

医薬品・医療機器等の費用対効果評価における分析ガイドラインの改定に資する研究
分担研究報告書

医療経済評価の分析ガイダンス改訂に関する研究

研究代表者 福田 敬 国立保健医療科学院保健医療経済評価研究センター
研究分担者 白岩 健 国立保健医療科学院保健医療経済評価研究センター

医療経済評価を行う際には、その透明性や比較可能性を担保するために、共通の枠組みに基づいて分析を行うことが望ましい。多くの国々では、政策応用にあたって基本的な分析手法を定めた医療経済評価ガイドラインが医療技術評価機関などによって作成されてきた。我が国においても、費用対効果評価が2019年4月に制度導入されるのにあわせて、『中央社会保険医療協議会における費用対効果評価の分析ガイドライン 第2版』（2019年2月20日 中医協総会了承）が本研究班の前身となる厚生労働科学研究費において作成され、以後診療報酬改定時における制度改正あるいは最新の学術的な状況に対応するため、2年に1回の頻度で改定されている。本研究班においても、各国における医療経済評価ガイドライン、学術的な状況の進展（学会、学術誌等による最新の報告など）、我が国における費用対効果評価制度の改定方針、公的分析班等からのインプットを受けて、研究班員により改定方針について検討を行い、改定後のガイドラインを作成した。

A. 研究目的

医療経済評価を行う際には、その透明性や比較可能性を担保するために、共通の枠組みに基づいて分析を行うことが望ましい。多くの国々では、政策応用にあたって基本的な分析手法を定めた医療経済評価ガイドラインが医療技術評価機関などによって作成されてきた。例えば、イギリス National Institute for Health and Care Excellence (NICE) では、“NICE health technology evaluations: the manual” が2022年1月末に改正されている。また、欧米諸国

のみではなくアジア諸国においても、政策応用が進んでいる韓国・台湾・タイ等においてはすでに医療経済評価ガイドラインが公表されている。

我が国においても、費用対効果評価が2019年4月に制度導入されるのにあわせて、『中央社会保険医療協議会における費用対効果評価の分析ガイドライン 第2版』（2019年2月20日 中医協総会了承）が本研究班の前身となる厚生労働科学研究費において作成され、以後診療報酬改定時における制度改正あるいは最新の学術的な状況に対応するため、2年

に1回の頻度で改定されている。

B. 研究方法

各国における医療経済評価ガイドライン、学術的な状況の進展(学会、学術誌等による最新の報告など)、我が国における費用対効果評価制度の改定方針、公的分析班等からのインプットを受けて、研究班員により改定方針について検討を行った。研究班員による自由議論の後、ドラフトとなる改正案をまとめ、それに基づき対面での会議、あるいは電子メール上等でのやり取りを繰り返すことにより、研究班員間での合意を得た。これらのドラフトについては、業界団体(製薬協、PhRMA、EFPIA等)からのレビューも受けて、最終案は2024年1月27日に中医協総会において了承された。

C. 研究結果

改定されたガイドラインは以下の14省からなる、すなわち「1. ガイドラインの目的」「2. 分析の立場」「3. 分析対象集団」「4. 比較対照技術」「5. 追加的有用性」「6. 分析手法」「7. 分析期間」「8. 効果指標の選択」「9. データソース(費用を除く)」「10. 費用の算出」「11. 公的介護費用・生産性損失の取り扱い」「12. 割引」「13. モデル分析」「14. 不確実性の取り扱い」である。

前半からの大きな改定点は以下の通りである。

- ・対象患者数の推計方法について基本となる考え方を示した。(3.3節)

- ・比較対照技術の選定方法につき、中医協における議論も踏まえ、明確化した。(4.1節)

- ・追加的有用性を検討する際の観察比較のあり方について、より議論を精緻化した。(5.7節)

- ・ICERを算出する際に、比較対照技術と比して費用も効果もほぼ同等であるものに関する対応について追加した。(6.3.3節)

- ・QOL値の測定について特にビニエツト法を用いた場合の調査方法について明確化した。(8.6節)

- ・QOL値について複数の測定値がある場合の優先順位づけを行った。(8.8節)

- ・小児のQOL値測定に関する記述を追加した。(8.9節)

- ・費用対効果の推計をあたって想定する治療プロセスについて明確化を求めた。(9.1節)

- ・マイクロシミュレーションなどの結果が確率的に変動するモデルに関する間考え方をまとめた。(13.7節)

- ・感度分析における閾値分析の取り扱いを追記した。(14.6.1節)

一方で、中医協費用対効果評価専門部会で大きな議論となった公的介護費用の取り扱い等については、特例に基づいて実施されるレカネマブの分析結果などを見てからの検討が望ましいと考えられたため、現時点では改正等を行っていない。

D. 考察

費用対効果評価における分析ガイドラ

インは、社会情勢の変化や最新の学術的動向によって、適時改定されていくことが望ましい。本研究により、より学術的に適切な分析が製造販売業者、公的分析班ともに提出されるようになることを期待している。

なお、本ガイドラインはWEB上で下記から入手可能である。

(日本語版)

https://c2h.niph.go.jp/tools/guideline/guideline_ja_2024.pdf

(英語版)

https://c2h.niph.go.jp/tools/guideline/guideline_en_2024.pdf

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況

なし

参考文献

なし

中央社会保険医療協議会における費用対効果評価の分析ガイドライン 2024 年度版

1 ガイドラインの目的	5
2 分析の立場	6
3 分析対象集団	7
4 比較対照技術	9
5 追加的有用性	10
6 分析手法	15
7 分析期間	17
8 効果指標の選択	18
9 データソース(費用を除く)	23
10 費用の算出	25
11 公的介護費・生産性損失の取り扱い	28
12 割引	30
13 モデル分析	31
14 不確実性の取り扱い	33
用語集	35
略語一覧	44

1 ガイドラインの目的

1.1 本ガイドラインは、中央社会保険医療協議会において、評価対象として選定された医薬品・医療機器・再生医療等製品(以下、評価対象技術)の費用対効果評価を実施するにあたって用いるべき分析方法を提示している。

1.2 本ガイドラインは製造販売業者により提出される分析と公的分析を対象としている。

2 分析の立場

2.1 分析を行う際には、分析の立場を明記し、それに応じた費用の範囲を決めなければならない。

2.2 費用や比較対照技術、対象集団などについて公的医療保険制度の範囲で実施する「公的医療の立場」を基本とする。

2.2.1 「公的医療の立場」以外の立場から分析を実施する場合でも、「公的医療の立場」の分析を実施しなければならない。

2.2.2 一部の予防技術(検診やワクチン等)など厳密には公的医療保険制度に含まれないものでも、それに準ずる医療技術であれば、公的医療保険制度に含まれるものと同様の取り扱いをした分析を実施する。

2.3 公的介護費へ与える影響が、評価対象技術にとって重要である場合には、「公的医療・介護の立場」の分析を行ってもよい。

2.4 評価対象技術の導入が生産性に直接の影響を与える場合には、より広範な費用を考慮する立場からの分析を行い、生産性損失を費用に含めてもよい。

3 分析対象集団

3.1 品目の指定時点において、評価対象技術の適応症となる患者を分析対象集団とする。

3.1.1 品目の指定から分析枠組み決定までの間に適応症が追加される場合は、それらの適応についても原則として分析対象に含める。

3.1.2 「3.1.1」で定める期間より後に新たな適応症が追加され、評価結果に影響を与えと考えられる場合、当初の評価終了後に改めて評価を実施する。

3.2 複数の適応がある場合、あるいは同一疾患内においても治療成績や使用方法・用法用量、比較対照技術が異なる主要な集団がある場合は、各集団についてそれぞれ分析を実施することを原則とする。

3.2.1 ただし、「3.2」を実施することが困難な状況下で、患者数や疾患の性質等を勘案して結果に与える影響が限定的な場合には、協議における両者(製造販売業者と国立保健医療科学院/公的分析班：以下同様)の合意のもとで、分析対象集団から一部集団を除外できる。

3.3 各集団において、評価対象品目に関する現在の臨床実態や使用実績等を考慮して、長期的な観点からの患者割合(特許期間の累積患者数に基づく患者割合等)を推計することを原則とする。

3.3.1 推計が困難な場合は、上市から一定期間後の安定した状態における断面

の患者割合を用いてもよい。

3.3.2 短期間における無視しうる影響(例えば評価対象技術の導入期における待機患者の影響など)は含めない。

4 比較対照技術

4.1 評価を行う際の比較対照技術は、評価対象技術が分析対象集団への治療として導入された時点で、臨床的に幅広く使用されており、評価対象技術によって代替されると想定されるものとする。うち、治療効果がより高いものを一つ選定することが原則的な考え方である。

4.1.1 「臨床的に幅広く使用されている」とは、使用患者数のシェアで一律に決めるものではなく、診療ガイドラインに記載があるなど臨床的に標準的な治療法として用いられていることを意味する。

4.1.2 「治療効果がより高いもの」を検討するにあたっては、既存の公表された費用対効果評価における追加的有用性の評価についても参照する。

4.2 「4.1」において、一意に決めることが難しい場合は、無作為化比較試験(Randomized controlled trial: RCT)等における比較対照技術、価格算定上の類似技術、費用対効果の程度等も考慮して最も妥当と考えられる比較対照技術を両者の協議により選定する。

4.3 比較対照技術としては無治療や経過観察を用いることもできる。

4.4 「4.3」の場合を除いて、比較対照技術は原則として公的医療保険で使用が認められているものとする。

4.5 比較対照技術として選定した理由については十分に説明する。

5 追加的有用性

5.1 費用対効果を検討するにあたっては、評価対象技術の比較対照技術に対する追加的な有用性がデータによって示されているか、示されていないかをまず評価する。

5.2 追加的な有用性を検討する際に、「4.」で選定した比較対照技術に対する RCT のシステマティックレビュー(Systematic review: SR)を実施し、追加的有用性の有無を評価する。適切なものであれば公開されていない臨床研究や治験の結果等を含めてよい。

5.2.1 SR を実施する際には、リサーチクエスチョン(Research question: RQ)を明確に提示する。例えば PICO(P: 患者(Patient), I: 介入(Intervention), C: 比較対照(Comparator), O: アウトカム(Outcome))などで構造化された RQ を定義する。

5.2.2 評価対象技術あるいは「4.」で選定した比較対照技術とアウトカムが同等であると考えられる同一作用機序あるいは同一機能区分等の類似技術について、それらを SR における介入(I)あるいは比較対照(C)に含めてもよい。

5.2.3 「5.2.1」におけるアウトカム(O)指標は、臨床的な有効性・安全性・健康関連 QOL の観点のうち、評価対象技術の特性を評価する上で、適切なもの(真のアウトカム指標など)を用いる。

5.2.4 PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses)声明の原則に従い、研究の組み入れ基準・除外基準、使用した

データベース、検索式、論文選択のプロセス(情報フロー図を含む)等を記載する。

5.2.5 信頼できる既存の SR が存在する場合、それらを活用することは可能であるが、RQ に合致するか、最新の文献まで含まれているか等を検討した上で、そのまま使用するあるいは追加的な SR を実施して研究を追加する等の検討を行う。

5.2.6 実施することが適切な場合は、対比較のメタアナリシスを用いて各試験を統合した結果を提示する。その場合、用いた統計手法、異質性の評価、フォレストプロット、統合した値とその信頼区間等を明らかにする。

5.2.7 SR における文献検索終了時点は、分析枠組みが決定された後から製造販売業者による分析提出までの一時点に決める。

5.2.8 「5.2.7」の文献検索期間終了後に費用対効果評価において重要と考えられる臨床研究(症例数の多い主要な研究結果、現在の知見と異なる結果が得られた信頼性の高い研究等)が公表等された場合、公的分析においてはその影響等を考慮し、それらを含めることを検討する。

5.3 「5.2」の SR の結果、適切なものが存在しない場合、「5.2」のプロセスに基づき、アウトカムを比較した非 RCT(観察研究等)の SR を実施し、追加的有用性を評価する。

5.3.1 非 RCT は RCT と比べて種々のバイアスが入りやすいので、研究の質(研究デザイン、群間での患者背景の差異、統計解析手法、患者数や施設数等)について十分に説明しなければならない。

5.3.2 レセプトやレジストリーなど既存の大規模なデータベース(リアルワールドデータ)を用いた研究については、研究の質にばらつきがあることから、データベースの性質、国内外における医療環境の差異、変数やイベントの定義、その妥当性やバリデーションの有無、解析手法、日本への結果の外挿可能性等について十分に説明しなければならない。

5.4 評価対象技術の有効性の評価において、一定の質を有する RCT が存在する場合、非 RCT による結果は RCT の結果を補完するものであることが基本的な考え方である。

5.5 「4.」で選定した比較対照技術との RCT は存在しないが、その他の RCT が存在する場合、SR の結果を用いて、間接比較により追加的有用性の評価をしてもよい。

5.6 単群試験しか存在しない場合は、評価対象技術と比較対照技術それぞれについての SR に基づき、間接比較を実施する。

5.7 間接比較を行う場合、以下を原則的な考え方とする。

5.7.1 個人レベルのデータを用いることができる場合、MAIC(Matching-adjusted indirect comparison: マッチング調整された間接比較)等を用いて患者背景の違いなど適切な補正を行う。

5.7.2 個人レベルのデータを用いることができない場合、あるいは手法としてより望ましいと考えられる場合は、複数の RCT を用いたネットワークメタアナリシスなどを実施する。

5.7.3 「5.7.1」および「5.7.2」においては、アンカーのある分析手法に基づくことが望ましい。

5.7.4 間接比較を行う場合は、間接比較を可能とする仮定や前提条件(疾患、重症度、患者背景等の異質性や試験の同質性など)についても十分に説明しなければならない。

5.7.5 個人レベルデータも比較試験結果も利用できない場合、その他の代替手段が存在しなければ naïve indirect comparison(単純な間接比較)を用いて評価をすることもできる。この場合、群間の比較可能性について、より詳細な説明が必要である。

5.7.6 複数の分析手法が実施可能な場合、実施した方法を選択した理由(仮定のあてはまりなど)について十分に説明し、必要に応じて異なる手法による感度分析を実施する。

5.7.7 ネットワークメタアナリシスを実施する場合、以下の点についても留意し、設定した理由を十分に説明する。必要に応じて、下記の設定について感度分析を実施することを推奨する。

- (A) ネットワークの広さ(治療群の数)
- (B) 各群の設定(薬効群等でまとめるか個別製品ごとに設定するか)
- (C) 投与量や投与方法の違う介入の取り扱い
- (D) 日本における未承認薬等の取り扱い

5.7.8 ネットワークメタアナリシスを実施した場合、第三者が再現可能なデータと解

析プログラムを可能な限り提出することとする。

5.8 追加的有用性の評価に用いた各研究(SR で特定したものやネットワークメタアナリシスに組み込んだものなど)については、文献情報を報告書中に提示する。

5.9 「5.3」から「5.7」までの手法において得られた結果が、研究の質に課題があると判断されるものの治療効果が劣っているとは考えられない場合には、評価対象技術のアウトカムが比較対照技術と同等であるという前提で、「6.」の分析を実施する。

5.10 ヒトを対象とした適切な臨床研究が存在しない場合、適切ならば、医薬品医療機器総合機構からの薬事承認をもって、比較対照技術と同等であるという前提で、「6.」の分析を実施する。

5.11 「5.2」から「5.7」までの結果、アウトカムが比較対照技術と比べて劣ると判断される場合は、費用対効果の分析は実施しない。

6 分析手法

6.1 効果を金銭換算せず、費用と効果を別々に推計する費用効果分析を分析手法として用いることを原則とする。

6.2 「5.」の分析に基づき、追加的有用性が示されていると判断される場合には、各群の期待費用と期待効果から増分費用効果比(Incremental cost-effectiveness ratio: ICER)を算出する。

6.3 ただし、以下の場合については、各群の期待費用と期待効果のみを提示して、ICER は算出しないこととする。

6.3.1 比較対照技術と比べて効果が高く、かつ費用が安い場合。このとき、ICER を算出せずに優位(dominant)であるとする。

6.3.2 「5.」の分析により、追加的有用性が示されていない場合には、比較対照技術と費用を比較する(いわゆる「費用最小化分析(Cost-minimization analysis: CMA)」)。このとき、結果は費用削減(同等含む)あるいは費用増加とする。

6.3.3 「5.」の分析により、追加的有用性が示された場合でも、増分効果がわずかに正で、増分費用も点推定値としては正負あるものの、ほぼゼロとみなせるときに、結果の不確実性が大きく、パラメータの軽微な変更等によっても見かけ上 ICER が大きく変動することがある。このとき、結果は費用及び効果同等とする。

6.4 「3.2」あるいは「3.2.1」により、分析対象となる複数の疾患や、同一疾患内での複数の分析集団がある場合、ICER は疾患や集団ごとに算出する。

6.5 評価対象技術に関する既存の費用効果分析や主要な諸外国の医療技術評価機関において公表された評価結果が既に存在する場合は、それらもあわせて示す。

7 分析期間

7.1 評価対象技術の費用や効果におよぼす影響を評価するのに十分に長い分析期間を用いる。

7.2 費用と効果は、原則として同じ分析期間を用いる。

7.3 分析期間については、そのように設定した理由を説明すること。

8 効果指標の選択

8.1 効果指標は質調整生存年(Quality-adjusted life year: QALY)を用いることを原則とする。

8.1.1 QALY を算出することが困難であり、かつ CMA を実施する場合、適切であれば、QALY 以外の評価尺度を使用することもできる。

8.2 QALY を算出する際の QOL 値は、選好に基づく尺度(preference-based measure: PBM)により測定したものをを用いることを原則とする。

8.2.1 QOL 値は日本における一般集団の選好が反映されたものをを用いる。

8.2.2 使用する PBM としては、日本語版の EQ-5D-5L を第一選択として推奨する。

8.3 PBM により QOL 値を測定する場合には、対象者本人が回答することが原則である。

8.3.1 PBM を用いる場合、対象者本人から回答が得られない場合に限り、家族や介護者等による代理の回答を用いてもよい。

8.3.2 PBM を用いる場合、医療関係者による代理回答は、対象者本人の回答と乖離する可能性があるため、推奨しない。

8.4 QOL 値は、「8.2」を満たすものがある限り、国内での調査結果を優先的に使用することを推奨する。

8.4.1 ただし、国内における研究がないあるいは不十分で、海外で質の高い研究がなされている場合は、海外で測定されたものを使用してもよい。

8.4.2 海外で得られた PBM の回答については、日本における換算表を用いて QOL 値を集計するなど「8.2.1」の原則と一致するよう対応する。

8.4.3 個人レベルのデータが利用できず、「8.4.2」が困難でかつ結果に大きな影響を与える場合は、日本における QOL 値に対応しているか十分な説明を行い、必要に応じて日本における QOL 値に補正するなどを検討する。

8.5 「8.2」に該当するデータが存在しない場合、その他の適切な健康関連 QOL(Health-related quality of life: HRQOL)データから QOL 値へマッピングしたものを使用してもよい。

8.5.1 マッピングを実施する場合は、マッピング関数により得られる QOL 値が「8.2.1」の観点から妥当なものであるかを検討する。

8.5.2 QOL 尺度間に概念上の重なりが乏しいあるいは存在しない場合は、マッピングを実施することの適切性について検討する。

8.5.3 使用するマッピング式の開発過程や性質については、MAPS(MApping onto Preference-based measures reporting Standards)声明のチェックリスト等を利用して詳細な報告が必要である。

8.5.4 詳細な報告ができないマッピング関数を使用することは推奨しない。

8.6 対象者本人から回答を得ることが困難な場合などには、仮想的な健康状態のシナリオを用いて、一般の人々を対象に SG(Standard gamble: 基準的賭け)法や TTO(Time trade-off: 時間得失)法、DCE(Discrete choice experiment: 離散選択実験)法などの手法を用いて測定してもよい(ビニエツ法)。

8.6.1 ビニエツ法で測定された QOL 値は、提示する健康状態のシナリオにより結果が大きな影響を受けることに留意が必要である。

8.6.2 評価に用いる仮想的な健康状態のシナリオについては、患者から直接得られたものではないことから、作成方法やその妥当性を十分に説明する。また、患者や臨床専門家のインプットを受けることが望ましい。

8.6.3 結果を活用する場合は、評価に使用した実際のシナリオを提示すること。

8.6.4 「8.6.1」から費用対効果のデータソースとして過剰な推計とならないよう同一の健康状態においては同一の QOL 値を用いるなどの留意が必要である。

8.6.5 WEB を用いた SG 法と TTO 法の調査については、得られる結果にバイアスが入ることが知られていることから、対面での調査を行うことが望ましい。

8.6.6 ビニエツ法については、国内外の差異について検討が困難である一方で、調査実施が比較的容易であることから、国内データを用いることを推奨する。

8.7 「公的医療・介護の立場」からの分析の場合、実際のデータがあれば家族等の介護者や看護者に与える QOL 値への影響について考慮に入れてもよい。

8.8 QOL 値については、測定研究の質が同等である場合、「8.2」の原則に基づき、以下の順序で使用を検討していくことを推奨する。下記があてはまらない場合は、理由を十分に説明すること。

- (A) 日本語版の EQ-5D-5L を用いて収集したもの（「8.2」）
- (B) 日本での換算表が存在する EQ-5D-5L 以外の日本語版の汎用的 (generic)PBM を用いて収集されたもの
- (C) 日本での換算表が存在する日本語版の疾患・病態特異的 (condition specific)PBM を用いて収集されたもの
- (D) 海外で PBM を用いて収集され、個票データが利用可能なもの。尺度の優先順位は(A)-(C)の考え方による。（「8.4」）
- (E) 海外で PBM を用いて収集され、個票データが利用できないもの。（「8.4」）
- (F) マッピングを用いて QOL 値へと変換したもの。変換先の尺度やマッピング式の国内外差の取り扱いについては(A)-(E)に準じる。（「8.5」）
- (G) 日本での換算表が存在しない PBM により収集されたもの
- (H) ビニエツ法を用いて収集したもの（「8.6」）
- (I) その他の手法についてはやむを得ない事情や、学術的な妥当性が示されていない限り、使用を推奨しない。

なお、日本において開発された性質のよいマッピング式などは(F)よりも優先的に使用しうる可能性がある。

8.9 小児における QOL 値測定については、「8.2」「8.3」「8.8」の原則を現時点では適用し

ない。特に選好に基づく尺度を用いた場合の「8.3」についての対応は、個別の状況による。

8.9.1 小児における QOL 値測定については、使用した測定方法とその理由を十分に説明する。

9 データソース(費用を除く)

9.1 費用対効果におけるアウトカムや費用の推計を行う際には、想定する当該疾患の治療プロセスについてその根拠とともに示す。

9.1.1 上記の治療プロセスは、日本における標準的なあり方を反映していること。

9.2 費用対効果を推計するにあたって使用する有効性・安全性・QOL 値等のデータ(モデル分析を実施するにあたって使用する推移確率等のパラメータも含める)については原則として、研究の質やエビデンスレベルが高く、かつ日本における現実の臨床成績を反映しているものを優先的に使用する。

9.2.1 有効性・安全性・QOL 値等のデータ選定においては、国内外の臨床研究のSR に基づくことを推奨する。適切なものであれば公開されていない臨床研究や治験の結果等を含めてよい。

9.2.2 原則としてエビデンスレベルの高いデータの使用を優先すべきであるが、研究の質や分析における対象集団、結果の外的妥当性等を勘案して適切なものを使用することを推奨する。(例:RCT の結果が、実際の臨床成績と大きく乖離している可能性があるなど)

9.2.3 適切な場合は、既存の臨床研究やデータベース等を再解析した未公表の結果を活用してもよい。その場合は、患者背景や解析手法等について詳細を明らかにする。

9.3 国内外でデータに明確な異質性が存在する際には、国内データを優先して使用する。

9.4 評価対象技術と比較対照技術における値の差の有無が結果に大きな影響を与えるパラメータについて、統計学的に有意な差が認められていない場合は、両群で同じ値を用いることを原則とする。

9.4.1 上記について統計学的な検出力が不足している状況下で、両群で同じ値を用いない場合は、支持するその他のデータや理由、治療効果の大きさ(臨床的に意味のあるものか)などについてあわせて検討を行い、妥当性を説明すること。

9.5 医療機器等の評価において、科学的に信頼できる定量的なデータがある場合は、協議における両者の合意のもとで、上記のデータに基づく分析とは別に、いわゆる習熟効果(経験の蓄積による治療効果等の改善)や製品改良による効果を反映した分析を基本分析とあわせて提出してもよい。

10 費用の算出

10.1 「公的医療の立場」においては公的医療費のみを費用に含める。

10.2 各健康状態の費用は、評価対象技術によって直接影響を受ける関連医療費のみを含め、非関連医療費は含めないことを原則とする。

10.3 各健康状態の費用の推計においては、日本における平均的な使用量や標準的な診療過程等が反映されている必要がある。

10.4 各健康状態の費用の推計において、適切な場合には、「10.3」の観点から実臨床を反映した国内におけるレセプトのデータベースを用いることを推奨する。ただし、レセプト上で健康状態の定義が困難である、評価時点においてデータの十分な蓄積がないなど、推計の実施が困難な場合はその限りではない。

10.4.1 レセプトデータを用いて推計する場合、各健康状態の定義とその根拠を示さなければならない。

10.4.2 費用を推計するために用いた手法(外れ値処理や非関連医療費の除外などを含む)については、用いたものとその根拠を示さなければならない。

10.5 各健康状態の費用の推計において、レセプトデータベースを用いることが困難と判断される場合、あるいはより適切であると判断される場合は、標準的な診療プロセス等に基づき積み上げで算定してもよい。

10.5.1 積み上げで算定する場合は、「10.3」の観点から、その算定根拠について示すこと。関連する資源消費項目と消費量について、レセプト等のデータベースを用いて同定するなどしてもよい。

10.5.2 費用を積み上げで推計する場合は、医療資源消費量と単価を区分して集計、報告することを原則とする。

10.5.3 「10.5.2」の資源消費量の集計において、注射剤のバイアルについては、残量廃棄を原則とする。

10.6 評価対象技術や比較対照技術の費用のみでなく、有害事象や将来の関連する合併症等の費用も含めて推計する。

10.7 公的医療費については、保険者負担分のみならず公費や患者負担分も含めて費用として取り扱う(公的医療費の全額)。

10.7.1 「2.2.2」の原則に応じて、検診やワクチン等の公的医療費に準じる費用も費用として含める。

10.8 単価は可能な限り最新時点の診療報酬点数表や薬価基準等を使用する。特に評価対象技術あるいは比較対照技術については必ず最新時点の価格を用いなければならない。

10.8.1 既存の疾病費用分析やレセプトデータを用いた分析などの場合、単価は医療資源が消費された時点ではなく、分析実施時点にそろえたものを用いる。その

際に、診療報酬改定率を乗じる等により調整してもよい。

10.8.2 結果に与える影響が無視できる程度である場合には、分析実施時点に調整しないことも許容する。

10.9 比較対照技術に後発医薬品(バイオシミラーを含む)が存在する場合は、それらの価格を用いた分析もあわせて提出する。

10.10 評価対象技術あるいは比較対照技術の費用が包括支払いの対象となっている場合は、出来高で費用を算出する。

10.11 将来時点に発生する費用も、現時点における医療資源消費や単価に基づき推計したものをを用いる。

10.12 海外データを用いる際には、医療資源消費量について、国内外における医療技術の使用実態等の違いに配慮する必要がある。単価は国内のものを反映させなければならない。

11 公的介護費・生産性損失の取り扱い

11.1「公的医療・介護の立場」では、基本分析に加えて、公的介護費を含める追加的分析を実施することができる。なお、公的介護費は国内の知見に基づき推計されたものを用いる。

11.2 公的介護保険の費用を分析に含める場合は、要介護度・要支援度別に費用を集計することを推奨する。

11.3 要介護度ごとの公的介護保険の利用額は、対象疾患等における実際の資源消費量に基づくことが原則であるが、測定することが困難な場合は平均的な受給者 1 人当たり費用額等を用いてもよい。

11.4 生産性損失を含めた分析を行う場合、基本分析に加えて、追加的分析を実施することができる。ただし、生産性損失を含めることができるかどうかは、疾患の特性等による就業可能性を考慮しなければならない。また、生産性損失は国内の知見に基づき推計されたものを用いる。

11.5 生産性損失の減少は、

(A) 医療技術に直接起因するもの(治療にともなう入院期間の短縮等)

(B) アウトカムの改善(病態の改善や生存期間の延長等)を通じて間接的に生じるもの

に分けて考えることができる。

生産性損失を分析に含める場合には、原則として(A)のみを費用に含めることとする。

11.6 生産性損失は、人的資本法を用いて推計することを基本とする。これは、当該疾患に

罹患していなければ、本来得られたであろう賃金に基づき推計する方法である。

11.6.1 生産性損失を推計する際に単価として用いる賃金は、公平性等を考慮して、最新の「賃金構造基本統計調査」(賃金センサス)に基づき、全産業・全年齢・全性別の平均あるいは全産業・全性別の年齢階級別の平均を用いることとする。

11.6.2 生産性損失を推計するにあたっては、対象となる集団において就業状況を調査し、実際に仕事等に従事できなかった日数や時間を測定する。これに全産業・全年齢・全性別の平均賃金を乗じて生産性損失を推計することが原則である。

11.6.3 「11.6.2」の実施が困難な場合、対象集団において仕事等に従事できないと推計される日数(休日は除く)や時間に全産業・全年齢・全性別の平均賃金を乗じて生産性損失とする。ただし、この方法は 18 歳以上の就業率を 100%と仮定するものであり、生産性損失を過大推計する可能性のあることに留意が必要である。

11.7 家族等による看護や介護のために本人以外の生産性が失われることが明らかな場合は、本人の生産性損失と同じ条件・取り扱いのもとで費用として含めてもよい。

11.8 仕事等の減少とは無関係な時間費用等については含めないこととする。

12 割引

12.1 将来に発生する費用と効果は割引を行って、現在価値に換算しなければならない。

12.1.1 ただし、分析期間が 1 年以下、あるいは短期間でその影響が無視できる程度であるときは、割引を行わなくてもよい。

12.2 費用・効果ともに年率 2%で割引を行うこととする。

12.3 割引率は、感度分析の対象とし、費用・効果を同率で年率 0%から 4%の範囲で変化させる。

13 モデル分析

13.1 「7.」の原則に基づき、予後や将来費用を予測するために決定樹モデル、マルコフモデル等を用いたモデル分析を行ってもよい。

13.2 モデル分析を行う際には、そのモデルの妥当性について示さなければならない。例えば、

(A) 内的妥当性：なぜそのような構造のモデルを構築したのか、病態の自然経過を十分にとらえられているか、使用しているパラメータは適切なものか等

(B) 外的妥当性：既存の臨床データ等と比較して、モデルから得られた推計が適切なものであるか等

13.3 モデルを構築する際に使用した仮定については明確に記述する。

13.4 モデルを構築する際に使用したパラメータとそのデータソースについてはすべて提示する。

13.5 使用したモデルや計算過程については電子ファイルの形式で、第三者の専門家が理解でき、かつ原則としてすべての主要なパラメータ(推移確率、QOL 値、費用)を変更できる形で作成し、提出する。

13.5.1 費用については、その全体額のみでなく、それらの詳細(積み上げの場合は各資源消費量と単価など)を変更できるようにしておくことが望ましい。特に評価対象技術や比較対照技術の単価については、第三者が変更できるようにしなけれ

ばならない。

13.6 マルコフモデルを用いて解析する場合、1 サイクルあたりの期間が長く、結果に影響を与える場合は、半サイクル補正を実施する。

13.7 マイクロシミュレーションなど確率的に ICER の値が変動し、一意に定まらないモデルを用いる場合は、以下の点について留意する。

- (A) 結果の再現性が担保できるように乱数のシード値などを設定する。
- (B) 乱数のシード値によって分析結果が大きく変わらないことを説明する。意思決定に使用することが困難であるほどに、試行によって結果のばらつきが大きいモデルは、受け入れられないことがある。
- (C) 確率的な変動により ICER の値が変わっても、乱数の生成によってのみ生じる ICER の誤差については、極端な外れ値を用いるなど恣意的に結果が採用されていない限り、論点としない。
- (D) 結果のばらつきが大きい場合は、単回の試行結果ではなく、複数回の試行結果の平均値等を用いることを推奨する。
- (E) 通常の事務用機器において、1 回の試行に長時間を要するモデルは、レビュー等の実施が困難となるため使用しない。

14 不確実性の取り扱い

14.1 分析の設定等が一意に定まらず、それらの違いが結果に影響を与える可能性がある場合は、複数のシナリオに基づいた分析を行う。

14.2 分析期間が長期にわたり不確実性の大きい状況では、臨床研究のデータが存在する期間を分析期間とするなど、より短期の分析もあわせて行う。

14.3 「5.」において比較対照技術との比較試験が存在しない場合、特に単群試験の結果同士を比較した場合は、不確実性が大きいので十分に広い範囲での感度分析を実施する。

14.4 推定値のばらつきの大きなパラメータ、実際のデータではなく仮定に基づき設定したパラメータ、諸外国のデータで国内のデータと異質性を有する可能性があるパラメータ等については、感度分析の対象とする。

14.5 推定値のばらつきのみが問題となる場合(パラメータの不確実性を扱う場合)、感度分析で動かす幅としては、95%信頼区間などを参考に設定する。

14.6 結果に大きな影響を与えるにもかかわらず、不確実性の大きなパラメータについては、基本分析において用いた値の妥当性について十分な説明が必要である。また、その不確実性の結果として、意思決定へ与える影響について検討する。

14.6.1 不確実性の大きなパラメータが意思決定に与える影響について検討する上で、閾値分析が有用な可能性がある。

14.7 確率的感度分析(Probabilistic sensitivity analysis: PSA)もあわせて実施することが望ましい。その場合、使用したパラメータの分布についても明らかにするとともに、費用効果平面上の散布図と費用効果受容曲線(Cost-effectiveness acceptability curve: CEAC)を提示する。

用語集

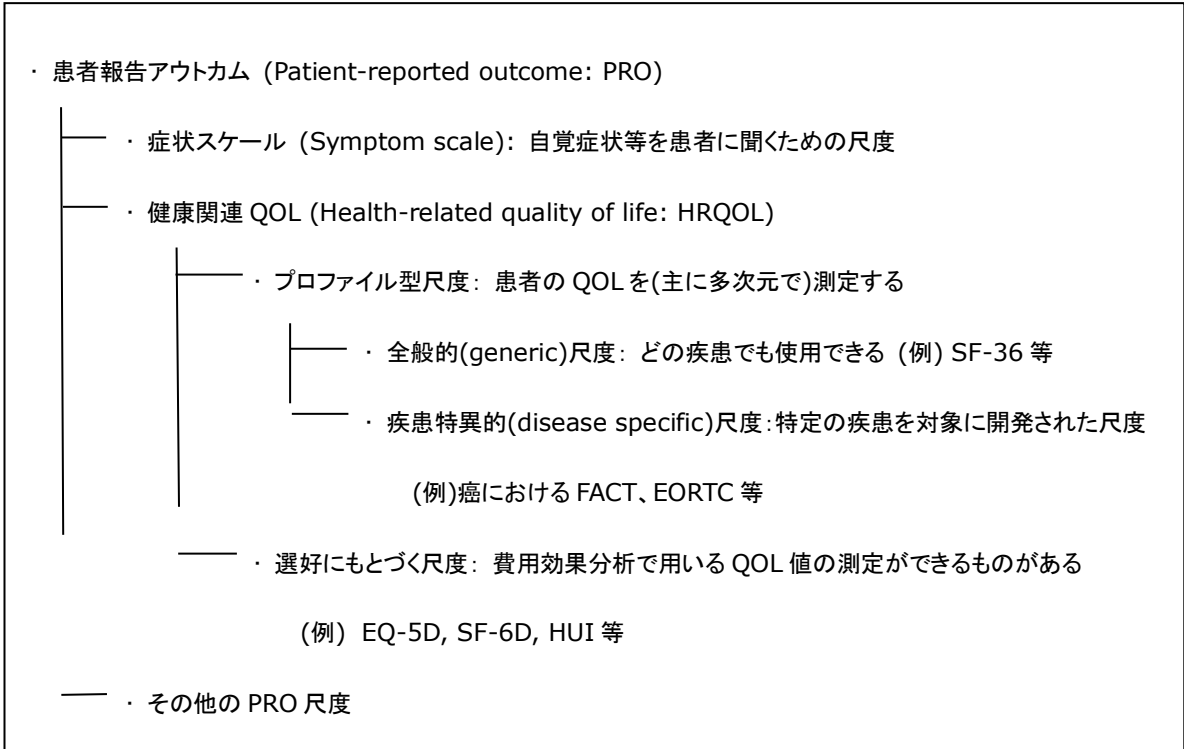
・ QOL 値

人々の健康状態(から得られる価値)を 0(死亡)から 1(完全な健康状態)に基準化して、一次元で数値化したもの。ただし、「死ぬより悪い」健康状態として負のスコアも取りうる。

QOL 値の測定方法には大別して、仮想的な(あるいは本人の)健康状態に対して、その状態の QOL 値を一般の人々を対象に質問する「直接法」(基準的賭け(SG)法、時間得失(TTO)法など)と、QOL 質問票により得られた回答からスコアリングアルゴリズムを用いて QOL 値を算出する「間接法」が存在する。なお、TTO と SG で測定された値には、系統的な差が存在する可能性があるため留意が必要である。

すべての患者報告アウトカム(Patient-reported outcome: PRO)や QOL 尺度での測定値から費用効果分析で使用される QOL 値が算出できるわけではないことに注意が必要である。費用効果分析で利用できるのは、下記のように QALY を算出するために開発された選好にもとづく (preference-based)尺度で測定したもののみである。

現在のところ、日本でスコアリングアルゴリズムが開発されている尺度としては EQ-5D(EuroQol 5 dimension)、HUI(Health Utilities Index)等がある。



・ エビデンスレベル

エビデンスレベルには様々な分類法が存在するが、Minds(Medical Information Network Distribution Service)では以下のように定めている。

I	システマティック・レビュー/RCT のメタアナリシス
II	1 つ以上の RCT による
III	非ランダム化比較試験による
IV a	分析疫学的研究 (コホート研究)
IV b	分析疫学的研究 (症例対照研究、横断研究)
V	記述研究 (症例報告やケース・シリーズ)
VI	患者データに基づかない、専門委員会や専門家個人の意見

ただし、RCTのような実験的研究が現実の臨床成績と乖離している可能性はしばしば指摘されている。エビデンスレベルの高いものの使用を原則としつつも、状況に応じた適切なデータを採用する必要がある。

・ 確率的感度分析

確率的感度分析は、モデルのパラメータに分布を当てはめること等により、増分費用や増分効果、ICERの分布を得るための手法である。確率的感度分析の結果は、費用効果平面上に散布図をプロットし、CEACとして $f(\gamma) = Pr(\gamma \cdot IE - IC > 0)$ を書くことが一般的である（IC: 増分費用、IE: 増分効果、 γ : 支払意思額）。

・ 間接比較

例えば臨床試験によって“A vs. B”と“A vs. C”の結果が得られているとき、これらの結果から直接比較のない“B vs. C”の結果を推測することを間接比較(indirect comparison)と呼ぶ。適切な比較対照を用いて分析しようにも直接の比較試験がない場合、間接比較を適用することができる場合がある。

間接比較が成り立つためには“A vs. B”の結果が“A vs. C”の集団にも適応できること、逆に“A vs. C”の結果が“A vs. B”の集団にも適応できることが条件となる。このことを同質性(similarity)の仮定と呼ぶ。間接比較を行うにあたっては、このような仮定に関する検討や、適切な統計手法(例えば、単純な(naïve)間接比較ではなく調整された(adjusted)間接比較)を使用することが必要である。また、ネットワークメタアナリシス(あるいは多群間治療比較(multiple treatment comparison: MTC)とも呼ばれる)のようなより高度な手法を用いた分析についても検討しうる。

・ 感度分析

不確実性が存在する場合等に、パラメータの値を変化させることにより、結果への影響を見ることを感度分析という。1つのパラメータを変化させる一次元感度分析、2つのパラメータを同時に動かす二次元感度分析、複数のパラメータの不確実性を同時に取り扱うPSA(確率的感度分析の項を参照)などがある。また、閾値を超過するパラメータの値を算出することを閾値分析(threshold analysis)と呼ぶ。

・ システマティックレビュー

システマティックレビュー(SR)とは特定の課題について文献等を網羅的に探索し、可能な限りバイアスなくその結果や評価報告する手法である。Mindsによれば『実際の作業面から定義すると、SRとは「クリニカルクエスチョンに対して、研究を網羅的に調査し、研究デザインごとに同質の研究をまとめ、バイアスを評価しながら分析・統合を行うこと」とされる。

しばしばSRとメタアナリシスが混同して使用されることもあるが、SRにおいて得られた結果は必ずしも統計的に統合する必要はなく、このようなものを「定性的システマティックレビュー」と呼ぶこともある。結果を統合することが適切な場合は、SRの結果に基づき、メタアナリシスを実施することとしている。

なお、SR(メタアナリシス)の報告様式としては、国際的にPRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)声明が標準的に使用されている。

・ 質調整生存年

質調整生存年(QALY)は、生存年にQOL値を乗じることにより得られる。QOL値が1は完全な健康を、0は死亡を表す。QOL値0.6の健康状態で2年間生存した場合、生存年は2年だが、 $0.6 \times 2 = 1.2$ QALY(完全に健康な状態で1.2年生存したのと同じ価値)と計算される。時間とともにQOL値が変化する場合、図のようにQOL値の経時変化をあらわす曲線下面積が獲得できるQALYとなる。

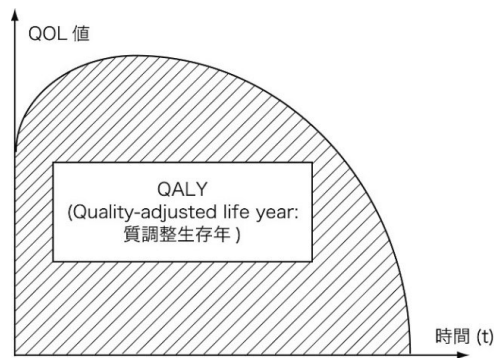


図 QALY の概念図

・ 人的資本法

生産性損失を、本来得られたであろう賃金に基づき推計する方法を人的資本法(human capital method)と呼ぶ。しかし、完全雇用が実現されていない状況下では、その人が働けなくてもかわりの誰かが働くはずであり、長期的には必ずしも生産性が失われるわけではない。そのため、社会における生産性がもとの水準に回復するまでにかかる摩擦費用(friction cost)のみを含めるべきという意見もある。賃金は、疾病により実際に労働できなかった期間を調査することにより推計すべきであるが、そのような作業が困難な場合は、家事労働を含めたデータが存在しないため就業率を 100%として取り扱ってもよい。また、賃金の単価については、公平性の観点から、実際の単価にかかわらず全産業・全年齢・全性別の平均賃金を用いる。

・ 生産性損失

病気が原因で仕事や家事ができなくなることによる社会的な損失(あるいは早期に回復できることによる社会的な便益)は生産性損失(productivity loss)として、分析の立場によっては費用に含めることができるが、基本分析では含めないこととしている。また、生産性損失の範囲としては、本人のみならず家族等による看護や介護(インフォーマルケア)について検討してもよい。ただし、本ガイドラインでは、効果との二重計上になる恐れがあること等から病態の改善や生存期間の延長などアウトカムの改善を通じた間接的な生産性損失は含めないこととし、入院期間の短縮

など医療技術に直接起因するもののみを含めてもよいこととしている。

・ 増分費用効果比

増分費用を増分効果で割ったものを増分費用効果比(ICER)という。以下の式により、治療 B と比較した場合の治療 A の ICER が算出される。

$$ICER = \frac{IC}{IE} = \frac{C_A - C_B}{E_A - E_B}$$

(IC: 増分費用、IE: 増分効果、C_A: 治療 A の期待費用、C_B: 治療 B の期待費用、E_A: 治療 A の期待効果、E_B: 治療 B の期待効果)

効果指標 1 単位獲得するあたり、いくらかかるかを表す指標であり、値が小さいほど費用対効果はよい。

・ 追加的有用性

費用効果分析を実施するにあたっては、増分費用効果比を算出する前に、比較対照技術と比して追加的有用性等を有することを示す必要がある。追加的有用性を示す際の効果指標については、必ずしも費用効果分析における効果指標と同一である必要はないが、臨床的に意味のあるものを選択する必要がある。追加的有用性を有すると判断される場合は、費用効果分析により ICER を算出する。一方で、追加的有用性を有すると判断できない場合は、いわゆる CMA により費用の比較を実施する。

・ 非関連医療費

医療費は、評価対象技術によって直接影響を受ける関連医療費(related medical cost)と生命予後の延長等により間接的に影響されるもの、あるいは当該疾患と関連しないもの等の非関連医療費(unrelated medical cost)とに分類できる。例えば、高血圧治療によって心血管疾患や脳卒中が減少すると、期待余命が延長して、非関連医療費(例えば認知症や糖尿病、腎透析

など)が増大する可能性がある。このような非関連医療費は原則として費用に含めないこととして
いる。

・ ビニエツト法

仮想的な健康状態の記述(ビニエツト, vignette)を回答者に読ませて、その健康状態を想起させることにより、QOL 値の評価を行う方法。評価にはいわゆる直説法とも呼ばれる SG 法、TTO 法、DCE 法などが用いられる。記述された健康状態は患者から直接得られたものではない一方で、得られる QOL 値に大きな影響を与えることから、恣意的なシナリオとならないよう慎重な配慮が必要である。

・ 費用効果分析

医療技術の経済評価は下記の 4 パターンに分類されることが多い。(a) アウトカムを同等と
いて費用のみを検討する「費用最小化分析 (Cost-minimization analysis: CMA)」、(b)
QALY 以外の種々のアウトカム指標(生存年、イベント回避など)を用いる「費用効果分析
(Cost-effectiveness analysis: CEA)」、(c) QALY を用いる「費用効用分析 (Cost-utility
analysis: CUA)」、(d) アウトカムを金銭化して評価する「費用便益分析 (Cost-benefit
analysis: CBA)」。

しかし、CMA、CEA、CUA は費用とアウトカムを別々に推計するという点では、同種の分析で
あるとも考えられるので、本ガイドラインではこれらの手法をまとめて費用効果分析と呼んでいる。

・ 不確実性

費用効果分析を行う上では、様々な分析の不確実性(uncertainty)がともなう。

異質性(heterogeneity)は、広義の不確実性の一様であり、比較対照技術や診療パターン、
対象患者等が一意に定まらない状況を指す。これは、次に説明する狭義の不確実性とは異なり、

統計学や医療経済学上の技術的な問題ではなく、現実が多様であることに起因する。このような異質性が存在する場合は、複数のシナリオ設定に基づいた感度分析を行うことを推奨している。

狭義の不確実性は、大きく(a)モデルの不確実性と(b)パラメータの不確実性に分けることができる。前者のモデルの不確実性は、さらに(a)-1 方法論上の不確実性や(a)-2 モデルの構造・仮定等に起因するものがある。

(a)-1 方法論上の不確実性は、割引率や生産性損失の推計方法、QOL 値の測定方法等が理論的には一意に定められないために生じる。これら为了避免するためには、標準的な共通の手法に従って分析を行うことが重要であるが、割引率など結果に大きな影響を与える場合には、一次元感度分析によってその不確実性の大きさを評価する。

(a)-2 モデルの構造・仮定に起因する不確実性は、健康状態や治療プロセスのモデル化法、モデルに組み込むパラメータの選択、観察期間を超えて長期的な予後を予測するための仮定等によって生じる。感度分析等によって評価する。

(b)パラメータの不確実性は、パラメータの推定値が持つ不確実性によって生じる。例えば、ある臨床試験の中で 100 人中 10 人にイベントが起こったとしても、真のイベント発生率(母イベント発生率)は $10/100=0.1$ ではないかもしれない。このような統計的推測に起因する不確実性に対処するには、通常感度分析に加えて PSA を行うことも有用である。

・ マッピング

選好にもとづく尺度による測定値が存在しない場合、患者報告アウトカム(PRO)での測定結果から、費用効果分析で使用する QOL 値を算出することが有用な場面もある。このような尺度間のスコア変換をマッピング(mapping)と呼ぶ。他のデータが存在しないときなどにマッピングは改善の手法として許容されうるものの、統計学的な妥当性などを十分に検討した上で実施すべきである。

・ メタアナリシス

システマティックレビューで得られた結果を統計的手法によって統合し、統合した値やその信頼区間を求める手法である。異質性が小さい場合は、固定効果モデル、異質性が大きい場合は変量効果モデルやベイズモデルを用いるのが一般的である。結果はフォレストプロットを用いて表記する場合が多い。治療間の対比較ではなく、複数の治療の比較を同時に行う場合は、ネットワークメタアナリシスと呼ばれる手法が用いられる(→間接比較)。

・ 優位・劣位

評価対象技術が比較対照技術と比して費用が安く効果も同等以上である場合、その医療技術は「優位(dominant)」であるという。一方、評価技術が対照技術と比較して費用が高いが効果は同等以下である場合、その医療技術は「劣位(dominated)」と呼ばれる。

・ 割引

費用効果分析においては、将来に発生する(あるいは得られる)費用とアウトカムを現在価値に換算するため、一定の率で割引くことが一般的である。年単位で割引を行ったあとの現在価値に換算された費用 C_p は、 i 年後の費用 C_i と割引率 d を用いて

$$C_p = \frac{C_i}{(1+d)^{i-1}}$$

によって計算することができる。効果についても同様である。

略語一覧

- ・ CBA: Cost-benefit analysis, 費用便益分析
- ・ CEA: Cost-effectiveness analysis, 費用効果分析
- ・ CEAC: Cost-effectiveness acceptability curve, 費用効果受容曲線
- ・ CMA: Cost-minimization analysis, 費用最小化分析
- ・ CSPBM: Condition specific preference-based measure, 疾患・病態特異的尺度
- ・ CUA: Cost-utility analysis, 費用効用分析
- ・ DCE Discrete choice experiment, 離散選択実験
- ・ EQ-5D: EuroQol 5 dimension
- ・ HRQOL: Health-related quality of life, 健康関連 QOL
- ・ ICER: Incremental cost-effectiveness ratio, 増分費用効果比
- ・ MAIC: Matching-adjusted indirect comparison, マッチング調整された間接比較
- ・ MTC: Multiple treatment comparison, 多群間治療比較
- ・ PBM: Preference-based measure, 選好に基づく尺度
- ・ PRO: Patient-reported outcome, 患者報告アウトカム
- ・ PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- ・ PSA: Probabilistic sensitivity analysis, 確率的感度分析
- ・ QALY: Quality-adjusted life year, 質調整生存年
- ・ RCT: Randomized controlled trial, ランダム化比較試験
- ・ RQ: Research question, リサーチクエスション
- ・ SG: Standard gamble, 基準的賭け法
- ・ SR: Systematic review, システマティックレビュー
- ・ TTO: Time trade-off, 時間得失法

Guideline for Preparing Cost-Effectiveness Evaluation to the Central Social Insurance Medical Council

The English version is a translation of the original version in Japanese. The Japanese version is preferentially applied in cases of discrepancy between the two versions.

1 Objectives	47
2 Perspective	48
3 Target population	50
4 Comparator	52
5 Additional benefits	54
6 Methods of analysis	60
7 Time horizon	62
8 Choice of outcome measure	63
9 Sources of clinical data (except costs)	68
10 Calculation of healthcare costs	70
11 Public long-term care costs and productivity loss	74
12 Discounting	77
13 Modeling	78
14 Uncertainty	80

1 Objectives

1.1 This guideline presents the standard methods to perform cost-effectiveness evaluations of medicines, medical devices, and regenerative medicine products selected by the Central Social Insurance Medical Council (selected product).

1.2 This guideline is applied to manufacturers' submissions and Academic Technology Appraisal Groups' analysis (review of submission and academic analysis).

2 Perspective

2.1 The perspective of the analysis should be specified. In particular, the perspective determines the costs to be included.

2.2 “Public healthcare payer’s perspective” is the standard setting. It uses costs, comparators, and target populations to reflect the public healthcare insurance situation in Japan.

2.2.1 Even when an analysis is conducted from a perspective other than the “public healthcare payer’s perspective,” an analysis from the “public healthcare payer’s perspective” should also be submitted.

2.2.2 There are products that are not covered by public healthcare insurance but are publicly funded, such as prophylactic procedures (e.g., health checkups and vaccinations). Analyses that consider these products can be submitted from the “public healthcare payer’s perspective.”

2.3 If consideration of public long-term care costs is important for the selected product, it is acceptable to submit an analysis from the “public healthcare and long-term care payer’s perspectives.”

2.4 If the introduction of a selected product influences productivity, then it is

acceptable to submit an analysis that includes productivity losses from a broader perspective.

3 Target population

3.1 Patient groups for whom the selected product had indications at the time designated as an item for cost-effectiveness evaluation should be considered as the “target population.”

3.1.1 In the case that a new indication is approved between the designation of the product and determining the framework of analysis, these patients are also included in the target population for cost-effectiveness evaluation.

3.1.2 New indication is added after the time when “3.1.1” is defined, and further evaluation is performed after the first evaluation is completed if it may influence the initial results.

3.2 If the product has multiple indications or subpopulations that differ in outcomes, doses, administration methods, or comparators, an analysis should be conducted for each population.

3.2.1 However, if “3.2” is difficult to achieve and the impact is limited, a part of the population can be omitted from the evaluation, considering the number of patients or disease characteristics. The exemption is determined based on an agreement between the manufacturer and the National Institute of Public Health/Academic

Technology Appraisal Group in consultation.

3.3 The proportion of patients in each population should be estimated from a long-term perspective (based on the cumulative number of patients during the patent period). The current clinical status, utilization patterns, and other empirical data on the selected products should be considered.

3.3.1 If it is difficult to estimate, the proportion of patients may be calculated at a certain cross-sectional time point where the utilization pattern becomes stable.

3.3.2 Short-term negligible effects should be excluded (e.g., the impact of elective patients just after the selected product is introduced to the market).

4 Comparator

4.1 The comparator should be selected from candidates that are widely used in clinical practice and expected to be replaced by the selected product when the product is newly introduced to treat the target population. Among them, the product that results in the best outcome should be selected.

4.1.1 Products that are “widely used in clinical practice” are not decided only by the amount of actual consumption; rather, the products are used as a clinical standard, as specified in the clinical guidelines.

4.1.2 When deciding whether a product is “the best outcome,” refer to the report on additional benefit published in the cost-effectiveness evaluation process.

4.2 If a single comparator cannot be determined based on “4.1,” the selection of a comparator should be considered following the comparators in randomized controlled trials (RCTs), similar products for the official pricing and cost-effectiveness of candidate products, based on agreement in consultation.

4.3 Non-treatment or watchful waiting can also be used as a comparator.

4.4 Except for the cases described in “4.3,” a comparator should be selected

from products that are reimbursed by public healthcare insurance.

4.5 The reasons for the selection of a comparator should be sufficiently explained.

5 Additional benefits

5.1 Whether the selected product has additional benefits to the comparator should first be evaluated using empirical data.

5.2 The additional benefits to the comparator, as defined in Section 4, should be evaluated through a systematic review (SR) of RCTs. Results of unpublished clinical studies and trials may also be included in the SR.

5.2.1 When an SR is conducted, research questions (RQs) must be clearly presented. For example, RQs structured according to PICO (P, patient; I, intervention; C, comparator; and O, outcome) should be provided.

5.2.2 There may be products with action mechanisms (drugs) or function categories (medical devices) similar to the selected technologies or the comparator, which are expected to show equivalent outcomes. These products can also be included as interventions (I) or comparators (C) in the SR.

5.2.3 As outcome (O) in "5.2.1," the most appropriate clinical outcomes (e.g., a "true outcome") should be selected from clinical effectiveness, safety, and health-related quality of life (HRQOL).

5.2.4 A description of the inclusion/exclusion criteria, databases, search strategy formula, and research selection process (inclusion flow diagram) is required, in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

5.2.5 It is acceptable to use existing reliable SR. In such cases, an existing review is used either directly or in combination with a new literature search. The existing literature is consistent with the RQs and covers the most recent literature.

5.2.6 If appropriate, the pooled results from pairwise meta-analyses should be presented. In such cases, the statistical method, heterogeneity, forest plots, pooled results, and confidence intervals should be reported.

5.2.7 A time point between determining the framework of analysis and the manufacturer's submission should be used as the cutoff date for the literature search in the SR.

5.2.8 There may be cases in which the results of new clinical trials are published after the cutoff date defined in "5.2.7." If evidence is important for cost-effectiveness evaluation (e.g., clinical trials with large sample sizes or reliable results different from current studies), the inclusion of these trials in the SR should be considered through academic technology appraisal group analysis.

5.3 When no studies are available based on the result of the SR described in “5.2,” additional benefits are evaluated through an SR of comparative non-RCT (e.g., observational) studies based on “5.2.”

5.3.1 More various biases tend to occur in the non-RCTs than in the RCTs. Sufficient explanation of research quality is required (e.g., study design, differences in background factors between groups, statistical analysis, sample sizes, and number of institutions).

5.3.2 Research quality is not necessarily high in studies that retrospectively use large-scale databases (real-world data), such as health insurance claims and registries. Therefore, adequate explanation should be given regarding the characteristics of the database, differences in the healthcare environment between Japan and other countries, definitions of variables and events, validity of the definition, validation survey on the definition, analytical methods, and extrapolation of the results to Japan.

5.4 As a general rule, the results of non-RCTs should be used to complement the results of an RCT to evaluate the benefits of the product.

5.5 When there are no RCT studies using the same comparator selected in Section 4 but there are RCT studies that compare the selected product to others, additional benefits can be evaluated through indirect comparison

using SR results.

5.6 When there are only single-arm clinical studies on selected technologies, an indirect comparison should be performed based on the SR results of the selected technologies and the comparator.

5.7 When indirect comparison is performed, the following items should be considered:

5.7.1 When individual patient data are available, the difference of background factors should be adjusted using appropriate method, such as matching adjusted indirect comparison (MAIC).

5.7.2 A network meta-analysis (NMA) can be conducted when patient-level data cannot be used or when an NMA is a preferable method.

5.7.3 For analyses in "5.7.1" and "5.7.2," an anchored method is desirable.

5.7.4 Sufficient explanation should be given on the assumptions for indirect comparisons (e.g., heterogeneity of disease, severity and demographics, and the similarity of studies).

5.7.5 A naïve indirect comparison may be used when comparative

study results are not available and when there is no other choice; in this case, intergroup comparability should be further explained in detail.

5.7.6 If more than one analytical method can be used, the reasons for selecting the methods (e.g., suitability of the assumption) should be explained; when necessary, sensitivity analyses with different methods should be performed.

5.7.7 If an NMA is performed, the following points should also be considered and explained. It is recommended to perform a sensitivity analysis as necessary.

- (a) Network size (number of treatment groups)
- (b) Treatment groups (drug classes or individual products).
- (c) Handling different doses and modes of administration
- (d) Handling drugs not approved in Japan

5.7.8 When a network meta-analysis is performed, data and programs reproducible by third parties should be submitted if possible.

5.8 Literature information should be presented and used to evaluate additional benefits (e.g., those identified in the SR or incorporated into an NMA).

5.9 There may be cases in which the detected studies based on "5.3" to "5.7"

have serious problems with insufficient quality. However, the selected product is not expected to be inferior to the comparator. In such cases, the analysis described in Section 6 can be performed, assuming that the outcome of the selected product is equivalent to that of the comparator.

5.10 When there are no available clinical data on the selected product in humans, the analysis described in Section 6 can be performed, assuming that the outcome of the selected product is equivalent to that of the comparator, if appropriate. It is assumed that products approved by the Pharmaceuticals and Medical Devices Agency (PMDA) have such an outcome.

5.11 When SR results obtained using the methods in "5.2" to "5.7" show that outcomes of the selected product are inferior to that of the comparator, no cost-effectiveness analysis is performed.

6 Methods of analysis

6.1 A cost-effectiveness analysis should be conducted. In this guideline, cost-effectiveness analysis is defined as an analysis that calculates cost and effectiveness separately, without converting effectiveness into monetary units.

6.2 If the evaluation described in Section 5 reveals additional benefits, the incremental cost-effectiveness ratio (ICER) should be calculated from the expected cost and effectiveness in each treatment group.

6.3 In the following cases, only the expected cost and effectiveness in each group need to be presented. The ICER should not be calculated.

6.3.1 In cases where the product has better effectiveness and lower cost, the product is considered “dominant.”

6.3.2 A cost comparison (the so-called cost minimization analysis or CMA) should be performed if the evaluation described in Section 5 does not demonstrate additional benefits. In such case, the results are “cost saving” (or equivalent) or “cost increase.”

6.3.3 Even if an additional benefit is demonstrated by the evaluation in Section 5, the ICER may have large uncertainty. Minor changes in the parameters can manifest as significant changes in the ICER when

the incremental effectiveness is only slightly positive and the incremental cost is almost zero. In these cases, the results are “equivalent in terms of cost and effectiveness”.

6.4 If the selected product has multiple indications or subpopulations defined in “3.2” and/or “3.2.1,” the ICER should be calculated for each indication or subpopulation.

6.5 If a cost-effectiveness analysis of a selected product published in an academic journal or reports by a health product assessment (HTA) agency are available, the results should also be presented.

7 Time horizon

7.1 The time horizon should be sufficiently long to evaluate the influence of the product on cost and effectiveness.

7.2 The same time horizon should be applied for cost and effectiveness.

7.3 The reason for setting this time horizon should be specified.

8 Choice of outcome measure

8.1 Quality-adjusted life years (QALY) should be used in principle.

8.1.1 When it is difficult to calculate the QALY and CMA is applied, other outcome measures can be used, if appropriate.

8.2 When the QALY is calculated, the QOL scores (utilities) should be measured using a preference-based measure (PBM).

8.2.1 The QOL scores should reflect the preference of the general population in Japan.

8.2.2 The Japanese version of the EQ-5D-5L is recommended as the initial choice for the PBM.

8.3 When the QOL scores are measured by the PBM, responses of the patients should be used.

8.3.1 Proxy responses (e.g., a family member or caregiver) may be used when the patients cannot respond.

8.3.2 Proxy responses from a healthcare professional should be avoided because of possible discrepancies from patients' responses.

8.4 If the QOL scores that satisfy “8.2” is available, the use of Japanese scores are preferentially recommended.

8.4.1 If Japanese research is insufficient but high-quality research is performed in other countries, the use of data collected in other countries is acceptable.

8.4.2 PBM data collected in other countries should be handled in a manner consistent with “8.2.1”; for instance, the Japanese value set is applied to these responses.

8.4.3 If patient-level data are not available, “8.4.2” is difficult to perform. If the results are significantly affected, explain whether the QOL scores are consistent with the Japanese scores. When necessary, consider adjusting the scores to the Japanese scale.

8.5 If data corresponding to “8.2” are unavailable, mapping of other appropriate HRQOL data is acceptable.

8.5.1 If mapping is implemented, the QOL scores obtained from the mapping function should be consistent with “8.2.1.”

8.5.2 If there is insufficient conceptual overlap between the QOL instruments, consider whether the use of mapping is appropriate.

8.5.3 The development process and characteristics of the mapping formula should be reported in detail using, for instance, the Mapping onto Preference-based measures reporting Standards (MAPS) checklist.

8.5.4 It is not recommended to use a mapping function on which details cannot be reported.

8.6 If it is difficult to collect QOL scores directly from patients, it is acceptable for the general population to evaluate the assumed health scenario using a standard gamble (SG), time trade-off (TTO), and discrete choice experiment (DCE) (the vignette method).

8.6.1 It should be noted that the QOL scores measured using the vignette method are significantly affected by the assumed health scenario provided to respondents.

8.6.2 Since the assumed health scenario is not obtained directly from the patients, its validity should be sufficiently explained. It is also desirable to seek input from patients and/or clinical experts.

8.6.3 If the results from the vignette method are used, the actual scenario shown to respondents should be presented.

8.6.4 Considering “8.6.1,” scores obtained from the vignette method may lead to overestimation of cost-effectiveness, the limitations of these methods must be considered. For example, the same QOL score should be used for the same health state.

8.6.5 Web-based SG and TTO surveys are known for systematic biases in results; therefore, face-to-face surveys are recommended.

8.6.6 Scores using the vignette method present difficulties in investigating the differences between Japan and other countries. It is relatively easy to collect data; therefore, the use of Japanese scores is recommended.

8.7 In the case of analysis from the public healthcare and long-term care payer’s perspectives, the QOL scores influenced by informal caregivers may be considered if actual data exist.

8.8 It is recommended to consider the QOL data with similar quality in the following order based on the principle in “8.2.” When this order is not applicable, the reasons for this should be explained.

- (a) Data collected using the Japanese EQ-5D-5L (“8.2”).
- (b) Data collected using a generic Japanese PBM with a Japanese value set other than the EQ-5D-5L.
- (c) Data collected using a condition-specific Japanese PBM with a Japanese value set.

(d) Data collected in other countries using a PBM with patient-level data available; the priority of instrument is based on the concept from (a) to (c) ("8.4").

(e) Data collected in other countries using a PBM with patient-level data unavailable ("8.4").

(f) Data converted to QOL scores using a mapping function; for which, follow (a) to (e) in terms of the mapped instruments and mapping formula ("8.5").

(g) Data collected using the PBM with no Japanese value set.

(h) Data collected using the vignette method ("8.6").

(i) Other methods are not recommended unless there is an unavoidable reason and/or scientific justification.

The high-quality mapping formula developed in Japan may be preferable to (f).

8.9 The principles in "8.2," "8.3," and "8.8" shall not be applied to the QOL in children and adolescents. Particularly regarding the issue in "8.3," it will depend on the situation as to who should respond to the PBM.

8.9.1 The method and reason for the selection of a QOL instrument in children and adolescents should be explained.

9 Sources of clinical data (except costs)

9.1 When estimating outcomes and costs for the cost-effectiveness analysis, the treatment process used in the analysis should be presented together with its rationale.

9.1.1 The above treatment process should reflect the standard practice in Japan.

9.2 Calculations of the ICER should preferentially use effectiveness, safety, and QOL data (including parameters such as transition probability for model analysis) derived from high-quality research, with a high level of evidence reflective of clinical practice in Japan.

9.2.1 The selection of effectiveness, safety, and QOL data based on the SR is recommended. This review will also include unpublished clinical study/trial data, if appropriate.

9.2.2 Data with a high level of evidence should be preferentially used. Consideration of research quality, target population, and external validity is also recommended when applying data (e.g., it is possible that the results of an RCT may differ markedly from practical clinical results).

9.2.3 Data using additional analysis of existing studies and/or registry

data can be used for this evaluation. In such cases, detailed information regarding the patient's background and statistical methods should be provided.

9.3 Japanese data should be used preferentially if there is evident heterogeneity between Japanese data and those of other countries.

9.4 Pooled data of both groups should be applied, if there is no statistically significant difference between groups.

9.4.1 If the same parameter values are not used in both groups owing to a lack of statistical power, show other supportive data and reasons as well as the size of the treatment effect (whether clinically important) and explain that the difference is interpretable.

9.5 If there are reliable and quantitative data on medical devices, analysis reflecting "learning effect" (i.e., improvement of treatment effect through the accumulation of clinicians' experience) or "product improvement effect" can be submitted in addition to analysis not considering the effects, upon agreement in consultation.

10 Calculation of healthcare costs

10.1 Only public healthcare costs should be included in the analysis from the public healthcare payers' perspective.

10.2 Healthcare costs of each health state include only related costs directly affected by the selected product. Do not include unrelated costs.

10.3 Healthcare costs of each health state should reflect the average resource consumption and standard clinical practices in Japan.

10.4 Claims databases established in Japan, which reflect actual clinical practice based on "10.3", should be used to estimate the costs of each health state. Cases in which it is difficult to define health states using only information from claims data, or insufficient data have been accumulated in the database, were excluded.

10.4.1 Definition of each health state and its rationale is required when claims data are used for cost estimation.

10.4.2 The methods and rationale for the method to estimate costs (including handling outliers and unrelated costs) should be presented.

10.5 Micro-costing (e.g., by medical fee schedule) based on the definitions

of the standard clinical process can be used if it is difficult to estimate the costs of each health state using a claims database, or if micro-costing is more appropriate.

10.5.1 In the application of micro-costing, the rationale for costing should be shown based on "10.2." In this case, the claims database can be used to identify relevant items and/or estimate the amount of resource consumption.

10.5.2 When micro-costing is used, the amount of resource consumption and unit costs should be reported separately.

10.5.3 For the estimation of resource consumption in "10.5.2," the injection products should be calculated by the number of vials rather than by patient dosages.

10.6 The estimation should include not only the costs of the selected product and comparator but also the costs of adverse events and related future events.

10.7 Public healthcare costs should include not only the portion of costs paid by the insurer but also those paid by the government and patients (i.e., total public healthcare expenses).

10.7.1 Based on the principle in "2.2.2," the analysis should include the costs of health checkups, vaccinations, or similar procedures that

are funded publicly but not reimbursed by public healthcare insurance in Japan.

10.8 Unit costs should be derived from the latest medical fee schedule, drug price lists, or similar resources. It is particularly essential to use the latest unit costs for a selected product or comparator.

10.8.1 Even if existing cost-of-illness studies or analyses of claims data are used, the unit costs at the time of this evaluation and not at the time of the study should be applied. It is acceptable to make adjustments, such as multiplication by the revision rate of the medical payment system.

10.8.2 Such adjustments may be omitted if the influence on results is minimal.

10.9 If generics of a comparator are already on the market, analysis using these costs should be also submitted.

10.10 Even if the costs of a selected product and/or comparator are included in bundled payment, the estimation should be based on a fee-for-service payment.

10.11 Future costs should also be estimated on the basis of current resource consumption and unit costs.

10.12 Calculations of resource consumption based on data from other countries will require attention regarding the possible differences in healthcare technology use between Japan and other countries. The unit costs in Japan should be considered in the analysis.

11 Public long-term care costs and productivity loss

11.1 Analysis from the public healthcare and long-term care payer's perspectives can include public long-term care costs. It is acceptable to include public long-term care costs only if they can be estimated using Japanese data.

11.2 When public long-term insurance care costs are included in the analysis, it is recommended that they be calculated based on the care level.

11.3 The amount used under public long-term care insurance should be based on the actual resource consumption. If consumption is difficult to determine, it may be acceptable to use the average costs per beneficiary.

11.4 Analysis including productivity loss can be performed in addition to the base-case analysis. However, regarding the appropriateness of including productivity losses, the disease characteristics that lead to the possibility of working are considered. Including productivity losses in additional analyses is acceptable only if they can be estimated using Japanese data.

11.5 Decreases in productivity losses may be classified as follows:

- (a) Decreases arising directly from selected products (e.g., treatment-related shortening of hospital stay).
- (b) Decreases arising indirectly from outcome improvements (e.g., improvement in illness and survival extension).

Only (a) is included when productivity loss is included in the analysis.

11.6 Productivity losses should be estimated using the human capital method. This method estimates the loss using the expected earned wage in the absence of illness.

11.6.1 The unit wage used for estimations of productivity loss should be the average wage across all industries, ages, and both genders or the average wage for each age group in all industries and both genders derived from the latest “Basic Survey on Wage Structure” (Wage Census).

11.6.2 Estimations of productivity loss require an actual investigation of the employment status of the target population (i.e., a measure of days or hours of work missed). The actual measured days or hours should then be multiplied by the average wage across all industries, ages, and both genders to estimate the productivity loss.

11.6.3 If the method described in “11.6.2” is difficult to apply, productivity loss should be calculated by multiplying the expected number of days (excluding holidays) or hours of work missed by the average wage across all industries, ages, and both genders. A 100% employment rate should be assumed for those aged 18 years and older. However, this method may overestimate productivity losses.

11.7 If other individuals (e.g., family members) experience productivity losses because of informal care, it is acceptable to count these productivity losses under the same conditions and methods as those used to calculate the patient's productivity loss.

11.8 Time costs that are unrelated to a decrease in work should not be included.

12 Discounting

12.1 Future costs and effectiveness must be discounted and converted into present values.

12.1.1 Discounting is not required if the time horizon is one year or less or is otherwise sufficiently short to ignore the influence of discounting.

12.2 Both cost and effectiveness should be discounted at a rate of 2% per year.

12.3 The discount rate should be subjected to sensitivity analysis and should be changed at the same rate of 0%–4% per year for both costs and effectiveness.

13 Modeling

13.1 To predict the prognosis and future expenses, it is acceptable to conduct a model analysis using a decision analytic model, Markov model, and/or other models in accordance with the principles described in Section 7.

13.2 Validity of the model should be presented. For example:

(a) Internal validity: Why has a model with a given structure been created, whether the natural history of illness has been sufficiently evaluated, and whether the parameters are appropriate?

(b) External validity: Whether the estimation yielded by the model is appropriate in comparison with other existing clinical data.

13.3 The assumption used for the model should be clearly described.

13.4 All parameters and data sources used for model should be presented.

13.5 The model should be submitted using electronic files. The model must be easily understood by third-party experts, and all main parameters (transition probability, QOL score, and healthcare costs) must be modifiable.

13.5.1 It is better that not only the total costs but also the breakdown (in the case of microcosting, medical resource consumption, and unit costs) can be changed. The Academic Technology Appraisal Group

must be able to change the unit costs of the selected product and the comparator.

13.6 Half-cycle correction should be used in the Markov model if the length of the Markov cycle is long and its influence is not negligible.

13.7 The following should be noted when using a model in which the ICER fluctuates probabilistically and the ICER does not always produce the same value, such as microsimulation.

- (a) Set seed for the generation of random number to ensure that the results are reproducible.
- (b) Demonstrate that the results do not vary significantly depending on the seeds. Models with large variability may not be acceptable if the results are difficult to use for decision-making.
- (c) Random errors in the ICER arising from probabilistic variations, such as the use of extreme outliers, must be accepted, unless the results are arbitrarily adopted.
- (d) If the results vary widely, it is recommended to use the mean values from multiple trials instead of single-trial results.
- (e) Do not use a model that requires longer time for a trial with a normal-performance PC because it makes reviews difficult to conduct.

14 Uncertainty

14.1 If the analysis setting has multiple scenarios and could affect the results, scenarios analysis should be conducted.

14.2 In the case where the uncertainty is large, resulting from a long time horizon, a shorter-time analysis is necessary, such as when the time horizon is limited to periods when clinical study data are available.

14.3 If no head-to-head studies are detected according to Section 5, particularly when indirect comparison data between single-arm studies are used, a sensitivity analysis with a sufficiently wide range is required to deal with the large uncertainty.

14.4 Sensitivity analyses are required for parameters with large variances, those based on assumptions rather than actual data, and those with heterogeneity between Japan and other countries.

14.5 When the variance of the estimator should be considered (parametric uncertainty), the range in the sensitivity analysis can refer to the 95% confidence interval of the estimator.

14.6 The validity of the parameters used in a base-case analysis must be explained if they involve large uncertainties and have a significant impact

on the results. The impact of such uncertainty on decision-making should also be examined.

14.6.1 A threshold analysis may be useful when examining how decision-making is affected by parameters with large uncertainty.

14.6 A probabilistic sensitivity analysis (PSA) is desirable. In such cases, the distribution of the parameters used for the analysis, scatter plots of the cost-effectiveness plane, and cost-effectiveness acceptability curves (CEAC) must be presented.

