(新治療)による獲得QALYあたりの追加的費用はWTPよりも高いため、linezolidの使用は費用対効果に優れない、であり、この場合EVPIはシミュレーション結果のうち、WTPを示す直線(以下直線と記載)より下に位置するプロットと、直線との間の垂直の距離の総和になる。WTPが増加するに従い直線下のプロットの数が増加すること、プロットと直線との距離が増加することから単調に増加している。EVPIはWTPが\$10,129の時点で極大となり、一旦減少に転じる。WTPが\$10,129を超えると、判断はlinezolidの使用が費用対効果に優れる、と逆転する。このため、シミュレーション結果のうち直線よりも上に位置するプロットと直線との間の垂直の距離の総和がEVPIとなる。このため、直線の傾きが増加することにより直線よりも上に位置するプロットの数が増加すること、また(第1象限における)直線とプロット間の距離が短くなることから、EVPIは減少する。しかしながら、WTPが更に大きな値を取った場合EVPIは増加に転じていることが分かる。

この原因は、WTP が相当に大きくなった場合、追加的 QALY が負の領域（第 2、第 3 象限）では WTP が増加するにつれ直線より上のプロットと直線との間の垂直の距離は増加することの影響を受けている。

EVPI による WTP が \$10,129 近傍であった場合に、そのパラメータについて行われる研究への追加投資の価値は大きいものと評価される。すなわち、「linezolid と vancomycin の費用対効果について、更なる研究投資を行い、どちらが費用対効果に優れているかを判断することに大きな価値がある」とみなす。しかしながら、WTP と ICER の点推定値が近いということは、「双方の治療には費用対効果の観点から大きな差が無い」と判断することが妥当であると考えられ、その判断から考えると更なる研究投資を行うことには大きな価値は無く、linezolid と vancomycin のいずれを使うかではなく、他の研究に注力するほうが望ましい、と考えることができる。EVPI はこのように、治療間で甲乙つけがたい状況においての意思決定が論理的整合性を失う場合がある。

この課題を解決するため、本研究では頻度論に基づく EVPI（通常の EVPI と区別するため、頻度論 (frequentist) の頭文字を添え  $EVPI_f$  と表記する) を定義し、その性質を調べるとともに (ベイズ流に基づく) EVPI との比較を行う。仮想例によるシミュレーションによりその基本的性質を明らかにし、更に実際の薬剤経済評価における確率的感度分析への適用を行い、その性質を実際に確認する。

シミュレーションを行った際、いずれの試行も真の値であるかのように扱うことから、EVPI はベイズ統計学を基礎とした尺度である。一方で頻度論は唯一の真値が存在することを前提に、それぞれの試行は真値に基づく分布からのサンプルに過ぎないと見なす。医学研究において通常示される  $p$  値や信頼区間、仮説検定は頻度論に基づくものが主流であり、薬剤経済評価はそれらの値をもとに行われることがほとんどであるため、EVPI においても頻度論に基づく方法を検討することは方法論の一貫性という観点からも一定の価値があると考えられる。

## 3.2 方法

頻度論に基づく EVPI (EVPI<sub>f</sub>) を定義し、ベイズ流に基づく通常の EVPI との相違を論じる。その後シミュレーションによって EVPI との比較を行う。

### 3.2.1 頻度論に基づく EVPI の定義

頻度論の考え方を基本にすると、新治療と対照治療の費用や効果の差は実際にはばらつきをもった量として観測されるが、それらには唯一の真値が定義できる。

薬剤経済評価を含む医学研究においては臨床試験や観察研究のデータとして観測された値から真値を推定している。この考え方に基くと、シミュレーションを行った結果は真値に基づく分布から得られる試行を繰り返し観測しているに過ぎない。

表 3-1 の例をもとに、図 3-3 を用いて頻度論に基づく EVPI (EVPI<sub>f</sub>) を図示する。このシミュレーションに用いた分布においては臨床試験から得られた知見をもとに効果 (e) と費用 (c) の真値の推定値に標本平均( $\hat{\mu}_e$ ,  $\hat{\mu}_c$ )を用い、それぞれ 0.25QALY、\$10,000 算出された (図 3-3 の菱形の点”Mean”)。頻度論は 5 つの試行を全てこれらの分布から抽出されたサンプルであるとみなすため、この考え方に従うと、正しい意思決定はいずれの試行においても「新治療を採用する」であり、この結果得られる効用は、真値の推定値を示す点から支払意思額\$50,000 を示す直線までの Y 軸方向の距離\$2,500 であると考えられる。試行 2, 4 の場合においてはこの効用を獲得することはできない一方で、パラメータ $\theta$ の情報により  $\hat{\mu}_e$ ,  $\hat{\mu}_c$  が正しく推定されれば、5 つの試行すべてにおいてこの効用を獲得できることになる。つまり頻度論によるとパラメータ $\theta$ の EVPI<sub>f</sub>はシミュレーションの結果のうち、当該薬剤経済評価の結論と異なる、誤った意思決定が得られた割合と、正しい意思決定を行った場合に得られる効用 (これはパラメータ $\theta$ の情報が所与の下で得られた  $\hat{\mu}_e$ ,  $\hat{\mu}_c$  から求められる) の積と定義される。よって EVPI<sub>f</sub>は以下のように定義される。

$$EVPI_f = \begin{cases} \frac{\sum_{i=1}^n \#(\lambda e_i < c_i)(\lambda \hat{\mu}_e - \hat{\mu}_c)}{n} & (\text{if } \lambda \hat{\mu}_e - \hat{\mu}_c > 0) \\ \frac{\sum_{i=1}^n \#(\lambda e_i > c_i)(\hat{\mu}_c - \lambda \hat{\mu}_e)}{n} & (\text{if } \lambda \hat{\mu}_e - \hat{\mu}_c < 0) \end{cases}$$

ただし、 $\#(\cdot)$ は括弧内の式が真の場合に 1 を、偽の場合 0 を示す指示関数とする。

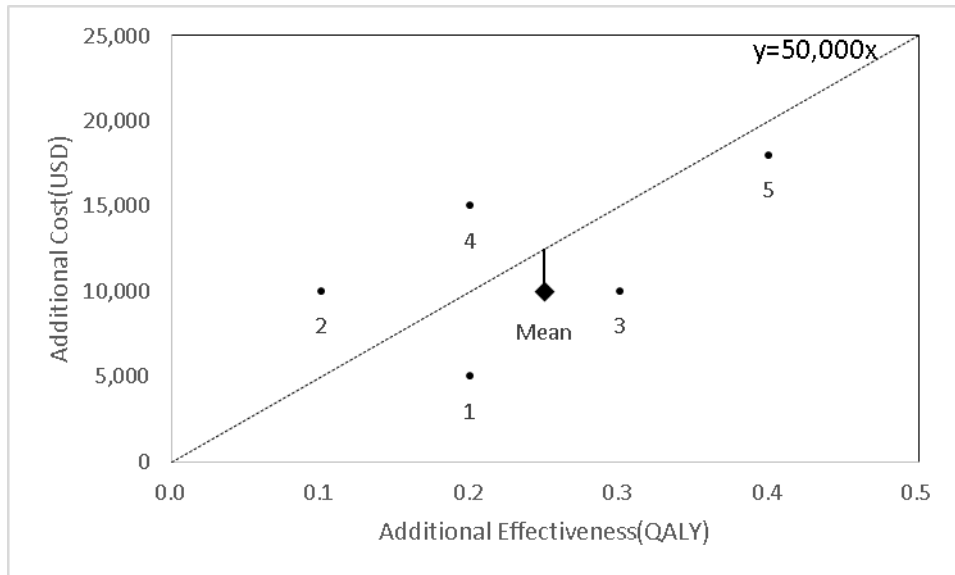


図 3-3 頻度論に基づく EVPI (EVPI<sub>f</sub>) の図示

### 3.2.2 EVPI と EVPI<sub>f</sub>の比較

薬剤経済評価及びそれに基づく意思決定における EVPI と EVPI<sub>f</sub>の比較を行った。簡単のため最も単純な例として、パラメータ $\theta$ の不確実性の結果、新治療の採用による NB が平均 $\mu$ 、分散 $\sigma^2$ の正規分布に従うとし、平均、分散をそれぞれ変化させた時の EVPI と EVPI<sub>f</sub>の変化を示す。平均の変化は NB の（分散で調整された）大きさを、分散の変化はパラメータ $\theta$ の不確実性による影響の大きさに相当する。 $\mu > 0$ の場合に、EVPI は追加的効用のうち負の値となった部分の期待値であるから、

$$\frac{E(\lambda e - c | \lambda e - c < 0)}{\Phi\left(\frac{-\mu}{\sigma}\right)}$$

で示される。一方 EVPI<sub>f</sub>は追加的効用が負の値となった割合と、得られるはずであった効用 $\mu$ の積



$$\mu\Phi\left(\frac{-\mu}{\sigma}\right)$$

となる。ただし $\Phi\left(\frac{x-\mu}{\sigma}\right)$ は平均 $\mu$ 、分散 $\sigma^2$ の正規分布に従う確率変数 $x$ の累積密度関数を示す。

### 3.3 比較結果

標準偏差を1に固定し、平均を変化させた場合の結果を図3-4に示す。EVPIでは $\mu$ が0の近傍になるにつれて情報の価値は高くなるが、EVPI<sub>f</sub>では逆に情報の価値は失われる。頻度論の立場では費用の真値が効果と支払意思額から算出される閾値に近い場合には新治療を採用するか否かの意思決定により得られる効用に大きな差がないことから、正しく判断することの価値が低い状態とみなされることによる。

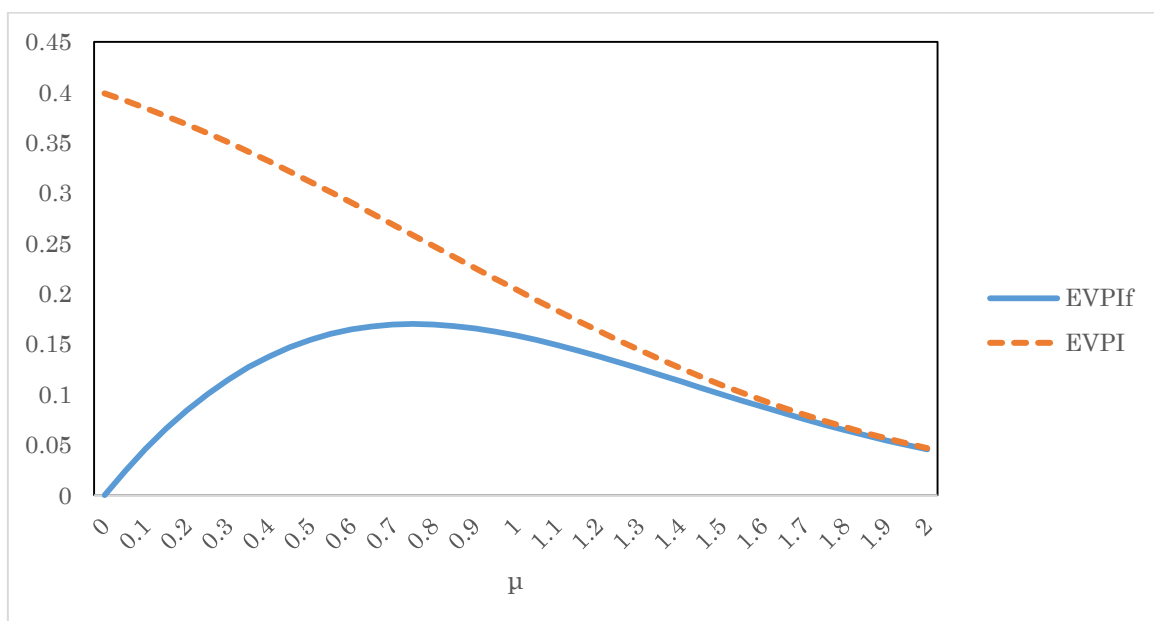


図 3-4 標準偏差一定のもとでの平均に対する EVPI と EVPI<sub>f</sub> の変動

平均を1に固定し、標準偏差を変化させた場合の結果を図3-5に示す。EVPIとEVPI<sub>f</sub>は共に標準偏差の増加に応じ単調に増加している。これはパラメータ $\theta$ の変動の影響が大きいほどその情報の価値は高くなるという点で論理的に一貫性を保つ結果であるといえる。

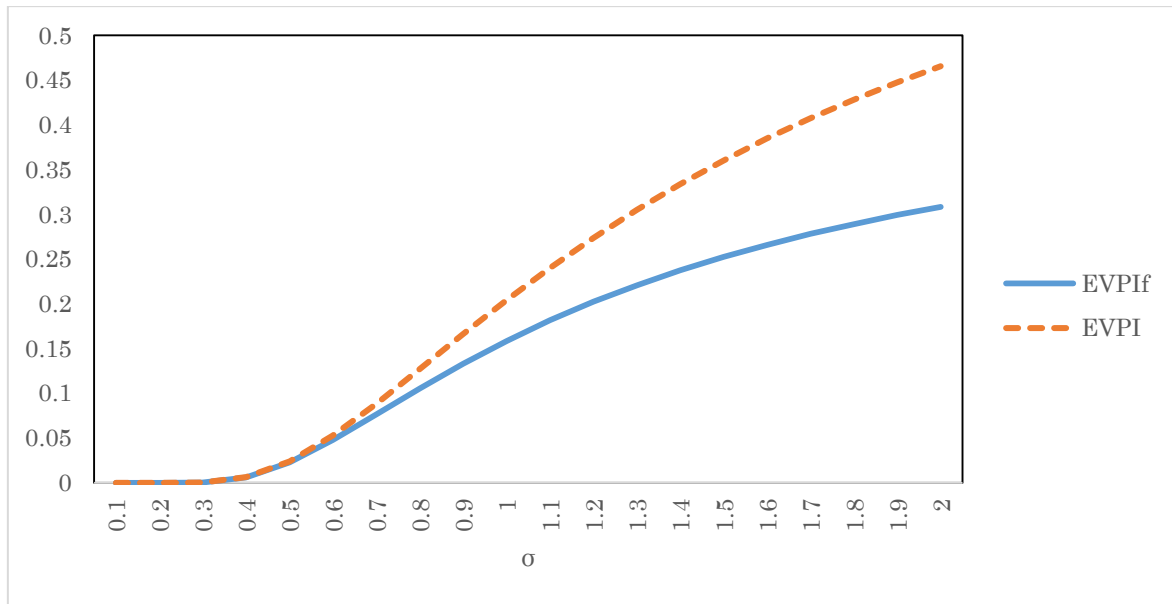


図 3-5 平均一定のもとでの標準偏差に対する EVPI と EVPIf の変動

### 3.4 EVPI<sub>f</sub>の例

Colins ら<sup>64)</sup>の例を用いて、EVPI<sub>f</sub>を算出し、EVPI との比較を行なった結果を図 3-6 に示す。EVPI<sub>f</sub>は WTP の増加と共に一旦は EVPI<sub>f</sub>が増加することが分かるが、\$10,129 近傍においてはむしろ減少している。これは、図 3-4 の $\mu$ が 0 の近傍にあるときに確認された、ICER が閾値に近い場合にむしろ情報の価値が減少し、EVPI<sub>f</sub>は低下するという性質と一致する。\$10,129 において EVPI<sub>f</sub>は 0、すなわちどちらの治療を選んだとしても得られる効用は同じである（ただし、linezolid のほうがより多い QALY を得られ、より多額の費用を支払っている）、という状態に到達する。この後 WTP の増加に従い EVPI<sub>f</sub>は増加する。

WTP が\$10,129 の点において、EVPI は\$652 である。これはすなわち、WTP 所与のもと、このパラメータの不確実性が解消されることにより、患者 1 人あたりで平均的に\$652 の費用が節減されると期待されるという意味である。この節減は、各試行において linezolid と vancomycin のより良いほうを選択できることにより発生する。一方で EVPI<sub>f</sub>はこの WTP において\$0 となるが、これは linezolid と vancomycin の真の費用対効果の差には差がないため、どちらを使用したとしてもまったく差が現れないということを示す。

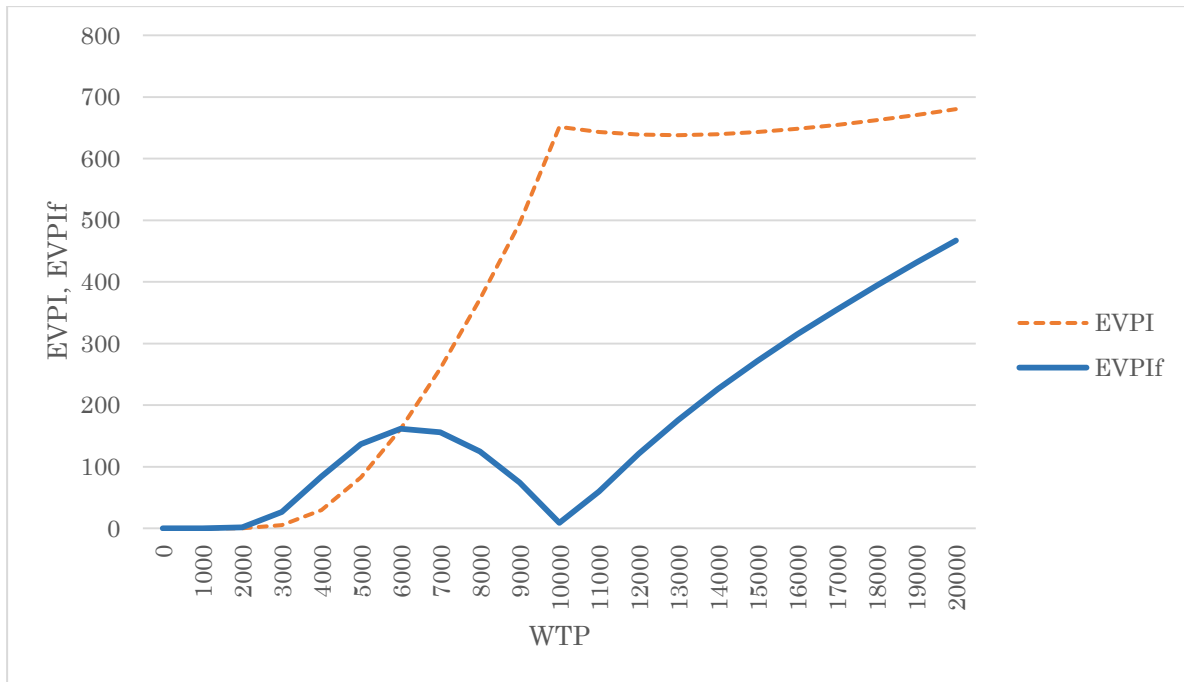


図 3-6 Collins らの例をもとにした EVPI と EVPIf の比較

### 3.5 考察

新治療採否の判断が難しい状況、すなわち手持ちのデータから推定された追加的効用が 0 近傍に分布している場合に、パラメータ  $\theta$  のばらつきを解決することで採否の判断をはっきりと行うことの価値を高いとみるか低いとみるかが、EVPI と EVPI<sub>f</sub> の最も大きな差異であるといえる。この差異について以下のように考察する。

Value of Information の理論は、もともとは景気動向の予測と株式、債券のいずれに投資するか判断のような場合に利用される。景気が改善するのであれば株式に、悪化するのであれば債権に投資したいが、景気動向を予測することは困難である、という状況が想定し、景気が改善するか悪化するかはわからないが、その確率  $p_{boom}$ 、 $p_{slump}$  を知ることは可能であり、さらに景気動向ごとに投資のもたらす収益 ( $m_{stock,boom}$ 、 $m_{stock,slump}$ 、 $m_{bond,boom}$ 、 $m_{bond,slump}$ ) が既知であるとした場合に、得られる収益の期待値が最大になるものに投資するとする。すなわち、この状態で得られる最大の収益は

$$\max(p_{\text{boom}}m_{\text{stock,boom}} + p_{\text{slump}}m_{\text{stock,slump}}, p_{\text{boom}}m_{\text{bond,boom}} + p_{\text{slump}}m_{\text{bond,slump}})$$

のように、株式と債権のいずれかに投資したいずれかの利益しか得られない。

一方で仮に景気が改善するか悪化するかを確実に予測する手段があった場合、この予測がもたらす追加的な収益の期待値をとると

$$p_{\text{boom}}\max(m_{\text{stock,boom}}, m_{\text{bond,boom}}) + p_{\text{slump}}\max(m_{\text{stock,slump}}, m_{\text{bond,slump}})$$

のように、将来の景気動向に応じた投資先の選択が可能となり、利益の期待値は情報が無い状態よりも高くなるであろう。この差を取ったものが EVPI である。

この例の場合、意思決定の段階でまだ景気動向がどうなるかは決定していないこと、景気改善/悪化の確率とそれぞれの投資のもたらす収益は独立であると設定されることが特徴であり、このようなケースにおいてベイズ流の考え方に基づく EVPI は受け入れやすいであろう。

新治療を採用するか否かの意思決定の命題にこの例をあてはめると、「景気動向」が「観測された追加的効用の正負」、「景気が改善するか悪化するかを確実に予測する」ことが「パラメータ $\theta$ のばらつきを解決する」ことに、「投資先の選択」が「新治療の採否」に読みかえられる。新治療採否の命題が景気判断の命題と異なる点は、追加的効用の大きさと新治療の採否の確率の間に相関がある点であるといえる。

入手したデータをもとに薬剤経済評価を行い、その結果新治療と対象治療の間に追加的効用の差が大きい、という結果が出た場合は、追加にデータを取り、いずれの治療が良いのか甲乙をつけたとしても大きな価値は無いであろう。前節での例を用いると、仮に WTP が \$10,129 近傍であった場合、現段階でのエビデンスによると vancomycin と linezolid のいずれを使用しても同等の費用対効果を示すことになる。この場合に、どちらの薬剤の方がより費用対効果に優れているかを決定する目的で追加データを獲得することに投資することに価値があるかを評価する場合、EVPI はそこに極大値を持つため、高い価値を見出すが、一方で  $EVPI_r$  はまったく価値を見出さず、追加データを獲得せず両者のうち任意の薬剤を活用することを推奨する。ICER の点推定値と、WTP ごとにそれぞれの指標でどのような意思決定につながるかを図 3-7 にまとめた。色が濃い箇所が検討価値が高いことを示す。

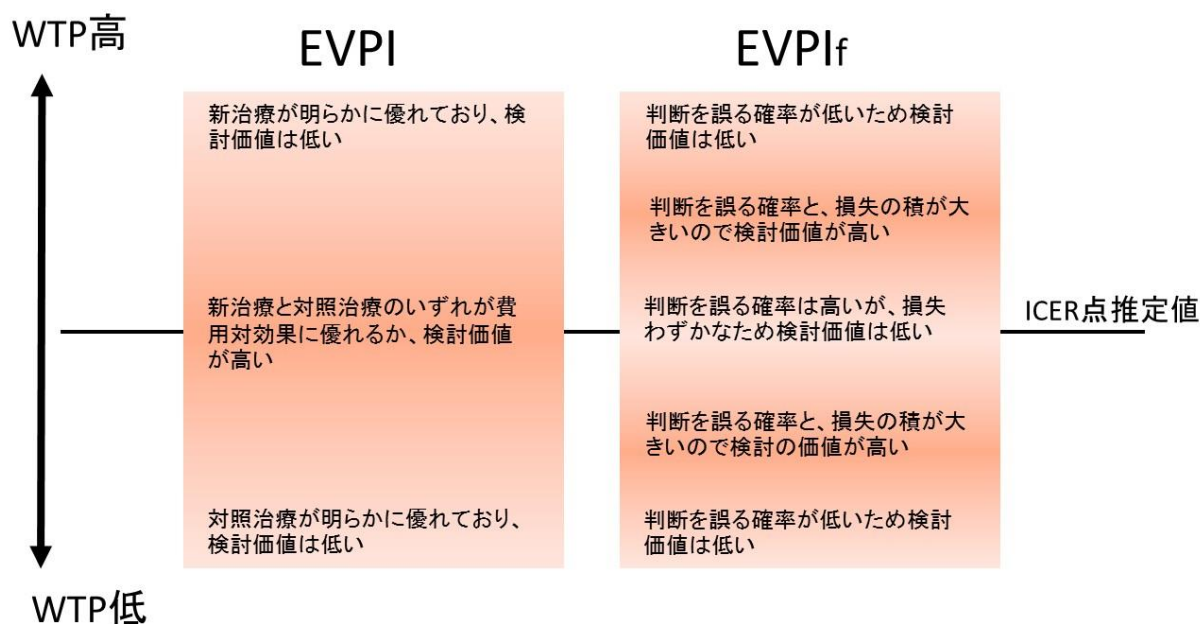


図 3-7 EVPI と EVPI<sub>f</sub> の差異

EVPI、EVPI<sub>f</sub>はいずれも、あるパラメータの不確実性が解消された場合、患者 1 人あたりどれだけの価値につながるか、を示すため、日本（あるいは世界）の患者数との積を求めると、パラメータの不確実性を解消するための研究投資が、得られる価値に見合ったものであるかを判断することができる。

ただしこれは少なくとも現時点での医療技術に関しての、経済的価値への貢献という観点からであり、研究投資が将来にわたり新たな知見をもたらす可能性は価値には含まれていない。EVPIにはいくつかの活用法が考えられるが、本研究にて提案する EVPI<sub>f</sub>は「追加データに価値があるか」という命題への回答を得るにあたり、例で示した WTP が ICER の点推定値近傍であった場合のように、比較対照薬と新薬に費用対効果上大きな違いが無い場合に、「費用対効果としてはいずれを使ったとしても大差が無い場合、これ以上研究投資を行い費用対効果に優れた方を選択し、使用したとしても大きなメリットはないであろう」という結論を導くことができることから、EVPI と比べ優れた指標であると考えられる（対照的に、

EVPIはこのような状況において、「甲乙つけることは重要であるからより多くの研究投資を推奨する」という結論に至る）。

なお、本研究において“不確実性が解消する”ということは、当該のパラメータにおいてばらつきが全くなくなる（分散が0になる）という仮の状況を想定した。現実的には分散が0になることはなく、追加的エビデンスにより分散が小さくなる、という状況を考えることが自然である。このため、元の状況と、分散が小さくなった状況の効用の期待値の差を取る Expected Value of Partial Perfect Information (EVPPPI)という指標も考案されている<sup>65)</sup>。本論分では簡単のためEVPIに着目したが、EVPPPIにおいても一般性を失わずEVPI<sub>f</sub>と同様の指標の構成が可能であろう。

### 3.6 結果のまとめ

本研究では薬剤経済評価における情報の価値の尺度であるEVPIについてその特性を考察し、頻度論に基づくEVPI<sub>f</sub>を提案、その性質の評価を行なった。治療間で追加的効用に差が少ないという結果が得られている場合において従来のEVPIでは更なる情報の獲得に大きな価値があると判断するが、EVPI<sub>f</sub>より評価を行なった場合、価値は大きくないと判断する。これはベイズ統計学と頻度論の立場の違いによるものであり、特性を理解した使い分けが重要である。

## 4 考察

本研究は、薬剤経済評価に伴う不確実性に着目し、その発生のプロセスごとにその評価法、対応についてまとめた。また、薬剤経済評価の不確実性を評価するひとつの方法である ICER の信頼区間について、本邦および海外にて行われている研究における使用状況を調査、整理した。医学研究が対象とするヒトが薬剤や治療に対する反応は均一ではなく、一般に医学研究は不確実性を完全に排除することは困難である。そのため、ランダム化比較試験や症例数設計など、不確実性の存在を認めながらも最善のエビデンスを導くための臨床試験の方法が考案されてきた。薬剤経済評価を目的とした介入試験が行われることがまれであり、既存の臨床試験の結果を用いる場合が多いが、それら臨床試験は薬剤経済評価を目的として設計されたものではないため、薬剤経済評価は通常の医学研究と比べ、研究の設計により不確実性を制御することがより難しいと考えられる。さらに、薬剤経済評価には有効性、安全性に加え費用のパラメータや、割引率などの設定が必要であり、それらに起因する不確実性はさらに増加する。

臨床研究はどの治療が有効性、安全性の観点から最善であるかを検討するものであるのに対して、薬剤経済評価はその時点において利用可能な情報を集約、評価することで有限の資源（医療費の予算）の配分に関する意思決定を行う学問であり、不確実性を内包するからといってその活用を躊躇するのではなく、不確実なデータからも現在考えられる最適の意思決定を導くことが重要となってくる。この観点から、薬剤経済評価の領域における不確実性に関する議論は今後さらに深める必要がある。

## 5 本研究の限界

本研究において示したとおり、薬剤経済評価に伴う不確実性はさまざまな原因で発生する。不確実性を（例えば ICER の信頼区間のような）単一の指標で評価することは難しく、多くの評価指標を用いて最終的な意思決定につなげることが必要不可欠である、というのが本研究で得られた知見である。しかしながら、多くの評価指標を用いた場合、単一の指標を用いた場合と比べ意思決定への活用の基準や手順は複雑となる。複数の指標が、それぞれ単体では異なる結論を示唆する場合であっても最終的にはいずれかの結論を採用し、意思決定を行う必要があるためである。本研究では不確実性を評価するための個別の方法論について論じたが、多くの評価指標をまとめ上げるプロセスについては考察を行っていない。これは薬剤経済評価がどのように利活用されるかなど、外的要因の影響を強く受けると考えられるが、今後日本における HTA の制度導入に伴い更なる研究が必要な領域である。



## 6 結論

本研究では、以下の2項目を目的とし、薬剤経済評価における不確実性について整理、調査および新たな方法の提案を行った。

1. 不確実性の評価が日本における研究にてどのように行われているかを調査し、海外のそれと比較する
2. 既存の評価法における課題を整理しその解決策として新たな方法を提案する

第1の目的において、邦文で出版された国内における薬剤経済評価においては、NICE等が推奨する方法であるPSAの使用状況は英文で出版されたものより少ないことから、今後国内における評価で不確実性の評価がより詳細に行われるため、ガイドライン等の整備が望まれることが明らかになった。第2の目的では、不確実性の評価指標として使用されているEVPIを頻度論のアプローチで定義した $EVPI_f$ を提案し、その性質を調べ、WTPがICERの点推定値近傍であるときの意思決定の結論においてEVPIと逆転することを確認した。

以上のことより、薬剤経済評価の不確実性の評価は複数の方法を併用、総合的判断を行うことで、それぞれの方法の長所、短所を補完して判断を行うことが推奨されることが結論付けられた。

## 7 謝辞

本研究の実施や本稿作成にあたり、懇切なご指導並びにご支援をいただきました国際医療福祉大学大学院薬学研究科の先生方に深甚の謝意を表します。

## 8 倫理上の配慮・利益相反

本研究は、公表論文および公開資料に基づく研究であり、倫理審査の必要性には該当しない。筆者は製薬企業である中外製薬株式会社の社員であり、本課程について2014年度以降の学費が中外製薬株式会社から支出されている。内容は筆者個人の見解であり、所属する団体・企業の見解を代表するものではない。

## 9 略称一覧

AAA: Abdominal Aortic Aneurysm

AIC: Akaike's Information Criteria

AMCP: Academy of Management Care Pharmacy

ASMR: Amélioration du Service Médical Rendu

BCEA: Bayesian Cost-Effectiveness Analysis

BIC: Bayesian Information Criteria

CBA: Cost Benefit Analysis

CDF: Cancer Drugs Fund

CEA: Cost Effectiveness Analysis

CEAC: Cost-Effectiveness Aceptability Curve

CEPS: Comité Economique des Produits de Santé

CMA: Cost Minimization Analysis

CUA: Cost Utility Analysis

CVD: Cardiovascular Disease

ECOG PS: Eastern Cooperative Oncology Group Performance Status

EQ-5D: EuroQol-5 Dimensions

ERG: Evidence Revision Group

EVAR1: The UK Endovascular Aneurysm Repair Trial 1

EVMI: Expected Value of Model Improvement

EVPI: Expected Value of Perfect Information

EVPPPI: Expected Value of Partial Perfect Information

HAS: Haute Autorité de Santé

HTA: Health Technology Assessment

HUI: Health Utility Index

ICER: Incremental Cost-Effectiveness Ratio

MCMC: Markov Chain Monte Carlo

NB: Net Benefit

NHS: National Health Service

NICE: the National Institute for health and Care Excellence

PSA: Probabilistic Sensitivity Analysis

SAS: Statistical Analysis System

SF-6D: Short Form-6 Dimensions

SMR: Service Médical RenduService

PBAC: Pharmaceutical Benefits Advisory Committee

QALY: Quality-Aadjusted Life Year

QOL: Quality of Life

QW: Quality Weight

UNCAM: Union Nationale des Caisses d'Assurance Maladie

VoI: Value of Information

WTP: Willingness to Pay

## 10 参考文献

<sup>1</sup>European Network for Health Technology Assessment Common questions. What is Health Technology Assessment (HTA). <http://eunethta.eu/about-us/faq> 2016/9/5

<sup>2</sup>東 美恵. 医療技術評価 (HTA) の諸外国での利用状況と課題 主要な新薬創出国 (米国、英国、ドイツ、フランス) における HTA/費用対効果評価の利用について. JPMA News Letter No.153. 2013.

<sup>3</sup>Cancer Research UK. Cancer Drugs Fund. <http://www.cancerresearchuk.org/about-cancer/cancers-in-general/cancer-questions/cancer-drugs-fund> 2016/9/5

<sup>4</sup><http://www.cochranelibrary.com/> 2016/9/17

<sup>5</sup>Drummond MF, Sculpher MJ, Torrance GW, et.al. Methods for the economic evaluation of health care programmes 3rd Edition. Oxford University Press. 2005.

<sup>6</sup>National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. <https://www.nice.org.uk/process/pmg9/chapter/foreword> 2016/9/5

<sup>7</sup><http://www.euroqol.org/> 2017/1/30

<sup>8</sup><http://www.healthutilities.com/> 2017/1/30

<sup>9</sup><https://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d> 2017/1/30

<sup>10</sup>医療経済評価研究における分析手法に関するガイドライン. 厚生労働科学研究費補助金政策科学総合研究事業「医療経済評価を応用した医療給付制度のあり方に関する研究」(研究代表者: 福田敬) 平成 24 年度総合研究報告書. 2013.

<sup>11</sup>白岩健. 「医療経済評価研究における分析手法に関するガイドライン」の解説. 保健医療科学. 2013; 62(6): 590-8.

<sup>12</sup>日本製薬工業協会データサイエンス部会 2014 年 タスクフォース 3. 医薬品の価値の科学的な評価ーデータサイエンス担当者のための費用対効果評価の現状と手法の解説ー. [http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/pdf/2014ds\\_tf3.pdf](http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/pdf/2014ds_tf3.pdf) 2016/9/15

<sup>13</sup>Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*. 2000; 17(5): 479–500.

- <sup>14</sup>National Institute for Health and Care Excellence. Lumacaftor–ivacaftor for treating cystic fibrosis homozygous for the F508del mutation. Appraisal consultation:1 Committee papers. <https://www.nice.org.uk/guidance/TA398/documents/committee-papers-2016/10/20>
- <sup>15</sup>Jackson CH, Thompson SG, Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*. 2009; 172(2): 383–404.
- <sup>16</sup>Strong M, Oakley JE. When is a model good enough? Deriving the expected value of model improvement via specifying internal model discrepancies. *SIAM/ASA Journal on Uncertainty Quantification*. 2014; 2(1): 106–25.
- <sup>17</sup>Akaike H. "Information theory and an extension of the maximum likelihood principle", in Petrov, B.N.; Csáki, F., 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akadémiai Kiadó. 1973; 267-81.
- <sup>18</sup>Schwarz G. Estimating the dimension of a model. *Ann. Statist.* 1978; 6(2): 461–4.
- <sup>19</sup>EVAR Trial Participants. Endovascular aneurysm repair versus open repair in patients with abdominal aortic aneurysm (EVAR trial 1): randomised controlled trial. *Lancet*. 2005; 365(9478): 2179–86.
- <sup>20</sup>Epstein DM, Sculpher M, Manca A, et al. Modelling the long-term cost-effectiveness of endovascular or open repair for abdominal aortic aneurysm. *Br J Surg*. 2008; 95(2): 183–90.
- <sup>21</sup>The Comprehensive R Archive Network. <https://cran.r-project.org/> 2016/10/23
- <sup>22</sup>Package 'survival'. <https://cran.r-project.org/web/packages/survival/survival.pdf> 2016/10/23
- <sup>23</sup>Edmonson JH, Fleming TR, Decker DG, et al. Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal residual disease. *Cancer Treat Rep*. 1979; 63(2): 241-7.
- <sup>24</sup><http://ecog-acrin.org/resources/ecog-performance-status> 2017/1/30
- <sup>25</sup>Buckland ST, Burnham KP, Augustin NH. Model selection: an integral part of inference. *Biometrics*. 1997; 53(2): 603–18.

- <sup>26</sup>Burnham KP, Anderson DR. Model selection and multi-model inference: a practical information- theoretic approach. *New York: Springer*. 2002.
- <sup>27</sup>Strong M, Pilgrim H, Oakley J, et al. Structural uncertainty in health economic decision models. *SchHARR Occasional Paper*. 2009.
- <sup>28</sup>Caro JJ, Möller J. Decision-analytic models: current methodological challenges. *Pharmacoeconomics*. 2014; 32(10): 943–50.
- <sup>29</sup>Strong M, Oakley JE, Chilcott J. Managing structural uncertainty in health economic decision models: a discrepancy approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2012; 61(1): 25–45.
- <sup>30</sup>National Institute for Clinical Excellence Four commonly used methods to increase physical activity: PH2. Technical Report. National Institute for Clinical Excellence, London. 2006.  
<https://www.nice.org.uk/guidance/ph2> 2016/09/17
- <sup>31</sup>Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics*. 2006; 24(11): 1055–68.
- <sup>32</sup>Raiffa H, Schlaifer R. Probability and Statistics for Business Decisions. 1959.
- <sup>33</sup>Pratt J, Raiffa H, Schlaifer R. Statistical Decision Theory. *Cambridge MA*. 1995.
- <sup>34</sup>Thompson KM, Evans JS. The value of improved national exposure information for perchloroethylene (perc): a case study for dry cleaners. *Risk Analysis*. 1997; 17(2): 253-71.
- <sup>35</sup>Howard RA. Information value theory. *IEEE Transactions on Systems Science and Cybernetics SSC2*. 1966; 2(1):122-26.
- <sup>36</sup>Hammit JK, Cave JAK. Research planning for food safety: a value of information approach.: *RAND publication series*. 1991.
- <sup>37</sup>Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics*. 2008; 26(9): 781–98.
- <sup>38</sup>Obenchain, RL. Resampling and multiplicity in cost-effectiveness inference. *Journal of Biopharmaceutical Statistics*. 1999; 9(4): 563–82.



- <sup>39</sup>Glick HA, Doshi JA, Sonnad SS, et al. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press; 2007.
- <sup>40</sup>Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, West Sussex: John Wiley & Sons Ltd. 2004.
- <sup>41</sup>Baio G. *Bayesian Methods in Health Economics*. CRC Press. 2013.
- <sup>42</sup>O'Hagan A, Stevens J W. Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. *Statistical Methods in Medical Research* 2002; 11(6): 469-90.
- <sup>43</sup>Baio G. BCEA: An R package to perform Bayesian Cost-Effectiveness Analysis. <https://sites.google.com/a/statistica.it/gianluca/bcea> 2014/3/2
- <sup>44</sup>汪 金芳, 田栗 正章, 手塚 集ら. 統計科学のフロンティア 11 計算統計 I 確率計算の新しい手法 2003.
- <sup>45</sup>Shiroiwa T, Sung YK, Fukuda T, et al. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? *Health Economics* 2010; 19(4): 422-37.
- <sup>46</sup>大日康史, 菅原民枝. 1QALY 獲得に対する最大支払い意思額に関する研究. *医療と社会* 2006; 16(2): 157-65.
- <sup>47</sup>東 美恵. 日本における費用対効果の政策利用と評価方法に関する示唆. 平成 26 年度国際医療福祉大学審査学位論文 (博士) 薬学研究科博士課程.
- <sup>48</sup>池田 俊也, 小野塚 修二. 医薬品の価格算定と薬剤経済学 -応用への道筋-. 医薬産業政策研究所リサーチペーパー・シリーズ No.19. [http://www.jpma.or.jp/opir/research/rs\\_019/paper\\_19.pdf](http://www.jpma.or.jp/opir/research/rs_019/paper_19.pdf) 2016/12/10
- <sup>49</sup>Australian Government Department of Health and Aging. Guidelines for preparing submissions to the pharmaceutical benefits advisory committee (Version 4.3). 2008. <https://pbac.pbs.gov.au/content/information/archived-versions/pbac-guidelines-v4-3-2008.pdf> 2017/1/30

- <sup>50</sup>Department of Economics and Public Health Assessment. Choices in methods for economic evaluation. 2012. [http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices\\_in\\_methods\\_for\\_economic\\_evaluation.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf) 2017/1/30
- <sup>51</sup>Academy of Managed Care Pharmacy. A format for submission of clinical and economic evidence of pharmaceuticals in support of formulary consideration Version 3.1. 2012. <http://www.amcp.org/practice-resources/amcp-format-formulary-submissions.pdf> 2017/1/30
- <sup>52</sup>医療経済評価研究における分析手法に関するガイドライン.厚生労働科学研究費補助金（政策科学総合研究事業）「医療経済評価を応用した医療給付制度のあり方に関する研究」（研究代表者：福田敬）平成 24 年度総合研究報告書.2013.
- <sup>53</sup>Okubo I, Kondo M, Toi M, et al. Cost-effectiveness of Letrozole versus Tamoxifen as First-line Hormonal Therapy in Treating Postmenopausal Women with Advanced Breast Cancer in Japan. *癌と化学療法*. 2005; **32**(3): 351-63.
- <sup>54</sup>Amemiya S, Takao H. Computed tomographic coronary angiography for diagnosing stable coronary artery disease: a cost-utility and cost-effectiveness analysis. *Circulation Journal*. 2009; **73**(7): 1263-70.
- <sup>55</sup>Tamayama K, Hoshi SL, Kondo M, et al. Cost-effectiveness analysis of pre-seasonal medication for cedar pollinosis in Japan. *薬学雑誌*. 2010; **130**(12): 1725-36.
- <sup>56</sup>池田 俊也, 小川 節郎, 細川 豊史, 他. 末梢性神経障害性疼痛に対する Pregabalin の費用対効果分析. *薬剤疫学*. 2011; **16**(1): 1-9.
- <sup>57</sup>井上 裕智. 中学 1 年生および高校 3 年生に対する麻疹ワクチン追加接種の経済評価. *医療経済研究*. 2011; **22**(2): 141-57.
- <sup>58</sup>駒場 大峰. 重度の二次性副甲状腺機能亢進症に対する副甲状腺摘出術における術前 99mTc-MIBI シンチグラフィの費用効果分析. *日本透析医会雑誌*. 2012; **27**(2): 361-2.
- <sup>59</sup>Briggs, AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*. 2000; **17**(5): 479–500.

<sup>60</sup>Howard RA. Information value theory. *Systems Science and Cybernetics, IEEE Transactions on*, 1966; 2.1: 22-6.

<sup>61</sup>Tappenden P, Chilcott JB, Eggington S, et al. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. *Health Technol Assess*. 2004; **8**(27):iii, 1-78.

<sup>62</sup>Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics*. 2006; **24**(11): 1055-68.

<sup>63</sup>Strong M, Oakley EJ. When is a model good enough? Deriving the expected value of model improvement via specifying internal model discrepancies. *SIAM/ASA J. Uncertainty Quantification*. 2014; **2**(1): 106–25.

<sup>64</sup>Collins CD, Schwemm AK. Linezolid versus vancomycin in the empiric treatment of nosocomial pneumonia: a cost-utility analysis incorporating results from the ZEPHYR trial. *Value in Health*. 2015; **18**(5): 614-21.

<sup>65</sup>Oostenbrink JB, Al MJ, Ooep M et al. Expected value of perfect information: an empirical example of reducing decision uncertainty by conducting additional research. *Value in Health*. 2008; 11(7):1070-80.