

医薬品経済性評価指針改正案の策定委託研究

2019年12月

健康保険審査評価院
ソウル大学校産学協力団
慶尚大学校産学協力団

提出文

健康保険審査評価院長殿

この報告書を「医薬品経済性評価指針改正案の策定委託研究」課題の最終報告書として提出します。

2019年12月

■ 主管研究機関名 ソウル大学校産学協力団

慶尚大学校産学協力団

■ 研究責任者 イ・テジン

(ソウル大学校保健大学院)

ベ・ウンヨン (慶尚大学校薬学大学)

研究員 カン・ウンジョン (順天郷大学校)

(カナダ順) ペ・スンジン

(梨花女子大学校薬学大学)

アン・ヒョンジン (高麗大学校医科大学)

チョ・ミンウ (蔚山大学校医科大学)

ハン・ソギョン (ソウル大学校医科大学)

ホン・ジヒョン (嘉泉大学校)

研究補助員 イ・ヨンシル

(ソウル大学校保健大学院)

目次

1. 序論	12
ア. 研究の背景及び目的.....	12
イ. 研究内容及び方法.....	13
2. 指針改正の範囲	18
ア. 経済性評価小委員会のワークショップの論議結果.....	18
イ. 国内外の指針改正現況.....	24
ウ. 既に提出経済性評価報告書の検討.....	27
エ. 利害関係者の意見取りまとめ結果.....	28
オ. 改正範囲	33
3. 項目別改正案の策定.....	34
ア. 観点	35
イ. 分析期間	44
ウ. 分析対象人口集団.....	53
エ. 分析技法	61
オ. 比較対象の選定.....	68
カ. 資料源（間接比較）	81
キ. 費用	99
ク. 効用、健康に関連する生活の質.....	94
ケ. 統計的な考慮事項.....	157
コ. 割引率.....	182
サ. モデル構築	190
シ. 診断検査随伴薬物.....	203
ス. 不確実性	210
セ. 付録：専門家の意見導出方法.....	220
4. 結論.....	228
付録1.相互依存医療技術に対する細部指針（オーストラリア）	230
付録2.診断検査の経済性評価と関連したチェックリスト	235
付録3.利害関係者アンケート調査用紙.....	241
5. 参考文献.....	253

要約

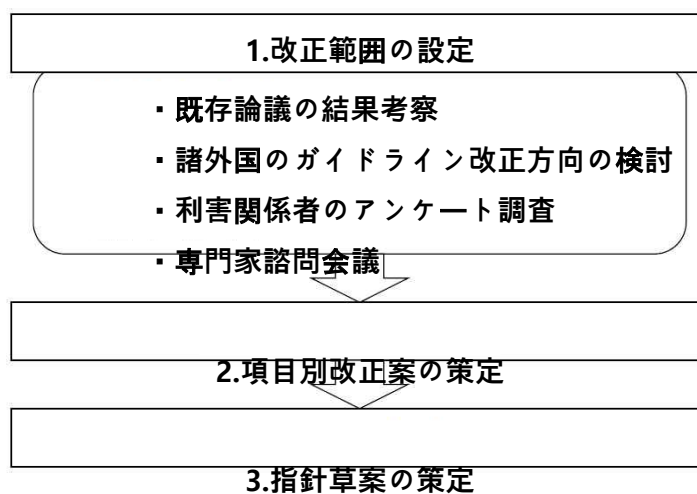
1. 序論

ア. 研究の背景及び目的

- 2011年の改正以降に行われた経済性評価方法論の発展の姿と評価経験を反映するために指針の改正が必要であるという意見が提起される
 - これまでに、指針の制定・改正の主な参照国であったイギリス、オーストラリア、カナダも2011年以降ガイドラインを全面改正したことがある
 - 国内でも評価過程で新たに登場したり争点化された事項を反映するために指針を最新化する必要があるという意見が利害関係者から提起されたことがある
- これについて、これまでに意思決定過程に適用されてきた細部評価基準を公式化し、改正が必要な事項は改正するなど、指針全般の内容を最新化して具体化しようとする

イ. 研究内容及び方法

1) 研究内容



[図]指針の改正過程

2) 研究方法

- 指針の改正案を策定するために、まず、これまでに国内で行われた論議現況と諸外国のガイドラインを検討して暫定的な改正範囲を選定した後、利害関係者の調査及び専門家の諮問会議の結果を反映し、最終の改正範囲を選定する
 - 既存の論議では、2015年12月以降、6回にわたって行われた経済性評価小委

- 員会のワークショップの結果と 2017-2018 年間の製薬業界と健康保険審査評価院間で行われた TF の活動を整理した報告書を検討した
- 諸外国のガイドラインとしては、カナダ、オーストラリア、イギリスのガイドラインを検討した
 - 利害関係者のアンケート調査
 - 調査期間：2019 年 7 月末～9 月初め
 - 調査対象：製薬会社（韓国グローバル医薬産業協会（KRPIA）、韓国製薬バイオ協会、韓国バイオの薬品協会）、患者/市民団体（韓国消費者団体協議会、グリーン消費者連帯、韓国消費者院、健康保険加入者フォーラム団体、韓国患者団体連合会）、研究者（韓国保健医療技術評価学会（KAHTA））
 - E メールを通して質問用紙の配布+回答の回収。現指針を要約して提示し、改正が必要であると思う項目にチェックする。改正が必要であると答えた項目は、その理由と希望する改正方向、根拠を記述する。
 - 最終的に製薬会社 9 部、患者/市民団体 0 部、研究者 3 部が回収される
 - 専門家諮問会議
 - 指針による経済性評価を行った経験がある専門家らを対象に諮問会議を開催し、指針の改正範囲と方向を論議した。
 - 最終的に改正が必要であると確認された項目については、次の過程を経て改正案を策定する
 - 現指針に対する分析を通して改正が必要な部分を確認する
 - 既存の提出資料に対する分析を通して指針の遵守程度、方法論適用の変移、提出資料の透明性、妥当性、不確実性のレベルなどについて調査し、指針の変更あるいは詳細指針が提供される必要のある部分を確認した
 - 既存の提出資料としては、2013 年から 2018 年間に経済性評価小委員会 で検討された薬剤 50 件の最初の提出資料を検討及び考察した。
 - 諸外国のガイドラインの分析
 - カナダ、オーストラリア、イギリスのガイドラインの他に EUnetHTA パートナー国の指針を分析して整理した EUnetHTA 報告書、その他の項目により他の国のガイドラインを分析した
 - 経済性評価指針の他、ISPOR 発刊文書など、経済性評価方法論上の争点と関連したその他の文献も共に考察した
 - 利害関係者の意見検討
 - これまでに行われた論議報告書、利害関係者のアンケート調査で提示された意見を基に利害関係者の意見を整理した
 - 専門家諮問会議
 - 項目別の改正案を策定する過程で、関連する専門家らが参加する諮問会議を開催する

- 暫定の改正方向について、関連する専門家らが参加する拡大専門家会議を経て改正方向に対する意見を求める

2. 指針改正の範囲

- 上記の過程を経て、次の項目について改正案を策定することを決定する。

<表>改正案提示項目

項目	改正の有無	項目	改正の有無
観点	V	費用	
分析期間	V	費用項目	V
分析対象人口集団	V	費用算出	
分析技法 ¹⁾	V	結果	-
分析結果の提示	-	結果指標	-
報告	-	効用、健康に関連する生活の質 ²⁾	V
比較対象の選定	V	割引率	V
資料源（間接比較）		モデル構築	V
資料の検索		不確実性	V
資料の選定	V	一般化	-
資料の抽出及び評価		公平性	-
メタ分析		財政影響	削除
長期効果の推定（外挿）、クロスオーバー法、診断検査の随伴薬物に対する評価指針			追加

注：1) CMAに対する指針;

2) Evidenceのヒエラルキー、tariff、mappingに関連する細部指針、直接測定資料の評価方法、

3. 項目別改正案の策定

ア. 考慮事項

-

各項目別の改正案を策定するにあたり、研究陣が考慮した事項は次のとおりである。

- 指針の具体性、明瞭性の確保
- 2011年の改正以降に行われた方法論上の進歩と合意レベルを反映
- 経済性評価資料を用意しなければならない国内企業と研究者の受容態勢を考慮
- 一貫した意思決定を裏付けられること
- 評価結果をめぐる不確実性の最小化

-

提示された案は、指針改正の方向を示したもので、具体的な指針（案）は別途提出する

- 指針（案）では、～する、～しなければならない、～推奨する、～選好する、～できる、などの指示的表現を 2011 年の改訂版で定義したのと同じ意味で使用する
- 指針の改正方向と関連して異見があり得る内容であるが、今後、より積極的な意見の取りまとめが必要な部分は、1 案と 2 案に区分して提示した

イ. 項目別改正案

- 各項目別の改正案策定の具体的な過程と結果は本文を参照する。
- 項目別の改正方向
 - 分析観点
 - 現在の制限的な社会的観点を保健医療システムの観点から変更し、直接費の医療費用を基本分析から除外することを骨子とする
 - 分析期間
 - 選定の原則は維持するが、分析期間選定時の考慮要素、モデルシミュレーションを通して決定された分析期間の不確実性の問題に言及する
 - 分析対象人口集団
 - 細部集団分析に関連する細部指針を提供する案である
 - 分析技法
 - 費用-効果分析より費用-効用分析を選好することを明確にする。費用最小化分析については、二つの改正案を提示する
 - 比較代案
 - 比較可能な代案のうち、市場占有率が最も高い代案ではない他の代案が比較対象として考慮されるかもしれない場合について言及する（現指針維持案と変更案をすべて提示）
 - 間接比較
 - 現指針では間接比較の遂行方法について別途出版された指針に言及しているだけで具体的な指針は提供していないが、改正案では妥当な間接比較であることを明らかにするために提出しなければならない資料を概略的に提示し、現在使用されている間接比較指針との関係性に言及
 - 費用
 - 費用及びリソースの使用量に対する資料源のうち、専門家の意見が具備すべき要件に言及し、投薬廃棄分（wastage）を考慮することに対して提案する
 - 効用
 - 効用測定法の優先順位、HRQoL ツール及び tariff の選定、算式（mapping）に関連する細部指針、直接測定に関連する細部指針を提案する
 - 効用測定法と関連しては、選好に基づいた一般ツールとして、臨床試験に含まれた患者の健康状態を測定し、この資料に国内の一般大衆を対象として導出された tariff を適用して質の加重値を算出する間接方式を選好することを明示する。
 - 選好に基づいた一般ツールで測定された資料がない場合には、他の方法も受容可能であるが、これについて十分な正当化及び説明、不確実性の検討が必要であることを強調

- 統計関連の 이슈
 - 外挿に対する一般的な内容や生存分析モデルの構築、クロスオーバー補正、非劣性評価などの統計的な 이슈に対する考慮事項に言及
- 割引率
 - 現在は費用と結果の両方に 5%の割引率を適用しているが、今後は記載部の予備妥当性の調査に使用する割引率（4.5%）と一致させる案を提案する
- モデルの構築
 - 現指針と基本方向は変わらないが、AdviSHE などの既存の妥当性評価ツールなどを参照した体系的な妥当性評価を誘導する
- 診断検査随伴薬物
 - 比較代案、考慮する要素など、診断検査に伴うオーダーメイド型医薬品に対する指針の提示
- 不確実性
 - 構造的な不確実性と母数的な不確実性を区分し、母数的な不確実性の場合、確率的な感度分析を行うことを要求する（以前の指針では推奨事項）

4. 結論

- 本研究で提案した内容、そしてこの報告書と共に提示する指針草案は、今後の様々な利害関係者との意見の取りまとめ過程を経て最終改正指針として確定する予定であり、この報告書は、今後発表される指針のバックグラウンドを提供する役割をする。
- 指針の改正案を策定する過程で一部追加研究が必要な部分が確認された場合、今後これらの研究が行われれば、これにあわせて部分改正が必要になる。
- 変化する状況により柔軟に対処するためには、全面改正だけではなく、随時、改正又は進める必要がある。
- この指針の改正と共に関連指針もあわせて改正する必要があり、費用マニュアルや統計方法論に対するマニュアルも追加で提供する必要がある。

1. 序論

ア．研究の背景及び目的

現在使用している医薬品経済性評価ガイドラインは2006年に発表され、選別収載制度の導入以降、申請薬の経済性を評価する方法論的な指針として活用されている。

初版の場合、主に諸外国のガイドラインと経済性評価方法論に対する文献的な根拠に基づいてその内容が開発された。しかし、初版発行以降、国内の経済性評価に関連する環境が急変し、これにより国内の使用経験を反映したガイドライン改正が要求され、2011年に改訂版が発刊された。

しかし、その後も全世界的に効果と費用を推定する方法論の発展があり、ガイドラインの制定と改正において重要な参考国であったイギリス、オーストラリア、カナダも2011年以降にガイドラインをアップデートした。国内でも様々な薬剤の効果と費用に対する評価経験が蓄積されるにつれ、ガイドラインを最新化する必要が利害関係者から提起されたことがある。

改訂版の発行以降、利害関係者から問題提起があったり、評価過程で争点として浮上した部分のうち、一部については、経済性評価小委員会で内部的に評価方向について合意し、実際の評価過程でこれを適用したりもした。当時に合意した内容は、大部分が既存のガイドラインの枠組み内で細部指針を明確にする性格のものであった。

しかし、最近提起された割引率の変更などは、内部評価基準を整備することから一歩進んで、ガイドライン自体に対する改正を必要とする事項である。したがって、これまでに適用されてきた細部評価基準を公式化し、改正が必要な事項は改正するなど、ガイドラインを全般的にその内容を最新化して具体化する必要性がある。

これについて、本研究では、まず最初に改正が必要なガイドライン項目を選定し、二番目に項目別に国内の評価経験、理論的考察、諸外国の経験、専門家の意見などを取りまとめて改正案を策定することを目的とする。

イ．研究内容及び方法

1) 研究内容

□ 改正が必要な項目の選定

主要国（イギリスNICE、オーストラリアPBAC、カナダCADTH）の最新の経済性評価ガイドラインを検討して現ガイドラインとの類似点及び差異点を確認し、これまでに給付決定申請書類の一部として提出された経済性評価資料に対する経済性評価小委員会の論議結果を検討した。

また、経済性評価制度の改善TFTと経済性評価小委員会のワークショップの結果に基づいて既存の指針の修正や細部指針が必要な項目を掘り起こし、指針の使用者（製薬会社、研究者、審評院の検討者）を対象に指針の改正が必要な部分及びその理由などに対する意見を取りまとめて改正が必要な項目を選定した。

□ 項目別の指針改正案の策定

改正が必要なものとして選定された項目は、観点、分析期間、分析対象人口集団、分析技法、比較対象の選定、資料源、費用項目、効用（健康に関連する生活の質）、割引率、モデル構築、不確実性、財政影響であった。これらの項目について、国内外の指針及び文献検討、利害関係者の意見の取りまとめを経て項目別に改正案を策定した。主な改正項目と概略的な改正方向は下記のとおりである。

○ 効用

効用の推定をめぐる不確実性を考慮して、効用推定関連の指針をより明確にしようとする。効用の改善の有無に対する判断基準、選好する効用測定ツール、それぞれ異なる効用資料に対する選好の偽計、基本分析に使用する tariff、マッピング資料に対する評価指針、最小限の改善度、臨床根拠-効用-モデルの連係問題、その他、効用資料の使用時に留意する指針などを提示した。

○ 割引率

現指針で使用している割引率は、2000年代の中盤に決定されたもので、その後、割引率の決定に関連する条件が変化したこと、割引率の修正が必要であるという意見が持続的に提起されたことがある。これについて、割引率に対する理論的な検討と専門家の諮問などを経て、適正な割引率を導出した。

○ 効果の推定

長期の効果の推定に伴う統計学的な 이슈などを検討して、関連指針を整備した。長期生存推定のための外挿法、効果の統計的な補正（クロスオーバー補正）方法、間接比較、観察期間以降の効果推定方法などに対する指針を提示した。

○ 不確実性の評価

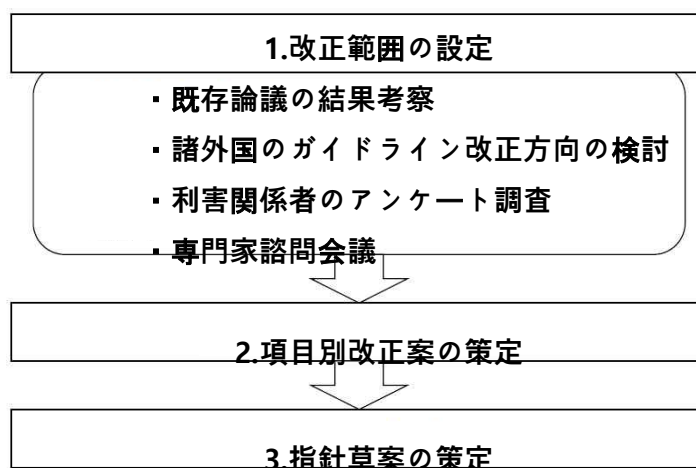
不確実性の評価及び不確実性の評価結果を提示する方法に対する指針を提示した。

○ その他

その他の比較対象、分析期間、費用の推定、モデルの構築などについて、細部指針の必要の有無を確認して改正案を策定し、現行のガイドラインで提示された文献根拠を最新化した。

項目別に専門家の諮問を経てガイドラインの改正方向に対する案を提示し、これを総合したガイドラインの改正方向について専門家の諮問を経て修正過程を経た後、最終的なガイドラインの改正方向を提示した。

以上の指針改正プロセスを図に示せば、下記のとおりである。



[図1]指針の改正過程

2) 研究方法

□ 文献考察

主な外国（イギリスNICE、オーストラリアPBAC、カナダCADTH）の経済性評価ガイドラインを比較考察した。項目別に関連文献の根拠収集及び考察を行い、ISPOR good practice guidelineとEUnetHTAガイドラインを参考にした。

□ 外国の主な機関の経済性評価資料の検討現況の調査

- 関連機関のウェブサイトを通して基本的な制度の運営現況の調査（イギリス、オーストラリア、カナダ）
- イギリス NICE、DSU（decision support unit）の訪問及び運営現況の調査
- オーストラリア PBAC（Pharmaceutical Benefit Advisory Committee）と PBAC 傘下の経済性評価小委員会（Economic Sub-Committee）対象の運営現況の調査（Eメール調査）

□ 既存の論議結果の考察

2015～2016年の経済性評価小委員会のワークショップの論議内容を回次及び争点別に分けて考察した。

□ 既に提出した経済性評価報告書の検討

これまでに給付決定申請書類の一部として提出された経済性評価資料（報告書）50件に対する情報を審査評価院から提供を受けて検討し、項目別の提出現況あるいは傾向を考察した。

□ 利害関係者の意見の取りまとめ

○ 経済性評価制度の改善 TF の運営結果

2017～2018年に運営された経済性評価制度の改善 TF の論議内容を検討した。

○ 利害関係者のアンケート調査

2019年7月末から9月初めまで、製薬会社¹⁾、患者・市民団体²⁾、研究者³⁾を対象に指針の改正範囲に対する意見調査を実施した。関連団体を通してEメールで質問用紙を配布して研究補助員のEメールで回答を回収する方法で進め、現在の医薬品経済性評価指針を下記のとおり要約して提示し、改正が必要であると思う項目にチェックするようにした。改正が必要であると答えた項目の場合、改正が必要な理由と希望する改正方向、根拠を共に記述するようにした。質問用紙の全文は付録に収録した。

<表1>既存の医薬品経済性評価指針案の要約

番号	項目	内容
評価の一般事項		
1	観点	制限的・社会的観点
2	分析期間	主な臨床結果を確認できるほど十分に長くすることを推奨
3	分析対象 人口集団	・費用と結果が導出された人口集団を明示 ・給付対象人口集団として一般化可能か根拠を提示
4	分析技法	CEA又はCUA。最終指標の使用推奨
5	分析結果の提示	ICER.総費用と総効果を別途提示
6	報告	段階別の結果と分析に含まれた資料を具体的に提示
評価方法		
7	比較対象の選定	取載医薬品がある場合は、これらの中で最も多く使用されるもの。同一成分の医薬品が多数ある場合は、市場占有率を利用した加重平均値を比較薬の単位費用として利用
8	資料源	・歪みの可能性が低い資料源（例：RCT）を優先選択
9	資料の検索	・申請薬と比較薬を直接比較した資料源の選好 ・直接比較資料源がない場合、共通対照群を利用した間接比較も許容 包括的検索。検索戦略を詳細に提示

- 1) 韓国グローバル医薬産業協会（KRPIA）、韓国製薬バイオ協会、韓国バイオ医薬品協会の協力を通して行われた。
- 2) 韓国消費者団体協議会、グリーン消費者連帯、韓国消費者院、健康保険加入者フォーラム団体、韓国患者団体連合会に協力を要請した。
- 3) 韓国保健医療技術評価学会（KAHTA）の会員らを対象とした。

10	資料の選定	資料の選定作業開始前を含む・排除基準の確立。報告書に詳述。排除文献リスト及び排除根拠の提示
11	資料の抽出及び評価	最終選定資料に対して各資料の特徴及び結果を要約・提示。質の評価結果も提示
12	メタ分析	異質性の考慮
13	費用	費用項目の列挙 → リソース消費量の測定 → 単位費用の適用 → 最終費用の算出
14	費用項目	<ul style="list-style-type: none"> ・制限的・社会的観点を取ることによる直接保健医療費用、患者及び家族（又は介護人）の費用、その他の社会の他の部門で発生する費用を含む ・生産性損失費用及び当該疾病の治療と関連しなかった未来の医療費用は基本分析に含めない ・保険者の観点から分析する場合、非給付項目は基本分析から除外
15	費用の算出	各項目の使用量の測定と単位費用の算出過程を詳しく記述。単位費用とリソース使用量の資料源は、国内資料を利用
結果		
16	結果指標	最終結果（final outcome）の使用推奨
17	効用、健康に関連する人生の質	<ul style="list-style-type: none"> ・QALY を使用。質の加重値は選好度を反映したもの。外国で開発されたツールを使用する際は、国内で妥当性が検証されたものを使用。選択ツールを指針の別添 3、表 4 に基づいて正当化し、感度分析を通して質の加重値の不確実性が結果に及ぼす影響を検討。 ・基本分析では選好度に基づいた一般ツールの使用が推奨される。健康状態の選好度測定対象は、国内の代表性のある一般大衆を選好。既存の文献から質の加重値の値を引用する場合は、資料源の包含・排除基準を詳細に説明して根拠を提示するなど、選定過程を透明に報告。また、当該健康状態に適用可能かどうかを論議して不確実性を評価
18	割引率	5%基本、0%、3%、7.5%の感度分析
19	モデル構築	<ul style="list-style-type: none"> ・モデルに対する詳述 ・モデル分析段階ごとに当該疾病状態にある者（コホートシミュレーションの場合）の数、費用などを記述 ・他の国で開発されたモデルの場合、国内の状況にもこれを適用できるか検討
20	不確実性	<ul style="list-style-type: none"> ・経済性評価過程で使用したすべての仮定とその他の不確実性を招く源泉について、臨床的、統計的に妥当な範囲にわたって一元的に感度分析を実施。母数に関連した不確実性は、PSA を推奨。 ・給付を制限して申請する場合、食薬処の許可事項まで対象が拡大する場合に対する感度分析を追加提示
21	一般化	<ul style="list-style-type: none"> ・単位費用及びリソース使用量の資料は、国内資料を利用 ・国内有病率や診療パターンが外国と大きく違いが生じ、結果の国内適用が問題となる可能性がある場合は、国内の資料に基づいて分析実施し、これを臨床試験資料に基づいた分析結果と共に提示 ・他の国が臨床試験の場合、内国人の被験者数が統計的に意味のある結論を下すことができるほど十分であれば、内国人の被験者の試験結果を基本分析、全体の被験者の試験結果として感度分析を実施。そうではない場合は、全体の被験者の臨床試験の結果を基本分析として実施。
22	公平性	公平性に関連してどのような仮定をしたのか記述。給付決定の恩恵集団の記述
23	財政影響	導入後、3～5年間の保険財政に及ぼす影響の分析

○ 専門家諮問会議

ガイドラインの項目別に専門家の諮問グループを運営し、項目別ガイドラインの開発過程で諮問を求めた。また、ガイドラインの使用経験が多い専門家らを中心に専門家のパネルを構成し、ガイドラインの開発段階別（改正範囲の設定、改正方向の草案決定）で意見

を照会した。

2. 指針改正の範囲

ア．経済性評価小委員会のワークショップの論議結果

2015年12月から全6回にわたり経済性評価小委員会のワークショップが開催された。ワークショップの目的は、既存の審議事例に対する分析と争点の整理を通して経済性評価資料の検討及び審議の一貫性を確保することであった。これは、経済性評価小委員会が作られた後、初めて試みられた作業であった。出版可能な形態の判例集が作られたわけではなかったが、これまでの審議過程で争点として浮上した部分をすべて整理し、小委員会で内部的に審議方向を再確認する意味のある席であった。最初の3回のワークショップは、2015年12月から2016年2月まで行われ、1次の統計的争点、2次の健康に関連する生活の質の推定、3次の相対的、臨床的な有用性など、その他の争点に分けて議論が進められた。以降は、さらに細部的な争点事項を中心にワークショップが行われたが、クロスオーバー補正、国内EQ-5D価値評価研究、割引率を単一イシューとして会議が行われた。各ワークショップで議論された内容を簡単に見れば、次のとおりである。

1) 統計的争点 (第1次、2015.12.29.)

長期効果の推定のための外挿、クロスオーバーが発生した臨床研究での効果推定、効果推定と相対的、臨床的な有用性の判断、間接比較による効果推定などが議論された。

長期効果の推定のための外挿の場合、臨床研究計画による最終結果までに確認することが望ましく、最新の資料まで確認することが妥当である。外部資料源は、主な資料源ではなくても参考の根拠となることができ、外部資料源と比較したexternal validityを確認して受容可否を判断する過程が活性化される必要がある。群間の生存曲線が交差する現象が確認された場合、以降の時点から申請薬の追加的な生存利益を仮定する外挿は、過大推定される恐れがある。

母数的な分布関数を通して効果を推定する場合、現在最も多く使用されている方法は、肉眼検査とcriteria統計量 (AIC/BIC) であり、log cumulative hazard plotなどを通してproportional hazardの仮定を満たすかどうかを確認する過程を追加して選定する方法で妥当性を増加させることができる。

クロスオーバーの場合、cross over又は交差投与という用語より混同がない処置転換という用語に区別して使用することが提案された。現在としては、補正しなかった結果と補正した結果をすべて参考にする他はないが、今後繰り返し提起されたり頻繁になり得る争点であるので、深い論議のために外部の専門家が参加する論議が考慮された。

治療効果の推定時の治療剤間の相対的、臨床的な有用性に対する判断が共に議論されなければならないものであり、特に間接比較による効果の推定時に改善又は非劣性認定の可否に対する論議が重要なことでもある。間接比較時の認定基準に対しては、相対的に厳格な基準と緩和された基準を提示することができる。

2) 健康に関連する生活の質の推定 (第2次、2016.1.13.)

費用-効用分析の対象と健康に関連する生活の質の測定が議論された。費用-効用分析の対象は、人生の量又は生活の質の改善がなければならず、最も望ましくは、当該臨床研究で選好度基盤の健康に関連する生活の質 (HRQoL) ツールで生活の質を測定した結果を持つ資料源で改善の可否を判断し、その根拠レベルと適用方法の適切性を評価することができるように提示することが望ましい。

異常反応、症状、投薬便宜性、服薬順応度の改善などを生活の質の改善として主張する場合には、臨床研究で当該指標の改善があり、最低限、この指標が健康に関連する生活の質 (HRQoL) に影響を及ぼすという関連性がなければならないものである。

健康状態のシナリオを構成して効用を測定するとき、シナリオ構成の根拠に様々な資料源が使用されるが、妥当性を提示しなければならず、特に介入群間の差を誇張しないように注意しなければならない。

内国人の生活の質の資料とは、直接法で効用を測定した場合だけでなく、臨床研究から間接法で測定した結果に内国人のTariffを適用することを含む内容である。文献引用方式で提出する場合、体系的な文献考察と感度分析を強調する必要がある。マッピング方式の場合、諸外国の指針などを参考にしてマッピングを提示する際に具備しなければならない最小要件を定義することができる。

3) 相対的、臨床的な有用性など (第3次、2016.2.16.)

相対的、臨床的な有用性を判断する際には、相対的な効果 (comparative effectiveness) と相対的な安全性 (comparative safety) を総合的に考慮しなければならない。効果と安全性をそれぞれ優越 (superior)、非劣性 (noninferior)、不確実 (uncertain)、劣性 (inferior) に区分して経済性評価の分析類型を分類している外国の事例を参考にして (オーストラリア)、国内指針の運営においても分類することができる。

例えば、効果は優れるが安全性が劣性なとき、費用効用分析の対象として見るか、効果が不確かであるが安全性に優れる場合、費用効果/効用分析の対象として見るかなどを議論することができる。過去事例の検討の結果、主な効果指標の非劣性の可否が不明な場合、費用最小化分析の対象として見るのが困難なものや、必要に応じて疎明内容と根拠資料を確認することができる。

比較代案を選定する際は、比較するに足りる収載医薬品を候補群にするので、許可範囲の超過や非給付の代案は考慮しないが、抗がん剤の場合、寛解後療法は含んでいる。学会

の意見などを通して選定された比較代案がextended dominanceとして見える場合、選定背景に対する根拠を確認する。

モデルの仮定は、一般化して原則を論議しにくい争点であり、臨床的な現実と符合するかどうかの確認、外部の妥当性を強調する必要がある。外国で開発されたモデルを内国人の患者に適用するとき、疫学的な特徴、治療形態などを確認しなければならない。

4) クロスオーバー補正

クロスオーバー発生時の生存期間に対してITT分析を行う場合、新たな治療法の便益が過小推定されるbiasが発生して適切ではない。イギリス、オーストラリアなどで発表されたガイドラインなどでもクロスオーバー補正の必要性を明示しており、イギリスの最近の評価事例でも補正方法が認められた事例が多数該当している。⁴⁾⁵⁾

これまで、クロスオーバー補正方法の不確実性を考慮して保守的な観点であるITTを基本分析として、クロスオーバー補正結果を意味のある感度分析として提示してきたが、今後は、提示した補正方法の適切性及び結果の提示、その他の使用可能な補正方法の結果（適用が適切ではない場合、当該事由を提示）に対して明確な疎明が提示される場合、補正方法を基本分析として受容することがより望ましいものと議論された。

資料提出と関連しては、今現在としては、意思決定時の内部的な判断基準や資料提出リストを細部的に定めることが難しく、件ごとに資料を判断しているが、検討経験が蓄積され、今後はガイドラインを提示できるはずである。具体的な検討方向を提示するよりは、オーストラリア

のガイドライン（2016）を参考にして、製薬会社で補正方法の妥当性を立証する資料を提示するようにすることが適切なものと見られる。

製薬会社で適切であると主張する補正方法以外にも、simple methodsの中ではITT分析、censoring分析、complex methodsの中ではIPCW、RPSFT、IPE、two-stageの補正方法は最低限含めて感度分析として提出しなければならない。各補正方法が基づいている仮定と関連して選定した方法の適切性を明らかにし、選定しなかった方法が適切ではないと判断した事由についても十分に明らかにしなければならない。

per protocol分析は、selection biasが明確に発生するので、一般的に基本分析よりは参考として考慮することができる。

クロスオーバーと関連して資料補完を要請するとき、実務でチェックしなければならない事項は下記のとおりである。

-
- 4) Latimer NR et al. Adjusting survival time estimates in the presence of treatment switching. NICE DSU Technical Support Document 16, 2014
 - 5) Guidelines for preparing a submission to the pharmaceutical benefits advisory committee. Version 5.0 September 2016.

- クロスオーバー許容基準、それぞれの基準に該当するクロスオーバー患者数
- 時点別のクロスオーバー患者比率
- 臨床プロトコルで事前に定義された補正方法の有無
- 感度分析の結果（ITT 分析、censoring 分析、IPCW、RPSFT、IPE、two-stage 分析を含む）をすべて提示。この他に新たな補正方法が使用された場合、補正方法の特性に対する詳細な説明が必要
- 交絡変数の不在、同一治療効果、two-stage methods で必要な 2 次ベースラインの存在など、補正方法の主な仮定を評価して、適切なあるいは適切ではない事由（仮説の検定方法に対して臨床試験又は外部資料源に対する考察と臨床諮問などを含む）を十分に明らかにする

5) 国内のEQ-5Dの価値評価の研究

国内のtariff研究間において、健康状態それぞれの効用値は多少差があるように見えるが、経済性評価で実際に活用される健康状態間の効用値差（difference）は、イギリスのtariffに比べて比較的一貫して示されている。すなわち、二つの健康状態に対するひとりの評価は、研究間である程度一貫した差が見られ、国内研究間ではそれなりのcorrelationがあるように見え、「医薬品経済性評価指針」で明示されたとおり、国内の代表性のある一般大衆から算出された選好度の結果が選好される。

EQ-5D-

3Lの価値評価の研究の中では最も標本の代表性があり、他の研究に比べて多くの健康状態（101個）について評価したLee et

al. (2009) 研究を基本分析とすることが妥当である。EQ-5D-

5Lの価値評価の研究の中では、EuroQoLグループが定めた標準プロトコルによりvaluation studyを行い、より標本の代表性があるKim et

al. (2016) を基本分析とすることが妥当である。

6) 割引率

「医薬品経済性評価制度の改善TFT」の推進により、経済性評価の検討手順及び経過の透明性向上案の策定及び迅速な評価を図るために、医薬品経済性評価制度の全般に対して業界の要請事項のうち、重点検討項目⁶⁾の選別及び深層検討のための根拠研究及び意見を共有し、「医薬品経済性評価指針」で割引率の変更の必要性及び可能性を検討した結果、割引率の変更に対する部分は、指針改正時に当該項目を含めて関連する専門家らの十分な意見の取りまとめを基に手順を進めることが妥当であるものとして議論された。

2006年から適用してきた現指針の割引率は、その調整が必要であると判断され、割引率の全般的な部分に対して理論的な考察及び現実的な適用問題に対する論議の必要性がある。

6) 比較薬剤、ICERのしきい値の範囲、効用（生活の質）、割引率

イ．国内外の指針改正現況

現指針の改正当時に参照した主要国は、2011年以降、すべてガイドラインを改正した。オーストラリアは2016年、カナダは2017年に、そしてイギリスは2013年にそれぞれガイドラインを改正した。イギリスの場合、現在、全面改正作業が進行中で、2020年7月頃に改正ガイドラインの草案が回覧される予定である。3か国の最新ガイドラインに対する検討を通して、現指針で改正が必要な部分、暫定的な改正方向を確認した。

□ オーストラリア

- Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee version 5.0 (2016年9月)
- 詳細な資料提出指針。Part A (section 1、2、3、4)、B、付録で構成。Part Bの section 2 は、clinical evaluation、section3 は、economic evaluation
- 2006年の version 4.0 の発表以降も version 4.1～4.5 まで発表 (minor update)。

□ イギリス

- Guide to the methods of technology Guide to the methods of technology appraisal 2013 (2013年4月)
- 付属文書、あるいは Decision Support Unit の報告書の形態で方法論に対するガイドが続いている。本ガイドライン自体も持続的にアップデートされる
- 現在、改正作業が進行中で、2020年の夏に草案の発表予定である。

□ カナダ

- Guidelines for the Economic Evaluation of Health Technologies:Canada 4th Edition (2017年3月)
- 韓国のガイドラインとフォーマットが類似。付属文書多数。

イギリス、オーストラリア、カナダの改正ガイドラインを検討した結果、概して以前のバージョンに比べて指針の具体性が強化されたものと見られる。ガイドラインの具体性が際立って見えた代表的な国はオーストラリアで、基本方向に対する説明だけでなく報告書に含めて提出する内容を表の形態で例示するなど、非常に詳細な指針を提供していた。また、方法論だけでなく、申請書の提出と関連した諸般の事項をガイドラインの中に含んでいた。内容的には、カナダとイギリスが類似の指針を提供しており、多少理論的な説明が多い方だとすれば、オーストラリアは、これまでの提出の件に対する検討経験が多く反映されたものと見られる。例えば、イギリスとカナダは費用-最小化分析に対する言及自体がないが、オーストラリアは費用-最小化分析の理論的な限界は認めながらも、費用-最小化分析の件に対して別途の指針を提供していた。分析の観点も、3国がすべて類似しており、制限的・社会的観点を採用した韓国とは異なり、保健医療システムの観点を採択

していた（表現は少しずつ異なるが）。分析期間の選定と関連した原則的な表現は同一であるが、オーストラリアは分析期間の不確実性について、さらに憂慮する表現が多かった。効用と関連して、3国がすべて間接測定を選好することを明示したが、選好するツールとtariffを特定したイギリスと異なり、カナダ、オーストラリアは、特定ツールの選好を提示しなかった。不確実性は、3国がすべて強調する領域や、イギリスとオーストラリアが確率的な分析を要求して決定論的な感度分析に対して否定的に言及したが、オーストラリアは依然として決定論的な感度分析を重く言及することを確認することができた。次の表は、指針の一部を比較したもので、この票に含まれていない詳細な指針の内容は、後続の項目別の改正論議で確認することができる。

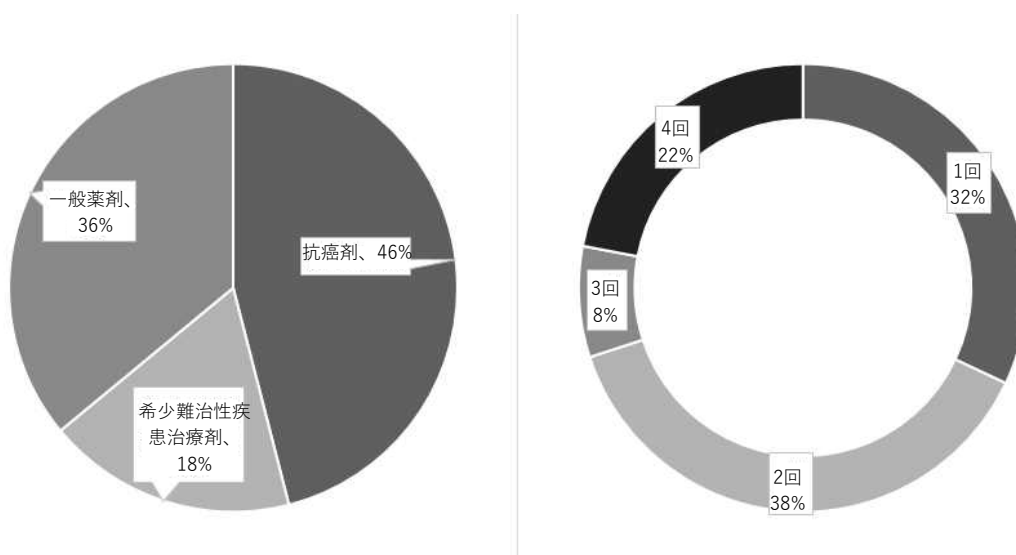
<表 2>国内外の指針の比較

	韓国	イギリス	オーストラリア	カナダ
分析期間	主な臨床結果を確認できるほど十分に長くすることを推奨	sufficiently long to reflect all important differences in costs or outcomes	captures all important differences in costs and outcomes...but does not extend unnecessarily beyond this	long enough to capture all relevant differences in the future costs and outcomes
比較代案	比較するに足りる収載医薬品がある場合は、これらの中で最も多く使用されるもの	As listed in the scope. comparatorの適切性を評価するとき、診療慣行を考慮	収載薬がある場合、代替の可能性が最も高い薬。比較薬選択のアルゴリズムあり。薬理機序、投与経路を考慮。	現在使用中で、代替される可能性がある薬、近い将来に進入する薬も考慮。収載薬が主に考慮される
健康状態の測定	言及なし	患者	言及なし	言及なし
健康状態選好度の測定対象	国内の代表性がある一般大衆	イギリスの一般大衆	オーストラリアの一般大衆	カナダの一般大衆
直接vs.間接測定	選好の言及なし	間接測定推奨（EQ-5D選好）	（一次的）間接測定推奨	間接測定推奨
直接測定時の方法	特定ツールに対する選好なし。代替選択に基づいたツールの使用	推奨しない（間接測定不可時に制限的に利用）	間接測定代案として提示。段階別に詳しく記述。偏向の最小化が必要。	間接測定法推奨

tariff	特定のtariffに対する言及はなし。 Tariffがいくつか存在する場合、最新値の使用を提案	単一UK EQ-5D tariff	特定のtariffに対する言及はなし	特定のtariffに対する言及はなし
既存値の引用	必要に応じて使用可能。 含む/排除基準を明確に明示。 透明に報告。 不確実性の評価	必要に応じて使用可能か、正当化及び不確実性の検討必要	必要に応じて使用可能か、正当化及び不確実性の検討必要	必要に応じて使用可能。目的、信頼性、一貫性の側面から選択。選択過程の詳細な記述が必要。
算式 (Mapping)	言及したが、大部分の限界及び問題を指摘	EQ-5D資料ない場合、マッピングを通してEQ-5Dに変換可能 (十分な説明、感度分析が必須)	MAUI使用しなかった理由の提示。 マッピングを通してオーストラリアの一般大衆の選好を反映。	推奨しない。マッピングを通してのみ効用の推定が可能な場合は、不確実性を十分に分析して提示
不確実性	一元の感度分析必須。 多変量分析の推奨	確率的分析を勧告 (決定論的な感度分析は勧告しない)	univariate sensitivity and scenario analyses. (すべて) & relevant multivariate analyses and any probabilistic sensitivity analysis	確率的分析を勧告 (決定論的な感度分析は勧告しない)

ウ．既に提出経済性評価報告書の検討

2013年から2018年の間、経済性評価小委員会で検討された薬剤50件の最初の提出資料を検討及び考察した。適応症を調べてみると、抗がん剤23件（46%）、希少難治性疾患治療剤9件（18%）、一般薬剤18件（36%）で抗がん剤の提出件数が最も多かった。小委員会審議の回数は平均2.2回で、1回検討16件（32%）、2回検討19件（38%）、3回検討4件（8%）、4回検討が11件（22%）であった。項目別の詳細な事項は「3.項目別改正案」パートで各項目別に扱った。



[図2]分析対象提出資料の適応症及び審議回数の分布

エ．利害関係者の意見取りまとめ結果

1) 経済性評価制度の改善TFの運営結果

2017.6～2018.12に審評院と業界の代表を中心に構成されて運営された。検討項目別の業界、審評院の意見は下記のとおりである。

比較薬剤の選定基準と関連して、業界では「古い薬と比較しないで欲しい」、「治療的な位置が同等な薬剤同士を比較して欲しい」、「使用可能な根拠がある薬剤vs.最も多く使用される薬剤」、「特許切れ薬の場合、最初の収載価格で比較して欲しい」、「First in class (高いICER適用、古い薬の比較対象から除外、最初の収載価格と比較、原価基盤の算定など)」、「代替薬剤がない先端バイオ医薬品(別途トラック)」などの意見を提起した。審評院側の検討意見は「国内外の指針と違わないため、変更事由なし」であった。

ICERの範囲は、論議対象から除外された。

割引率と関連しては、割引率の下方調整(外国の割引率を参照する場合、1.5～5%まで可能。社会的な割引率の水準を参考にして低く設定する場合、2.5%内外の割引率も可能)が製薬会社側で提起された。審評院側の検討意見は、「予備妥当性研究で使用する社会的な割引率4.5%を考慮して調整の必要性は認める」であった。

効用と関連しては、国内のtariffに対する長短所の比較及び優先順位に対する意見(資料源、ツール、測定法、測定対象及び国内の有無など)が提起され、製薬会社側では様々な状況を考慮して柔軟に対応することを要求した。審評院(経済性評価小委員会)の検討意見は、「tariffは国内の代表性のある一般大衆から算出されたものを選好して、EQ-5D-3L価値評価の研究の中ではLee et al. (2009)を、EQ-5D-5L価値評価の研究の中ではKim et al. (2016)を基本分析として」であった。

2) 利害関係者のアンケート調査の結果

ア) 製薬会社

合計9部が回収され、国内企業が1か所、外資系が8か所であった。改正が必要な項目としては、「割引率」が最も多く挙げられ(7社)、その次には「比較対象の選定」(5社)、「費用項目」、「効用/健康に関連する生活の質」(以上4社)。「観点」(3社)に対する意見が多く、「分析期間」、「分析対象人口集団」、「資料源」(以上2社)、「財政影響」(1社)の順であった。各項目に対して改正が必要であると答えた理由と希望する改正方向は、下記のとおりである。

割引率は、経済性評価の導入当時の割引率が10年を超えて適用されており、最近の低成長傾向を勘案する際、概略3%の水準に引き下げを希望する意見が多かった。

比較対象の選定の場合、会社の立場からは市場占有率が正確に分かりにくく、同じ治療的な位置を持つ薬がなかったり、直接的な臨床的根拠を確保することができない場合があり、古くなった比較薬剤の場合、薬価が非常に安い場合があり、会社が使用可能な資料範囲を考慮して欲しいという意見があり、古くなった比較薬剤は除外したり価格の引き下げ

分の補正、原価基盤の算定、ICER

thresholdの弾力的な適用などを検討して欲しいという意見もあった。

費用項目の場合、疾患の特性により、生産性損失費用、副作用管理費用などが重要な場合があるにもかかわらず、反映されないのので、これを含む方向で改正を希望する意見が多かった。

効用（健康に関連する生活の質）の場合、特定ツール及びtariffを適用するようにすることに対する反論が多く、概して柔軟なツール及びtariffの適用を選好したが、tariffによる不確実性が大きいので、かえって国内のtariffを一元化して欲しいという意見もあった。その他には、mappingの結果を基本分析として受容して欲しいという意見と、直接測定を行う際のシナリオについて例を挙げて欲しいという意見があった。

観点に対する意見は、費用項目と類似していた。疾病の特性により、重要な費用項目が制限的・社会的観点によって含めることができない場合があるので、観点を拡大して欲しいという意見が大多数であった。

分析期間項目の改正が必要な理由は、概して現在の審評院が勧告する分析期間に対する反論であったが、希望する改正方向は柔軟な設定（短い方向で）を望む方向とlifetimeを基本分析として望む方向に分かれた。

分析対象人口集団に関しては、人口集団を制限する場合、RCTの根拠を活用しにくい状況が発生するので、このような場合について指針でガイドを提示して欲しいという意見があった。

資料源の場合、RCTの資料を活用できない場合、観察研究、real world資料を使用するようにして欲しいという意見と、間接比較を行う方法について言及がないので、受容の可能性を明示して欲しいという意見があった。

財政影響の項目は、薬剤決定申請書にも記載するようになっているので、経済性評価報告書では削除して欲しいという意見があった。

それ以外に、その他の意見として、ICERのしきい値の適用に対する意見と実務指針の必要性を主張する意見、指針改正周期に対する意見などがあった。

イ) 患者・市民団体

約2週間5団体に協力を要請したが、回収された質問用紙はなかった。

ウ) 研究者

合計3部が回収され、大学所属が2名、研究所所属が1名であった。改正が必要な項目としては、「観点」と「割引率」が最も多く挙げられ（3名）、その次には「効用/健康に関連する生活の質」、「財政影響」（以上2名）、「分析期間」、「分析結果の提示」、「費用項目」、「不確実性」（1名）の順であった。各項目に対して改正が必要であると答えた理由と希望する改正方向は、下記のとおりである。

観点と関連しては、制限的・社会的観点という用語と非医療費用を含むことに対する疑

問が提起された。保健医療システムの観点から改正を希望する意見が多かった。

割引率は、現在の割引率を高いと思う意見が多く、予備妥当性調査の割引率（4.5%）と一致させたり、3%程度に引き下げる意見が提示された。

効用の場合、現在の不確実性が大きく、研究者の間でも混乱があるので、選好するツールと推定方法を明示したり、算出方法及び資料源に対してさらに明確に説明しようという意見が提示された。

財政影響の項目は、患者数の推定など、分析時点で資料の確保が難しいこともあり、今後の価格交渉時に再び議論されるので、経済性評価指針では削除したら良いという意見が多かった。

分析期間は、現在の指針の表現が曖昧なので、さらに具体的に明示して欲しいという意見があった。分析結果の提示の場合、ICER自体の不安定性があるので、net monetary benefitを追加記述する意見が提示された。

費用項目の場合、韓国保健医療研究院の費用算出マニュアルと費用項目を一致させて欲しいという意見が提示された。

不確実性の場合、現指針の不確実性の項目が母数の不確実性にのみ集中しているので、異質性、構造的な不確実性などに対する言及が拡大記述されたら良いという意見が提示された。

それ以外に、その他の意見として、reference case、教育用指針、内部指針などを別途作成することを検討して欲しいという意見と、ICER thresholdの範囲を提示して欲しいという意見、方法論的に使用可能なものなどの範囲を提示して欲しいという意見などがあった。

エ) 小結

利害関係者を対象に指針の改正範囲に対する意見を調査した結果、製薬会社の意見と研究者の意見間に共通する部分もあったが、概して関心を持っている項目が異なるものと見られた。

割引率は、製薬会社と研究者の両方すべてで最も改正が必要な項目に選ばれたが、引き下げ水準は、製薬会社側で期待する引き下げ幅がさらに大きいものと見られる。

効用/健康に関連する生活の質と観点も、両側の多数で改正が必要であると回答されたが、観点の場合、研究者は用語の変更（制限的・社会的観点→保健医療システムの観点）に言及したことに対して、製薬会社の意見は、観点を拡大して欲しいという要求が多かった。

効用/健康に関連する生活の質にも希望する改正方向が異なって示された。製薬会社側は、柔軟なツール及びtariffの選定を希望する意見が相対的にさらに多かったが、研究者の調査結果では、むしろもっと明示して欲しいという意見があった。

分析期間、費用項目、財政影響は、多数ではないが、両方で改正が必要であるものと言

及された。財政影響の項目の改正が必要であると答えた場合、概して当該項目の削除を提案した。

比較対象の選定は、製薬業界では改正を要求する声が多かったが、研究者対象の調査では、改正が必要であると答えた回答者はいなかった。分析対象人口集団と資料源も製薬会社の調査でのみ言及され、分析結果の提示と不確実性は、研究者対象の調査でのみ言及された。

3) 専門家の諮問会議

8月初めの経済性評価研究の遂行経験がある学界の研究者を対象に指針の改正範囲に対して専門家の諮問会議を行った。会議で提案された意見の要約は、下記のとおりである。

指針の基本的な方向性と関連して、記述の程度に対する 이슈があった。ガイドラインを原則中心に簡単に記述し、ディテール的な部分は分離すべきであるという意見があり、とても大きな変化を与えれば混乱するので、現在の枠組みを維持すべきという意見があった。

観点の場合、外国では「制限的・社会的観点」という表現をあまり使用しないので名称変更の考慮が必要であるという意見があり、保険者の観点と非給付費用に対して検討する必要があるという意見があった。

分析期間の場合、「十分に長く」という表現に対して検討する必要があるという意見があり、比較対象の選定は「最も多く使用する薬」という定義が曖昧なので、最大限に具体的であって欲しいという意見があった。

資料源の場合、臨床試験の結果の使用に対する 이슈があったが、臨床試験で導出された結果は、最大限使用するようにして欲しいという意見があった。data sourceと関連しても 이슈があったが、公式的でpublishな資料だけを受け入れるなど、受容可能なdata sourceを明示して欲しいという意見があった。

費用の場合、NECAと外国の事例を参照して、費用項目の名称修正、整理を考慮して欲しいという意見があった。

効用の場合、選好する効用測定法を明示することと関連した 이슈があった。ガイドラインで最もvalidな効用測定法から順番どおりに順序を定める必要があるという意見があった反面、優先順位を定めることには反対するという意見があった。また、現指針は無条件に一般人を対象に測定されたものが先であり、臨床試験から出た値はあまり選好されないように理解できるため、文面の修正が必要であるという意見があった。特定ツールの勧告と関連しては、EQ-5D-3L以外のツールもリストアップして欲しいという意見と、特定ツールを勧告することは適切ではないという意見があった。

割引率の場合、予備妥当性調査の割引率(4.5%)と同じにするか検討が必要であるとい

う意見があった。

モデル構築の場合、外国で開発されたモデルに、値だけ韓国の値に変えて入れる場合を考慮して指針に内容を入れるべきだという意見と、ブラックボックスモデルの可能性があるので不確実性のチェックをさらに徹底して行わなければならないという意見があった。

不確実性の項目は、現在あまりに母数の不確実性中心に行われているため、より広範囲な不確実性テストに対する内容を入れて欲しいという意見があった。

財政影響の場合、引き続きガイドラインに入れるべきか疑問であり、継続して入れるのであれば、さらに深くdevelopして、除外してもかまわないのであれば除外する方が良いという意見があった。

その他の意見として、体系的な文献考察の範囲をガイドラインで提示して欲しいという意見があった。資料(SASコードなど)提出の範囲/レベルをあらかじめ定めて欲しいという意見もあったが、これを指針に明示することは、さらに熟考しなければならないという意見もあった。診断に附随する医薬品の経済性評価の問題を改正指針に含めて欲しいという意見もあった。

オ . 改正範囲

上述した内容を基に設定した指針の改正範囲は、下記のとおりである。観点、分析期間、分析対象人口集団、分析技法、比較対象の選定、資料源（間接比較）、費用項目、効用（健康に関連する生活の質）、割引率、モデル構築、不確実性の項目を改正範囲として選定し、財政影響は項目の削除を勧告する。また、既存の指針にはなかった統計的な考慮事項（長期効果の推定、クロスオーバーなど）と診断検査に随伴する薬剤に対する評価指針を新たに追加する。

<表3>選定された改正範囲

項目	改正の有無	項目	改正の有無
観点	○	費用	
分析期間	○	費用項目	○
分析対象人口集団	○	費用算出	
分析技法	○	結果	-
分析結果の提示	-	結果指標	-
報告	-	効用、健康に関連する生活の質	○
比較対象の選定	○	割引率	○
資料源		モデル構築	○
資料の検索		不確実性	○
資料の選定	○	一般化	-
資料の抽出及び評価		公平性	-
メタ分析		財政影響	削除

新たに追加される項目

- ・ 統計的な考慮事項（長期効果の推定、クロスオーバーなど）
- ・ 診断検査の随伴薬剤に対する評価指針

3. 項目別改正案の策定

この章では、今回の改正範囲に含まれる項目について、文献検討及び利害関係者の意見聴取結果などを基に改正方向を提示する。各項目別の改正案を策定するにあたり、研究陣が考慮した事項は次のとおりである。

最初に、経済性評価の遂行者と検討者が各自の業務を行うにあたり、実質的な基準となる明瞭な指針を提供する。

二番目に、2011年の改正以降に行われた経済性評価方法論の進歩と合意レベルを十分に反映するが、国内での受容の可能性も考慮する。

三番目に、評価結果をめぐる不確実性を最小化する。四番目に、一貫した意思決定を裏付けなければならない。

以上の原則を基に、各パート別に、まず現指針で改正が必要な部分を検討し、続いて既に提出された申請の件を分析して、国内の経済性評価の遂行現況を調べた。引き続き、カナダ、オーストラリア、イギリス、そしてその他の国の指針とその他の文献検討を通して、現在、全世界的に通用して受容される方法論、資料要求レベルを確認し、最後に利害関係者が提出した意見を考慮して改正案を策定した。

各項目別の改正案の策定過程と結果は次のとおりである。

ア．観点

1) 現指針

ア) 主な指針

医薬品に対する経済性評価を行うにあたり、社会的観点に基づいて費用と結果を評価するが、理論的に定義された社会的観点に比べてその領域及び包括性は制限的である。ただし、保険収載と関連した意思決定者の判断を助けるために、保険者の観点に立った分析結果を追加で提示することを選好する。

イ) 指針内の追加説明

…本指針で「社会的観点ではあるが制限的」として用語を明確にしたものは、資料作成者に不必要な混同を減らすためであり、結果的に初版と比較して含まれる項目は変更されない。どのような観点下で分析を行っても分析全般にわたって一定の観点を維持しなければならない。資料提出者が保険者の観点から追加で分析結果を提出することは望ましい。

現指針では、初版の発行当時に言及した社会的観点が実際の基本分析に含まれる費用の範囲と一致しておらず混乱を引き起こしているという点を勘案して、社会的観点の前に制限的という修飾語を付けた。もちろん、含まれる費用の範囲は初版と同一である。しかし、意思決定者の主観審査が保健医療分野に及ぼす影響という点を考えれば、社会的観点、保健医療システムの観点でもない現在の観点を継続維持するかどうかに対する検討が必要である。

2) 既存の提出資料の検討結果

現在適用されている指針が制限的観点を標榜しており、含まれる費用項目として直接医療費用と非医療費用をすべて含んでいるが、提出資料中では直接非医療費用を省略している場合が多かった。合計50件の提出資料のうち、交通費用を含むものは44件、時間費用を含むものは16件、介護費用を含むものは7件であった（表4）。

<表4> 交通費用、時間費用、間接費用を含むかどうかによる分類

	交通費用	時間費用	介護費用
含む	44	16	7
含まない	6	34	43
計	50	50	50

交通費用を提示した場合、訪問回数は主に仮定をするか専門家のパネルを通して導出し、件当たりの交通費用は、2005年の国民健康栄養調査報告書を参照した後、これに統計庁の交通部門の消費者物価指数を適用して算出した場合が大部分であった。時間費用の場

合、2004年の健康保険審査評価院の外来医療利用者の利用時間調査の結果、あるいは2000年の国民保健医療実態調査を活用して利用時間に対する情報を求めた場合が多く、時間当たりの賃金としては、大部分を労働部の雇用形態別の勤労実態調査報告書で提示された値を使用した。これに提出の件により、統計庁の資料を通して雇用率を掛けた場合があった。その他、3期の国民健康栄養調査報告書、1次の韓国医療パネル調査報告書などを参照した場合もあった。介護費用を提示した場合、介護日数は仮定や様々な資料源を参考にし、一日当たりの介護費用は、介護協会で発表する介護料資料を引用するが多かった。三項目の算出内訳を全体的に調べれば、リソース利用量と単位費用を推定するほどの直接的な資料がなく、既存の調査資料を活用したが多かったが、それすら最新の調査資料がなく、過去の資料に消費者物価指数を適用して現在値を求めた場合が多かった。概して費用推定過程の不確実性が大きい方であると言える。

3) 外国の指針

各国が採択する評価観点は、大きく保健医療の観点と社会的観点に分類することができる。本指針の改正過程において、主な参考資料として活用するオーストラリア、カナダ、イギリスは、具体的な表現は異なるが、すべて保健医療の観点を採択していると見ることができる。保健医療部門内で発生する直接的費用と患者/保護者で発生する健康便益は含んでいるが、その他の費用は基本分析に含まない。もちろん、介護費用や政府の他の分野で発生する費用の場合、その大きさを測定して別途提示するように許容しているが、基本分析には含んでいない。その他の国のうち、オランダやスウェーデンのように社会的観点を標榜して、生産性の損失まで含むことを明示した国もある。主要国の指針内容は次のとおりである。

ア) カナダ

カナダは、公費負担医療の支払者 (Publicly funded health care payer) の観点を採択している。費用としては、公費支払者に発生する費用、成果としては、患者と非公式の介護人のすべてに意味のある健康効果を反映するようにしている。また、対象人口集団外で発生する波及 (spillover) 費用や効果があれば、これをnon-reference case分析として含むようにしている。

指針の表1では、採択観点に伴う費用と成果の例を提示しているが、その内容は次のとおりである。

- 費用：
 - 医薬品、医療機器（材料）、施術（Drugs, medical devices, procedures）
 - 設備、施設、間接費（Equipment, facilities, overhead）
 - 保健医療提供者（Health care providers）
 - 病院サービス（Hospital services）
 - 診断、検査、検診サービス（Diagnostic, investigational, and screening services）
 - 非公式介護人の保健医療費用（Informal caregivers' health care costs）
 - 施設あるいは家庭でのリハビリ（Rehabilitation in a facility or at home）
 - ホームケアや社会的支持のような地域社会基盤のサービス（Community-based services, such as home care, social support）
 - ナーシングホームでの長期療養（Long-term care in nursing home）
- 健康効果：
 - 健康関連の人生の質（Health-related quality of life）
 - 生存期間延長（Life-years gained）
 - 臨床的罹患（Clinical morbidity）

イ）オーストラリア

オーストラリアは、保健医療システムの観点（Health system perspective）を採択している。オーストラリアの指針では、費用としては患者と公共あるいは民間保健医療提供者で発生するすべての費用、成果（outcome）は、患者と関連したものなどを含むようにしている。また、申請された介入が、医薬品を提供される患者の健康成果と保健医療システムを越えて重要な社会的な含意を有しているのであれば、補充的な分析を提示できるとも言及している。

オーストラリアの指針で、保健医療リソースの項目で考慮に言及したものは次のとおりである：

- 医薬品（治療の直接費用と副作用を治療するための医薬品費用）
- 施術を含む医療サービス
- 病院サービス
- 診断及び検査サービス
- 地域社会基盤のサービス
- その他の直接医療費用

また、オーストラリアでは、費用マニュアルを別途提示しているが、これに含まれる項目は次のとおりである。

<表5> オーストラリアの費用マニュアルに含まれる項目

項目	含まれる細部項目
Medicines, medicinal preparations or vaccines	proposed medicine, medicinal preparation or vaccine comparator and co-prescribed medicines or medicinal preparations over-the-counter medicines medicine delivery systems medicine administration costs
Medicinal services, including investigative, diagnostic and allied health services	
Hospital services	admitted patient services nonadmitted hospital care emergency department services
Community-based services	residential care home care home nursing ambulance services other community-based services

ウ) イギリス

イギリスは、NHS と個人・社会サービスの観点（NHS and personal and social service; NHS and PSS）を採択している。NHS と PSS 以外の部門で発生する費用は、事前に保健省がその費用を含むものとして決定したとすれば、non-reference case analysis で提示できるが、評価方法が明瞭に提示されなければならない。そして、これら費用は、NHS と PSS 費用と区別して提示し、漸増的な費用対効果（incremental cost-effectiveness ration; ICER）には含まない。

患者が負担する費用の場合、この費用がNHSやPSSによって償還される場合には含め、償還率が地域間、患者間で差があるときは平均値を使用する。NHSやPSSにより償還されない患者負担費用があれば、これら費用は別個で提示されなければならない。生産性費用は排除する。

家族介護の場合、家族ではなく、NHSやPSSが当該サービスを提供したものであれば、介護に所要する時間費用を考慮することができる。ただし、基本分析とは別個に当該項目を提示しなければならない。そして、選択した価値評価（valuation）方法は明確に記述し、選択した方法と異なる方法を利用した結果も感度分析として提示しなければならない。

エ) その他の指針

EUnetHTA（2015）が発表した勧告案は、パートナー国間の経済性評価の移転の可能性を高めることを目的として作られたもので、パートナー国の経済性評価指針を分析して共通の勧告案を提示している。EUnetHTA 報告書を見れば、ヨーロッパの国の場合には、保健医療システムの観点を採択した国もあり、社会的観点を選んだ国もある。EUnetHTA では、国別に採択した観点が異なることを示した上で、経済性評価は、最低限、保健医療の観点で行われなければならないという点を強調した。また、保健医療を越えて社会の他の部分で発生する費用の場合、補足的な分析形態で提示することにより、それぞれ異なる

観点を採択した国で便利に活用できるようにする必要があると助言する。そして、与えられた観点下でリソースの使用に対する情報は、できるだけ詳細に提示することを勧告している。

<表6> EUnetHTAパートナー国の分析：費用と結果の観点

国	費用に対する観点	結果に対する観点
オーストリア	The choice of perspective must be justified.	The choice of perspective must be justified.
ベルギー	Health care payers	Effects on patients primarily
クロアチア	Perspective of the Croatian Institute for Health Insurance (public payer)	All health effects on individuals. Health effects in informal caregivers and/or family members can be reported separately
チェコ	Health care payers	Na
デンマーク	Socio-economic perspective	Na
イギリス (England)	The NHS and personal social services	All direct health effects, whether for patients or where relevant for caregivers
エストニア&ラトビア	Health care	Na
フィンランド	Societal	Na
フランス	Collective perspective. All the resources used in the production of interventions.	Collective perspective. All the health effects relevant of the individuals concerned (patients, informal carers, general population)
ドイツ	SHI insurants primarily according to Social Code	SHI insurant primarily according to Social Code
ハンガリー	Na	Na
アイルランド	The perspective of the publicly-funded health and social care system	All health benefits accruing to individuals
イタリア	Health care	Effects on patients primarily
オランダ	Societal	Societal
ノルウェー	Societal	Societal
ポーランド	Two variants are required: Public health care payer and public health care payer and the patient.	The patients. Societal perspective (health effects to other members of the society) in specific cases
ポルトガル	Societal. This means considering the costs for the patient, for his or her family and also for third parties, i.e. public and private payers in particular. Society's perspective should be broken down into other relevant points of view, with special attention to the third payers if they	Societal. This means considering the consequences for the patient, for his or her family and also for third parties, i.e. public and private payers in particular. Society's perspective should be broken down into other relevant points of view, with special attention to the third payers if they

	are the users of the study.	are the users of the study.
ロシア	No priority for any perspective; it is only recommended that researchers declare clearly the perspective of the study (any: societal, healthcare, etc). For drugs submitted into the lists either societal or health care perspective are recommended.	No priority for any perspective; it is only recommended that researchers declare clearly the perspective of the study (any: societal, healthcare, etc). For drugs submitted into the lists either societal or health care perspective are recommended
イギリス (Scotland)	A healthcare perspective is required, but a societal perspective can be explored through sensitivity analysis	A healthcare perspective is required, but a societal perspective including effects on other individuals than patients (principally carers) can be explored through sensitivity analysis
スロバキア	Health care payers	Societal
スロベニア	Health insurance; societal perspective can be performed as well	Patients only
スペイン	Societal and third-party National Health System (NHS) Societal and that of the decision-maker (OSTEBA) The perspective of the financier – CATALUT and as a complement a societal perspective (CATSALUT)	Societal and third-party National Health System (NHS) Societal and that of the decision-maker (OSTEBA) The perspective of the financier – CATALUT and as a complement a societal perspective (CATSALUT)
スウェーデン	Societal	Societal
スイス	Health care	Only health effects on patients (not utility)

出典: EUnetHTA (2015a)

Na: No information available, NHS:National health service, SHI:Social health insurance

<表7> 分析観点に関連する海外の指針内容

国	指針内容原文	要約
カナダ CADTH 指針 (4版) (2017)	In the reference case, the perspective should be that of the publicly funded health care payer. The perspective of the economic evaluation should be related to the decision problem. (p18)	公共保健医療支払者 (publicly funded health care payer)
オーストラリア PBAC 指針version 5 (2016)	The PBAC's preferred health care system perspective includes health and health-related resource use (costs and cost offsets), and health-related outcomes. Costs include those incurred by the patient, and the public or private health care provider; outcomes are those associated with the patient. Do not include costs and outcomes that are not specifically related to 'health and/or provision of health care' in the base case (p.65)	保健医療システムの 観点

イギリスNICE 指針 (2013)	For the reference case, the perspective on outcomes should be all direct health effects, whether for patients or other people. The perspective adopted on costs should be that of the NHS and personal and social services. (p34)	NHS and personal and social services 観点
EUnetHTA (2015)	the economic evaluations should at minimum be conducted from a health care perspective. However, several countries recommend a societal perspective. Thus, presenting resource use related to other sectors of society in complementary analysis may increase the usefulness of an evaluation to other countries. Within the chosen perspective(s), it is recommended to present the resource use in as detailed a manner as possible	

4) その他の文献の考察結果

国内では、韓国の保健医療院が発刊した費用算出マニュアルである「保健医療分野における費用算出方法」(キム・ユンヒなど、2013)で分析観点を保険者の観点、保健医療システムの観点、社会的観点到分類して各観点下に含まれる費用項目を提示した。目につくのは、保健医療システムの観点を採択したときに含まれる費用として、時間費用、交通費用だけでなく非公式な介護費用まで非医療費用に含む点である。結果的にこの分類は、現在の経済性評価指針が採択している制限的・社会的観点和一致する。

<表8> 分析観点別の費用項目

	保険者の観点	保健医療システムの観点	社会的観点
医療費用	公式な医療費用 給付内保険者負担費用 給付内 本人負担費用(非給付本人負担費用)	公式な医療費用 給付内保険者負担費用 給付内 本人負担費用(非給付本人負担費用) 非公式医療費用	公式な医療費用 給付内保険者負担費用 給付内 本人負担費用(非給付本人負担費用) 非公式医療費用
非医療費用		交通費用 患者費用 長期療養サービス費用 介護費用	交通費用 患者費用 長期療養サービス費用 介護費用
生産性損失費用			罹患費用 早期死亡費用

資料源: キム・ユンヒなど(2013)、保健医療分野における費用算出方法、韓国保健医療研究院。

Zhao et

al.(2018)のガイドラインの分析結果によれば、14種のガイドラインで社会的観点を推奨し、19種が保健医療の観点を推奨した。先立って調べたイギリス、オーストラリア、カナダは、すべて保健医療システムの観点を採択しており、社会的観点を採択した代表的な国はオランダとスウェーデンである。社会的観点を採択する場合、生産性損失と未来の保健医療費用が費用項目に含まれる。

Neumann(2009)は、Tufts Medical CenterのCEA

Registryを分析し、発表された費用-

効果分析研究に使用された観点を分析したが、分析の結果、発表された研究らの費用推定方法に相当な変移が存在したことを確認した。多くの研究者が、彼らの研究が社会的観点を取っていると主張するが、実際には、保健医療支払者の観点で分析を進めたことを確認することができたと言う。

<表9> 発表されたCUA研究に使用された観点 (1976-2005年)

	As stated by author	As stated by reviewer
Societal	473 (40.6%)	341 (29.3%)
Health care payer	382 (32.8%)	799 (68.6%)
Not stated/could not be determined	298 (25.6%)	16 (1.4%)
Other	11 (0.9%)	8 (0.7%)
Total	1164	1164

資料源:Neumann (2009)

US panel on

CEAが発刊した指針の1版では社会的観点を勧告した、改訂版である2版では保健医療システムの観点と社会的観点をすべて勧告している (Neumann et al., 2017)。多数の研究が実用的な理由から保健医療システムの観点を採択することを勧告したものである。ここで、保健医療システムの観点で分析するということは、保健医療部門で発生する費用だけを含むということで、第3支払者や患者の本人負担で支出する医療費用をすべて含む。社会的観点下で分析する場合は、impact inventoryを作成するよう勧告しているが、これは保険介入の健康、非健康の影響をすべて羅列したものである。

5) 利害関係者の意見

□ 意見調査の結果

○ 製薬会社の意見

- 疾患の特性により、保護者の介護及び保護が必須の場合、これを幅広く認定しなければならないという意見が陳述された。
- ただし、観点自体を社会的観点に変更すべきという意見よりは、疾病特性により流動的に評価に含む範囲を決定できるようにすべきという意見であった。特に、疾病による生産性損失費用が大きい場合、経済性評価モデルにこれを反映するようにすべきという意見が提示された。

○ 研究者の意見調査

- 研究者の意見調査では、社会的観点到拡大すべきという意見より、現在使用している制限的・社会的観点的の代わりに保健医療システムの観点到変更すべきという意見が提示された。

- その理由としては、制限的・社会的観点よりは保健医療システムの観点がより広く通用する用語であるという点で変更意見を提示した場合もあり、制限的・社会的観点で扱われる交通費、時間費用などの不確実性が大きいので、HealthCare Sector を基本として Societal 又は制限的・社会的観点を追加的に提示すべきという意見もあった。

□ 諮問会議

- 「制限的・社会的観点」という用語が広く通用する用語ではないので、名称の変更が必要であるという意見があった。
- しかし、非医療費用を除外することについては、一部反対する意見もあった。
- また、保険者の観点と非給付費用の処理について考える必要があるという意見もあった。

6) 指針の改正方向の提言

- 健康保険給付決定という意思決定状況と現実的環境を考慮して分析観点を改正する必要がある。
 - 健康保険給付決定者の関心は、意思決定が保健医療分野に及ぼす影響であるので、保健医療分野の他の分野で発生する影響は一次的な関心対象でない。
 - 既に提出された評価の件では、直接非医療費用を含む場合もあったが、含まなかった場合も多く、評価の一貫性の側面から再検討が必要である。また、これら項目の場合、使用量などを裏付けできる根拠も不足しており、不確実性の側面でも問題になり得る。
- これに 1 案としては、既存の制限的・社会的観点を保健医療システムの観点に変更して、直接非医療費用に該当する項目を基本分析から除外することを提案する。
 - 費用は直接医療費用を中心に含むが、介護の場合、公式部門である長期療養保険で給付される場合であれば含むものとし、効果は患者に現れる健康便益を反映するものとする。
 - 現指針と比較すると、交通費用、時間費用、保護者の介護費用などが基本分析から除外される。
 - ただし、疾患や治療法の種類によって非医療費用の比重が大きい場合は、生産性損失費用と同様にこれら費用もまた別途資料で提示するようにする。
- 2 案は現指針を維持する案である。

イ . 分析期間

1) 現指針

ア) 主な指針

分析期間は、主な臨床結果を確認できるほど十分に長くすることを推奨する。モデルの構築が必要な場合には、分析結果と共にモデルが必要な理由とモデルの構造などに対する詳細な説明を提示する。万一、モデルを利用して長期間外挿した場合には、原資料に基づいた分析結果を別途提出することを推奨する。

イ) 指針内の追加説明

...一般的に経済性評価のための分析期間は、主な臨床結果を確認できるほど十分に長い方がいい。疾病の発生と進行、終結が短い期間で行われた急性期疾病の場合には、最終結果を臨床試験期間内に確認できるが、そうではない疾病の場合には、通常は臨床試験期間内に疾病の経過が完了しない場合が多い。この場合、モデルの構築を通して臨床試験期間以降の経過を推定することができるが、疾病によっては患者の生存期間中、経過を追跡しなければならないこともある。しかし、分析期間を拡張するほど不確実性もまた共に大きくなる問題点があるので、観察期間が原資料より長くなる場合には、外挿する前の原資料に基づいた分析結果を追加で提出することを推奨する。このとき、外挿した結果との差を分析して記述することを推奨する。

現指針では、主な臨床経過を確認できるほど分析期間を十分に長くすることを勧告している。ただし、外挿を通して分析期間が長くなるほど不確実性が大きくなることを考慮して、観察期間が原資料より長くなる場合には、外挿する前の原資料に基づいた分析結果を追加で提出することを推奨している。しかし、依然として分析期間をめぐる論議が続いており、外挿期間が長くなるほど不確実性が大きくなる問題も提起されているので、分析期間の設定時に考慮する要素などに対してさらに具体的な指針提示が必要である。

2) 既存の提出資料の検討結果

□ 提出資料から確認された傾向

多くの提出資料は、対象患者の大部分が死亡する期間を分析期間として選択し、具体的な期間の算定方法は、当該年齢帯の期待余命を考慮したり、モデルを分析して申請薬群あるいは対照薬群に含まれた患者の00%が生存する時点を考慮するなど、多様であった。

指針では、観察期間が原資料より長くなる場合には、外挿する前の原資料に基づいた分析結果を追加で提出することを推奨するとしたが、実際の原資料に基づいた分析結果を提出した場合は非常に稀であった。明らかに観察期間の間の資料で感度分析をした件は2件に過ぎず、大部分の基本分析において、使用された期間の前後に一定の間隔を置いて感度分析を行った。一部、1~3年を分析期間とした感度分析の結果を提示した場合もあったが、観察資料を基に感度分析をしたということよりは、モデルの分析期間を多様にするとい

う次元で結果が提示された。

相対的に短い分析期間を選択した場合は、痛みの軽減や症状消失など、比較的短期間に観察可能なイベントを主な結果指標として使用した場合、あるいは残存する期待余命自体が短い場合であった。

分析期間を設定するにあたり、その根拠として提示された資料は、疾病疫学、臨床研究、専門家の意見など多様であったが、代表的な例は次のとおりである。

- 申請薬の臨床研究（生存率、平均年齢、生存差） - 延長研究含む
- 疫学資料：発病年齢、疾患5年/10年生存率、自然死亡などを考慮した対象患者の期待余命
- モデルから推定された生存率、臨床専門家の意見

□ 個別の審議事例

- 申請薬と比較薬の分析期間をそれぞれ異なる設定にした場合
 - 分析期間を「生涯期間」として設定するが、生涯期間は「モデルで95%が死亡するときまで」として定義した場合があった。このように分析期間をモデル内で死亡率が一定の%になるときまでと定める場合、申請薬と比較薬の追跡期間に差が発生する。すなわち、同じ時点で生存した患者であっても、比較群で余命がさらに短く反映されるという問題点がある。この件に対して、経済性評価小委員会では、申請薬と比較薬の分析期間を異なるように設定することは妥当ではなく、死亡率がさらに早く減少する比較薬を基準として分析期間を定める保守的なアプローチを選好するという意見を提出した。
- 生存曲線が交差する場合
 - A薬の場合、提出された分析結果では、比較薬剤に対して14年間の持続的なOS gainがあると推定しているが、経済性評価小委員会では、36か月の時点で二群の生存曲線が交差する点や対象疾患の生存期間などを考慮したとき、効果が過大推定された可能性があることを憂慮した。これに経済性評価小委員会では、二群の生存曲線の交差以降も治療効果が同一であったという仮定（hazard ratio=1）を適用して分析したり、生存曲線が交差した時点までの効果を反映した資料の提出が必要であると勧告した。
- 期待余命に対する判断
 - B薬は、特発性肺線維症の治療剤として、プラセボに対する分析として、分析期間を19年に設定した。これは当該患者の平均年齢での期待余命を根拠としたものである。しかし、経済性評価小委員会では、特発性肺線維症患者の期待余命が一般人口と同一であるという根拠が十分でないことを理由に、分析期間をこれより短期間にすることを勧告した => 最終的に臨床研究の追跡期間よりは長期間であり、一般人の期待余命よりは短期間に該当する15年を基本分析期間として採択した。
- 臨床専門家の意見を考慮

- C 治療剤の場合、慢性疾患である多発性硬化症の (i) 疾病特性を反映して、(ii) 臨床試験資料の外挿による不確実性を低くするために、20 年を基本分析期間として設定した。しかし、経済性評価小委員会では、期待余命に対する臨床専門家の意見と臨床試験参加者の平均年齢を勘案して SOT 群は 25 年、RES 群は 30 年を分析するものとした。
- 資料の不確実性を問題視した場合
 - D 治療剤は「局所ステロイド治療に反応しない成人の再発性慢性重症手湿疹 (PGA で評価)」患者を対象として、分析期間を 10 年に設定した。しかし、経済性評価小委員会では、モデルに使用された仮定が本分析結果に及ぼす影響を確認できず、臨床試験期間以降の観察資料が無かったため、分析期間を臨床資料源に基づいた期間に変更するということが妥当であるという意見を提出した。
- 類似疾患に対する審議事例
 - 類似疾患に使用された薬剤の場合、分析期間の設定において、既存の審議事例を参考としたが、代表的な例は次の<表 10>のとおりである。

<表10> 類似疾患の審議事例

	薬L	薬M	薬O
適応症 (小細胞肺癌)	以前の白金ベース化学療法に失敗したstageⅢB以上		以前のEGFR-TKIの投与後、疾病が進行した局所進行性又は転移性
諸外国の評価結果	5年(オーストラリア、カナダ) 20年、end of life(イギリス)	5年(カナダ) 20年、end of life(イギリス)	5年(カナダ) 15年、end of life(イギリス)
臨床試験の結果 (申請品)	(非扁平細胞癌) mOS: 19.9か月 mPFS: 5か月	mOS: 15.8か月 mPFS: 5.2か月	mPFS: 10.4か月
経済性評価モデルの申請品の死亡率	(7年基準) 扁平細胞癌95.2% 非扁平細胞癌83.1%	(7年基準) 90.8%	-
経済性評価小委員会の会議結果 (7年)	基本分析で提示された分析期間7年と関連し、第I相試験で最大7年まで観察された資料が追加で提示され、既存の審議の類似薬剤の事例と経済性評価モデルで7年時点の生存率などを考慮したとき受容可能である。	製薬会社の追加根拠資料及び同時に審議薬剤の根拠資料を土台とした再論議で、申請品より相対的に観察期間が長い資料が提出された同一機序の薬剤と申請品の第III相試験の同一観察期間で生存効果が類似の傾向を見せる点、及び既存の審議の類似薬剤の事例の分析期間などを考慮して、分析期間を7年まで認定可能である。	申請品は、むしろOSに対して観察された臨床資料が不足し、相対的に以降の効果が不確かであるという制限点があるが、臨床専門家の諮問会議で1次治療剤の5年生存率などを基に申請品の効果を推定したとき、分析期間として7年ほどが適切であるという意見及び既存の審議の薬剤らと対象患者の状態が類似している点を考慮して7年を基本分析とする。

□ 審議事例の総合

既存の事例を総合するなら、概して次のような要素が分析期間の算定で考慮されたことが分かる：患者年齢（コホート進入年齢）、一般人口集団の期待余命、当該疾患の生存率、類似薬剤の分析期間、臨床試験成績（median OS）、追跡観察期間、臨床諮問、不確実性

3) 外国の指針

外国の指針も、分析期間は比較代案との間の差を捕捉できるほど十分に長くなければならないという点について、共通して言及している。ただし、外挿期間が長くなることによる不確実性をどのように見るのかに対する指針が日増しに具体化されつつある。

ア) カナダ

比較対象となる介入案らと関連したすべての未来の費用と成果を捕捉できるほど分析期間が十分に長くなければならないと明示している。しかし、QALY の増加分の相当部分が治療中断以降に発生したとすれば、分析期間と関連した不確実性に対しても論議すべきであると述べている。

イ) オーストラリア

オーストラリアの指針もまた他の国と基本原則と変わらないが、分析期間の延長に対して注意を喚起する内容が多く含まれている。指針では、介入案と比較代案間の費用と成果の面でのすべての重要な差を捕捉できるように分析期間を設定すべきであると明示している。しかし、不必要に分析期間を拡張することはしてはならないという点も強調しているが、分析期間が非常に長く、外挿された便益が占める比重が大きくなれば、それだけ不確実性が大きくなることになり、この場合、PBAC はその結果に対して確信を持つことができない可能性があるという点について言及している。

その他の介入案が死亡に影響を及ぼさずに一時的な健康改善、生活の質の改善効果だけが見られるのであれば、相対的に短い分析期間が適切であるということにも言及している。また、現実的ではない投入資料（inputs）は、現実的ではない分析期間につながることから、最終分析期間に到達するまで、モデルがどのように曲線を外挿したかを含めて分析期間の妥当性（plausibility）に対する評価が必要であるという点も強調している。

ウ) イギリス

指針では、分析期間は、費用と効果の面で比較対象である技術のすべての重要な差を反映できるように十分に長くなければならないと明示している。

しかし、観察期間以降に資料を外挿する場合、「差」が発生する期間や「差」の大きさにおいて不確実性が大きくなる。特に外挿された便益の大きさが総便益の相当部分を占める場合であれば、不確実性の問題はより一層重要となる。この場合、疾病の自然死（疫学資料、患者の生存期間）、患者の年齢、一般人口集団の期待余命、類似医療技術との比較

(類似薬剤の分析期間)、臨床試験成績、その他の分析期間を裏付ける根拠などを考慮して、基本分析に使用する分析期間を決定し、分析期間を変えて感度分析を実施すべきであると提示している。

委員会の審議時には、既存の審議事例などを検討する。このとき、臨床試験期間中に直接観察した資料に基づいた感度分析結果も共に提示する。資料を観察期間以降に外挿する際には、効果の持続と関連した相異なる仮定に基づいた感度分析結果を提示する。治療の中断以降、効果の持続に対する仮定では、以前の効果差が持続するという仮定、治療中断以降、効果差が次第に減るという仮定、治療中断と共に効果差がそれ以上発生しないという仮定に対して、それぞれ感度分析を実施する。特に治療中断以降、これ以上効果差を認めない仮定を含めて分析結果を提出するようにしている。

エ) その他の指針

ヨーロッパの国のガイドラインを考察して共通勧告案を策定した EUnetHTA (2015) の文書によれば、25 か国中 18 か国のガイドラインにおいて、分析期間は、費用と結果との重要な差が発見できるほど十分に長くなければならないと勧告している。しかし、感度分析を通して様々な期間に対する分析結果を提示することを求める場合もあり、例えば、スコットランドのガイドラインでは、観察期間が終了した時点、5 年が過ぎた時点、その後 5 年間隔で分析結果を提示する案に言及している。ドイツは他の国とは異なるが、ドイツは疾病の性格により、適切な分析期間を選定することができるが、最低限 RCT の期間にしなければならないと言及している。

最近、意見の取りまとめ中である EUnetHTA の「経済性評価時の考慮事項 (practical considerations when critically assessing economic evaluations)」では、外挿された結果の不確実性に注目して、分析期間と関連して次のような点などを考慮することを求めている。

- 費用と効果をモデリングする期間は、同一でなければならない。
- 分析期間は、現実的な要素（長期資料、疾病の自然死）を考慮して決定されなければならない。シミュレーションにだけ基づいて決定されてはならない。シミュレーションを通して得られた生涯の分析期間は、その妥当性 (plausibility) を立証しなければならない。
- パラメーターは、実際にはシミュレーション期間にわたって変化できるが、全体の分析期間にわたって固定された値を適用する場合がある。このような単純化を通しては、今知っているものに対する不確実性の分析は可能であるが、我々が知るこのできない未来のイベントに対する不確実性の分析は可能ではない

<表11> EUnetHTA パートナー国の分析：分析期間

国	選好する分析期間
オーストリア	Choice of time horizon depends on research question and study subject
ベルギー	Sufficiently long to reflect all important differences in costs and outcomes
クロアチア	Sufficiently long to reflect all important differences in costs and outcomes

チェコ	Sufficiently long to reflect all important differences in costs and outcomes
デンマーク	Sufficiently long to reflect all important differences in costs and outcomes
イギリス (England)	Sufficiently long to reflect all important differences in costs and outcomes
エストニア& ラトビア	Not indicated, but it is stated that "...modelling techniques can be applied when trial data provide too short a time frame..."
フィンランド	Sufficiently long to reflect all important differences in costs and outcomes
フランス	Sufficiently long to reflect all important differences in costs and outcomes
ドイツ	At least length of RCTs, yet the appropriate time horizon depend on the nature of the disease
ハンガリー	Sufficiently long to reflect all important differences in costs and outcomes
アイルランド	Sufficiently long to reflect all important differences in costs and outcomes
イタリア	Sufficiently long to reflect all important differences in costs and outcomes
オランダ	Sufficiently long to reflect all important differences in costs and outcomes
ノルウェー	Sufficiently long to reflect all important differences in costs and outcomes
ポーランド	Sufficiently long to reflect all important differences in costs and outcomes
ポルトガル	Sufficiently long to reflect all important differences in costs and outcomes
ロシア	Na
イギリス (Scotlan d)	Sufficiently long to reflect all important differences in costs and outcomes. Results (in net cost per QALY gained) should also be reported at different time horizon intervals e.g. at end of study follow-up, at 5 years follow-up and at 5-year intervals thereafter
スロバキア	Sufficiently long to reflect all important differences or 5 years
スロベニア	Time horizon must be specified in economic analysis
スペイン	Sufficiently long to reflect all important differences in costs and outcomes (Spanish recommendations, CatSalut) A shorter horizon which only includes primary data and a longer horizon that includes modelling (Osteba) It is recommended to do a complimentary analysis with a time perspective of 3-5 years (CatSalut)
スウェーデン	Sufficiently long to reflect all important differences in costs and outcomes
スイス	Na

出典: EUnetHTA (2015a)

Na: No information available, RCT: randomized controlled trial

<表12> 分析期間に関連する海外の指針内容

国	指針内容原文	要約
カナダ CADTH 指針 (4版) (2017)	In the reference case, the time horizon should be long enough to capture all relevant differences in the future costs and outcomes associated with the interventions being compared. (p18)	未来の費用と結果の差を捕捉できるほど十分に長くなければならない。
オーストラリ	...the time horizon captures all important differences in	分析期間は、介入案と

<p>ア PBAC 指針 version 5 (2016)</p>	<p>costs and outcomes between the intervention and the comparator, ...but does not extend unnecessarily beyond this The assessment of plausibility should also apply to how the model extrapolates the curves to reach this time horizon</p> <p>As a modelled time horizon extends – in absolute terms and relative to available data – it is associated with increasing inherent uncertainty. Therefore, economic claims based on models with very extended time horizons and predominantly extrapolated benefits will be less certain and are likely to be less convincing to the PBAC. (p68)</p>	<p>比較代案間の費用と結果の重要な差を捕捉できること、観察期間越しに効果を外挿する際に不確実性が大きくなる可能性があることを明示</p>
<p>イギリス NICE 指針 (2013)</p>	<p>...should be sufficiently long to reflect all important differences in costs or outcomes between the technologies being compared</p> <p>....When the impact of treatment beyond the results of the clinical trials is estimated, analyses that compare several alternative scenarios reflecting different assumptions about future treatment effects using different statistical models are desirable.</p> <p>These should include assuming that the treatment does not provide further benefit beyond the treatment period as well as more optimistic assumptions. (p35)</p>	<p>比較代案間の費用と結果の重要な差を反映できるほど十分に長くなければならない。観察期間以降で外挿する際に外挿仮定を異にした分析結果を提示</p>
<p>EUnethTA (2015)</p>	<p>it is recommended that the time horizon for the reference case analysis should be sufficiently long to reflect all relevant differences in costs or outcomes between the technologies being compared. The choice concerning any alternative time horizon for the reference case analysis should be clearly justified and described.</p>	<p>比較代案間の費用と結果の差を反映できるほど十分に長いこと</p>

ガイドラインそれ自体では、分析期間に対して似た指針を提示しているが、実際には同じ薬が互いに異なる国で互いに異なる分析期間で審議されたことを確認することができる。もちろん、製薬会社が自発的に互いに異なる分析期間を提示したものであるか、でなければ審議機関で他の分析期間を要求したのかは明らかではないが、このような差は、不確実性に対する各国の態度と関連したものと見られる。表10を見れば、同じ薬に対して、オーストラリア、カナダ、イギリスが最終選択した分析期間が異なることを確認することができる。

4) その他の文献の考察結果

既存の教科書や研究論文では、費用-効果分析の分析期間は、代案間の費用と効果の面で意味のある差を捕捉できるほど十分に長いものとし、多くの場合、この期間は lifetime であり得ると説明する (O'Mahony et al., 2015; Drummond et al., 2015; Neumann et al., 2017)。

Kim et al. (2017) は、2005 年で 2014 年間、Tufts Medical Center CEA Registry とアメリカで行われた支払者の観点を使用した費用-効果分析に対する体系的な考察を通して、研究特性と分析期間の間の関連性を調べた。各研究を分析期間が 5 年以下か、それを超えるか

によって短期と長期に区分して調べた結果、2013-2014年間では、対象研究の40%（235件中94件）が短期あるいは複数の分析期間を採択した。反面、2013年以前には、単に25%だけが短期観点（547件中135件）を採択した。そして、診断や予防接種、検診のような介入を評価した研究は、圧倒的多数（90%）が長期分析期間を採択した反面、施術や機器、手術を評価した研究の場合、64%が5年を超える分析期間を使用した。

短期と長期の分析期間をすべて報告した研究を分析した結果、多数の研究で分析期間が長くなるほど費用効果比が低くなる、すなわちより費用-効果的な結果が見られることが分かった。著者は、分析期間を延長するほど費用効果比が良くなる傾向が現れることは、費用は初期に集中的に発生して、健康便益は少しずつ発生する介入の典型的な特性のためであると解釈した。そして、短期分析期間を採択した理由を、観察範囲越しの結果に対する情報不足のためであると見た。しかし、著者は、分析期間の延長が費用効果比に相当な影響を及ぼす場合、感度分析でも長期分析の結果を提示して、ICERが分析期間にどのくらい影響を受けるのか、そして分析期間を延長する際に使用される外挿仮定にどのくらい依存するのかを示す必要があると主張した。

5) 利害関係者の意見

□ 製薬会社の意見調査

- 製薬会社別に共通した意見が提出されることはなかった。ただし、個別的に提出された意見を見れば、
- 類似薬剤の経済性評価資料で設定した期間があることから、申請薬剤に妥当ではない長い期間を設定することは妥当ではない。経済性評価モデル及び疾患の一般的な経過を考慮し、当該薬剤に妥当な分析期間を設定するようになければならない。
- lifetime分析を基本分析として、感度分析で疫学資料を参考にした分析を行うようにする。

□ 研究者の意見調査&専門家の諮問会議

- 研究者は、さらに具体的な指針が必要であるという意見を提示した。現指針では、「主な臨床結果を確認できるほど十分に長くすることを推奨」しているが、実際の検討過程でさらに短い分析期間を要求する場合もあることから、さらに具体的な指針を提示することが望ましいという意見であった。特に、「十分に長く」という表現に対する再検討が必要であるという意見であった。

6) 指針の改正方向の提言

- 現在の指針は、「主な臨床経過を確認できるほど十分に長い期間」を要求しており、基本的にこの方向は維持されるものである。
- しかし、現実ではシミュレーションを通して導出した対象人口集団の大部分が死

亡する時点を分析期間として設定する事例が多いので、導出された分析期間の妥当性に対する考慮が必要である。どんなモデルを適用するかにより、そしてどんな資料が投入されるかより、対象人口の大部分が死亡する時点は変わり、時に現実的ではない期間が導出されることもある。したがって、シミュレーション過程全般の不確実性を勘案することと共に、疫学資料など、他の外部要素を通して分析期間の適切性を検討する必要がある。

- 特に外挿された便益が全体の便益の相当部分を占める場合、その影響に対する考慮が必要である。分析期間による不確実性を考慮するようにして、観察資料に基づいた感度分析をさらに強いトーンで要求するようにする。
- また、もう一つ重要なことは、観察期間以降の効果についてどのような仮定をするかである。漠然と、観察期間中に確認された効果が、その後も持続するものであると仮定することは望ましくない。観察期間以降、さらに多くの効果が見られなかったり、一定の期間にわたって効果が縮小したり（waning effect）、観察された効果が以降も持続するものであるとする仮定に対して感度分析を実施し、その結果を提出する必要がある。
- 他の国のガイドラインとこれまでに提出された資料、経済性評価小委員会の審議内容を考慮する際、分析期間の決定時に次の事項を考慮する必要がある。
 - 適用人口の年齢、期待余命、疫学資料に基づいた生存率
 - 臨床試験で観察された生存期間の中央値、比較群との生存率差、観察期間
 - モデルに投入される母数の不確実性
 - その他の類似薬剤の評価事例や諸外国の評価事例、専門家の諮問などを考慮

ウ．分析対象人口集団

1) 現指針

ア) 主な指針

経済性評価の結果を提示する際には、費用と結果が導出された人口集団を明示し、分析対象人口集団から導出された結果が実際の研究結果を適用する人口集団（給付対象人口集団）で一般化することができるかの根拠を提示して論議する。

イ) 指針内の追加説明

...対象人口集団は、患者の疫学的特性（例：性、年齢、社会経済的な状態など）、随伴傷病（co-morbidities）や危険要因（risk factors）保有の有無、疾病の重症度（severity）などを利用して定義することができる。分析対象人口集団から給付対象人口集団への一般化の可能性について、臨床的、人口学的、疫学的特徴に基づいて論議する。

分析対象人口集団と給付対象人口集団の定義を明確にして、分析対象人口集団から給付対象人口集団への分析結果を一般化することができるかについて検討することを明示している。しかし、現実では、給付基準と分析対象人口集団が一致しない場合がしばしば発生し、これについて一般化の可能性を判断するさらに具体的な指針を提示する必要がある。

2) 既存の提出資料の検討結果

□ 一般化の可能性に対する検討

既に審評院に提出された 50 件の経済性評価資料と、これに対する小委員会の論議の結果を見れば、経済性評価に使用された臨床資料が包括する人口集団（分析対象人口集団）と給付対象となる人口集団（給付対象人口集団）が異なる場合があった。この場合、分析結果を給付対象人口集団として一般化することができるかについて、小委員会の判断が行われることとなる。これまでに経済性評価小委員会（以下、経評小委）の審議では、臨床試験に含まれた患者の人口学的特性や重症度などを給付範囲と比較して不一致の程度を確認したが、差が発生した原因をいくつかに分類すると次のとおりである。

- 最初から臨床試験自体が制限的に行われた場合：
 - 臨床試験が給付対象人口集団の多数の患者を含めて実施されたが、一部の少数患者に対する資料が不足した場合
 - 臨床試験の介入法/患者群が給付範囲と一致しない場合
- 給付基準が事後に決定されることにより、許可範囲と給付範囲が変わった場合：
 - 給付基準を通して他の治療に失敗した患者などに治療段階を制限した場合
 - 細部の患者群に給付範囲を制限した場合：給付基準設定において、細部の患者群の効果資料が必要であるが、使用可能な資料がない場合
- 製薬会社が許可範囲より範囲を狭めて分析して給付申請をした場合。

- 例えば、申請薬剤の効果が目立つ下位集団（重症度が高い集団）に対して分析し、給付を申請した場合。
- 一部の適応症に対する分析結果を他の適応症に一般化しようとする場合

経評小委では、最終給付範囲と分析対象人口集団が一致するかを検討し、一致しない場合、臨床諮問を基に不一致の程度と一般化の可能性を検討した。また、臨床試験内で細部集団別に効果と費用の発生傾向がどのように変化するかを確認し、一般化の可能性を裏付けることができる外部資料源があるかを確認した。また、適応症別に費用-効果性が検討されなければならないという原則を有しており、当該薬剤が収載された場合、給付範囲を逸脱して使用される可能性（以降の給付範囲の拡大申請の可能性）なども考慮した。

□ 細部集団分析

合計 50 件の申請中、10 件で細部集団分析を行った。大部分が事前に計画されたり給付基準により行われたものとして見られるが、細部集団分析を実施した根拠を提示しなかった場合も 2 件あった。

細部集団と全体集団の分析結果が異なる場合は、概して全体集団の分析結果を採択したが、給付基準により細部集団の分析が実施された場合は、給付基準による分析結果を採択した。経評小委では、細部集団より全体集団の分析結果を好んだ理由として、全体の集団分析が無作為性が維持された状態での観察結果であるためでもあり、今後の給付基準の拡大を予想しているものとして言及した。

□ Personalized medicineの場合

選別検査を経た後、その結果により薬物治療を実施する場合、分析対象を既に選別された人口集団として見るか、又は、選別前の人口集団として見るのかが問題となった。経評小委の審議過程、そして以降に行われたワークショップでは、全体の患者を対象に費用-効果性を評価しなければならないという結論を下した。当該個人のオーダーメイド医薬品が導入されていなかったとすれば、遺伝子検査なしに全体の患者群に標準治療剤を使用したものであるため、全体の患者群を対象に比較しなければならないということである。

3) 外国の指針

外国の指針では、共通して対象人口集団に対する詳細定義を提示するよう要求しており、細部集団別に費用と効果に有意な差がある場合、このような差が現れた理由に対する説明と共に、これを裏付ける資料を提示し、細部集団の分析結果を全体集団の分析結果と共に基本分析として提示するようにしていた。また、事後、資料を探す可能性を警戒した。

ア) カナダ

□ 対象人口集団

指針では、対象人口集団を明確にすることを要求している。そして、細部集団別に費用と効果の差が変わるのであれば、このような差を作る要因について明示すると共に、各細部集団の分析結果を基本分析として提示するようにしている。

□ 細部集団分析

細部集団は、ベースラインの人口学的特性（例：年齢、性、社会経済的な状態）や疾病の重症度、疾病段階、随伴傷病、危険要因、治療関連要因（例：地域社会あるいは病院のセッティング）、地理的位置、通常の服薬順応度（usual adherence rates）あるいは典型的治療パターンによって定義することができる。

指針では、経済性評価は、全体集団を反映しなければならないという点を明確にしているが、細部集団別の異質性を引き起こす要因についても点検することを強調している。また、細部集団分析で留意する事項についても比較的具体的に提示している。

カナダの指針では、細部集団間の差に対する頑強（robust）な根拠を提供する資料があるかを確認するよう勧告するが、これは特に事後に資料を探す（dredging）可能性を避けるために重要であると強調している。このために、投入資料（data inputs）での差に対する表面的な妥当性（face validity）の評価を推奨している。

また、細部集団別の層化分析が行われたが、意思決定者が細部集団別に異なる意思決定をできないのであれば、最終結果は、全体集団を対象とした平均結果（例：ICER）でなく、各細部集団別の推定値をそれぞれの有病率を加重値として決定することが適切であると提示した。

イ) オーストラリア

□ 対象人口集団

オーストラリアの指針では診断、症状、予後、その他のイシューに対する詳細事項を含めて申請薬で治療できる疾病あるいは状態に対する概要を提示することを要求する。

また、当該適応症を持つ全体患者ではない一部の細部集団に使用を制限して申請するのであれば、その細部集団と全体集団が、疾病の通常的な経路や使用可能な治療オプションの面で異なるかどうかを確認しなければならないという。

資料は、オーストラリア人を含むことを選好し、「医療利用量算出のための資料源（Sources of data for use in generating utilisation estimates）」で羅列されたような公信力のある資料源からの資料を利用して、オーストラリアで当該疾病の発生率と有病率を提示することを要求する。

オーストラリアの参加者を含む資料源がない場合、ここで提示された人口特性がオーストラリアの環境を代表できるかを論じ、可能であれば、この資料に対して比重 (percentage) と平均、そして四分位範囲、標準偏差、範囲のような不確実性の推定値を含めるようにとも言及している。

□ 細部集団分析

○ 細部集団使用の正当化

PBAC は、無作為割付臨床試験が包括する全体集団に基づいた申請書を選好する。万一、臨床試験内の特定細部集団に対して医薬品の収載申請をしようとする場合、なぜ臨床試験で細部集団ではなく、さらに広範囲な人口集団を登録させたのか、なぜ当該細部集団以外の残りの集団には申請薬を提供してはならないのかを明確にするよう要求する。

○ 細部集団分析をする場合、次の情報を提供することを要求する。

- 細部集団別の治療効果の差を説明する薬理学的、生物学的あるいは臨床的根拠 (rationale)。
- 細部集団分析があらかじめ指定されているか、無作為割付が細部集団別に層化されているかを確認できる臨床試験計画書の関連パート。細部集団を定義するために使用されたしきい値についても明瞭に正当化すること
- もともと実施された細部集団分析の数、そして多重比較のための統計的な補正

○ 細部集団分析の結果

- 結果の指標別に全体の臨床試験対象人口、細部集団、そして当該細部集団を除外した残りの集団に対して、それぞれ相対的、絶対的な効果測定値を提示すること (例示表も提供)。
- 治療効果と細部集団を定義する共変量間の交互作用の評価。
- 万一、細部集団が連続変数によって定義されるのであれば、特に細部集団があらかじめ定義されていなければ、細部集団を定義するために選択されたしきい値に対して感度分析を提示すること。
- 可能であれば、統合資料に対して random effect meta-analysis を利用すること。

ウ) イギリス

□ 対象人口集団

イギリスは、評価範囲の設定 (scoping) 段階で対象人口集団 (target population) が明示される。このとき、対象人口に対する情報は具体的に提示し、臨床効果や費用-効果が全体の対象人口集団と異なる可能性がある細部集団があれば、これについても評価範囲の設定 (scope) 段階で確認しなければならない。

□ 細部集団分析

細部集団分析が必要な場合、細部集団別の効果と費用効果推定値を基本分析として提示するが、細部集団を構成する患者特性を明確に定義し、細部集団別の差を裏付けできる妥当な生物学的な機序、社会的な特性、その他に他の正当化できる要因があれば共に提示するものとしている。他の国と同様に、細部集団効果を求めて事後的に資料を探す (dredging)

ことは避けるように強調している。また、特定の健康結果のベースのリスクの差により細部集団を区分する場合、ベースの資料を確認する方法を詳細に提供することが重要であるとも言及している。

細部集団分析を行う場合、その方法を詳細に記述し、細部集団の推定値の正確度を不確実性分析に反映するようにしている。また、多数の細部集団が報告される時、偶然に細部集団間に差が現れる可能性が高いという点にも言及し、研究計画段階で細部集団をあらかじめ指定（Pre-specification）することで、細部集団分析の信頼度を増加させることができると強調する。

細部集団別の母数推定のためには、個別の患者資料を使用することが選好されるが、分析の質や使用可能な根拠が臨床診療をどのくらいよく代表するか、そして意思決定問題にどのくらい関連性があるかという点などを考慮して、そのような資料の適切性を評価するよう明示している。

エ) その他の指針

EUnetHTA 報告書で提示されたパートナー国の細部集団分析に対する指針は、次の表のとおりである。大部分が細部集団分析に対する指針を持っていることを確認することができる。2019年にEUnetHTAが整理して回覧中の「経済性評価時に考慮する事項」の中にも細部集団分析に関連した勧告案を見ることができるが、細部集団の区分、細部集団分析時に留意する事項などに対する言及が含まれている。また、事前計画されずに事後に行われた細部集団分析の限界についても言及している。

<表13> EUnetHTAパートナー国の分析：細部集団分析

国	細部集団分析に対する指針があるか？
オーストリア	Yes. Recommended for populations with high heterogeneity
ベルギー	Yes. If the intervention's safety, effectiveness, costs and/or baseline risk for events differ between subgroups, separate subgroup analyses should be performed.
クロアチア	Yes. Estimates of clinical and cost effectiveness separately for each relevant subgroup of patients
チェコ	Subgroup data may be presented additionally in case of potentially important differences in clinical effectiveness or costs
デンマーク	Na
イギリス (England)	Yes. Subgroup analyses should be presented separately for each relevant subgroup where appropriate.
エストニア&ラトビア	Yes. Subgroup data may be presented additionally in case of potentially important differences in clinical effectiveness or costs
フィンランド	Yes. A separate evaluation should be prepared for each indication
フランス	Yes. Subgroups analyses may be necessary in case of documented heterogeneity of the health effects or the costs
ドイツ	Yes. There can be subgroup analyses, which need to be documented
ハンガリー	Yes. When the clinical effectiveness or cost-effectiveness in particular patient subgroups differ significantly
アイルランド	Yes. Stratified analysis of sub-groups is appropriate when there is biological or clinical support for heterogeneity in the target population
イタリア	Yes. Subgroup analyses should be derived from proven differences in the parameters
オランダ	Yes. With respect to assumptions in the discount rate, unit costs, subgroups, patient characteristics and possible model structures, it is possible to conduct an extra

	analysis.
ノルウェー	Yes. When/if the intervention is expected to differ significantly in cost and/or efficacy for different groups
ポーランド	Yes. If the analysis of subgroups has been carried out, the cost-effectiveness in the sub-group should be indicated in comparison to the total population.
ポルトガル	Yes. The target population can be divided into subgroups
ロシア	Na
イギリス (Scotland)	Yes. A clear definition of subgroup analysis (when appropriate) and a justification of a differential effect within patient subgroups are required
スロバキア	Yes. Subgroup analysis should be performed
スロベニア	Na
スペイン	Yes. Use data that will determine whether differences in age, gender, disease severity, and risk factors have a significant impact on either effectiveness or costs (Spanish recommendations, Osteba). If there is clinical evidence that there are differences between subgroups, the results should be analyzed separately for these different subgroups (CatSalut).
スウェーデン	Yes. Separate calculations should be made for different patient groups where the treatment is expected to have different cost-effectiveness
スイス	Na

出典: EUnetHTA (2015a)
Na: No information available

<表14> 対象人口集団に関連する海外の指針内容

国	指針内容原文	要約
カナダ CADTH 指針 (4版) (2017)	In the reference case, the target population(s) for the intervention and its expected use should be specified, and should be consistent with the decision problem... A stratified analysis with results presented for each subgroup should be provided in the reference case if factors are identified to support the consideration of distinct subgroups. Otherwise, the analysis should be for the entire target population. (p18)	目標集団は、意思決定問題と一致しなければならない。 細部集団分析の結果は、細部集団間の差を引き起こす要因と共に分析結果を基本分析として提示

	<p>Describe the setting of the model and the demographic and patient characteristics for the modelled population. Translate the important applicability concerns associated with the clinical data and identify any remaining uncertainty. (p71)</p> <p>Provide an overview of the disease or condition that can be treated by the proposed medicine.</p> <p>If the medicine is proposed for use in a subgroup(s) of the Australian population with the disease or condition, indicate whether the usual course of the disease or condition – or the available treatment options for that subgroup(s) – differs from that of the whole population.</p> <p>Describe the Australian population who would be treated with the proposed medicine, Provide data (preferably include Australian datasets or studies involving Australian participants). Summarise the incidence and prevalence of the disease or condition in Australia using data from a reputable source Where data sources involving Australian participants are not available, discuss whether population characteristics presented here are likely to be representative of the Australian setting. (12-13)</p>	<p>モデルに含まれた人口集団の特性について記述、適用の可能性の問題も探求。</p> <p>薬が対象適応症を持つ患者の一部に使用することが提案された場合、全体集団と当該細部集団の差に対して記述</p> <p>オーストラリアの患者の特性に対して記述</p>
イギリスNICE指針(2013)	<p>The scope defines the population for whom the technology is being appraised as precisely as possible. When the technology is a medicine, the marketing authorisation will generally specify the therapeutic indications. The scope may highlight potential subgroups of the population for whom the clinical or cost effectiveness of the technology might be expected to differ from the overall population, or who require special consideration. (p17)</p>	<p>評価範囲の設定段階で目標集団と細部集団を決定</p>
EUnetHTA(2015)	<p>it is recommended to perform subgroup analyses in the economic analysis when there is a clinical rationale to believe that the cost-effectiveness of the assessed technologies may vary between sub-groups. It is important that the choice of subgroups is clearly justified and described.</p>	<p>細部集団間で費用-効果に差があるとき、細部集団の分析を行う。</p> <p>細部集団の選択の正当化</p>

4) 利害関係者の意見

この項目については、利害関係者から特別な意見は陳述されなかった。ただし、対象人口集団の不一致問題は、比較対象の選定とも直結する問題なので、比較対象の選定と関連しては様々な意見が提示された。

5) 指針の改正方向の提言

- 人口学的特性、重症度、その他の臨床的特性の面から給付対象人口集団と分析対象人口集団が一致するかを確認し、比較表を提示するようにする。万一、一致しない場合は、このような差が結果に及ぼす影響を検討するようにする。
- 給付範囲が全体患者の細部集団に制限された場合は、全体集団の分析結果と細部集団の分析結果を別途に提示し、細部集団別に費用と効果がどのように変わるかを確認するようにする。また、収載後に給付範囲以外の使用が行われる可能性についてもあわせて検討するようにする。
- 細部集団分析に対する具体指針
 - 細部集団分析をする際には、細部集団を構成する患者の特性を具体的に提示する。
 - 細部集団を分けた根拠（細部集団別に治療効果や費用、効用が変わるという妥当な生物学的な機序、臨床的な根拠、社会的な根拠など）と根拠の強度を提示する。連続変数を基準として細部集団を定義した場合、細部集団の定義に使用されたしきい値を正当化して感度分析を実施する。
 - 細部集団分析の結果を提示する際には、細部集団分析が（臨床試験段階で）あらかじめ計画されたものか、あるいは事後の基準設定により細部集団分析をすることになったのかをあわせて提示するようにする。多数の細部集団分析をしてみると、偶然に細部集団間の差が現れる可能性があるためである。前者の場合、臨床試験計画書などの事前計画を裏付けできる根拠を提示する。また、事前計画された細部集団分析の結果は、すべて報告するようにする。
 - 細部集団分析をした場合、全体集団分析の結果と細部集団分析の結果を共に提示し、細部集団分析の結果についても不確実性評価の結果を提示するようにする。

エ . 分析技法

1) 現指針

ア) 主な指針

比較対象薬と収載申請薬の効果が同等であるということを証明できれば費用最小化分析を実施し、比較対象薬と収載申請薬の効果が互いに異なるのであれば、費用-効果分析あるいは費用-効用分析を実施する。

治療にあたり、生活の質が重要であったり、健康結果が様々な指標として提示されて、一つの指標で反映しにくい場合には、費用-効用分析を推奨する。このとき、QALY (質調整生存年、Quality-Adjusted Life Years) を費用-効用分析の指標として使用する。万一、生活の質が重要ではない指標や QALY の使用が適切でないのであれば、費用-効果分析を推奨する。分析者は、選択した分析技法の根拠を提示する。費用-効果分析を行う場合、可能であれば、最終指標を利用することが推奨され、中間結果を利用する場合は、本指針「2-4-1 結果指標」の項目に基づいて報告する。

イ) 指針内の追加説明

...費用-効果分析を行う場合、可能であれば、最終指標を使用して効果の一単位の増加がどのような臨床的価値、経済的価値を持つのかについて十分に記述するようにする (2-4-1 結果指標参考) 。

...QALY は、概念の限界、公平性反映の限界、仮定の限界などを持っているが、まだ QALY を凌駕するほどの適切な代案が存在しないという現実のために、様々なガイドラインで QALY の使用を勧告している。本ガイドラインもまた、異なる研究結果の比較が容易であるという点から、QALY を費用-効用分析の結果指標として使用することを提案する。

...費用-便益分析のみを実施することは、現在としては望ましくない。

現指針では、生活の質が重要であったり、健康結果が様々な指標で提示される場合、費用-効用分析を推奨するという表現を使用しているが、他の分析技法に対する説明と併記することによって、費用-効用分析を優先的な分析技法として考慮するという点が明確に伝えられていない。また、比較代案間の効果が同等であるということを証明できるとき、費用-最小化分析を実施するよう言及しているが、効果の同等性に対する具体的な判断基準や分析方法などに対する具体的な指針がなく、この点が補強される必要がある。

2) 既存の提出資料の検討結果

既に経済性評価資料を提出した50品目のうちでは、2件が費用-最小化分析を実施した。このうち1つの品目は、間接比較を通して代理アウトカム (surrogate outcome) の非劣性さを見せた後、費用-最小化分析を実施し、また他の件は優越性の検証を目的に設計された直接比較臨床で有意な差を示すことができず、事後的に非劣性マージンに対する体系的な文献考察を実施し、これを直接比較臨床の結果に適用して非劣性を主張した場合であった。費用-効果分析を実施した場合は1件で、EFS (event-free survival) が代理アウトカム指標として使用された。評価過程で根拠の不確実性が議論になったが、使用可能な資料の限界などを考慮して受容された。

<表15> 提出資料の分析技法

分析技法	品目数	備考
費用-最小化分析	2	
費用-効果分析	1	結果指標：EFS
費用-効用分析	47	このうち1件は今後、費用-最小化分析に切り替えた
費用-便益分析	0	
計	50	

上の三品目を除いた残りの品目は、すべて QALY を結果指標として費用-効用分析を実施した。ただし、この中には当初から費用-効用分析の結果を提出したが、委員会で効果の改善をめぐる不確実性が大きいと判断して、今後の費用最小化分析に変更された 1 件が含まれている。費用-効果分析あるいは費用-効用分析を行う場合、先に臨床的な有用性の改善を立証しなければならない。特に費用-効用分析は、生活の質や生存期間延長の側面での効果を前提とする。提出の件の中では、臨床試験を通して人生の量や健康に関連する生活の質 (Health-Related Quality of Life, HRQoL) の改善を直接立証した場合もあったが、代理指標の改善を見せた後、既存の文献値などを活用して QALY の変化を推定した場合も多かった。

比較薬と比較した分析資料を提出しなかったり、効果指標が適切でなかった場合、そして主な効果指標に対する改善の有無が不明であった場合は、経評小委の審議過程で臨床的な有用性の改善を認めなかった。例えば、骨粗しょう症治療剤である薬 A は、BMD を代理指標として使用して最終成果を推定したが、経評小委の審議過程で (臨床諮問結果などを基に) BMD でない股関節の骨折回数を結果指標として使用することが勧告された。薬 B は、根拠文献に含まれた患者が少数であり、資料源及び対象患者により効果 (疾病進行期間、無進行期間など) の変移が大きく、結果値の不確実性が大きいという点、そして反応率の改善が最終指標 (PFS) に及ぼす影響が確認できない点などにより、改善主張が受

容されなかった。

効果の他に副作用を検討した品目は、合計 50 品目のうち 21 品目であった。

3) 外国の指針

国別に、指針で費用-効果分析と費用-効用分析を分離して言及した場合もあり、一つで言及した場合もあったが、概して QALY を結果指標とした費用-効用分析を基本分析として提示するようにしていた。費用-最小化分析の場合、主な3か国中では、オーストラリアだけが別途指針を提示し、カナダは改正指針で費用-最小化分析を推奨しないという立場を明示している。しかし、EUnetHTA (2015) でパートナー国の指針を整理したものによれば、依然として費用-最小化分析を分析類型の一つとして提示する国が多数あり、費用-結果分析 (cost-consequence analysis; CCA) を分析の一類型として提示した国もいくつもあった。国別指針の内容は、次のとおりである。

ア) カナダ

カナダの指針では、費用-効用分析を適切な評価方法として提示している。費用-効用分析をした後、付加的に費用-効果分析の結果を提示することは可能であるが、基本的には費用-効用分析を提示しなければならぬと明示している。また、たとえ主な結果指標の面で代案が互いに同一であるということを示す根拠があるとしても、費用-最小化分析は推奨しないことを明らかにしている。

イ) オーストラリア

オーストラリアの指針では、費用-効果分析と費用-効用分析を区分する表現を使用せず、経済性評価を完全費用-効果分析 (full cost-effectiveness analysis; full CEA) と費用最小化分析に区分した後、別途の項目を通して分析指針を提示している。

オーストラリアの指針によれば、臨床的評価結果の申請薬が、主な比較代案に比べて治療的にさらに優秀であるが、費用もまたさらに必要とされるとき、あるいは比較代案に比べて治療的に劣性であるが、費用がさらに少なく必要とされるとき、full CEA が適切で、治療的非劣性 (あるいは優秀性) があつたり、安全性プロファイルが同等であつたり優秀なとき (性格や大きさのすべての面で)、そして申請薬の使用が保健システムにおいて同等であつたり、さらに少ない費用を発生させるものと期待されるとき、費用最小化分析が適切であると言及している。

□ 費用-最小化分析 (オーストラリアのガイドラインでは別sectionで提示)

オーストラリアの指針で費用最小化分析を認める場合は、効果と安全性がすべて非劣性（あるいは優越）しているものと認められるときである。効果は非劣性であっても、副作用プロファイルが全く異なるとしたら、費用-効果分析や費用-効用分析がさらに適切になるということである。

また、費用-最小化分析に主に含まれる費用を i) 申請薬と比較薬を処方したり投薬するのに必要とされた費用、ii) 申請薬と比較薬に関連した副作用をモニターしたり管理するのに必要とされる費用、iii) その他のリソースの使用に影響を及ぼしかねないものなどに分類し、費用-最小化分析の結果を提出する際には、次に対する情報をあわせて提供するようにしている

- 費用最小化分析の核心要素と仮定に対する要約
- 同等効果の用量の推定：使用過程で用量変更をすることになる場合は、「安定状態」の用量（用量 titration が完了した後、そして中断患者を排除した後の平均用量）の比較が適切。資料が製品許可情報で勧告するところ（用量、titration 方法）と一致するか確認
- 追加費用そして/あるいは費用削減：申請薬と比較薬の投薬プロファイルを比較して差を確認し、このような差が追加的な費用や費用削減につながるかを提示
- 結果：費用-最小化分析の結果を提示；関連文書と原資料源の写本を添付し、本文中で相互参照すること

<表16> オーストラリアのガイドラインで提示された費用-最小化分析の核心仮定と要素

要素	主張あるいは仮定
治療的主張：効果	効果は[非劣性/優越]であるものと仮定
治療的主張：安全性	安全性は[非劣性/優越]であるものと仮定
根拠	[直接無作為比較臨床試験/無作為臨床試験の間接比較]
同等-効果用量 (Equi-effective doses)	申請薬[dose/day/course記述]と比較薬[dose/day/course記述]
直接医薬品の費用	[低い/同等/高い]; [申請薬の費用] vs [比較薬の費用] (費用は還付性あるいは自己制限がある治療の場合、患者当たりcourse当たりとして、慢性あるいは持続治療の場合、患者当たりの年間費用)
他の費用あるいは費用削減分	[はい/いいえ] [「はい」の場合；簡単に記述 - 例：副作用関連費用、モニタリング費用、投薬費用..]

ウ) イギリス

イギリスの指針では、基本分析 (reference case) として費用-効用分析を選好し、健康効果は QALYs で表現することを明示的に要求している (費用-効果分析であると表現しているが、内容、上費用-効用分析を指す)。ただし、QALY の底辺にある仮定 (例えば、constant proportional trade-off と健康状態の間の additive independence) が不適切な場合、二

者択一の測定値を利用した分析を non-reference-case analysis として付加的に提示することができると言及している。

エ) その他の指針

EUnetHTA 報告書で提示された国別に選好する分析類型は次の表のとおりで、これを基に EUnetHTA では費用-効果分析と費用-効用分析を勧告するが、代案間の差がなければ、費用-最小化分析で十分なこともあると言及している。

<表17> EUnetHTAパートナー国の分析：選好する分析類型

国	選好する分析類型
オーストリア	No preferred type
ベルギー	CUA, CEA or CMA.
クロアチア	CUA or CEA
チェコ	CUA
デンマーク	Not explicitly stated. CEA and CUA seem to be accepted
イギリス (England)	CUA (Technology Appraisals and NICE Diagnostics Assessment Programme) CCA (NICE Medical Technologies Evaluation Programme Methods Guide)
エストニア&ラトビア	CEA or CMA
フィンランド	CUA, CEA, CMA or CBA
フランス	CUA and CEA
ドイツ	CEA (several endpoints= several efficiency frontiers)
ハンガリー	CUA, CEA or CMA,
アイルランド	CUA or CEA
イタリア	CUA or CEA
オランダ	CUA, CEA or CMA
ノルウェー	CUA, CEA or CMA
ポーランド	CUA (preferred according to the regulation), CEA or CMA and a CCA. CBA is possible only as an additional analysis (according to the guidelines).
ポルトガル	CUA, CEA, CMA or CBA (CUA is preferred)
ロシア	CEA or CMA (Ministry of health) CMA, CEA, CUA or CBA (ISPOR Russian HTA Chapter)
イギリス (Scotland)	CUA or CMA
スロバキア	CUA, CEA or CMA
スロベニア	CUA, CEA or CMA and Cost Analysis
スペイン	CUA, CEA, CMA or CBA. CUA is preferred. (Spanish recommendations, Osteba) CUA, CEA or CMA (AETSA) CUA or CMA (CEA only if a CUA cannot be conducted) (CatSalut)
スウェーデン	CUA, CEA, CMA or CBA
スイス	CEA

出典: EUnetHTA(2015a)

CBA: Cost-benefit analysis, CCA: Cost-consequence analysis, CEA: Cost-effectiveness analysis, CUA: Cost-utility analysis, CMA: Cost-minimization analysis.

<表18> 分析技法に関連する海外の指針内容

国	指針内容原文	要約
カナダ CADTH 指針（4版） （2017）	In the reference case, the economic evaluation should be a cost-utility analysis (CUA) with outcomes expressed as quality-adjusted life-years (QALYs). Any departure from this approach should be clearly justified. (p17)	CUA
オーストラリア PBAC 指針version 5 （2016）	State whether cost-effectiveness will be estimated using a CEA and/or a cost-utility analysis (CUA). Identify the incremental health outcomes (as nominated for the CEA, or as quality-adjusted life years [QALYs] for a CUA) and incremental health costs. Other economic evaluations (eg cost-benefit analyses or cost-consequences analyses) should not be presented as base-case analyses (p63)	CEA or CUA.
イギリス NICE 指針 （2013）	For the reference case, cost-effectiveness (specifically cost-utility) analysis is the preferred form of economic evaluation. (p93) Health effects should be expressed in terms of QALYs. (p94)	CUA
EUnetHTA (2015)	To enhance the usability of the economic evaluations, it is recommended that results be presented in terms of both a cost-effectiveness analysis (CEA) and a cost-utility analysis (CUA). A costminimization analysis (CMA) is sufficient when it is demonstrated that there is no difference in effect between an intervention and its relevant comparators. If appropriate and adequately justified, a cost-consequence analysis (CCA) may be a useful alternative in cases where CEA and CUA cannot be undertaken.	

4) その他の文献の考察結果

分析類型に関連して文献で主に扱われたものは、費用-最小化分析の適切性に対する検討であった。費用-最小化分析無用論を主張した代表的な論者である Briggs & O'Brien (2001) は、費用と効果をめぐる不確実性のために、多くの場合には費用-最小化分析が不適切であると述べた。根拠の無いことが無いことの根拠ではないので（‘absence of evidence is not evidence of absence’）、治療の同等さを示すために（費用や効果の面で）特別に設計された研究でなければ、治療法間の効果や費用の差において有意ではないものとして観察されたという事実に基づいて費用最小化や成果最大化のような類型の分析を行うことは不適切であるということが彼らの説明である。Drummond et al. (2015) また、Briggs & O'Brien (2001) を引用して、費用と効果をめぐる不確実性のために、費用-最小化分析はあらかじめ決定できる固有の研究設計ではないという点を強調した。これらは、費用-最小化分析を適用できる唯一の場合は、以前の研究結果や専門家の判断に基づいて二つの選択肢が効果の面で同等であることを事前に判断できる場合であるが、これは非常に制限的なものとして判断した（例：同一 pharmacological class の薬）。費用-最小化分析の制限性に対する言及は、他の国の経済性評価指針にもよく登場し、これは、初期とは異なり、最近では費用-最小化分析を別途の経済性評価類型として分類しない場合が多い。

5) 利害関係者の意見

特別な意見は陳述されなかった。

6) 指針の改正方向の提言

- 費用-効用分析を選好することを明確にする。費用-効用分析が難しい場合、費用-効果分析も可能になり得るが、なぜ費用-効用分析が不適切、あるいは不可能なのかに対する明確な説明が提示されなければならない。
 - 費用-効果分析時に最終結果指標を使用し、効果の1単位改善の経済的な意味を共に提示し、意思決定に参照することができるようにする。
- 費用-最小化分析と関連して1案は、費用-最小化分析を提出できる対象、費用-最小化分析を行う際に考慮する費用項目と提出資料について、さらに具体的な指針を提示するものである。
 - 費用-最小化分析対象：効果が非劣性（あるいは優秀）の場合 & 安全性プロファイルを検討し、安全性の面で申請薬が劣性ではないことが認められる場合（非劣性検証過程を経なければならないという意味ではない）
 - 費用推定：同等効果用量（equi-effective doses）を決定した後、比較代案の投薬関連費用を提示、モニタリング費用、副作用費用、その他の関連費用を共に提示。
 - 各費用項目を提示する際には、これを裏付ける根拠を共に提示
 - 感度分析結果の提示
- 2案は、費用-最小化分析を分析技法から除外するものである。これは、効果の同等性、不確実性の問題をめぐる理論的な論議を反映した選択である。

オ．比較対象の選定

1) 現指針

ア) 主な指針

比較対象を選定するにあたり、比較するに足りる収載医薬品がある場合は、これらの中で最も多く使用されるものを比較対象とする。場合により、一つの医薬品だけでなく複数の医薬品と比較することもでき、比較するに足りる医薬品がない場合は、手術などの他の治療方法も比較対象とすることができる。

どの比較対象を選定したとしても、選定事由については報告書で十分に言及する。すなわち、当該疾病あるいは症状に使用できる医薬品や治療法のうち、申請薬に代わって使用できるものとしてはどのようなものがあり、なぜその中の特定医薬品（あるいは治療法）を比較対象として選定したのかなどを詳細に記述する。比較対象として選定された薬に対しては、用法、用量、剤形、適応症、副作用、治療期間、その他、一次薬剤としての使用が認められるかどうかなどを記述する。

比較対象として選定された薬の単位費用と関連して、同一成分の他の医薬品が収載されていない単独品目の場合は当該薬の価格を、同一成分の医薬品が多数ある場合は、市場占有率を利用した加重平均値を比較薬の単位費用として利用する。

イ) 指針内の追加説明

...原則的に ICER を計算する際には、他の代案に比べて非効率的なものとして評価された代案は除き、与えられたしきい値下で費用-効果的なものとしてみなされる代案と比較することが望ましいが...実際には、既収載薬のうち、どれが費用-効果的な薬なのか分からない場合が大部分である。したがって、多くのガイドラインでは比較可能な既収載医薬品の中で現在広く使用されているものを比較対象として勧告している。

現在広く使用されている医薬品とは、比較対象として適切な医薬品のうち、市場占有率が最も高いものをいうが、場合によっては市場占有率が高い一つの医薬品を選定することが困難な場合がある。例えば、同一効能群に含まれる 10 種類の薬のうち、比較的市場占有率が高い三つの薬、X、Y、Z の市場占有率がそれぞれ 20%、25%、23%であるとき、この中で市場占有率が最も高いのは Y であるが、X や Z もまた市場占有率の面で Y と大きく変わらない。したがって、この場合には、Y とだけ比較するより、X、Y、Z とすべて比較をすることが適切なこともある。

比較対象薬を選定するにあたり、市場占有率よりさらに重要なことは、比較対象として適切かどうかに対する判断である。たとえ市場占有率が高いとしても、対象患者群が申請薬と異なる場合は、比較対象として選定することが不適切なこともある。

万一、比較するに足りる医薬品がない場合は、手術やその他の異なる治療方法も対案になり得るが、どの比較対象を選定したとしても、選定過程と選択事由については、報告書で十分に言及すべきである。すなわち、既収載医薬品のうち、申請薬と比較可能な医薬品あるいは治療方法としてはどのようなものがあるかを具体的に羅列し、その中の特定医薬品、特定治療方法を比較対象として選定した事由を記述する...

ここで、比較可能な医薬品とは、申請薬が目標とする対象疾患あるいは症状を治療・管理する目的で今まで使用されてきたものとして、申請薬の給付に伴い代替が起こり得る医薬品を指すが、比較可能であるということが必ずしも作用機序の同一を意味するものではない。比較対象として選定された医薬品に対しては、用法、用量、剤形、適応症、副作用、治療期間、その他、一次薬としての使用が認められるかどうかなどの事項を記述する。

一旦、比較対象薬が決まれば、当該薬の費用と効果を推定するべきであるが、後発医薬品が収載されていなかった単独品目の場合は、その製品の価格がすなわち比較薬の単位費用になるが、後発医薬品が既に進入していて、その価格が選定された医薬品よりさらに低ければ、比較薬の単位費用として.....市場占有率を利用した加重平均値を利用することを勧告する...これは新しい薬が進入する場合、実際の代替が起きる薬の価格を反映するという点からも、それなりの根拠を持つ値と言える。

現指針では、比較可能な薬のうち、最も市場占有率が高い薬を比較代案として選定するよう提示している。市場占有率が類似の薬がいくつもある場合、複数の比較代案を選定することができ、比較可能な医薬品がない場合、手術などの他の治療法も比較代案になり得ることに言及している。指針で一貫して強調することは、実際の臨床での使用現況であるが、比較可能であることの判断基準も臨床での使用現況であり、作用機序が同一であることなどによって、比較範囲を制限しないことを明確にしている。また、市場占有率に対する比較的詳細で具体的な言及をしており、市場占有率が最も高い薬ではない場合、比較代案として考慮できる余地はほとんどないという特徴を持っている。これは、指針の明瞭さは高められるが、市場の動的な特性、根拠の質などを考慮したとき、考慮可能な比較代案を排除することになるという限界がある。

2) 既存の提出資料の検討結果

既に提出した資料で使用された比較代案を類型別に区分すれば、次の表のとおりである。比較可能な薬の範囲が、同一機序という狭い範囲に縛られていないことを確認でき、時にBSCが比較代案となったりもした。比較可能な薬の範囲が決定されれば、大部分の市場占有率を基準として比較代案が選定された。

<表19> 比較代案の種類の整理

区分	申請薬
比較薬は同一機序（ATC code 4桁基準）に属するか？	
同一機序	8品目
同一機序ではない	39品目
ATC code表示なし	3品目
新しい機序の薬か？	
はい	14品目
いいえ	23品目
BSCを比較代案として使用したか	
はい	3品目

また、今までは許可範囲の超過あるいは非給付薬は比較代案として考慮しておらず、選定された比較代案が劣性な代案（extended dominance 含む）として確認された場合、比較対象から除外した。

既存の治療法に失敗したが、他に代案がなく、既存の治療法を継続して使用する場合、申請薬の比較代案をプラセボと見るか既存の治療法として見るかに関しては、効果不十分でも継続して先行治療剤を使用していれば（別途の中断基準なし）、不応の先行治療剤という比較代案から排除することはないということが、今までの経評小委の判断であった。ただし、許可事項又は給付基準で中断を明示する場合であれば、比較代案から排除した。

標準治療あるいは通常の治療法のような単独治療法ではなく、様々な治療法で比較代案を構成する場合、臨床資料に基づいて比較代案を構成するのか、国内の使用現況を基にして比較代案を構成するのも争点となった。前者は分子（費用）と分母（効果）が同一資料源から導出されるという長所はあるが、国内の使用現実とは異なる問題点があり、後者は分子と分母が一致しない問題点がある。申請件の中では、国内臨床医のパネルを対象に調査した結果を基に BAT（best available therapy）を構成した件が2件あった（パネルの回答結果の平均値を適用）。これに対して、経評小委ワークショップでは、実際の処方事例と臨床試験での比較代案の構成を比較して適切性を検討する必要がある、パネル間の回答の偏差が大きい場合、平均値より中間値の使用が適切であるという検討意見を提示している。

選別検査後、陽性と確認された患者に申請薬を投与する治療戦略の場合、通常と比較代案の選定過程とは異なり、薬対薬ではない戦略対戦略の比較を必要とする。患者のオーダーメイド治療においてよく比較される戦略は、Test and Treat 治療戦略対 Treat all 治療戦略で、この場合、代案は次のように構成される。

代案1：検査後の陽性患者には、申請薬、陰性患者には標準治療剤を投与

代案2：すべての患者に現行治療剤を投与

経評小委のワークショップでは、これまでの審議結果を基に診断検査後の治療対象を選別する場合であれば、検査費用も含めて申請薬の費用-効果性を評価することが望ましいとの結論を下した。ただし、治療対象を選別する検査に該当するとしても、申請薬投与の有無にかかわらず行われる検査であれば費用から除外し、申請薬投与のために追加で行わ

れる検査であれば費用に含めることが望ましいとも言及した。

3) 外国の指針

他の国の指針を見ると、大部分が、何が現在当該適応症に最も広く使用されているかを基準に代案を選択していた。ただし、カナダの場合、新技術の進入の影響を受けるすべての潜在的代案を確認することを強調しており、オーストラリアの場合、比較代案選択のヒエラルキーを提示しているという点が特異である。また、他の国とは異なり、オーストラリアは多くの場合、比較代案が単独代案になるという点に言及している。しかし、実際のオーストラリアの評価事例を見れば、PBACの審議過程で追加的な比較代案を提示する場合もしばしば見られた。

ア) カナダ

□ 比較代案の確認

カナダは、比較代案を選定するにあたり、現在使用されていて（新技術が進入すれば）潜在的に代替されるすべての介入、そして近い未来に使用可能な介入まですべて確認することを要求する。そして、これには個別介入だけでなく介入戦略、管理までを含み、現在使用されている技術がBSCに比べて低いあるいは不確かな価値を持つと判断されるときには、BSCを比較代案として考慮することができると言及している。

□ 比較代案の選択

比較代案の包括的リストから適切な比較代案を選択する出発点は、何が現在の臨床診療（current care）を代表するか、あるいはどのような技術が代替される可能性が高いかを定めることである。指針では、比較代案は、現在給付になっていてよく使用されるものであるべきで、資料の可用性によって比較代案が決定されてはならないと言及している。

多数の適切な比較代案があるとき、比較代案の選択が複雑なこともある。指針では、この場合、すべての比較代案が考慮されなければならない、ある代案を除外することを選択したとすれば、なぜそのような選択をしたのかを正当化すべきであると明示している。

□ 比較代案の記述

最終選定された比較代案については、用法、用量、投与経路、他の治療法との併用の有無、臨床経路上での位置などを含む詳細事項を記述し、関連費用と結果を確認できるようにすべきであると明示している。

イ) オーストラリア

□ 比較代案の選択

オーストラリアの指針では、現在オーストラリアで使用する治療法として、申請薬が進出した際に代替される可能性が最も高いものを比較代案として選択すべきであると提示している。多くの場合には、単一比較代案が適切である点にも言及している。

そして、多くの場合、当該適応症について、現在 PBS に掲載された薬が比較代案となるだろうが、当該人口集団を対象とした既掲載薬がなかったり、申請薬が既存薬に代わるものではなく、既存薬に加えて使用される場合には、標準治療法 (standard medical management) も適切な比較代案であり得ると言及している。

□ 最も代替の可能性が高い薬の選択

PBAC は、何が起こるべきかというより、何が起きる可能性が最も高いかに基づいて主な比較代案を判断する。オーストラリアの指針では、適切な主な比較代案の選択のために、次のようなヒエラルキー (hierarchy) を提示している。

- 薬理学的作用が類似した薬 (pharmacological analogue)。申請薬と類似機序の薬が既に掲載されている場合であれば、これらの中で最も多数の対象患者に使用される PBS 掲載薬が通常は主な比較代案となるだろう。
- 申請薬が新たな治療群に属するが、対象患者に広く適用される他の掲載薬があれば、これらの中で最も多く処方される薬が主な比較代案となるだろう。
- 申請薬の投与方法。投与経路 (注射剤、点眼剤など) も比較代案を選ぶ上で重要な考慮要因となるだろう。

□ 市場近接比較代案 (Near market comparator)

オーストラリアの指針では、まだ掲載されていない薬であっても、条件付きの比較代案として考慮する必要がある薬を市場近接代案と呼び、これに対する指針を提示している。当該患者を対象に、近い将来また他の薬がオーストラリア市場に進出するという合理的な期待があるとすれば、そして同じ回数 PBAC 会議や隣接する PBAC 会議においてこれらの薬剤が検討される予定であれば、これらの薬が条件付き比較代案として考慮されるかもしれない。

□ 細部集団別、対象人口別の比較代案

また、オーストラリアの指針では、場合によっては異なる細部集団において異なる比較代案を使用することも可能であるという点も簡単に言及している。

ウ) イギリス

イギリスの指針では、比較代案を選定する際には、NHSでの確立された診療慣行を考慮することを明らかにしている。しかし、このような確立された診療慣行が他の使用可能な代案に対して合理的ではないと評価される場合には、他の代案を比較対象とした評価を追加的に進めることもできるという事実にも言及している。

評価範囲 (scope) で定義された適応症に対して市販許可を受けなかった技術も、それが NHS で確立された診療慣行の一部であれば、比較代案として考慮することができると言及しているが、許可を受けなかった使用根拠、特に安全性、有効性に対する根拠と根拠の質を考慮すべきであると強調する。

エ) その他の指針

EUnetHTA (2015) でも、ヨーロッパ各国のガイドラインを比較して比較代案選定の原則について言及した。EUnetHTA が検討したところでは、ほぼすべての国が (日常的) 臨床診療で使用されるものを主な比較代案として選定するよう勧告していた。また、一部のガイドラインでは、通常行われる選択であれば、「no treatment」もまた比較対案になり得ると言及しており、あるガイドラインは、最も安価であったり最も効率的な代案との比較を追加すべきであるとも言及しているという。これに EUnetHTA は、臨床現場で使用される最も関連する (relevant) 代表的な介入を比較代案として選択するべきで、選択背景が明確に提示されて正当化されなければならないとしている。

現在、EUnetHTA (2019) が意見の取りまとめ中である「経済性評価時の考慮事項 (practical considerations when critically assessing economic evaluations)」では、適切な比較代案を不適切に排除したり、費用-効果的ではない代案との不適切な比較が生じ得る問題点について指摘した。そして、日常的診療で使用する介入に対してのみならず、現在診療を代替できる他の新たな介入、(根拠に基づいた) off-label use、そして、less intensive treatment/screening intervals も考慮するよう勧告している。そして ICER は、効率境界線上にある最後の比較代案と比較して計算すべきであることを強調している。時には標準治療が費用-効果的ではないこともあるが、その際は、可能であればより費用-効果的な他の代案を比較代案として考慮することが重要であると強調している。そして、対象人口集団により比較代案が変わる場合には、それぞれの人口集団別に評価することが適切であるとも言及している。そして、使用可能な根拠がない代案の場合、経済性評価を行うことが難しいという点にも言及している。

<表20> EUnetHTAパートナー国の分析：比較代案の選択

国	比較代案の選択
オーストリア	Na
ベルギー	Selected by help of an efficiency frontier
クロアチア	Therapies routinely used in the Croatian health system, including technologies regarded as current best practice
チェコ	Therapies routinely used and reimbursed in the Czech health system. The comparator should be selected and justified properly
デンマーク	Na
イギリス (England)	Technologies or tests that are current practice or are recommended in current NICE guidance (Technology Appraisals and NICE Medical Technologies Evaluation Programme Methods guide)
エストニア&ラトビア	Standard treatment or the usual treatment in daily practice
フィンランド	Therapeutically the most appropriate alternative. Based on Finnish clinical practice
フランス	All interventions that compete with the intervention evaluated
ドイツ	All therapeutic alternatives
ハンガリー	Standard/ most common treatment
アイルランド	Routine care, i.e. the technology(ies) most widely used in clinical practice in Ireland
イタリア	Current practice
オランダ	Standard treatment
ノルウェー	The treatment (drug(s) or health program(s)) that the new pharmaceutical will most likely replace. If currently used treatment not cost-effective, the efficiency frontier
ポーランド	Existing practice – procedure that will likely be replaced by assessed health technology in medical practice (Guideline) Reimbursed technology that is the existing practice should be the first choice (Regulation)
ポルトガル	Portugal Current practice, i.e. the most common treatment
ロシア	Drugs that are already included in the reimbursement list or, if there is no such drugs, common drugs with similar indications
イギリス (Scotland)	Treatments considered to be in routine use or represent best practice in NHS Scotland, and are the treatments that are most likely to be replaced
スロバキア	The treatment that is most likely to be replaced by the new treatment or, in case of add-on treatments, the current treatment without the add-on product
スロベニア	The drug with the same therapeutic indication (other drugs can be included as well)
スペイン	The standard technology used in current health care practices (AETSA, Osteba, CatSalut, Spanish recommendation). If possible also the most effective alternatives (Osteba, CatSalut)
スウェーデン	The most appropriate alternative treatment in Sweden (e.g. the most used)
スイス	The current treatment standard in Switzerland

出典: EUnetHTA (2015a)
Na: No information available

<表21> 比較代案に関連する海外指針の内容

国	指針内容原文	要約
カナダ CADTH 指針（4版） （2017）	In the reference case, "current care" should be considered. In many cases, this may include more than one relevant comparator. The choice of comparator (s) should be related to the scope of the decision problem. As such, the comparators should reflect the target population of interest and the jurisdiction for which the decision is being made. (p18)	current care
オーストラリア PBAC 指針 version 5 （2016）	Select the comparator(s) in the context of the targeted Australian population, the current alternative therapies in Australia, and the therapies most likely to be replaced in clinical practice. A single comparator will be appropriate in most circumstances. (p13) Where there is more than one comparator, the main comparator should be the therapy that prescribers would most replace with the proposed medicine.	臨床診療での代替の可能性が最も高いもの
イギリス NICE 指針 （2013）	When selecting the most appropriate comparator(s), the Committee will consider: <ul style="list-style-type: none"> - established NHS practice in England - the natural history of the condition without suitable treatment - existing NICE guidance - cost effectiveness - the licensing status of the comparator The Committee will normally be guided by established practice in the NHS when identifying the appropriate comparator(s). (p61)	NHSでの診療現況の考慮
EUnetHTA （2015）	it is recommended that the comparator(s) reflect the most relevant alternative intervention(s) used in clinical practice. The choice of comparators should be clearly presented and justified.	

4) その他の文献の考察結果

比較代案に何を選定するかによって個別介入法の費用-効果が変わるため、費用-効果的ではない代案を比較対象とすることによって介入法の費用-効果が歪曲されることがあってはならないということが大枠で合意する比較代案の選定原則である（Neyt & Brabandt, 2011; Drummond et al., 2015; Neumann et al., 2017）。

しかし、現実では既存の代案の費用-効果性に対する情報を得ることができない場合が多く、意思決定による影響を調べるという点から、既存の代案の中で代替の可能性が高い代案を比較対象として選定して分析することを勧告する機会が多い。（Zhao et al., 2018; Ziouani et al., 2016; EUnetHTA, 2015）

Neyt & Brabandt（2011）は、相互排他的な介入に対する費用効果の分析を行う際には、効率境界線上に位置した比較代案と比較することが望ましいと言及し、費用-効果的ではない代案と比較した経済性評価結果は、非常に注意して考慮されるべきであることを強調

した。

Drummond et al. (2015) も、既存の治療を比較代案として考慮する際に、既存の治療法が現在評価対象である治療法がない状況において、それ自身が「最上」であるかを考慮しなければならないという点を強調した。他の使用可能な代案を考慮せずには、費用-効果的ではない代案と誤って比較することにより、誤った結論を下す危険があるためである。適切な代案として、費用-効果的 (worthwhile) な可能性があるすべてのものを含むことを注文した。

Neumann et al. (2017) は、既存の治療法が最も適切な比較代案ではないこともあるという問題点を避けるために、分析では best available option や do nothing option を含め、使用可能なすべての範囲の代案を考慮することを推奨した。

Zhao et al (2018) は、文献検索を通して確保した 40 種の経済性評価ガイドラインを調査して比較代案をどのように記述しているかを比較したが、3 種のガイドラインで使われた比較代案に対して詳細な説明をするようにとだけ提示しており、多数 (15) のガイドラインは、新技術の導入に伴い、代替が起きる可能性が最も高い技術を、11 種のガイドラインは、標準治療法が非代替案として選択されるべきであると明示していたという。2 種のガイドラインは、すべての治療的代案を含むことを要求し、また他の 2 種のガイドラインは、最も適切な治療 (例: 最もよく使用される治療、最小治療方法、あるいはいかなる治療もしないこと) を比較代案として勧告した。概して、臨床現場で通用する治療法を比較代案として勧告する傾向があることを確認した。

Ziouani et al. (2016) も 29 か国の経済性評価ガイドラインを分析し、比較代案の選定基準を確認したが、Zhao et al. (2018) の研究が発行主体とは関係なく発表したガイドラインを比較分析したが、この研究は、比較代案の選定基準に対する国別の比較に焦点が当てられている。最もよく使用される比較代案は、「国内診療環境で使用される標準治療 (standard of care for local practice) (86%)」であった。45% (13 か国) では、「watchful waiting or doing nothing」も考慮するよう勧告した。「申請薬が収載される場合、最も代替の可能性が高い治療法 (therapy that prescribers would most replace with the proposed drug)」は、38%のガイドライン (11 か国) で望ましい比較代案として言及されたという。「最も費用が低い代案 (lowest cost alternative)」を選択すべきとしたガイドラインは 24% (7 か国) で、臨床診療で使用されるが、「市販許可が下りなかった薬 (Unauthorized drug)」も考慮すべきであるとしたガイドラインは 21% (6 か国) であった。Zhao et al (2018) のレビューと同様に臨床現場で通用する標準治療が最も広く言及された選定基準であった。

5) 利害関係者の意見

□ 経済性評価制度の改善TF

- 経済性評価制度の改善 TF で製薬業界が出した意見は、次のとおりに要約することができる。
 - 古い薬と比較しないこと、
 - 治療的な位置が同等な薬剤同士を比較すること、
 - 使用可能な根拠がある薬剤と比較すること、
 - 特許切れ薬の場合、最初の収載価格を比較すること、
 - First in class に該当する医薬品の場合、高い ICER を適用したり、古い薬を比較対象から除外したり、最初の収載価格と比較したり、あるいは経済性評価でなく原価基盤で価格を算定すること、
 - 代替薬剤がない先端バイオ医薬品の場合、別トラックで評価すること。
- これに対して審評院側は、現行指針が他の国の指針とも変わらず、経済性評価の一般原則にも符合するので、変更事由がないという意見を提示した。

□ 意見調査の結果

- 個別の製薬会社の意見調査を通して TF で提示された意見の他に、次のような意見が追加的に提示された。
 - 市場占有率よりも治療的な位置と実質的な代替関係がさらに重要である。治療的な位置を考慮する際に、許可適応症やガイドラインを参考にすること
 - 市場占有率に対する正確な情報を分かりにくい。また、臨床診療指針及び教科書とは異なり、特異的に韓国だけで多く使用される治療があれば、実際の経済性評価モデルの構築ができないことがあり、臨床診療指針、教科書を通して比較薬剤を選定することを提案する
 - 経済性評価モデルの不確実性を最小化するために、会社が持っている使用可能な資料も共に考慮することを提案する
 - 他の国で評価した結果も考慮すること

□ 専門家諮問会議

- 専門家諮問会議では、選定基準自体の変更の必要性には言及されなかった。ただし、最も多く使用する薬という定義が曖昧なため、選定基準を最大限具体的に提示することを要求した。
- 審評院の実務陣との懇談会では、直接根拠がある薬を比較対象として選定する問題についての検討が必要であるという意見があった。

6) 指針の改正方向の提言

- 比較代案の選定問題は、長年製薬業界で提起されてきたイシューの中の一つである。しかし、国内外の指針と関連文献を調べた結果、古い薬を比較対象から排除するとか、新しい作用機序の薬を評価する際に、比較代案の選定原則と異なる場合は見られなかった。むしろ、実際の臨床現場で広く使用されているのであれば、

治療法の形態を区分せずに幅広く比較代案として考慮すべきであるということが、多くの指針で推奨される場所である。また、根拠の可用性ではなく、実際の使用現況を重視して考慮すべきであるということが一貫した指針である。

- ただし、韓国の指針と他の国の指針を比較してみれば、韓国の指針では代替可能な既存の薬のうち、最も市場占有率が高いものを選択すべきであると明示しているが、他の国の指針では、市場占有率が最も高い一つのものに限定する表現を使用しない場合が多い。また、オーストラリアの評価結果を見ると（単数の比較代案を選定する場合はほとんどであるとガイドラインでは言及しているが）製薬会社が一つの薬を比較代案として提示しても、国内の使用現況などを考慮して他の代案も **secondary comparator** として要求する場合が多い。これに対して1案では、「最も多く使用される」薬を比較代案として選定するという基本原則は堅持しつつ、占有率と関連した表現を緩和して、比較薬の選定過程で代替の可能性を優先的に考慮し、他の状況も共に考慮できる余地を置く。
- その他、これまでの評価過程で示されたいくつかの争点については、次のとおり整理することを提案する。
 - オーストラリアのように比較代案選択のヒエラルキーを提示することに対して -> オーストラリアの指針では同一機序の薬を比較代案として優先的に選択するという表現が登場する。しかし、オーストラリアや国内の評価事例を見れば、ATC level 4 レベルで同一機序の薬が比較代案として考慮されたケースは相対的に少数であった。そして、オーストラリアの他には、このようなヒエラルキーを具体的に提示している国は珍しかった。したがって、評価と指針を最大限一致させて不要な消耗的な論争を防止する次元で、比較代案選択のヒエラルキーは提示しないものとする。
 - 給付される薬で比較代案を制限するか -> 他の国の指針を見れば、申請薬に代わる可能性が非常に高い薬が同時に給付申請中であつたり、近い将来に給付予定であれば、これもまた市場近接代案として共に考慮するようにした場合がある。しかし、国内指針の一次的使用者が給付決定申請をする製薬会社で、この場合、経済性評価を行う時点で市場近接代案に対する情報を確保しにくいため、指針では収載医薬品の中から比較代案を選定するようにした当初の指針を引き続き固守するものとする。ただし、市場状況に関する情報は、製薬会社が提出した経済性評価資料と共に委員会の論議過程や交渉過程で検討できるようにする。
 - 代替の可能性が最も高い薬 vs. 直接根拠のある薬（臨床試験で使用した比較代案）
 - ・ 時として、市場占有率によって決定された比較薬剤と臨床試験で使用した比較薬剤が一致せず、間接比較を通して比較薬剤と申請薬剤の相対的効果を評価した場合があつた。間接比較の場合、直接比較に比べて不確実性が大きく、比較の可能性が問題となる場合もある。
 - ・ 根拠の質のそのものだけを考えるならば、直接比較資料のある薬を比較

薬として選定することが妥当であるが、経済性評価の目的を考えると（この薬が給付されることによって変わる状況に対する評価）、新薬が収載されることによってどんな薬が代替されるかを無視して、使用可能な資料があるかどうかだけで比較対象を選定するのも妥当ではない。

- ・ これにより、新薬が収載されることによって代替の可能性が最も高い薬を比較代案として選定するという既存の原則を守るものの、制限された場合に限り、良質の根拠が裏付けられる代案を共に考慮できる余地を置く。例えば、臨床試験で使用した比較代案が次に該当する場合、代替の可能性を基準とした比較代案と共に追加的に比較代案として考慮できるようにする：1) 標準治療法に該当する場合、そして 2) 市場占有率が最も高い治療法と申請薬を比較した直接根拠がなく、間接比較をするにおいても比較の可能性が問題となるなどの理由により不確実性が大きい場合、3) 臨床試験に使用した比較薬がたとえ占有率が最も高い治療法ではないとしても、国内で普遍的に使用する治療法ではある場合。
- 既存の治療法に失敗したが他に代案がなく、既存の治療法を継続して使用する場合、申請薬の比較代案は？ -> 不応の先行治療剤だと比較代案から自動排除されることはない。他の代案がなく、当該治療法を継続して使用しているのであれば、プラセボよりは既存の治療法を比較代案として選択することが適切である。ただし、許可事項又は給付基準で中断を明示する場合は、比較代案から排除するものとする。
- 比較代案が通常の治療法の場合、通常の治療法をどのように構成するかの問題（臨床研究基盤 vs.国内の使用現況基盤） -> 経済性評価の目的が申請薬が収載されることによって変わる状況に対する評価という点を勘案して、実際の国内の使用現況を基盤として通常の治療法を構成することを提案する。ただし、国内の使用現況により構成された通常の治療法と臨床研究で使用された通常の治療法が効果や副作用の面で有意な差があるものと判断されれば、臨床研究で使用した通常治療法を共に考慮することができる。
- 戦略対戦略で比較される場合 -> 診断検査を伴う場合のように、時として薬対薬の比較ではなく、治療戦略対治療戦略で比較することが望ましい場合がある。この場合、比較対象となる戦略を具体的に提示しなければならない。
- 2案は、意思決定の明確性を考慮して現指針を維持するものである。

カ．資料源（間接比較）

1) 現指針

ア) 主な指針

収載申請医薬及び比較対象の費用と効果を推定する資料源としては、歪みの可能性が低い資料源（例：RCTs）を優先的に選択するようにする。また、収載申請医薬と比較対象を直接比較した資料源（head to head trials）を選好して、二つの代案を直接比較した資料源がない場合には、他の第3の治療方法やプラセボを参照代案として収載申請薬と参照代案、比較対象と参照代案を比較した資料源を利用することができる。資料を検索して選定するすべての過程は、透明に提示されなければならない。

イ) 指針内の追加説明

...そして、主な臨床効果（主な適応症）に対して収載申請医薬と比較対象を直接比較した資料源（head-to-head trial）を選択することが最も望ましい。万一、直接比較した資料がない場合には、プラセボ（placebo）や他の第3の治療代案を参照対象として収載申請医薬と参照対象、参照対象と比較対象を比較した資料を利用することができる」

現指針では、主な臨床効果を推定する資料源として、収載申請薬と比較対象薬を直接比較した資料がない場合には、プラセボや他の第3の治療代案を参照対象として、収載申請薬と参照対象、そして参照対象と比較対象を比較した資料を利用することができるものと言及することにより、間接比較資料源を使用することができるという簡単な記述になっているが、上記のような簡略な言及の他に、具体的な指針が提示されていない状態である。

2) 既存の提出資料の検討結果

既に経済性評価資料を提出した50品目のうち、9個の薬剤で主な臨床効果の推定及び効果差の判定のために共通対照群を利用した間接比較又は単純比較による間接比較の結果を提示しており、提出資料の評価過程で製薬会社によって提示された結果の受容の妥当性を論議する過程で、各資料源での交換可能性の成立の有無や補正の必要性などが検討され、「臨床的有用性評価のための間接比較の遂行段階別の資料提出指針」（健康保険審査評価院、2014）で提示された指針に準じて論議が行われたことを確認することができた。

臨床的有用性評価のための間接比較の遂行段階別の資料提出指針は、間接比較を行い、臨床的効果差の有無を結論付けようとする際、必須の分析仮定に当該分析資料が背かないかを段階別に評価し、検討者が最終分析結果と共に段階別に評価内容を確認できるように提示するものとしている。その際、必須的に確認して提示しなければならない仮定は、間

接比較に含まれた研究が互いに交換可能であるかであった。すなわち、患者群の特性、研究設計、薬剤の使用方法及び結果変数の測定などの主な特性において、十分に類似しており、当該研究から提示された結果値を合成して導出される相対効果値が妥当であるとみなされるかどうかについて、特性別に整理して明らかにした後、資料を合成するようガイドしている。その際、間接比較で使用される共通対照群での結果値もまた研究間で十分に同質であるかを確認するよう勧告する。研究特性が十分に類似していたとすれば、同じ共通対照薬剤が使用されたときには、結局、結果値も同質でなければならないためである。交換可能性に対する論議と共に、経評小委中の製薬会社から提出された結果が受容されなかったケースは、上記の9件中2件に該当し、間接比較に使用された資料の不確実性を反映するための感度分析を追加で提出するよう勧告したケースがあった。

間接比較を主な臨床的効果の推定のための資料提出のために利用することはなかったが、直接比較の結果資料が適切ではないものと判断されたり、事前に非劣性として設計されなかった直接比較の結果に対する追加資料として間接比較を提出し、明らかにするように勧告されたケースが2件あった。間接比較に使用された方法としては、共通対照群が不在の状況で行われた単純比較による間接比較が行われた4件の薬剤評価結果のうち、定性的な特性評価による交換可能性が認められて結果が受容されたケースが2件あり、1件のケースで認められなかったが、その他の傾向スコアを利用して加重値を付与して補正した方法（IPTW）を使用して提出された結果について、方法論的な受容の可能性が認められたケースが1件あった。

現在の経済性評価指針における間接比較に対する言及は、先立って指摘したとおり、非常に簡単な言及にとどまっているが、審評院で提示している臨床的な有用性評価のための間接比較資料の提出指針が別途存在しており、間接比較の遂行を通して提出資料を準備する製薬会社は、当該指針に準じて資料を準備して提出することにより、審評院の検討及び経評小委の評価過程においても、当該指針が活用されているので、経済性評価指針においても、さらに細部的なガイドとして当該指針を準用するよう勧告する一貫した言及が必要で、具体的な間接比較資料の提出指針もまた、より向上した観点及び方法論に対する受容を考慮して、今後の改正が必要であると思われる。

3) 外国の指針

ア) カナダ

大きく「Guideline statement」と「Guideline in Detail」のパートに区分されて記述されたカナダの「Guidelines for the Economic Evaluation of Health Technologies」（CADTH、2017）では、細部指針（Guideline in Detail）の「8.Modelling」節で、「Incorporating Effectiveness Estimates」の関連事項と「9.Effectiveness」節の「Data Sources and Assessment」と「Parameter Estimation」関連の段落及び「Standard Reporting Format」においてネットワークメタ分析（Network Meta-analysis）に対する言及が行われている。

指針内の主な言及としては、ベースライン状態での結果と相対効果の母数値がネットワ

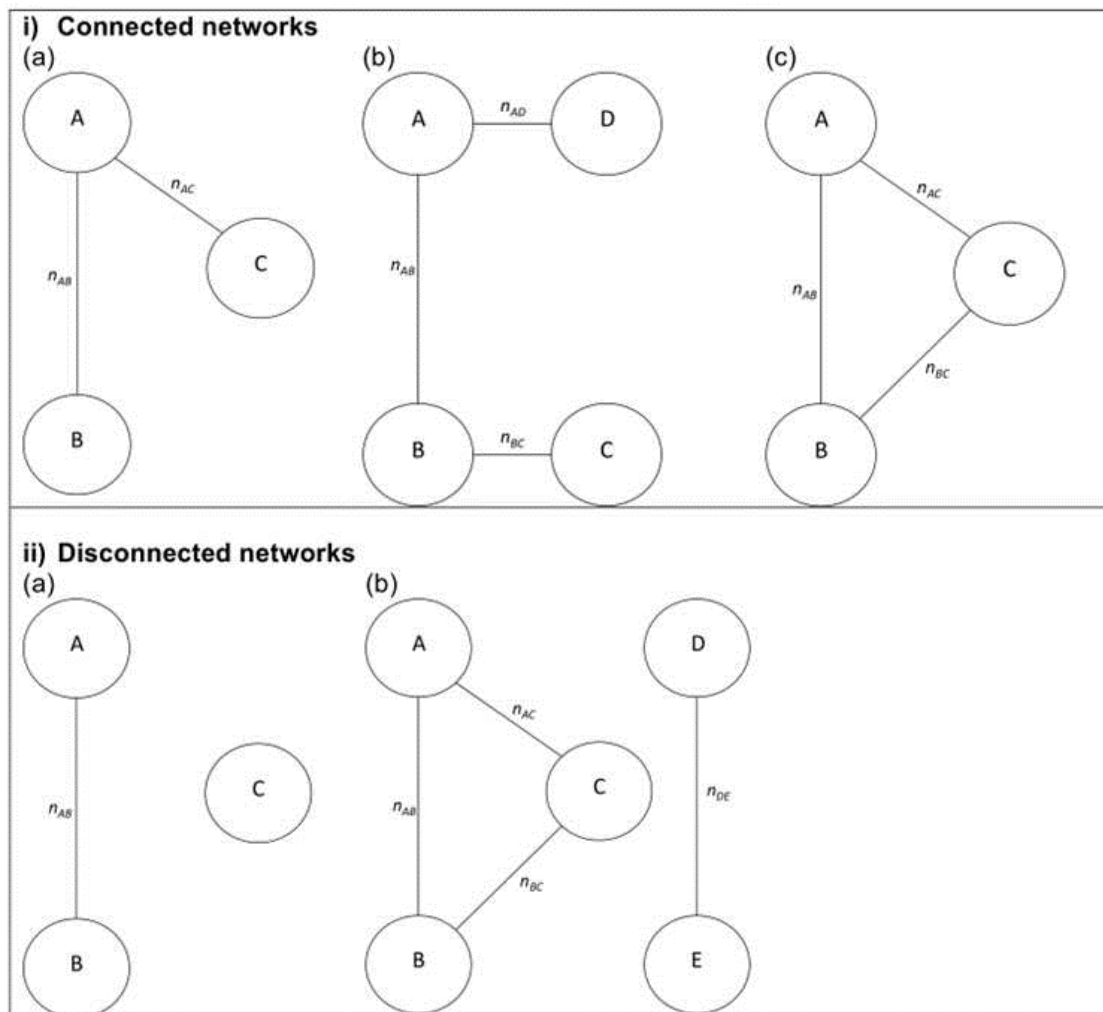
ークメタ分析の研究で推定される場合、資料が現在の文脈で対象母集団を十分に代表するか、評価目的に符合するかを考慮して正当化しなければならないということであり、資料源として既存のメタ分析やネットワークメタ分析を考慮する際に、既存の分析の目的と範囲及び選定/除外基準が適合して、現意思決定と関連しているか評価されなければならないということである。

本指針では、使用可能なすべての資料を結合できる方法がない場合には、目的、信頼性及び一貫性の側面から適合するものと評価された資料を合成することに集中して、メタ分析又はネットワークメタ分析の方法を使用する可能性があることに言及した。特定の意思決定に必要な情報を提供するための経済性評価と関連して、意思決定に二つ以上の比較代案が考慮されなければならなかったり、又は互いに直接比較されたことがない二つの比較代案が含まれることがあり、その際には、結果的に同時に様々な介入法に対する比較が可能となり、間接比較に基盤を置いたデータの統合を許容するネットワークメタ分析が対になって比較する（pairwise）メタ分析アプローチより選好されることを記述している。

本指針では、間接比較に対する遂行アプローチなどの細部事項を具体的に提示してはいないが、有効性の結果資料に対する資料源として、既存のメタ分析だけではなくネットワークメタ分析を並べて言及していることが 2017 年改訂版の新しい事項と言え、メタ分析だけではなく、ネットワークメタ分析も資料源としての適合性を明確に明らかにすることが明示的なガイドとして提示されている。具体的にどのような方法とアプローチを使用しなければならないかは、本指針においては言及して記述してはいないが、方法論的な参考指針として、Cooper et al. (2015) の文書を指針内で提示した。

本指針で付録として提示される標準報告形式（Standard Reporting Format）で提示される各報告書の節で提示されるべき事項のうち、有効性の結果に関連した方法論の提示でネットワークメタ分析のための別途の事項は言及されておらず、ただし、標準的な細部事項を言及するという点において、資料源の例示として臨床試験、メタ分析、ネットワークメタ分析を並べて言及しており、ネットワークメタ分析を資料源として「考慮」することを明示したと言える。

経済性評価指針において、具体的な方法論の指針として参考にするように提示される Cooper et al. の文献は、妥当なネットワークメタ分析はどのような者であり、どのような仮定を満たすべきかを詳細に提示し、様々なイシューを説明し、ケーススタディを例示している。本文献では、ネットワークメタ分析で考慮されるネットワークは、下図 3 の i) で見られるように接続できたネットワークであるべきで、ii) のように接続できなかったネットワークでのネットワークメタ分析は妥当ではないことを明示し、ネットワークに関して意思決定と関連がない介入が含まれる場合、説明すべきであることを明示した。また、妥当性の条件として、一貫性（Consistency）仮定が満たされなければ、ネットワーク分析で得られる介入法の間での介入効果が対によって異なることがあり、この場合、解釈が難しく、意思決定モデルに含むことが困難であることを明確にした。したがって、ネットワークメタ分析を行うときは、一貫性の仮定について明確にすべきであり、これに対する評価方法を提供すべきであることを強調した。



[図3] Use of network meta-analysis to inform clinical parameters in economic evaluations

出典: Cooper et al. (2015)

イ) オーストラリア

オーストラリアの「Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee」は、PBAC (Pharmaceutical Benefits Advisory Committee) に提出する資料を準備するための指針として、資料の提出準備のための具体的で細部的な事項を含み、間接比較と関連した事項は、以前の 4.3 版 (PBAC、2008) と 2016 年に改正された 5 版 (PBAC、2016) のいずれも、基本的に間接比較のワーキンググループの方法論報告書 (PBAC、2008) を参照して提示されている。ただし、以前の版では、共通対照群を利用した比較的単純な構造の中で非常に厳格な評価に基づく間接比較に限定して間接比較の資料提出を勧告している反面、最近の改訂版ではネットワークメタ分析や混合比較などに対する言及として間接比較の範囲について広く言及して提示している。根本的には、分析に必須の仮定の成立と分析の妥当性に対する具体的で細部的な疎明資料を明確に提出するよう勧告しており、基本的には単純な構造を選好して行わなければならないという点におい

ては、既存の勧告の枠組みから大きく外れてはいない。

オーストラリアの PBAC 指針は、全般的な資料の提出準備のためのガイドとして、毎評価別に非常に具体的な事項を提示していることが特徴であり、5 版の指針では「Section 2 Clinical evaluation」で具体的に間接比較と関連した勧告が行われている。臨床的な評価に対する資料として、PBAC は直接無作為割付試験をベースとした臨床及び経済評価を強力に選好することを優先して提示し、直接無作為試験が常に可能となるわけではないので、本指針は間接比較を考慮するための枠組みを提供することに言及している。これにより、間接比較に対する別途の準備指針が含まれているというよりは、全般的な資料提出のための段階において間接比較が含まれる場合、追加的に考慮しなければならない事項が言及される形となっている。例えば、臨床的な主張や評価の範囲を定めるとき、文献検索を行うとき、対象集団を定義するとき、非劣性評価のためのマージンを決定するとき、分析結果を提示するときなど、各段階で間接比較が関連している場合の要求事項を追加的に言及したり、具体的に提示する形を取っている。

臨床評価結果に対する提出資料に間接比較が含まれる場合について具体的な事項は、「2.6 Trial results:additional analysis」節で臨床評価の結果に小グループ分析が使用された場合、メタ分析が使用された場合、治療転換について補正が行われた場合、間接比較が使用された場合について、具体的な各節を提供した。

間接比較の評価に含まれた臨床試験に対する対象者のベースライン特性、治療の細部事項、結果変数の定義及び含まれた個別臨床試験に対する結果を標準的な方法によって提示するようにし、Bucher 法による一対 (pairwise) 比較法、MAIC (matching-adjusted Indirect Comparison) 法、シミュレーション介入比較 (Simulated Treatment Comparison)、ネットワークメタ分析又は混合比較 (Mixed Treatment Comparison) のような間接比較の遂行方法を記述することを勧告し、当該方法に対する利用の可能性を提起した。

しかし、以下のとおりの原則を提示して、厳格な仮定に対する不確実性が大きな分析結果は考慮しないことを明示している：

- 1) ネットワーク中に含まれた各可能な経路に対して介入のペアに対する一対比較を行って結果を提示することを基本として勧告し、Bucher 法を提案する
- 2) 複雑なネットワークメタ分析を行うことは、補助的な分析として提示すること
- 3) ネットワーク分析に非無作為研究を含むことは避けなければならない、万一含む場合は、当該研究を含む場合と含まない場合の結果を両方提示すること
- 4) 補正されていない単純比較や分析に含まれた研究の推移性 (transitivity) が満たされていない間接比較の結果は、PBAC の意思決定過程において結果の解釈に対して信頼を得ることが困難であり、少なくとも一つの研究で個別の患者資料 (IPD:Individual Patient-level Data) を使用できる場合、MAIC あるいはシミュレーション介入比較法を使用して研究間の異質性を補正すること

また、分析を再現できるだけの十分に具体的な分析過程とプログラムコードを提供することと、IPD が要求する分析方法を利用した場合には IPD の資料のスプレッドシートを共に提出し、そうでない場合は理由を明らかにすることを明示した。

間接比較を行うときに要求される推移性（transitivity）仮定は、間接比較内の介入間の比較の結果が知られている介入効果の修正因子（treatment effect modifier）の分布によって変わってはならないことを意味する。カナダの CADTH 指針で言及する一貫性（consistency）のような概念であり、PBAC の以前の 4.3 版の指針で重要視された交換可能性（exchangeability）の拡張された概念であると言える。本指針は、間接比較に対する推移性仮定に対して段階別に評価して提示するように、下図 4 で提示された表を通して枠組みを提供している。

Comparison	Issues to consider
A vs C direct randomised trials	<ol style="list-style-type: none"> 1. Assess the trials for factors that may cause heterogeneity of the A vs C comparative treatment effect 2. Assess the event rates in the medicine C populations 3. Assess the impact of the measure of comparative treatment effect for A vs C 4. Assess statistical homogeneity of the A vs C comparative treatment effect across trials
B vs C direct randomised trials	<ol style="list-style-type: none"> 1. Assess the trials for factors that may cause heterogeneity of the B vs C comparative treatment effect 2. Assess the event rates in the medicine C populations 3. Assess the impact of the measure of comparative treatment effect for B vs C 4. Assess statistical homogeneity of the B vs C comparative treatment effect across trials
A vs B indirect comparison	<ol style="list-style-type: none"> 1. Assess the sets of trials (ie the A vs C and the B vs C trials) for factors that may cause heterogeneity of the A vs B comparative treatment effect 2. Assess the event rates in the medicine C populations across the sets of trials 3. Assess the impact of the measure of comparative treatment effect for A vs B 4. Assess statistical homogeneity of the synthesised comparative treatment effect A vs B across the sets of trials (only possible if A vs B has been compared via multiple common references)

[図4] Steps to assess the transitivity assumption

出典: Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 5). (PBAC, 2016)

提示された表を通じた段階別の評価方法と共に各段階が何を意味するのかについて説明を付け加えて提示した。同様の概念と観点に基づいているだけに、基本的に以前のバージョンの指針における交換可能性の評価方法及び提示過程と非常に類似しており、ただし IPD が使用可能な場合に利用できる補正方法が追加的に言及されている点が異なると思われる。これには、以下のような内容が含まれる：

- 研究セット内あるいは研究セット間の特性差を要約して（下図 5 で提示されたようにメタ分析を行う際の異質性探索のための定性的評価方法として提示された潜在的な特性表を相互参照することを勧告）、効果差に与える影響に対して明確に記述すること

Category	Factor	Trial 1	Trial 2	Trial 3
Different quality of methods of trials	Adequate concealment of randomisation	[add]	[add]	[add]
	Blinding	[add]	[add]	[add]
	Duration of follow-up	[add]	[add]	[add]
	Loss to follow-up	[add]	[add]	[add]
	Crossover	[add]	[add]	[add]
Confounding factors in relation to participant populations	Age	[add]	[add]	[add]
	Sex	[add]	[add]	[add]
	Genetic variation	[add]	[add]	[add]
	Diagnostic workup	[add]	[add]	[add]
	Intensity of surveillance	[add]	[add]	[add]
	Severity of disease or condition	[add]	[add]	[add]
	Physiological reserve	[add]	[add]	[add]
	Stage or duration of disease or condition	[add]	[add]	[add]
	Previous therapy	[add]	[add]	[add]
	Coexisting disease or condition	[add]	[add]	[add]
	Background therapy of concomitant treatments/advances in standard of care	[add]	[add]	[add]
	Confounding factors in relation to circumstances	Health systems	[add]	[add]
Geography		[add]	[add]	[add]
Setting in hospital or ambulatory care		[add]	[add]	[add]
Date of trials		[add]	[add]	[add]
Different treatment	Dose	[add]	[add]	[add]
	Duration	[add]	[add]	[add]
	Timing	[add]	[add]	[add]
	Stopping or continuation criteria	[add]	[add]	[add]
Different outcome measures and methods of statistical analysis	Definition of outcome(s)	[add]	[add]	[add]
	Rating instrument	[add]	[add]	[add]
	Frequency of measurement	[add]	[add]	[add]
	Start point of measurement against duration or progression of disease or treatment, especially in time-to-event analyses	[add]	[add]	[add]

[図5] Example factors that might cause comparative treatment effect heterogeneity

出典: Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 5). (PBAC, 2016)

- 万一、間接比較に攪乱要因が含まれているのであれば、メタ回帰分析 (meta-regression) を適用することが適切で、変数当たり 10 個以上の研究が行われてこそ結果に安定性があると言及した。Butcher 法を利用した一対比較に加え、MAIC やシミュレーション介入比較を利用することができる
- 共通対照群でのイベント発生率 (event rate) を比較すること
- 比較効果を推定するための測定値が何かによって統計的異質性に影響を及ぼすかを評価すること

間接比較による結果を提示する方法は、改訂版以前のように二分型資料をベースに基づいて枠組みを提示しており、いくつかの追加的に注目する事項として以下の内容が言及さ

れた：

- 共通対照群がいくつか存在して複数の間接比較が可能な場合、これについて明確に提示して間接比較別の相対効果を比較して説明すること
- 推定された相対効果値において説明できない差があれば、解釈に困難があることを周知すること
- 適切な場合には、ネットワークメタ分析を補足的に行って提示すること
- 特定の理由で分析から除外された研究がある場合には、これを含んだり排除した状態での感度分析を行い、異質性を増加させる研究や研究セットを除いた感度の分析を行うこと

ウ) イギリス

イギリスの NICE (National Institute for Health and Care Excellence) で提示された「Guide to the methods of technology appraisal」(NICE、2013) は、改訂版でも以前のバージョンのように大きな枠組みと範囲に対する基本的な原則を提示する指針の性格を有しており、ネットワークメタ分析に関する内容は、基本分析に対する原則を提供する 5 章の「5.2 Synthesis of evidence on health effects」で体系的な文献考察に対する基本的なアプローチと共に「Indirect comparisons and network meta-analyses」という題名の節で構成されて言及されている。

ひとまず、基本分析は直接比較臨床試験に基づいて提示されるべきであり、単一 RCT 内で評価されなかった医療技術を比較する際は、RCT を利用した一連の対比較で分析を行い、適切な場合にはネットワークメタ分析と共に提示することを勧告している。ネットワークメタ分析は、RCT を利用した基本分析に追加分析として行われ、これに伴うすべての内容が完全に説明されて提示すべきであることに言及した。

評価委員会は、間接的な出典だけで導出された相対的な効果の推定を考慮する際、直接的な証拠の不足及びこれに関連した追加的な不確実性を評価に考慮し、一対のメタ分析で遵守しなければならないすべての標準的な原則は、補正された間接比較及びネットワークメタ分析を行う際にも従うべきであると言及した。

本指針で使用するネットワークメタ分析という用語は補正された間接比較を含み、混合比較のようなより複雑な証拠合成分析を指し、補正された間接比較は、関心のある医療技術が RCT を通して直接比較されなかったが、共通比較対照群を使用して間接的に比較された臨床試験からのデータの合成を意味し、混合比較には、評価対象介入と代案治療を直接比較した RCT 研究、又は評価対象の介入や代案介入中の一つを含む RCT 研究が両方含まれる分析として定義した。

混合比較や間接比較を行う際は、資料合成方法に対する明確な説明と RCT に対する検索と選定及び除外基準に対する理論的根拠を提示し、ネットワークメタ分析に含まれた個別の臨床試験の遂行方法と結果及び各試験に対するベースライン特性表を文書化して提示することを勧告した。

一対比較結果での異質性と直接比較及び間接比較の結果間の一貫性について報告するべきであり、ネットワークメタ分析内で非一貫性が発見されれば、このような不一致を明ら

かにして解決するための試みが行われるべきであることを勧告した。

間接比較又はネットワークメタ分析のアプローチを使用して根拠を合成するすべての場合において、臨床試験の無作為割付が維持されるべきであり、すなわち、互いに異なる無作為割付の臨床試験で得た単一治療群の結果を比較することは許されないことを明らかにしている。このような類型の比較が提示されれば、結果資料は本質的に観察的研究で得られた資料のレベルで扱われ、増加した不確実性を考慮することに言及した。

ネットワークメタ分析から得られる根拠は、表及びグラフを利用して提示することと、ネットワークメタ分析に含まれる直接及び間接資料に明確に区分して各比較に含まれる臨床試験の個数を明示するものとした。また、直接比較を使用した一对のメタ分析結果は、ネットワーク全体のメタ分析をベースとした結果と共に提示すべきであること明確にした。

一对のメタ分析又はネットワークメタ分析に含めるのに十分な資料がない場合には合成をやめ、個別研究を批判的に評価して結果を技術的に提示する定性的なアプローチに制限されるべきであり、評価委員会は結果を検討して、評価対象介入の治療代案に対する相対的な臨床効果に対する結論を導出する際に特に注意することに言及した。

不確実性の検討に対する指針関連の部分において、メタ分析に限らず、ネットワークメタ分析で特定研究の質や又は関連性が疑われる場合には、感度分析を行って提示することをあわせて言及することもあった。

指針の本文に付け加えて発刊された「Developing NICE guidelines:the manual – appendices」(NICE、2014)は、指針の各内容によって必要な基準やチェックリストなどの付録を提示しているもので、ネットワークメタ分析に対する結果を報告するときは当該付録で提示された PRISMA-NMA チェックリストの修正版の基準を満たすべきであることを明示した。PRISMA-NMA 修正版のチェックリストは、全体の PRISMA-NMA チェックリストのうち、特にネットワークメタ分析の結果を報告するために適用する項目で構成され、各項目に対する具体的な説明は、全体の PRISMA-NMA (Hutton et al., 2015) を参考にするようにした。

PRISMA-NMA 修正版のチェックリストで扱われる主な内容は次のとおりである：

- ネットワークメタ分析が行われた理由に対する言及
- 含まれた臨床試験の研究特性を報告
- ネットワークの構造及びこれと関連した潜在的な歪みを探索するために使用される方法を説明
- 結果を要約するために使用された測定値に対して記述
- データを処理して各ネットワークメタ分析に対する遂行及び結果の合成方法を説明
- ネットワークにおいて、直間接根拠の一貫性を評価するための統計的なアプローチ方法を記述
- 直接結果と間接結果に非一貫性が発見された場合、解決するために使用された統計的なアプローチ方法を記述
- ネットワークを視角化することができるようにグラフを提供
- 結果の制限点に対する論議
- 推移性及び一貫性のような仮定の妥当性に対する論議

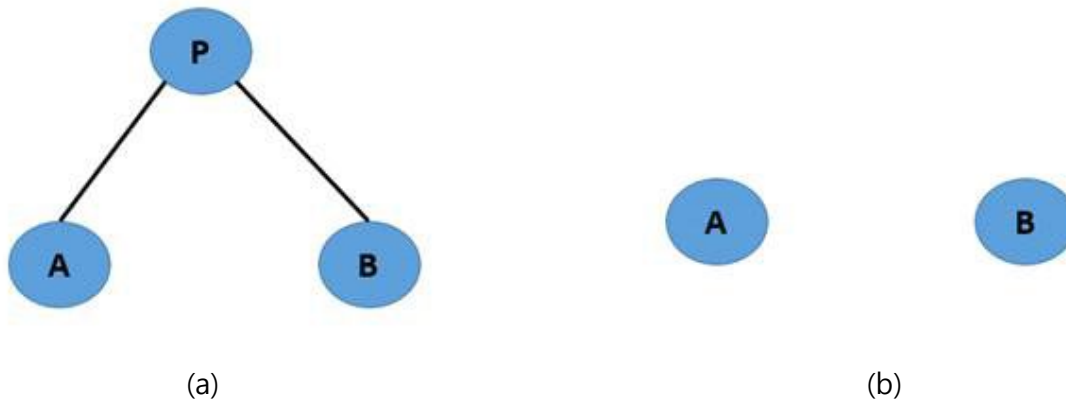
- ネットワーク形態に関する論議

エ) 各国指針の要約

国	内容
カナダ CADTH 4th Edition (2017)	<ul style="list-style-type: none"> ・本文中でネットワークメタ分析の使用について言及 ・直接比較されたことのない比較代案が含まれたいくつかの治療代案が同時に比較されなければならないときには、ネットワーク分析が選好される ・具体的な細部指針は提供されず、方法論文書 (Cooper et al., 2015) を参考にすることにガイド ・Cooper et al. (2015) には、接続できなかったネットワークでの分析は妥当ではないことを明示し (単純比較はこれに該当する)、一貫性の仮定について明確にし、評価方法を提供することを強調した
オーストラリア PBAC 指針 version 5 (2016)	<ul style="list-style-type: none"> ・ネットワークメタ分析、混合比較などの使用について言及 ・単純な構造の中で一対比較の遂行が基本であり、複雑なネットワーク分析は補助的な分析として提示すること ・間接比較に含まれた対象者のベースライン特性、治療の細部事項、結果変数の定義の個別の臨床試験結果を標準方法で提示 ・間接比較の遂行方法を具体的に記述 ・一貫性が満たされていない間接比較の結果は受け入れられないこと ・IPD が使用可能な場合、MAIC 又は STC 法などを使用して最大限補正すること ・分析の再現が可能なほど具体的な分析過程とプログラムコードを提出すること
イギリス NICE 指針 (2013)	<ul style="list-style-type: none"> ・間接比較を使用する場合、一連の一対比較が優先であり、適切な場合にはネットワークメタ分析を追加的に提示できる ・随伴するすべての内容が完全に説明されて提示されること ・直接比較及び間接比較の結果間の一貫性について報告すること、非一貫性が発見されれば不一致を明らかにする ・ネットワークメタ分析のアプローチを利用する場合、無作為割付が維持されること (カナダの場合、連結されなかったネットワーク分析は妥当でないというような言及。つまり、単純比較にはならない) ・ネットワークメタ分析を行うことによって資料を提出する際は、報告書は PRISMA-NMA 修正版 ・チェックリストにしたがって作成すること

4) その他の文献の考察結果

臨床的な効果評価において、評価対象薬剤 (A) と代案薬剤 (B) 間の直接比較臨床試験を通して行われた評価結果を優先するという事は、すべての指針において共通して明言されている事項であるが、直接比較臨床試験が行われたことがなかったり存在しない場合、二つの薬剤を間接比較する方法を選択することとなり、間接比較は、図6で示すように、ひとまず二つの薬剤が共有する共通の対照群 (P) を経て行われる間接比較である (a) の場合と単純比較である (b) の場合に大きく分けられる。



[図6]共通対照群を含む間接比較 (a) と単純比較 (b)

間接比較を行う際には、間接比較を通して得られた結果に妥当性があることを明らかにしなければならない。間接比較が行われる際には、様々な補正方法が考慮されるが、最近の改訂版の海外指針において補正方法として言及されているMAIC (Matching adjusted indirect comparison) 法とSTC (Simulated treatment comparison) について簡略に要約して提示する。

(a) の場合、P-A臨床試験とP-B臨床試験のうち、少なくとも一つ又は (b) の場合、A Bそれぞれの単一群の研究のうち、少なくとも一つの臨床試験に対する個別患者資料 (IPD) が存在するとき、MAICあるいはSTCを通して二つの臨床試験に含まれた参加者の特性差を縮めてAとBの比較性を高めることができることを期待するものである。

ア) Matching Adjusted Indirect Comparison (MAIC)

P-A 臨床試験と P-B 臨床のうち、一つの臨床試験に対する IPD が存在する場合に行える方法として、片方の母集団では IPD を使用でき、他方の集団で要約資料のみ使用可能な場合に適用される方法で、一種の傾向スコア加重値 (propensity score weighting) を適用する方法である。IPD が使用可能な母集団に属する個々に相対側の母集団での共変量の分布とバランスをとるために傾向スコアの逆加重値を適用する。IPD が二つの母集団のうち片方だけで利用できるため、既存の傾向スコアの利用方法とは異なり、傾向スコアの推定に対する新たなアプローチ方法が必要である。

基本アイディアは、Inverse Probability Weighted Estimator 法のように、いずれも同一の加重値を資料に再加重値として賦課し、相対効果の推定値を補正する方法である。IPD が存在する臨床試験に含まれた患者が要約資料 (aggregate data) のみが存在する臨床試験に含まれる確率を効果推定時に加重値で反映して値を推定する。IPD がある母集団の対象者のうち、要約資料のみ臨床試験の平均患者特性と類似する場合、高い確率を持ち、結局、二つの臨床試験の患者特性の分布が似たように調整されることになるものである。手順は下

記のように要約することができる。

- Step 1 : 分析に使用する標本の選択
 - P-A 臨床試験については IPD が存在し、P-B 臨床試験について要約資料が存在するとするとき、P-A trial の選定除外基準は、P-B 臨床試験の選定除外基準より広い範囲あるいは同じ範囲を持っていなければならない。また、片方は二重盲検研究であるが、片方は公開臨床試験などの臨床試験間の研究設計上の差があってはならない。
- Step2 : 資料を把握して matching に使用する患者のベースライン特性 (baseline characteristics) を選定
- Step 3 : 二つの臨床試験間の集団を matching
 - P-A 臨床試験の各患者を P-B 臨床試験で報告された平均患者の特性に matching させるために reweighting する。

Step 3のmatching方法は下記のように要約することができる。

- 各研究 (P 群は暫時く除外) の変数 : (X, T, Y)
 - X : vector of baseline characteristics (e.g. age, sex, baseline disease severity, etc.)
 - T : T=0 for A, T=1 for B
 - Y : outcome measure
 - データ : (x_i, t_i, y_i)
 - AのIPD : $(x_i, 0, y_i)$ 、Bのaggregate data : $(\bar{x}_1, 1, \bar{y}_1)$
 - 最終的に導出されるBに対するAの効果の大きさ :

$$\hat{\theta} = \frac{\sum_{i=1}^n y_i (1-t_i) w_i}{\sum_{i=1}^n (1-t_i) w_i} - \bar{y}_1$$

$$w_i = \frac{\Pr(T_i=1|x_i)}{\Pr(T_i=0|x_i)} : x_i \text{ 特性を持つ } i \text{ 番目の患者が } T=1 \text{ vs } T=0 \text{ 治療を受けることとなるdds}$$

- w_i を求める
 - logistic regression に従うことを仮定して $w_i = \exp(\alpha + x_i' \beta)$ で表現する
 - $\hat{\beta}$ を求める方法は、次の方程式を通して得ることができる :

$$0 = \frac{\sum_{i:t_i=0} x_i \exp(x_i' \hat{\beta})}{\sum_{i:t_i=0} \exp(x_i' \hat{\beta})} - \bar{x}_1$$

○ 分析にP群が含まれた場合

各研究のvariables : (X, T, Z, Y)

P群の場合はZ=0

Bのaggregate data : $\bar{x}_1^{(0)} \bar{y}_1^{(0)}$ when $Z = 0$, $\bar{x}_1^{(1)} \bar{y}_1^{(1)}$ when $Z = 1$

$\bar{x}_1 = \frac{x_1^{(0)} n_1^{(0)} + x_1^{(1)} n_1^{(1)}}{n_1^{(0)} + n_1^{(1)}}$, $n_1^{(0)}, n_1^{(1)}$: P-B trialのP群とB群に含まれた各患者数

→最終的に導出されるBに対するAの効果の大きさ :

$$\hat{\theta} = \frac{\sum_{i=1:n}^{t_i=0, z_i=1} y_i \exp(x_i' \hat{\beta})}{\sum_{i=1:n}^{t_i=0, z_i=1} \exp(x_i' \hat{\beta})} - \frac{\sum_{i=1:n}^{t_i=0, z_i=0} y_i \exp(x_i' \hat{\beta})}{\sum_{i=1:n}^{t_i=0, z_i=0} \exp(x_i' \hat{\beta})} - (\bar{y}_1^{(1)} - \bar{y}_1^{(0)})$$

さらに具体的な細部の理論的な説明と例示などをSignorovitch et al. (2010) 及びPhillippo et al.のNICE DSU方法論報告書 (Phillippo et al., 2016) 及び論文 (Phillippo et al., 2018) を通して参考にすることができる。

イ) Simulated Treatment Comparison (STC)

STC は、P-A 臨床試験と P-B 臨床試験のうち、一つのあるいは両方に対する IPD が存在する場合に行える方法である。ある臨床試験の IPD を使用して回帰分析を通してモデルを適合させて推定した後、このモデルに他の臨床試験の要約資料あるいは IPD を代入した予測を通して相対効果の推定値を得る方法である。

STC 法を適用するためには、P-A 臨床試験と P-B 臨床試験の母集団が十分に重ならなければならない。すなわち、共変量の分布及び二つの臨床試験の選定除外基準が十分に類似していなければならない。本方法を適用するためには、予後要因 (prognostic factor) と効果修正因子 (effect modifier) を区分して定義する必要がある。P-A 臨床試験については IPD が、P-B 臨床試験については要約資料が存在する状況を基準として、方法を次のとおり要約して説明することができる。

- Step 1 : P-A trialのIPDを利用してregression modelをfitting

$$g(Y) = \beta_0 + \beta_1^T X + (\beta_A + \beta_2^T X^{EM}) I(t = A)$$

X : 説明変数、 Y : 独立変数、 g : Link function (Identity link functionの場合、 $g(Y) = Y$)

t : 治療群

β_0 : intercept term

β_1 : a vector of coefficients for prognostic variables

β_A : $X=0$ で Pに対するAの relative effect

β_2 : a vector of coefficients for effect modifiers

- Step 2 : P-B trialで観察されるPに対するAの効果の大きさの推定

P-B trialのaggregate data : $\bar{Y}_{P(PB)}, \bar{Y}_{B(PB)}, \bar{X}_{(PB)}, \bar{X}_{(PB)}^{EM}$

$$\hat{Y}_{P(PB)} = g^{-1}(\hat{\beta}_0 + \hat{\beta}_1^T \bar{X}_{(PB)})$$

$$\hat{Y}_{A(PB)} = g^{-1}(\hat{\beta}_0 + \hat{\beta}_1^T \bar{X}_{(PB)} + \hat{\beta}_A + \hat{\beta}_2^T \bar{X}_{(PB)}^{EM})$$

$\therefore \hat{Y}_{B(PB)} - \hat{Y}_{A(PB)}$: pに対するAの効果

○ Step 3 : Aに対するBの効果の大きさ

$$\therefore \bar{Y}_{B(PB)} - \bar{Y}_{P(PB)} - (\hat{Y}_{A(PB)} - \hat{Y}_{P(PB)}) \text{で推定可能}$$

さらに具体的な細部理論的説明と例示などをCaro et al. (2010) 及びNICE DSU (NICE方法論報告書 (Phillippo et al., 2016) 及び説明論文 (Phillippo et al., 2018) を通して参考にすることができる。

ウ) 二つの方法の適用

最近よく議論される MAIC と STC のような方法を利用した間接比較は、一種の母集団補正間接比較 (Population Adjusted Indirect Comparison) 法と言える。NICE DSU の報告書は、二つの方法について非常に詳細なレビューと理論的説明、仮定、分析方法、適用例、問題点などと共に NICE 提出文書に臨床的効果を提示するために使用する際の勧告事項を提示している。母集団の補正介入効果の推定値を考慮することには明確なバックグラウンドがあるが、実際にどのように、いつ適用しなければならないのか、そして得られた結果が当該意思決定と関連があるかどうかについても明確ではなく、これは一つの分析で行った仮定が同一条件の今後の他の分析でも同様に適用されないリスクが大きくなるためである。したがって、当該報告書は、まだ追加で明確にされるべき部分が多く、現在は暫定的という但し書きの下で一連の方法適用に対する推奨事項を提示し、間接比較方法の選択及び実現方法及び結果の提示方法を要約した模式図と順序図を提供した。

5) 利害関係者の意見

ア) 専門家の意見

- なし

イ) 製薬会社の意見

改正が必要な理由として、共通の対照群の間接比較を行うことができない場合の方法論として、MAIC (matching adjusted indirect comparison) や他の間接比較の遂行方法としてネットワークメタ分析に対する言及がないということを指摘し、様々な間接比較の遂行方法に対する受容の可能性に対して明示することを改正方向として要請するものである。

6) 指針の改正方向の提言

現指針では、主な臨床効果を推定する資料源として収載申請薬と比較対象薬を直接比較した資料がない場合には、プラセボや他の第3の治療代案を参照対象とし、収載申請薬と参照対象、そして参照対象と比較対象を比較した資料を利用することができるという言及をすることにより、間接比較の資料源を使用することができるという簡単な記述はあるが、上記のような簡略な言及の他に具体的な指針が提示されていない状態であるため、さらに具体的な言及と方向性の提示が必要である。

現在の経済性評価指針での間接比較に対する言及は非常に簡単な言及にとどまっているが、審評院で提示している臨床的な有用性評価のための間接比較資料の提出指針が別途存在しており、間接比較の遂行を通して提出資料を準備する製薬会社は、当該指針に準じて資料を準備して提出しており、審評院の検討及び経評小委の評価過程でも当該指針が活用されているので、経済性評価指針においてもさらに細部的なガイドとして当該指針を準用するように勧告する一貫した言及が必要で、具体的な間接比較資料の提出指針もまた新たな方法論などに対する受容を考慮して、今後の改正が必要であると思われる。

既に提出された50件の資料のうち、間接比較を使用して臨床的有用性の資料を提出した9件の事例のうち、共通対照群の間接比較5件中1件が認定されず、共通対照群なしに行われた単純比較が4件あり、このうち1件が認定されなかったもので、3件の間接比較は単純比較として結果が認められた事例であった。共通対照群が不在の状況で行われた単純比較による間接比較が行われた4件の薬剤の評価結果のうち、認められなかった1件のほか、定性的特性評価による交換可能性が認められて結果が受容されたケースが2件あり、その他の1件は傾向スコアを利用して加重値を付与して補正した方法（IPTW）を使用して提出された結果について、方法論的な受容の可能性が認められたと報告書に記述されているもので、MAICのような方法を使用したものと思われる。したがって、現在の審評院の評価過程では、相当数の単純比較も検討されており、既にMAICのような方法が受容された場合もあると言える。

審評院の経済性評価指針で間接比較に対する言及はほとんどなかったが、別途の間接比較指針が準用されており、当該指針は2014年の改訂版で利害関係者の要請を受け入れ、単純比較に対する受容の可能性を既に開いているため、むしろ海外の指針よりもさらに緩和された基準を有していたとも考えられる。

先立って要約されたとおり、カナダ、オーストラリア及びイギリスの医療技術評価のための指針書のいずれにおいてもネットワークメタ分析に対する使用が言及されているが、実際に内容を見ると非常に制限的な状況に限定されているということが分かる。カナダとイギリスの指針は、事実上ネットワーク分析は補助的な資料としてのみ提出されるべきであり、単純比較は許容しないものとして記述されており、オーストラリアの場合も、ネットワーク分析は単純な構造の中で行い、IPDが使用可能であれば、MAICやSTCのような方法を使用して補正し、使用された方法論と妥当性について十分に明らかにすべきであると提示されている。

カナダの場合は、Cooper et al. (2015) の別途参考文献で方法論的な指針を提示しており、イギリスの場合には、NICE DSU 報告書がそのような役割をしている。Cooper et al. (2015)

では、連結されていないネットワークを含む分析は妥当ではないと提示しており、NICE DSU 報告書は、MAIC 又は STC などの補正方法について踏み込んだ長短所を提示した上で、依存するには不確かな点が多いことから、提出者は最大限妥当性を明らかにして行い、受け入れられるには困難があり得ることを提示している。

MAIC や STC は、基本的に IPD が使用可能なときに適用できる方法として、グローバル製薬会社の本社の患者資料を利用して分析可能で、当該方法論が状況に適切であれば試みることができるアプローチとして、現在の審評院の間接比較指針は、単純比較の結果提出について開かれているが、単純比較を利用した間接比較の結果を収載対象薬剤の主な効果を判別する根拠として認めるかどうかについては優先的な考慮が必要であるものと思われる。上記で要約したとおり、海外の指針は、ネットワーク分析及び単純比較について指針で言及はしているが、主な分析は、直接比較臨床試験の結果に基づき、ネットワーク分析の遂行は補助的であったり、ネットワーク分析を使用しても統計的仮定のチェックが可能な単純構造及び連結されたネットワークの使用に制限している。

現在、審評院の間接比較資料の提出指針においても、「単純比較の結果は共通対照群を通した間接比較の結果と同じ根拠レベルであると認定することはできない」という点と、「少数の研究で補正なしに行われた単純比較の結果は受け入れられないことがある」と既に言及されているが、このような事項を本指針でも明確に言及する必要がある。これと共に、現在、定性的に比較性（交換可能性）を評価して明らかにした上で行った単純比較の結果について、資料提出についても検討を考慮するものとして理解されている部分に対して、さらに厳格な適用がなされるように調整される必要があり、共通対照群がない場合、IPD を使用した適切な統計的な補正などが行われていない資料の提出は認めず、特定の補正方法を適用したとしても、適切性が十分に明らかにされていない場合は、結果が受容されないこともあり得るという方向で審評院の間接比較資料の提出指針もまた、明確に追加修正が行われることが適切であるものと思われる。

ネットワーク分析に関連しても、審評院の間接比較資料の提出指針で、既に 2014 年の改正時に混合比較又はネットワーク分析を考慮でき、「ネットワークを構成しているそれぞれの比較において本指針が提示する枠組みにより交換可能性の仮定を十分に明らかにできるのであれば、混合比較又はネットワーク分析を行うことができるが、十分に明らかにされて裏付けされていない複雑な設計の混合比較の結果は、補助的で追加的な情報として考慮することが勧告される」と言及し、現在の海外の指針と類似する基調の指針を提示しているにもかかわらず、利害当事者の意見では、ネットワーク分析を使用できることに對する言及がないことを指摘しているため、関連内容を本指針の中でも一貫して理解されるように明示することが必要である。

要約すれば、現指針の「プラセボや他の第 3 の治療代案を参照対象として、収載申請医薬品と参照対象、そして参照対象と比較対象を比較した資料を利用することができる」という言及については、二つの代案を直接比較した資料源がない場合には、妥当性を備えた間接比較を通して導出された資料を利用することができるという、さらに広い脈絡の意味に修正して提示した上で、指針説明においてさらに具体的な事項を提示し、提出資料の準備に審評院の間接比較資料の提出指針を参考とするように連係して言及することが必要で、

具体的な指針の追加説明としては、以下のような内容を含むことができる：

1. 臨床的効果に対する判断及び推定のための資料源としては、収載申請薬剤と代案薬剤間の無作為割付を通じた直接比較臨床試験を通して行われた評価結果を優先する
2. 直接比較臨床試験が行われたことがなかったり存在しない場合、二つの薬剤が第3の代案薬剤として連結されたネットワーク構成を通して分析された間接比較結果の資料を利用することができる
3. 代案薬剤間のネットワークを通じたネットワーク分析は、単純な構造の中で一対比較が基本的に提示されなければならない、様々な代案が含まれた複雑なネットワーク分析は、補助的な分析として提示されること
4. 共通対照群に繋がらない単純間接比較は、根拠として受容されず、十分な研究が存在して要約資料としての補正（例：メタ回帰分析）又は患者単位の資料が使用可能で、患者単位の資料を利用した細部的な補正方法（例：傾向スコアを利用した方法など）が適用され、これの適切さが認められる場合に限り考慮される
5. 間接比較に含まれた対象者のベースライン特性、治療の細部事項、結果変数の定義の個別の臨床試験結果を標準化された方法で提示し、間接比較の遂行方法が具体的に記述されていること
6. 資料の提出方法は、間接比較資料の提出指針を参考にして下記の内容が提示されること：
 - 間接比較が行われた理由及びネットワーク構造に対する説明
 - 含まれた臨床試験の研究特性の評価及び報告
 - 結果を要約するために使用された測定値に対して記述
 - 潜在的な歪みの可能性及び探索と補正のために使用された方法に対する具体的な記述
 - データの検索及び処理方法、分析の遂行及び結果の合成方法を説明
 - ネットワークで直・間接根拠間の一貫性を評価するための統計的なアプローチ方法を記述

キ．費用

1) 現指針

ア) 費用項目 (identification)

(1) 主な指針

どの項目を費用に含めるかは、分析の観点によって異なる。社会的観点を取れば、費用の支払い主体が誰であるかと関係なく、医薬品の治療と関連したすべての適切な費用を分析に含む。本指針では社会的観点を取るものの、その領域及び包括性は制限的であるため、分析に含まれる項目としては、直接保健医療費用、患者及び家族（又は介護人）費用、その他に社会の他の部門で発生する費用などがある。

罹患及び死亡による生産性損失費用（productivity cost）は基本分析には含まず、当該項目を別途提示する。当該疾病の治療と関連しない将来の医療費用（unrelated future medical

cost）もまた含むかどうかと関連して論議の余地があるが、基本分析には含まない。

臨床試験自体に必要とされた費用として、実際の診療過程では発生しない費用や移転費用などは概念上でも社会的費用に含まれないので、分析から除外する。

保険者観点で分析する場合、健康保険給付部門と本人負担金を含み、非給付項目は基本分析に含めずに別途提出する。

(2) 指針内の追加説明

直接保健医療費用としては、医薬品（治療薬及び副作用治療）、診療サービス、病院（入院）サービス、診断及び検査、その他の保健医療費用などが含まれる。患者及び家族（又は介護人）費用には治療を受けるために医療機関を訪問する過程で患者やその保護者が支出する交通費、訪問及び治療待機時間の間の時間費用、家族や介護人の介護と関連した時間費用などが含まれる。他の社会部門で発生する費用は、社会福祉サービスなどの保健医療部門以外の他の部門で発生する費用を意味する。

罹患及び死亡による生産性損失費用（productivity cost）は、概念上では社会的費用に該当するが、二重計算の問題、正確な規模を測定しにくいという問題などで、実際の分析に含めるかどうかは議論となる項目である。

...生産性損失費用を含むことに反対する理由のうちの一つは、費用-

効用分析をする場合、既に健康状態を反映する効用加重値（utility weights）に生産性損失に対する価値が反映されていると見ることができるので、これを再び費用計算に含むことになれば、同一の価値が分母、分子ともに集計される二重計算の問題が発生する可能性があるためである（Luce, et al., 1996; Garber, et al., 1996）。

...そして、生産性損失費用を含むとしても測定する方法が不完全であるという問題点がある。伝統的に人的資本アプローチ法（human capital approach）を通して生産性損失費用を測定するが多かったが、これは社会全体が完全雇用状態ではない場合、過大推定の問題が発生する。

分析対象疾病や治療と直接的関連がない未来の医療費用（unrelated future health care costs）の場合も含むかどうかが議論となる項目である。すなわち、疾病治療で延びた生存期間の間、現在の疾病治療と関係がない疾病が発生して、これによって保健医療リソースを消費するようになった場合、この部分を分析に含めるべきかどうかという問題である。

...本指針では、理論的には論議が繰り返されている事案ではあるが、多くの研究でこの部分を分析から除外しており、他の経済性評価指針でも対象費用の範囲から除外していることを勘案して、当該疾病の治療と関連しなかった未来の医療費用は、基本分析から除外する。

経済性評価に含む費用項目は、観点により異なる。今回の経済性評価指針の改正では、既使用していた制限的・社会的観点の代わりに保健医療システムの観点を提案し、基本分析には直接医療費用のみを含み、直接非医療費用（交通費用、時間費用、保護者介護費用）は除くことを提案した。非医療費用の比重が大きい場合には、生産性損失費用と同様に別途資料で提示できるようにした。費用項目に関連した経済性評価指針の改正は、このような変化を反映する線で一部表現が修正されるものと思われる。

イ）費用算出（measurement and valuation）

（１）主な指針

費用は、費用として挙げた各項目の使用量（自然単位）に単位費用を掛けて算出するものとする。各項目の使用量をどの単位で測定したか、そして単位費用をどのように算出したかについて詳しく記述する。費用は、分析時点を基準として、一貫して補正して算出し、単位費用とリソース使用量の資料源は、国内の資料を使用する。

費用算出の具体的な方法については『標準費用算出基準（案）』を参考にする。

（２）指針内の追加説明

費用を算出する場合には各項目の使用量をどんな単位で測定するのか、そして単位あたりの費用ではどんな値を使用するのかを決めなければならない。リソース使用量の資料源は、臨床研究、請求資料、臨床診療指針、そして専門家の意見などによることができる。測定時点により、医療利用形態、単位費用などが変わることがあるため、費用の算出は、分析時点を基準として、一貫して補正することが望ましい。また、単位価格及び医療利用の様相は、国間で差があるため、医療環境が全く異なる国で測定された結果を国内で一般化させることが懸念される。万一、研究目的に符合する国内の資料源が複数存在して資料提出者が選択できる場合、可能であれば、最近算出された代表性のあるサンプルから算出された値を選好する。これは、最近の資料が国内の医療利用形態及び単位価格を最も現実的に反映でき、代表性のある大規模資料から算出された場合、さらにそうであるからである。

市場賃金の資料がない人に対する時間費用（家族が患者を介護する場合）の処理には、大きく機会費用法（当該人材が労働市場に投入されたときに受けることができる賃金水準を基礎

とする)と代替費用法(提供するサービスの市場価値に基づく)がある...本指針では、機会費用法を使用することを提案する。

費用を算出する際、同じ項目で同じ使用量であっても単位費用をいくらにするかによって費用の算出値が変わることがある。したがって、「標準費用算出基準(案)」を参考にして費用算出過程を標準化し、異なる研究間の比較可能性を高める。費用算出指針では、項目別の費用算定方法と共にリソース使用量の評価単位及び単位費用に対する細部指針が提示されている。

経済性評価のための費用算出と関連して現指針では、費用各項目の使用量(自然単位)に単位費用を掛けて算出するが、費用項目それぞれの使用量をどの単位で測定したか、そして単位費用をどのように算出したかについて詳しく記述し、資料源は国内の資料を使用するようにする。リソース使用量の資料源には、臨床研究、請求資料、臨床診療指針、そして専門家の意見などがあるが、これら資料源の中でどれを使用することがより良いかについて検討が必要である。あわせて、抗がん剤などの使用が増えるにつれ、注射投薬後に残る量、すなわち廃棄分(wastage)に対する考慮の有無によって投薬費用が変わることがあるので、これについて検討が必要である。

2) 既存の提出資料の検討結果

ア) 費用資料源

既に審評院に提出された50件の経済性評価資料から直接医療費用のうち、薬価以外に検査費用、治療費用、副作用管理費用などの資料源を調べた。

検査費用の場合、健康保険の報酬ファイルを使用した場合(26件)が最も多く、専門家対象調査(7件)、病院調査(2件)、健康保険の標本資料やコホート資料を利用した場合(2件)などがあった。治療費用の場合、健康保険の報酬ファイルを使用した場合(17件)と審評院の統計資料を使用した場合(10件)が多く、専門家対象調査(5件)、先行研究(3件)、健康保険の標本資料やコホート資料を利用した場合(2件)などがあった。

副作用管理費用を含む抗がん剤は全14件であったが、審評院の統計資料を使用した場合(6件)が最も多く、健康保険の報酬ファイル(3件)、臨床研究(2件)、専門家対象調査(2件)の順であった。

<表22> 既に提出された経済性評価資料に使用された直接医療費用の資料源

区分	検査費用	治療費用	副作用管理費用 ¹⁾	その他の直接医療費
臨床研究	1	1	3(2)	2
審評院の統計資料	1	10	7(6)	5
報酬ファイル	26	17	4(3)	6
健康保険の標本、コホート資料	2	2	-	-
臨床ガイドライン	1	0	-	-

先行研究	0	3	-	7
専門家調査	7	5	2(2)	2
病院調査	2	1	-	1
その他(該当なし、不明確)	10	11	2(1)	27
計	50	50	18(14)	50

注：1) 副作用管理費用を含む研究は一部である。()内の値は抗がん剤の数字である。

イ) 投薬廃棄分 (wastage)

50件の経済性評価資料において投薬廃棄分 (wastage) に対する考慮の有無を把握した結果、当該事項がある23件中8件で廃棄分 (wastage) に対する考慮を行ったが、非抗がん剤よりは抗がん剤 (7件) が主となった。

<表23>既に提出された経済性評価資料における投薬廃棄分 (wastage) の反映現況

区分	抗がん剤	非抗がん剤	全体
廃棄分 (wastage) 考慮	7	1	8
廃棄分 (wastage) 未考慮	10	5	15
該当なし	6	21	27
計	23	27	50

3) 外国の指針

ア) カナダCADTH

□ 費用資料源

- リソース使用量と単位費用の資料を選択する際に使用可能な資料源について目的の符合度、信頼性及び一貫性の側面から長短所を評価しなければならない。
- リソース使用量の資料源として行政資料や臨床研究資料があり得る。
 - 行政資料は、現実的な情報 (real-world information) を提供するだけでなく、特定患者コホートのリソース使用を追跡観察する際に有用である。
 - 臨床研究資料を利用する際は、プロトコルによるリソースの使用を念頭に置かなければならない。臨床研究で示されたリソースの使用はプロトコルに従ったが、カナダの診療パターンを反映しなかったり、対象人口集団を反映しない場合は、他の資料源を求めなければならない。

□ 投薬廃棄分 (wastage)

- カナダの指針には、投薬廃棄分 (wastage) に対する内容は具体的には出ておらず、CADTH で発行した保健医療費用算出案内書に下記のように簡略に言及されている。
 - 医薬品の廃棄分 (wastage) が発生する場合 (例えば、処方された総投薬量が1バイアルに入っている量よりも少ない場合)、1バイアル全体の費用を基本分析に含めなければならない。

イ) オーストラリアPBAC

□ 費用資料源

- オーストラリアの指針では、リソース使用量を求める際、その根拠を正当化して資料源を明らかにするようにしている。そして、臨床研究の過程でリソースの使用のパターンを前向きに測定しなければならないが、その際、後向きに医務記録を検討したり請求資料と関係する方法、アンケート調査をする方法、日誌を活用する方法を適用することができる。
- 一次的な根拠から出たリソース使用量の資料と、一次資料として使用可能な範囲を超えたリソース使用量に対する外挿やモデリングの結果は区分しなければならない。

□ 投薬廃棄分 (wastage)

- オーストラリアの指針には、投薬廃棄分 (wastage) を考慮するよう明示されている。
 - すなわち、提供された薬やリソースの量が実際のリソース使用量よりさらに適切な経済的尺度であり、廃棄分 (wastage) が実際に費用を誘発するためである。

ウ) イギリスNICE

□ 費用資料源

- 医薬品については、公的な薬価リストの資料を基本分析に使用しなければならない。全国的に薬価の引き下げが行われた場合、例えば NHS Commercial Medicines Unit が二次診療の使用目的で交渉を通して購買契約した医薬品の場合には、NHS に適切な価格反映のために引き下げされた価格を基本分析に使用しなければならない。
- Healthcare resource groups (HRGs) は、共通レベルの保健医療リソースを使用する類似の臨床治療を標準グルーピングしたものであるが、これはリソース使用量に対する重要な資料源である。特定 HRG の全国平均単位費用は、イングランドのすべての NHS 機関で準拠費用 (reference cost) を毎年義務収集するものの一部として報告される。この費用を活用すれば、マイクロコストイング (micro-costing) の必要性を低減することができる。

□ 投薬廃棄分 (wastage)

- イギリスの指針に投薬廃棄分についての具体的な言及はない。

4) その他の文献の考察結果

□ Guidance Document for the Costing of Health Care Resources in the Canadian Setting: Second Edition (CADTH, 2016)

- 医薬品の廃棄分 (wastage) が発生する場合 (例えば、処方された総投薬量が 1 バイアルに入っている量よりも少ない場合)、1 バイアル全体の費用を基本分析に含めなければならない。

5) 利害関係者の意見

□ Scoping意見調査の結果

- 回答者は、主に経済性評価の観点と関連がある費用項目を中心に答え、費用資料源や投薬廃棄分 (wastage) に対して言及した事例はなかった。
- 製薬会社の回答者 (9 名) 中 5 名 (観点から関連意見を提示した 1 名を含む) が費用項目に対する意見を提示したが、費用の種類別に見れば、介護費用 (2 名)、生産性損失費用 (2 名)、副作用管理費用 (1 名) に関するものである。
 - 介護費用の場合、疾患の特性により保護者の介護及び保護が必然的な場合、これを含む必要があり、生産性損失費用の場合は、疾患の特性により雇用率が高い集団や小児患者に発病して早期に死亡に至る場合に含まれる必要があるとした。
 - 副作用管理費用の場合、申請薬の導入により副作用の発生が減少した際に費

用が反映されないとした。

- 研究者（3名、すべて経済性評価の遂行経験あり）のうち、1名が費用項目の名称を韓国保健医療研究院で発刊した費用算出マニュアル（保健医療分野で費用算出方法）での費用項目と一致させる必要性を提起した。他の1名は、交通費用、時間費用などでの不確実性も大きく、含める必要性がよく分からないとした。

6) 指針の改正方向の提言

経済性評価の観点を制限的・社会的観点から保健医療システム観点に変更することにより、基本分析に含めなければならない費用項目の変化が必要である。直接保健医療費用以外に患者及び家族（又は、介護人）費用、その他に社会の他の部門で発生する費用などは基本分析から除外し、必要な場合は別途提示できるようにする。

既存の指針では、費用やリソース使用量は国内資料を使用し、資料源は臨床研究、請求資料、臨床診療指針、そして専門家の意見などによることができるとしている。

既に提出された経済性評価を調べれば、上記の資料源のうち、専門家の意見が検査費用で7件（14%）、治療費用で5件（10%）を占め、少なくない比重を占めていることが分かる。

専門家の意見や病院の調査資料に基づいた費用及びリソース使用量は、データ基盤の資料源に比べて信頼性及び一貫性の面で限界がある。したがって、専門家の意見などを資料源とする場合には、公式な学会の意見を含めたり、専門家パネルの最小人員を定めて、信頼性及び一貫性を向上させる必要があり、これを改正指針に簡略に含むことを提案する。

既存の指針には、投薬廃棄分（wastage）に対する内容はないが、廃棄分もまた保健医療システム内で費用を発生させる要因であるため、これを含む費用を基本分析に使用するように提案する。

ア. 効用、健康に関する暮らしの質

1) 改正が必要な項目

現在、医薬品に関する経済性の評価・効用に関する指針の内容は、表24の通りである。健康保険の審査評価員、保険経済の専門家、薬剤師、市民団体などの意見を吸い上げ、改正の必要な項目を下記のように選定した。

- 効用測定方法のプライオリティ
- HRQoL（Health-Related Quality of Life）に関する道具及びtariffの選定を検討
- 算式（mapping）に関する詳細な指針
- 直接的な測定に関する詳細な指針
- 効用のMCID（Minimal Clinically Important Difference）を追加検討

<表24> 医薬品の経済性評価における効用に関する指針の内容と改正が必要な項目

項目	2011年の指針内容	改正方向の検討
健康状態の測定対象群	直接的なコメントなし（暗黙的な患者）	-
健康状態の測定資料元	コメントなし（効果資料元[RCTを嗜好]と同様に推定）	-
健康状態の効用測定		
健康状態の選好度測定対象群	韓国の一般大衆	-
健康状態の効用を測定する方法に関するプライオリティ	MAUI患者の資料に、既存のtariffを適用し、間接的に質の加重値を算出する方法を言及。健康状態のシナリオを構成し、提出者が直接、質の加重値を測定する方法をコメント。この両代案におけるプライオリティが明確に表現されていない。 他にも、既存の文献から提示した値の引用及び算式（mapping）による質の加重値算出を提示。	測定方法のプライオリティを提示
[間接測定] HRQoLの測定道具	選好度に基づいた一般道具の使用を推奨。国内では、妥当性の検証も行われており、公認された道具を用いるようにしている。選んだ道具に関する正当化や敏感度分析の必要性を明示。 *疾病別の道具、選好度に基づいていない道具を用いた結果は、追加的にのみ提出可能。	-

[間接測定] Tariffの選定	複数のTariffが存在する場合、特別な理由がない限り、代表性のある最新値の使用を明示	基本分析にて用いる（道具及び） tariffの選定可能性に関する可否を検討
[直接測定] 選好度の算出方法	選択に基づいた方法（TTO or SGなど）を推奨	-
[直接測定] 健康状態に関するシナリオ構成への適切性の評価	シナリオ構成及び内容評価における考慮事項に関し、別添にて簡単に提示	シナリオの構成方法及び適切性の評価に関する詳細な指針を追加
既存の文献を引用	必要に応じて使用可。資料元の包含・排除に関する基準を明示。透明に報告し、不確実性への評価が必要。	- (現在の指針を維持するが、様々な健康状態に関する質の加重値が必要な場合、複数資料元からの引用は望ましくないというコメントが必要、文献の考察及び選定過程に関する透明性及び正当化を強く強調)
算式（mapping）による効用の算出	選好度に基づいた一般道具の使用を基本にする。必要に応じて算式（mapping）を用い、選好度に基づいた値に差し替えた上、結果値の「追加的な」提出は可能。算式（mapping）に関する限界及び問題点を中心に記述	算式（mapping）に関する詳細な指針を追加
効用のMCID	コメントなし	追加的な検討
様々な健康状態が重なる際の効用を計算	コメントなし	-

この章では、各項目に関し、国内の製薬会社による提出資料の現況、主な国における経済性への評価指針、利害関係者の意見、他の関連した文献などを調べ、これを基に効用に関する指針の改正方向について述べたい。⁷⁾

1) QALY（quality-adjusted life years）及び効用の測定に関する理論的な背景は、既存の改正研究レポート（Bae Seungjinなど、2012）を参考する。

2) 効用測定方法のプライオリティ

ア) 現在の指針

(1) 主な指針

質の加重値の測定に用いられた道具は信頼性、妥当性が検証されたものでなければならない。とりわけ、外国にて開発された道具を用いる場合は、言語や文化的な差に留意し、国内にて妥当性が検証された道具を用いること。

道具により、結果値に差が生じかねないため、研究者は自ら選択した道具を当指針の別添3、表4に基づいて正当化すると共に、敏感度の分析などを通じ、質の加重値の不確実性が結果へ及ぼす影響を検討すること。

資料の提出者は、調査方法に関し、詳細に報告すること。1)基本的な分析では、選好度に基づいた一般道具の使用を推奨し、疾病別の道具、選好度に基づいていない道具2)を採用して測定した結果は追加的に提出することができる。

健康状態に関する選好度の評価対象を明示する。健康状態に関する選好度の測定対象は、国内の代表性のある一般大衆を選好する。

もし、既存の文献から質の加重値を引用する場合は、資料元の包含・排除に関する基準を詳しく説明し根拠を提示するなど、選定の過程を透明に報告すること3)。そして、該当の健康状態に適用可能かの可否を議論し、不確実性を評価する。

- 1) 別添5に基づく
- 2) 選好度に基づいた道具をもって変換 (transformation) し、この際、変換に向けた算式 (mapping function) の構築方法論、歪みの発生可能性に関して詳しく記述後、同方法論を正当化し、模型の予測性及び適合度に関して記述する。
- 3) 2-4-2の“既存値を引用する際、報告すべき項目”、別添4、表5の項目を順守する。

(2) 指針における追加的な説明

選好度に基づいた道具は、それぞれのメリット・デメリットがあり、どの道具が優れているかなどは報告されていない。ただ、直接測定道具の場合、視覚的アナログ尺度 (Visual Analogue Scale、以下VAS) 方法などのRS方法は、単独使用に懸念があると報告されている (別添2を参照)。従って、直接測定道具を用いる場合、標準機会選択法 (standard gamble)、時間交換法 (time trade-off) などの選択に基づいた (choice-based) 道具をもって測定することを推奨する。

既存の間接方法の道具は、殆どが我々と言語的・社会経済的・文化的な背景が違う国にて開発されたものであるため、注意して用いる必要がある。外観妥当性 (face validity)、内容的妥当性 (content validity)、翻訳の妥当性 (validity of the translations) のみならず、道具

これらの道具を用いる際は、国内にて妥当性への検証が行われ、公認された道具を採用するようにする。

個人の選好度が反映されないSF-36、Nottingham Health Profile (NHP)、Sickness Impact Profile (SIP) などのような道具及び選好に基づいていない疾病別の道具は算式 (mapping) を用いて選好に基づいた値に変換 (transformation) する方法がある。しかし、こうした算式 (mapping) は大体、予測可能性 (predictability)、モデル適合度 (model fit) において相当な変異があるという根本的な限界を持っていると報告されており ((Brazier et al, 2007)、Chancellorなど (1997) は算式 (mapping) を変換することが根本的に不可能だと提示しているなど、方法論的な問題が存在している。

QALY (quality-adjusted life year) の計算に用いられる質の加重値は、各健康状態が個人へ与える効用の程度を測定したものと、現在の指針としては個人の選好度を反映することを勧めている。指針では、質の加重値に関する測定方法として間接測定、直接測定、既存値の引用を提示しており、それぞれの方法の内容は下記のとおりである。参考として、これらへの相対的な選好度は明確に提示されていない。

- 間接測定：選好に基づいた一般道具を用いて患者の健康状態を測定し、この資料に一般大衆の選好を反映した既存のtariffを適用して質の加重値を算出する方法。
- 直接測定 (シナリオ基盤の効用を測定)：提出者が健康状態のシナリオ (vignette) を構成し、選択に基づいた (choice-based) 道具を用いてこの健康状態に関する質の加重値を直接導き出す方法。
- 既存値の引用：既存の文献から提示された質の加重値を引用。

また、現在の指針では選好に基づいていない疾病別の道具を用いて、患者の健康状態を測定した場合は算式 (mapping) を用いて、選好に基づいた値に変換 (transformation) することもできると述べているが、同方法に関しては主に限界を中心に記述している。

イ) 既存の提出資料の検討結果

2014~2011年の5年間で健康保険審査評価院 (以下、審評院) に健康保険への登載の目的で提出された経済性の評価の資料50件の中、費用-効用の分析 (cost-utility analysis、CUA) 47件を対象に、分析を実行した。分析には、抗癌剤21件、希少・難治性疾患の治療剤8件、その他の一般薬剤18件が含まれた。

(1) 質の加重値の測定方法

- (2) 表25から提示された通り、様々な質の加重値に関する測定方法が混在して用いられていた。計47件のCUAの中、既存の文献から提示された質の加重値を引用したケースが23件 (48.

9%)で最も高い頻度を記録、その次はシナリオに基づいた直接測定(9件)、間接測定(6件)、算式(mapping)の利用(6件)、その他(3件)の順であった。既存値を引用した理由としては、暮らしの質に関する資料の不在(8件)もしくは不十分(5件)、疾病別の道具使用(7件)、直接比較可能な臨床試験の不在(3件)などがあった。

- (3) 抗癌剤(21件)だけで絞って分析する場合、既存値の引用が23件(48.9%)から6件(28.6%)に大幅減し、他の測定方法が占める比重が相対的に増加した(表26)。直接測定と間接測定がそれぞれ5件、算式(mapping)の利用が3件、その他2件が用いられた。

<表 25> 2014~2018年、国内にて提出されたレポートにて用いられた質の加重値に関する算出方法

タイプ	N	百分率
間接測定	6	12.8%
EQ-5D-3L	4	-
EQ-5D-5L	2	-
直接測定	9	19.1%
TTO	7	-
SG	2	-
既存値の引用	23	48.9%
算式(mapping)	6	12.8%
その他	3	6.4%
総計	47	100.0%

※ 主な基本分析を基準に分類
SG: standard gamble, TTO: time trade-off

<表 26> 2014-2018年、国内にて提出されたレポートで用いられた質の加重値に関する算出方法 (抗がん剤)

タイプ	N	Percent
間接測定	5	23.8%
直接測定	5	23.8%
既存値の引用	6	28.6%

算式 (mapping)	3	14.3%
その他	2	9.5%
総計	21	100.0%

※ (最初の提出資料の) 主な基本分析を基準に分類

(4) 国内人の選好度 vs. 外国人の選好度

- (5) 国内の一般大衆の選好を反映し、質の加重値を測定した事例は20件 (42.6%) として、外国人の選好を反映して質の加重値を測定した事例 (27件) より、少し低いことが分かった (表 27)。これは、既存の文献を引用する場合、殆どが外国の文献を用いたり、国内の資料と混合して用いたりしたためである。既存値を引用した23件の中、ただ2件のみが韓国の一般大衆の選好を反映した質の加重値を用いた。
- (6) 算式 (mapping) の場合も、外国にて算式 (mapping) が開発されたケースが多いため、大体、外国人の選好を反映したEQ-5D点数に変換して分析に用いた。間接測定や直接測定を用いた場合は、大体、韓国の一般大衆の選好を反映して質の加重値を算出した。

<表 27> 2014-2018年、国内にて提出されたレポートに用いられた選好度の測定対象

タイプ	国内人	外国人	全体
間接測定	6件	-	6件
EQ-5D-3L	3件 (一般大衆 : Lee et al., 2009) 1件 (コメントなし)	-	4件
EQ-5D-5L	2件 (一般大衆 : Kim et al., 2016a)	-	2件
直接測定	8件	1件	9件
TTO	6件 (一般大衆)	1件 (イギリスの一般大衆)	7件
SG	2件 (一般大衆)	-	2件
既存値の引用	1件 (一般大衆 : Lee et al., 2009) 1件 (一般大衆 : Jo et al., 2008)	21件 国外若しくは国内外の 混合	23件
算式 (mapping)	2件 (一般大衆 : Lee et al., 2009)	イギリスの一般大衆 3件 1件は不確実	6件
その他	3件 (一般大衆 : Lee et al., 2009)	-	3件
総計	20件 (42.6%)	27件 (57.4%)	47件 (100.0%)

※ (最初提出資料の) 主な基本分析を基準に分類
SG: standard gamble, TTO: time trade-off

ウ) 外国の指針

カナダ (CADTH、2017)、オーストラリア (PBAC、2016)、イギリス (NICE、2013) などをはじめとするほとんどの国で、評価の一貫性、比較の可能性、資料収集の容易性などの理由で、間接測定を好んでいることが分かった (表28、表29)。

(1) カナダ (CADTH、2017)

カナダの指針では、一般道具を用いた間接測定方法を基本分析 (reference case) として採用することを推奨している。健康状態のシナリオに基づいた直接測定方法の場合、設計及び実行が困難な上、健康状態記述 (description) の妥当性により、質の加重値が大幅に変わり得ると述べた。算式 (mapping) を用いた質の加重値の算出は推奨しないが、算式 (mapping) のほかに代案がない場合は可能だと明示している。但し、この場合は算式 (mapping) の妥当性を裏付ける実証資料 (empirical data) を提出するようにしている。

間接測定が困難な場合、既存の文献から提示した値を引用することもできるが、この際はその過程と方法について詳しく記述するようにしている。

(2) オーストラリア (PBAC、2016)

オーストラリアの指針においても、MAUI (Multi-Attribute Utility Instrument) で測定した患者の健康状態に関する資料に、オーストラリアの一般大衆を対象に算出したtariffを適用し、質の加重値を算出する間接測定を推奨している。8) 一般道具のほか、PRO (patient-reported outcomes、被験者の主観的な指標) の道具が臨床試験にて用いられた場合、算式 (mapping) を用いて選好に基づいた一般道具値へ変換することを提示している。この他にも、健康状態のシナリオ (vignettes) に基づいた直接測定、臨床試験に含まれた患者らと類似な集団を別途に募集し、彼らを対象に間接測定で質の加重値を算出する方法 (population matching study)、既存値の引用などを代案と簡単に記述している。資料の提出者は使用した方法について、詳しく記述し正当化することを明示している。

(3) イギリス (NICE、2013)

イギリス・NICEの指針でも、一般道具をもって測定された資料にイギリスの一般大衆の選好を反映したtariffを適用して、質の加重値を算出する間接測定を推奨している。ただ、他の指針とは違い、評価の一貫性のために、大人を対象にする場合はEQ-5D-3Lの使用を選好すると明示した。EQ-5D-3L資料がない場合は、算式 (mapping) を用いてEQ-5D-3L点数に変換することもでき、透明かつ体系的な文献の考察により、既存の文献から提示した値を引用することもできると述べている。

2) オーストラリアの指針では、カナダやイギリスの指針のように間接測定を「reference」と明確に明示してはいないが、全般的に間接測定を推奨していると記述している。

NICEの本指針では、健康状態のシナリオ (vignette) に基づいた直接測定方法に関しては何も述べていないが、NICE DSU TSD 11 (Brazier & Rowen, 2011) では検証されたHRQoL道具に基づいていない直接測定はNICEの基準に合わないと明示した。ただ、検証された道具をもって測定された資料がない場合は直接測定が制限的役割 (limited role) を担当する可能性もあると述べている。

(4) その他

EUnetHTA (2015a) から、経済性評価の指針若しくは類似な情報を提供しているヨーロッパの25か国の指針を分析した結果、デンマークとスウェーデンを除いたほとんどの国が間接測定を 선호していることが分かった (表 29) 。さらに、最近、開発した日本の指針においても、選好度に基づいた一般道具と日本にて開発したtariffを用いて質の加重値を算出する間接測定を推奨しており、直接測定はあまり好まれない (less preferred) と明示している (Shiroiwa et al., 2017) 。

<表 28> 効用測定方法のプライオリティに関する海外の主な国の指針

国	指針内容の原文	サマリー
カナダ CADTH 指針 (4版) (2017)	In the reference case, researchers should use health preferences obtained from an indirect method of measurement that is based on a generic classification system (e.g., EuroQol 5-Dimensions questionnaire [EQ-5D], Health Utilities Index [HUI], Short Form 6-Dimensions [SF-6D]). Researchers must justify where an indirect method is not used. (p20)	一般道具に基づいた間接測定を基本分析に使用。他の方法を用いる場合は正当性が必要。
オーストラリア・PBACの指針 version 5 (2016)	If a MAUI has been used in an included study to estimate utility weights (as described in Subsection 2.4.3), state where and when the scoring algorithm was derived, and consider how applicable it is to the general Australian population. It is preferred that Australian-based preference weights are used in the scoring algorithm used to calculate utility weights. (p77) Where utility weights or QALY changes cannot be directly estimated from data collected in the clinical studies from Section 2, or there are significant concerns about the reliability and relevance of trial-based utility, transform the Section 2 health outcomes to estimate QALY gains. (p77) Describe the source(s) and method(s) (as described below) used to derive externally derived health state utilities, and justify their inclusion in the model. (p78)	MAUIが用いられた場合、点数体系のアルゴリズムがいつ、どこで算出されたかを記述。オーストラリアの一般大衆の選好を好む。 MAUI臨床資料の使用が困難な場合、QALYを推定するために他の資料を利用。 外部から算出された質の加重値を用いる場合、出典と方法を記述&正当化が必要。
イギリス NICE 指針 (2013)	For the reference case, the measurement of changes in health-related quality of life should be reported directly from patients and the utility of these changes should be based on public preferences using a choice-based method. The EQ-5D	間接測定を基本分析に使用。大人の場合、EQ-5Dを 선호。

	is the preferred measure of health-related quality of life in adults. (p40)	HRQoLの変化測定の対象は患者であるべき。
--	---	------------------------

表 29> EUnetHTAのパートナー国の分析：質の加重値測定に関し、好む方法と道具

国	選好する方法と道具（間接測定 vs. 直接測定）
オーストリア	間接測定（HUI、EQ-5D）
ベルギー	間接測定（EQ-5D）
クロアチア	間接測定（EQ-5D）
チェコスロバキア	間接測定（EQ-5D）
デンマーク	直接測定（TTO、SG）
イギリス（England）	間接測定（大人の場合、EQ-5D）
エストニア & ラトビア	間接測定（EQ-5D、HUI）
フィンランド	検証された一般HRQoL道具
フランス	間接測定（フランスのtariffに存在するEQ-5DとHUI-3）
ドイツ	好む方法なし
ハンガリー	間接測定（効用に基づいた（utility-based）HRQoLアンケート）
アイルランド	間接測定（EQ-5DやSF-6Dのような選好に基づいた一般道具）
イタリア	間接測定（EQ-5D）
オランダ	EQ-5Dを好むが、他の道具も十分に可能（EQ-5D-5Lの基本分析）
ノルウェー	間接測定（EQ-5D、SF-6D、15D）
ポーランド	間接測定（EQ-5D）
ポルトガル	ポルトガルにて検証された道具、選んだ道具に関する正当化が必要
ロシア	コメントなし
イギリス（Scotland）	間接測定（EQ-5D）
スロバキア	間接測定（EQ-5D）
スロベニア	コメントなし
スペイン	CatSalut：間接測定（EQ-5D、SF-6D） Osteba：直接若しくは間接測定
スウェーデン	直接測定（SG、TTO）、間接測定（EQ-5D） 十
スイス	該当なし

† 源の出典には、間接測定は含まず。ISPORサイト (<https://tools.ispor.org/peguidelines>) を基にアップデート
出典：EUnetHTA
(2015a)SG: standard gamble, TTO:
time trade-off

工) その他の文献考察の結果

先述のように、モデルから定義した健康状態の質の加重値を推定する方法は、次のように大きく4つに分けることができる。

- ① 間接測定
- ② 直接測定
- ③ 算式（mapping）を用いて選好度に基づいた値に変換
- ④ 既存の文献から提示した値の引用

間接測定は、multi-attribute utility instrument (MAUI) 9) を用いて、健康の領域別に患者の健康レベルを測定し、この測定資料に一般大衆を対象に前もって算出した健康領域及びレベルに関する選好度（点数体系）を適用し、患者が経験している健康状態の効用を単一の数値（質の加重値）で推定する方法である。この際、質の加重値は0（死亡）と1（完全に健康）を基準にするが、道具によっては死亡よりも悪い状態を意味するマイナスの値が算出されることもある。代表的なMAUIとしては、EuroQol 5-dimension (EQ-5D)、Short Form-6-dimension (SF-6D)、Health Utility Index (HUI-2、HUI-3)、Quality of well-being (QWB)、Assessment of Quality of Life (AQoL)、15Dなどがある (Brazier et al., 2017) 10)。

殆どのMAUIは標準化されているため、比較可能性 (comparability) が高く、政策の決定者が一貫した決定を下すのに役立つ上、使用が容易い側面がある。さらに、対象患者群が経験する効果と副作用を評価に直接反映することができると共に、国内の一般大衆の選好度を反映して客観性と一貫性を確保することができる。しかし、道具により算出された質の加重値に差が発生しかねないが、これは道具が記述している健康エリア及びレベルが異なるためでもあるが、選好度（点数体系）を導き出す際に用いた方法が異なるためでもある。たとえば、EQ-5Dの場合は通常TTOを用いて選好度（点数体系）を導き出すが、HUIの場合は通常SGを用いて選好度を導き出す。

直接測定は、文献の考察、臨床資料、専門家の意見などを基に、健康状態に関するシナリオ (vignette) を構成し、SGやTTOなどの選択に基づいた方法を用いて一般大衆を対象にこれに関する質の加重値を直接測定する方法である。一般人を募集し、別途の研究を実行しなければならないため、時間がかかると共に複雑である。また、健康状態がどのように記述されるか、どのような質の加重値の算出方法が用いられるかにより、質の加重値が大幅に変わることがある (Brazier & Rowen, 2011)。臨床試験に参加した様々な患者の経験を評価に直接反映できないことも短所とされる。

Arnold et al. (2009) は間接測定 (EQ-5D、SF-6D、HUI) と直接測定 (SG、TTO) により算出された

3) 選好に基づいた一般 HRQoL 道具と一般MAUIを混用して用いた。但し、指針や文献を引用する際は、元の文献にて用いられた用語をそのままに利用した。

4) この道具に関する詳細は、『3) HRQoL道具及びtariffの選定検討』の章を参照。

質の加重値を比較するために、両方法を全て用いた研究を対象に体系的な文献の考察を行った。計32の研究と、これに含まれた直・間接的な測定83件が最終的な考察に含まれた。直接測定の場合、SGが57件、TTOが60件でほぼ同水準で用いられ、間接測定の場合はEQ-5D 67件、HUI-3 37件、SF-6Dが13件、HUI-2 5件の順と調査された。平均的な質の加重値はSG 0.81、TTO 0.77、EQ-5D 0.59、HUI-3 0.68、SF-6D 0.63、HUI-2 0.75として、道具別に質の加重値への推定に差があった。研究チームは、間接測定より直接測定のほうが、質の加重値が相対的に大きく測定されるきらいがあると報告、その理由としてMAUIの硬直性（例えば、肯定的な面を追加で報告できない、自分の健康状態を無条件にいくつかの範疇の一つに分類）と選好度の測定対象の差を述べた。文献の考察に含まれた直接測定の研究では、大体、患者を対象に選好度を測定したが、間接測定の研究では一般大衆を対象に選好度を測定したため、こうした差が発生した可能性があるとして報告した。

Briggs et al. (2019) は、大規模の臨床試験から収集されたEQ-5D-3L資料をもって算出した質の加重値とこの臨床試験にて経験する患者の健康状態を反映したシナリオに関し、TTOを用いて算出した質の加重値を比較・分析した。健康状態のシナリオは副甲状腺機能亢進症で治療を受けている末期の心不全患者が経験する様々な症状を反映して作成、両方法共にイギリスの一般大衆の選好を反映した。分析の結果、直接測定よりEQ-5D-3Lを用いた間接測定にて質の加重値がより高く表れ、Arnold et al. (2009) が報告した結果とは違う傾向となった。

両研究における結果の差は明確には把握できないが、Briggs et al. (2019) では、直接測定は一般大衆を対象にしたため、当健康状態に適応した患者より時間の交換をしようとする傾向がより大きく表れた可能性がある。こうした結果の差にもかかわらず、両研究にて共通的に指摘する問題は道具間の質の加重値の差である。直接測定と間接測定においても差は発生したが、同じ測定方法においてもどの道具を用いたかにより、質の加重値の差が発生した。

直接測定と間接測定において、何を選択すべきかに関する指針を提供できるような普遍妥当な理論的な基礎はない (Gold et al., 1996; Arnold et al., 2009; Briggs et al., 2019)。ただ、直接測定を使用するには、質の高いシナリオを開発することが何より重要である。ISPOR Task Force (2016) では、直接測定の制限的な役割を認知しており、直接測定が用いられるには質の高いシナリオ (vignette) の開発に向けた漸進的な投資が行われなければならない上、これに関する指針開発の必要性も述べた。Arnold et al. (2009) はどの測定方法がより優れているかは述べていないが、両方法により算出した質の加重値の差が各範疇内にある道具により測定した質の加重値の差より大きい傾向があるため、リソースの割り当てに向けた意思決定では、この両方を混用しないことを主張した。Briggs et al. (2019) も、方法論的一致化 (methodological harmonization) に向けた研究がもっと必要であると述べた。

オ) 利害関係者の意見

(1) 専門家の意見 (scoping survey及び専門家の諮問内容)

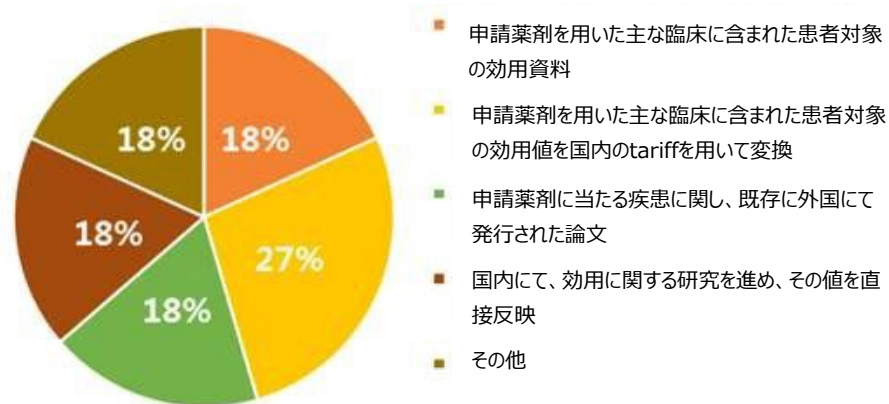
- 効用に関する不確実性を減らすべく、審評院にて選好する道具及び推定方法を明示する必要がある。
- 健康に関する暮らしの質の算出方法と資料元に基づいた質に関し、より明確に説明し、この中

で何を好むかについて明確に記述する必要がある。

- 指針においても、効用測定方法のプライオリティについて述べる必要がある。
- 臨床試験では、健康状態を測定するのではなく、効果を測定することである。これに関し、指針を改正する際は学問的に議論し明示してほしい。
- 測定方法別に、メリット・デメリットがあるため、状況に応じて選択させる必要がある。
- 現実的に、韓国人が多く含まれた臨床資料は多くない。審評院では、韓国人が殆ど含まれていない臨床試験から収集されたHRQoLの資料は受け入れられないため、間接測定の選好は不適切である（→事実関係の誤り）。

(2) 製薬会社の意見（scoping survey及びその他の会議内容）

- 最新の抗癌剤などの臨床試験において、該当患者からの暮らしの質の改善が報告される場合、これはとても有意義で良質な結果であるため、これを活用できるようにしなければならない。国内の一般大衆の選好度ではなく、外国人患者の選好度であっても活用できるようにしなければならない。
- 臨床研究にて、測定された暮らしの質若しくは効用値を活用することができれば、経済評価の際、これを直接利用するようにしなければならない。
- 現在、国内のEQ-5Dを基本分析し資料を要請しているが、臨床試験にて測定された値に関し算式（mapping）を用いて、EQ-5D値に変換した値も基本分析として受容れなければならない。
- 製薬会社へのアンケート（経済性評価制度改善のTFTレポート[健康保険審査評価院、2018]）：2017年7月25日～2017年8月4日の11日間実行し、計11社で応答（図7）。
- ☞ 著しく選好される項目なしに選んだ回答が算出された。製薬会社の立場：予め、効用測定方法のプライオリティを明確に決めるよりは、疾患の特性および状況に合わせて選択できるようにしてほしい。相互の合意による一貫性のある意思決定と共に、時間の流れに沿った変化を反映することも必要である。



[図 7] 効用のプライオリティに関する製薬会社（n=11）の回答結果

カ) 指針の改正方法に関する提言

間接測定の場合、対象の患者群が経験する効果と副作用を評価に直接反映することができ、標準化された道具の使用により評価の一貫性及び効率性を向上することができる。一方、直接測定の場合は、健康状態シナリオの妥当性により、質の加重値が大きく変わりかねないが、情報と時間の制約により、これを真面に評価できない可能性がある。そのため、多くの国のHTA指針同様に、指針の改正版には下記のように、間接測定の選好を明示することを提言する。

- 選好に基づいた一般道具をもって、臨床試験に含まれた患者の健康状態を測定し、この資料に国内の一般大衆を対象に算出された既存のtariffを適用し、質の加重値を算出する間接測定方法を選好すると明示する。
- 選好に基づいた一般道具をもって測定された資料がなかったり、制限的であったりする場合は健康状態のシナリオ（vignette）に基づいた直接測定、算式（mapping）による質の加重値を算出し、既存の文献から提示した値の引用なども受け入れることができるという案を提示する。この場合、一般道具使用の限界点と代案として用いられた方法に関し十分に説明し、敏感度の分析を通じて代案の使用に関する不確実性を十分に検討することを明示する。

3) HRQoL道具及びtariffの選定検討

ア) 現在の指針

(1) 主な指針

質の加重値の測定に用いられた道具は信頼性、妥当性が検証されたものでなければならない。とりわけ、外国にて開発された道具を用いる際は、言語・文化的な差に留意し、国内にて妥当性が検証された道具を採用する。

道具により、結果値に差が生じるため、研究者は自ら選んだ道具を本指針の別添3、表4に基づいて正当化し、さらに敏感度の分析などを通じ、質の加重値の不確実性が結果へ及ぼす影響を検討する。

資料の提出者は、調査方法について詳細に報告する。基本分析では、選好度に基づいた一般道具の使用が勧められ、疾病別の道具、選好度に基づいていない道具を用いて測定した結果は追加的に提出することができる。

(2) 指針における追加的な説明

現在、MAUIにおいて国内の一般大衆を対象にtariffが算出された道具はEQ-5Dが唯一であるが、研究別に値の差は大きい。当指針は、国内人のtariffが複数存在する場合、なるべく直近で算出された研究でありながらも、対象群が代表性のある研究にて算出された値を基本値として推奨する。これは、大体、後発研究は先行研究のデメリットを補うために行われ、その間の理論的な発展を反映する傾向があり、代表性が足りないサンプルから算出された値は国内人全体へと一般化させることに限界があるためである。もし、後発研究であっても研究方法の妥当性が先行研究より発展されなかったり、サンプリングの方法及びその規模が劣ったりする場合は先行研究を採用することができ、この際はその根拠を明示することである。さらに、研究に用いていないtariff値を用いて敏感度の分析を行うことを勧める。

現在の指針は、標準となるMAUI道具とtariffを提示していない。研究者は自分で選んだ道具に関し、指針にて提示する様式に基づき、正当化し敏感性の分析などを行い、不確実性を検討するようにしている。国内人のtariffが複数存在する場合は、なるべく直近で算出された研究でありながら、代表性のある対象群の研究にて算出された値を基本値として推奨している。

イ) 既存の提出資料に関する検討結果

2014～2019年の5年間で、審評院に健康保険への登載などの目的で提出された経済性評価の資料50件（CUA 47件）において、一般道具を用いて質の加重値を間接測定した事例は6件であった（表 30）。この中、EQ-5D-3Lを用いた事例が4件、EQ-5D-5Lを用いた事例が2件あった。EQ-5D-3Lの場合はtariffの出典を明かさない1件のほか、3件にて全てLee et al.（2009）のtariffを用いて質の加重値を計算した。EQ-5D-5Lの場合は、2件共に、Kim et al.（2016a）のtariffを用いた。敏感度の分析にて、他のtariffを用いたケースが計2件あったが、Jo et al.（2008）のtariffを用いた事例が1件、イギリスのtariffを採用した事例が1件あった。ICER（incremental cost-effectiveness ratio）の場合、Jo et al.（2008）のtariffを採用したケースで1.46%増加、他の事例でイギリスのtariffを用いて敏感度の分析を行った場合のICERは15.1%増えたと報告された。

<表 30> 2014～2018年、国内にて提出されたレポートにて用いられたMAUI及びtariff

事例	疾患	タイプ	健康状態、選好度の測定対象	Tariff
1	消化管間質腫瘍	EQ-5D-3L	患者、一般大衆	Lee et al. (2009) 敏感度分析：Jo et al. (2008)
2	多発性骨髄腫	EQ-5D-3L	患者、一般大衆	Lee et al. (2009)
3	全身性紅斑性狼瘡	EQ-5D-3L	患者、一般大衆	Lee et al. (2009)
4	非小細胞肺癌	EQ-5D-3L	患者、コメントなし	コメントなし
5	非小細胞肺癌	EQ-5D-5L	患者、一般大衆	Kim et al. (2016a) 敏感度分析：イギリス tariff
6	腎細胞癌	EQ-5D-5L	患者、一般大衆	Kim et al. (2016a)

※（最初に提出した資料の）主な基本分析を基準に、分

類

MAUI: multi-attribute utility

instrument

ウ) 外国の指針

道具による質の加重値の差が大きいため、一貫した政策を定めるべく、イギリスをはじめとした多くのヨーロッパ諸国はEQ-5Dを一つの道具として推奨したり、選好したりすると述べた。しかし、オーストラリアやカナダなど、一部の国では対象疾患に相応しい道具を選べるようにし、その代わり道具の選択を巡る根

拠を提示するようにした（表 31）。

(1) カナダ (CADTH, 2017)

カナダの指針では、選好度に基づいた一般道具の使用を推奨している。多く用いられている一般道具としてEQ-5D、HUI、SF-6Dなどを述べたが、どちらかを標準道具としては提示していない。ただ、測定しようとする健康状態の属性 (attribute) と健康状態の変化及び差を上手く捉えられる道具を選ぶように勧めており、道具の実現可能性 (feasibility)、信頼性 (reliability)、妥当性 (validity) などが十分に検証されたかを検討するようにしている。また、指針では質の加重値を測定するために用いられた資料の一貫性を強調しているが、なかんずく単一経済性の評価に使用する質の加重値を算出する際は同様な道具と人口集団から算出された資料を用いることを勧めている。

カナダの一般大衆の選好を反映した選好点数体系 (tariff) を考慮する際にも、研究者は資料が評価目的に合致するかの可否や資料の信頼性及び一貫性などを比較し、検討することを提示している。現在、使用できるカナダのtariffとしては、TTO及びDCE (discrete choice experiments) を用いて算出したEQ-5D-3L tariffとEQ-5D-5L tariffがあり、SGとVAS (visual analogue scale) を混用して測定したHUI2とHUI3のtariffがあるが、特定のtariffを標準値としては提示していない。

(2) オーストラリア (PBAC, 2016)

オーストラリアの指針では、よく用いられている一般MAUI (即ち、HUI2、HUI3、EQ-5D-3L、EQ-5D-5L、SF-6D、AQoL、CHU9D) 以外のMAUIやPRO (patient reported outcome) 道具が用いられた場合は、(1) この道具をもって測定しようとする暮らしの質、症状、機能領域 (2) 点数体系 (scoring method) (3) 妥当性 (validity) (4) 信頼性 (reliability) (5) 健康状態の差若しくは変化に対する敏感度 (responsiveness)、(6) 臨床的に有意義な最小限の差 (minimal clinically important differences, MCID) などに関する情報を追加的に提出するようにしている。また、一つ以上のMAUIが臨床研究にて用いられた場合は当道具により測定された結果を比較するようにしている。点数体系がオーストラリアの一般大衆を対象に算出されたものでない場合、可能であれば代案に関する敏感度の分析を行うことを提示している。

(3) イギリス (NICE, 2013)

イギリスのNICEでは、評価における一貫性のため、一つの道具を選定して用いることを選好している。指針では分析の対象が大人の場合、EQ-5D-3Lの使用を選好すると明示、同資料にイギリスの一般大衆を対象に算出されたEQ-5D-3L tariffを用いることを推奨している。指針では、EQ-5D-5Lも基本分析に用いられかねないと明示しているが、イギリスの一般大衆を対象に行われたEQ-5D-5Lの価値評価研究 (Devlinなど、2018) に関する質 (quality) の問題が提起 (Hernandez Alava et al., 2018) され、現在は使用を保留にしているという別途の意見表明の報告書 (position statement) を発表した (NICE, 2019)。同書では、現在EuroQolグループにてイギリスの一般大衆を対象に、新たにEQ-5D-5Lに関する価値評価の研究を行うようにしており、同研究が終わるとEQ-5D-5Lに関するNICEの立場を再検討する計画だと明示した。指針では、受容可能なEQ-5D-5L tariffが開発されるまで、EQ-5D-5Lが用いられた場合、van Hout et al. (2012) により開発された算式 (mapping) のアルゴリズムを用いてEQ-5D-3L点数に変換することを勧めている。

さらに、指針ではEQ-5Dが子供向けに開発された道具ではないということを明示している。7~12歳の児童を対象にHRQoLを測定するEQ-5Dバージョンも開発されたが、まだ、これに関するtariffが開発されていないことを述べている。この他にも、EQ-5Dの使用が不適切だと判断される場合は、EQ-5Dの内容的妥当性（content validity）が不足しているという主張に対し、質的な実証資料（qualitative empirical data）を提示するように明示している。EQ-5Dが特定の患者集団にて構成概念妥当性（construct validity）及び感受度（responsiveness）の面にて、真面に機能できないという根拠を提示しなければならないが、この際、根拠は相互審査文献（peer-reviewed literature）を総合して算出しなければならないとした。こうした根拠が十分に提示された場合は他のHRQoL道具が代案として用いられることがあるが、この際、代案として用いられた道具の妥当性と測定資料をはじめ、この道具の使用が質の加重値に及ぼし得る影響に関し、詳しく説明するようにしている。

(4) その他

EUnetHTA（2015a）によると、間接測定方法を推奨する多くのヨーロッパ諸国（例：ベルギー、クロアチア、チェコ、イギリス、イタリア、オランダ、ポーランド、スコットランド及びスロバキア）にて、EQ-5Dの使用を勧めている。オーストリア、バルト海（エストニア及びラトビア）及びフランスではEQ-5D若しくはHUIの使用を好み、アイルランド、ノルウェー及びスペインのCatSalutではEQ-5DもしくはSF-6Dの使用を勧めている（EUnetHTA、2015a）。ノルウェーの指針では、追加的に15Dも受容可能な方法と述べている。ISPOR Task Forceでは、経験的証拠（empirical evidence）に基づいた道具の実用性（practicality）、信頼性（reliability）、妥当性（validity）、感受度（responsiveness）と当HTA機関の選好を考慮して道具を選定することを推奨している（Wolowacz et al., 2016）。

<表 31> HRQoL道具及びtariffの選定に関する海外の指針内容

国	指針内容の原文	サマリー
カナダ・	The most common generic classification systems include the	健康状態の特性と

<p>CADTH 指針 (4 版) (2017)</p>	<p>EQ-5D, the HUI, and the SF-6D.78-80 (p45) The selection of a particular indirect generic method should be based on its fitness in terms of reflecting the health states of interest and their associated attributes, and in terms of capturing potentially important changes within and among states. The researcher should also assess the credibility of the instrument in terms of whether it represents an established instrument with demonstrated psychometric properties including feasibility, reliability, and validity.⁸³ Consistency with respect to the data used for estimating utility values is strongly recommended; in particular, it is recommended that utility data from the same instrument and population be used to estimate all the utilities in an economic evaluation. (p45)</p> <p>When considering Canadian-specific preference sets, researchers must weigh their fitness for purpose against any issues of credibility and consistency with the data used to inform other parameter estimates, to ensure that the specific instrument addresses the decision problem. (p46)</p> <p>Examples of Canadian-specific preference sets include the EQ-5D, which has been measured using the time-trade-off technique and a discrete choice experiment for both the EQ-5D-3L and EQ-5D-5L (three-level and five-level versions, respectively),^{81,82} and the HUI Mark 2 and Mark 3, which have been measured using a combination of the standard gamble and visual analogue scale from a sample of adults from Hamilton, Ontario.² (p45)</p>	<p>変化を最も有効的に捉えられる一般道具を選択。 研究者は、道具の実現可能性 (feasibility)、信頼性 (reliability)、妥当性 (validity) などの検証可否を確認しなければならない。</p> <p>カナダの選好点数体系を考慮する場合、研究者は資料が目的に相応しいかの可否と資料の信頼性 (credibility) 及び一貫性 (consistency) などを比較して評価。</p> <p>カナダのtariff : TTO及びDCEを用いて、測定したEQ-5D-3L & EQ-5D-5L SG及びVASを混用して測定したHUI2 & HUI3</p>
<p>オー ストラ リア・ PBA Cの 指針 versi on 5 (2016)</p>	<p>Several commonly used MAUIs for which a detailed discussion of the validity or reliability is not required are the Health Utilities Index (HUI2 or HUI3), the EQ5D-3L or -5L (EuroQol), the SF-6D (a subset of the Short Form 36, or SF-36), the Assessment of Quality of Life (AQoL) instruments, and the Child Health Utility 9D (CHU9D) index for children and adolescents. (p37)</p> <p>Where a patient-reported outcome measure is used, or a MAUI that is not listed above, provide, in an attachment, a discussion of (or reference supporting) the: domains of quality of life, symptoms or function that are covered by the instrument scoring method of the instrument validity of the instrument reliability of the instrument responsiveness of the instrument to differences in health states between individuals and to changes in health states over time experienced by an individual clinical importance of any differences detected by the instrument</p>	<p>よく用いられるMAUIに関しては、道具の妥当性及び信頼性に関し、議論する必要はない。</p> <p>しかし、PRO道具もしくは上記に記載されていないMAUIが用いられた場合は (1) 道具をもって測定しようとする暮らしの質、症状、機能領域、(2) 点数体系、(3) 妥当性、(4) 信頼性、(5) 健康状態の差もしくは変化に関する感受度 (responsiveness)、(6) MCIDを提示。</p> <p>一つ以上のMAUIが臨床研究にて用いられた</p>

	<p>(see Subsection 2.4.4 for guidance on minimal clinically important differences [MCIDs]). (p37)</p> <p>If more than one MAUI has been used in the included study, compare the results from the two MAUIs. (p42)</p>	<p>場合、両道具の結果を比較。</p>
	<p>If the scoring algorithm has not been derived from the general population in Australia, consider presenting a sensitivity analysis using alternative scoring algorithms, if possible. (p42)</p>	<p>点数体系がオーストラリアの一般大衆を対象に、算出されたものでない場合、できれば代案について敏感度の分析を実行。</p>
イギリス・NIC Eの指針 (2013)	<p>Different methods used to measure health-related quality of life produce different utility values; therefore, results from different methods or instruments cannot always be compared. Given the need for consistency across appraisals, one measurement method, the EQ-5D, is preferred for the measurement of health-related quality of life in adults. (p41)</p> <p>The standard version of the EQ-5D has not been designed for use in children. An alternative version for children aged 7–12 years is available, but a validated UK valuation set is not yet available. (p42)</p> <p>The EQ-5D-5L may be used for reference-case analyses. The descriptive system for the EQ-5D-5L has been validated, but no valuation set to derive utilities currently exists. Until an acceptable valuation set for the EQ-5D-5L is available, the validated mapping function to derive utility values for the EQ-5D-5L from the existing EQ-5D (-3L) may be used. (p42)</p>	<p>評価の一貫性のため、大人を対象にした場合、一つの道具 (EQ-5D) に関する選好を明示。</p> <p>7~12歳児童向けの道具はあるが、まだ検証されたイギリスのtariffはない。EQ-5D-5Lも基本分析に使用可。道具の技術体系は検証されたが、現在、受容可能なイギリスのtariffはない。受容可能なtariffが開発されるまで、EQ-5D-5Lを利用する場合、算式 (mapping) を用いてEQ-5D-3Lに変換。</p>
	<p>A set of preference values elicited from a large UK population study using a choice based method of valuation (the time trade-off method) is available for the EQ-5D health state descriptions. This set of values should be applied to measurements of health-related quality of life to generate health-related utility values. (p41)</p>	<p>TTOを用いて算出したイギリスのtariffはある。これを用いなければならない。</p>

工) その他の文献考察の結果

(1) MAUI道具

表32に提示されたように、現在、大人を対象に最も多く用いられている選好に基づいた一般道具もしくはMAUIでは、Quality of Well-Being Self-Administered(QWB-SA)(Kaplan & Anderson, 1988; Seiber et al., 2008), Health Utility Index three(HUI-3)(Feeny et al., 2002),

EQ-5D(Dolan, 1997), 15D(Sintonen & Pekurin, 1993), SF-6D(Brazier et al., 2002; Brazier & Roberts, 2004), Assessment of quality of life-8D(AQoL)(Hawthorne et al., 1997; Richardson et al., 2014) などがある。

同じ患者の健康状態であっても、どのMAUIを用いるかにより、かなり違った質の加重値が算出される。その理由は、各道具別の技術体系に含まれた健康領域及びレベルが異なり、健康状態に関する選好度の測定方法なども道具別に違うためである。例えば、SF-6Dにはエネルギー領域が含まれているが、EQ-5Dには含まれていない。選好度を算出する際、EQ-5Dは主にTTO方法を用いるが、HUI-3とSF-6DはSG方法を用いる。EQ-5DとHUI-3の場合は、死亡より悪い健康状態を許すため、負の効用値も存在するが、QWBとSF-6Dはこれを許さない。さらに、EQ-5D、SF-6D、QWB-SAは質の加重値を算出する際、加法 (additive) 関数を用いるが、HUI-3やAQoL-8Dは乗法 (multiplicative) 関数を用いる。

Richardson et al. (2015a) は、MAUI(EQ-5D, HUI-3, SF-6D, 15D, QWB, AQoL) による質の加重値を比較するべく、オーストラリア、イギリス、アメリカ、カナダ、ノルウェー、またドイツに住んでいる慢性疾患患者と健康な一般人を対象に、Multi Instrument Comparison (MIC) プロジェクトを行った。当道具の元のtariff (即ち、道具開発国のtariff) を適用して質の加重値を算出した結果、道具間で大きな差が生じることが分かった。なかんずく、15Dは質の加重値を相対的に高く、QWBは質の加重値を相対的に低く測定するきらいがあった。研究チームは、こうした差が道具の健康状態記述体系によるものか、それともtariffによるものかを確認するため、追加的な研究を行った(Richardson et al., 2015b) 。分析の結果、道具間における質の加重値の差は、主に記述体系とスケール効果 (scale effect) によるものとされたが、とりわけ、記述体系が占める比重は66%と報告した。質の加重値の範囲の差により発生するスケール効果も道具間における差の30.3%を説明することが分かった。

上記のように、道具の選択は、結果へ大きな影響を与えかねない一つの要因であるが、まだどの道具が最も優れているかに関する合意は行われたことがない。ただ、現在まで最も多く用いられた道具はEQ-5Dと報告された。2005年~2010年まで用いられた全体のMAUIにおいて、EQ-5Dの比重は63.2%であり、次はHUI-2/HUI-3 (14.4%) とSF-6D (8.8%) の順であった (Richardson et al., 2014) 。EQ-5Dの高い使用頻度はイギリス・NICEの影響を受けたと推定されるが、イギリス・NICE (2013) の指針では評価の一貫性のため、一つの道具、即ちEQ-5Dを選好すると明示した。Brazier et al. (2017) は、実用性 (practicality) 、信頼性 (reliability) 、内容妥当性 (content validity) 、構成概念妥当性 (construct validity) 、価値評価方法及び対象を考慮し、道具を選ばなければならないと主張した。

表 32> MAUI道具の比較

道具 (開発国)	領域	健康状態 概数	(元) 価値評 価方法 (質の 加重値の最小 値)
QWB-SA (アメリカ)	運動能力 (mobility)、身体的活動 (physical activity)、社会的機能 (social functioning)	945	RS(0.09)
HUI-3 (カナダ)	視覚 (vision)、聴覚 (hearing)、言語能力 (speech)、歩行 (ambulation)、器用さ (dexterity)、感情 (emotion)、認知 (cognition)、痛み (pain)	972,000	SG(-0.36)
EQ-5D (EQ-5D-3L &EQ-5D-5L) (イギリス)	運動能力 (mobility,)、自己管理 (self-care)、日常的活動 (usual activities)、痛み・不快感 (pain/discomfort)、不安・うつ病 (anxiety/depression)	243(3L) 3,125(5L)	3L : TTO(-0.59) 5L : TTO &DCE(-0.208)
15D (フィンランド)	運動能力 (mobility)、視覚 (vision)、聴覚 (hearing)、呼吸 (breathing)、睡眠 (sleeping)、食事 (eating)、言語能力 (speech,)、排泄 (excretion)、日常的活動 (usual activities)、精神的機能 (mental function)、不快感・症状 (discomfort/symptoms)、うつ病 (depression)、苦痛 (distress)、活力 (vitality)、性行動 (sexual activity)	310億	RS(0.11)
SF-6D (イギリス)	身体的機能 (physical functioning)、役割制限 (role limitation)、社会的機能 (social functioning)、痛み (pain,)、エネルギー (energy)、精神健康 (mental health)	18,300(SF-36) 7,500(SF-12) 18,750(SF-V2)	SG(0.301)
AQoL-8D (オーストリア)	自立生活 (independent living, 4項目)、感覚 (senses, 3)、痛み (pain, 3)、精神健康 (mental health, 8)、幸福 (happiness, 4)、自尊心 (self-worth, 3) 対処能力 (coping, 3)、関係 (relationship, 7)	2.37×1023	RS & TTO(-0.04)

出典 : Brazier et al. (2017) 参照

AQoL-8D: assessment of quality of life-8 dimensions, HUI-3: health utilities index-3, DCE: discrete choice experiment, MAUI: multi-attribute utility instrument, QWB-SA: quality of well-being self-administered, RS: rating scale, SG: standard gamble, SF-6D: short-form 6 dimensions, TTO: time trade-off

上記にて、取り上げた殆どのMAUIは主に大人向けに開発されたものであるため、内容的妥当性 (content validity) と表面的妥当性 (face validity) の面 (例: 健康領域及び単語表現 (wording)) では、児童に相応しくないという指摘がある (Stevens, 2009; Brazier et al., 2017)。それで、様々な道具が児童向けに修正及び開発されたが、その例としては HUI-2、16D、17D、AQoL-6D、EQ-5D-Y、CHU-9Dなどがある。これに関する詳しい説明は表33を参考にすることができる。

<表 33> 児童及び青少年向けのMAUI

道具 (開発国)	年齢	回答者	領域	健康状態個数
HUI2 (カナダ)	≥5、5-8 歳の場合、 代理者	自己報 告、面接 官の記入、	感覚 (sensation) 、運動能力 (mobility) 、感情 (emotion) 、認 知 (cognition) 、自己管理 (self- care) 、痛み (pain) ,	24,000

	回答	代理者の記入	(受胎能力 (fertility) ; 児童癌患者向けに開発され、副作用評価のために含む。現在は省略)	
16D (フィンランド)	12~15才	自己報告、面接官の記入、代理者の記入	運動能力 (mobility) 、視覚 (vision) 、聴覚 (hearing) 、呼吸 (breathing) 、睡眠 (sleeping) 、食事 (eating) 、言語能力 (speech) 、排泄 (excretion) 、学校・趣味 (school/hobbies) 、精神的機能 (mental function) 、不快感・症状 (discomfort/symptoms) 、うつ病 (depression) 、苦痛 (distress) 、活力 (vitality) 、外観 (appearance) 、友人 (friends)	1.5×1011
17D (フィンランド)	8~11才、8才未満の場合は代理者が回答	面接官の記入、代理者の記入	運動能力 (mobility) 、視覚 (vision) 、聴覚 (hearing) 、呼吸 (breathing) 、睡眠 (sleeping) 、食事 (eating) 、言語能力 (speech) 、排泄 (excretion) 、学校・趣味 (school/hobbies) 、学習・記憶 (learning/memory) 、不快感・症状 (discomfort/symptoms) 、うつ病 (depression) 、苦痛 (distress) 、活力 (vitality) 、外観 (appearance) 、友達 (friends) 、集中 (concentration)	7.6×1011
AQoL-6D 青少年 (オーストラリア)	青少年	自己報告	自立生活 (independent living) 、関係 (relationships) 、精神健康 (mental health) 、対処能力 (coping) 、痛み (pain) 、感覚 (senses)	7.8×1013
EQ-5D-Y (イギリス)	8~15才、4~7才の場合は代理者が回答	自己報告、代理者の記入	運動能力 (mobility) 、自己管理 (self-care) 、日常的活動 (usual activities) 、痛み・不快感 (pain/discomfort) 、不安・うつ病 (anxiety/depression) * 児童向けに文言を修正	243
CHU-9D (イギリス)	7~17才	自己報告、代理者の記入	心配 (worry) 、悲しみ (sadness) 、痛み (pain) 、疲れ (tiredness) 、迷惑 (annoyance) 、学校 (school) 、睡眠 (sleep) 、日常生活 (daily routine) 、日常的活動 (daily activities)	1,953,125

出典 : Brazier et al. (2017) 参照
AQoL-6D: assessment of quality of life-6 dimensions, CHU-9D: child health utility 9-dimensions, HUI-2: health utilities index-2, MAUI: multi attribute utility instrument.

(2) 国内におけるMAUIの価値評価研究

現在、国内の一般人向けに行われたMAUIの価値評価研究は、計7つある。EQ-5D-3L対象3つ（Kang Eunjung, 2006; Jo et al.,2008; Lee et al.,2009）とEQ-5D-5L対象2つ（Jo Minwoo, 2011;Kim et al.,2016a）があり、国内にて新たに開発されたHRQoL Instrument with 8 items (HINT-8) 対象1つ（Jo Minwoo, 2017）とChild Health-6D（CH-6D）対象1つ（HanHankyoreh & Kang Eunjung, 2015）がある。なかんずく、CH-6Dは国内にて開発された唯一の児童向けのものである。

□ 国内にて行ったEQ-5D-3Lに関する価値評価研究

EQ-5D-3Lは、ヨーロッパのEuroQolグループにて開発したHRQoLアンケート道具として、運動能力、自己管理、日常活動、痛み・迷惑、不安・うつ病の5つの領域と領域別の3段階の回答レベルからなる。簡単な構成と回答の容易さにより、暮らしの質に関する評価道具において、最も多く用いられている。表34のように、現在、国内では3つの価値評価研究があるが、研究ごとに健康領域の相対的な重要度及び算出された加重値の範囲が異なっている。

Jo et al.（2008）は、国内初で一般人口集団を対象に、EQ-5D-3Lに関する価値評価研究を行った。Seoul, Gyeonggi地域の20歳以上の大人500人を対象にTTO方法を用いて、健康状態の選好度を調べた。33333と無意識の状態を含めた43の健康状態を評価し、回答者1人当たりで13の状態を評価するようにした。最終的なモデルとしてLN（1-健康状態の質の加重値）を使用したが、この際の平均絶対誤差（mean absolute error、MAE）は0.074であった。

自己管理や日常活動が相対的に重要と評価され、質の加重値の範囲は、-0.562（33333）から1（11111）であった。

Kang Eunjungなど（2006）は、国民健康栄養調査にて用いた調査区にて7つの調査区を選定、302（分析基準287）人に、対面の面接方法でアンケートを行った。TTO方法を用いて、EQ-5D-3Lと定義される243の健康状態にて、17の状態に関する選好度を調べた。同研究では、N3とI2sq項を含んだモデルを最終モデルとして選んだが、N3はレベル3が一つでもある場合に活性化される項であり、I2sqはI2の二乗項として、I2はレベル2が二つ以上ある際、レベル2がある領域の個数から1を引いた値である。同モデルのR2は0.4321と報告され、Jo et al.（2008）の研究とは違い、運動能力が最も重要な領域、自己管理はこれより重要でない領域と評価された。質の加重値の範囲は、-0.243（33333）から1（11111）であった。

Lee et al.（2009）は、直近で行われた研究として、全国の20歳以上の大人1,307人を対象に価値評価研究を行った。EuroQolグループで提案した100の状態に33333状態を含めて測定し、各回答者には11111、33333、死亡状態を含んだ15の状態を割り当てて評価するようにした。先の二つの研究同様に、TTOを用いて質の加重値を算出した。N3を含んだOLS（ordinary least squares）モデルを最終モデルとして選定し、この際、修正されたR2は0.984と、高い説明力を表した。運動能力と日常活動が最重要の領域と評価され、質の加重値の範囲は-0.171（33333）から1（11111）であった。ちなみに、国民健康栄養調査にて調べたEQ-5D-3Lの点数は、同研究のモデルに基づいて算出されたものである。

<表 34> 国内にて行われたEQ-5D-3Lに関する価値評価研究

項目	Jo et al. (2008)	Kang Eunjungなど (2006)	Lee et al. (2009)																																																																																				
標本数	500人	302人 (287人を分析)	1,307人 (1,264人を分析)																																																																																				
標本特性	Seoul、Gyeonggi地域の満2才以上の男女	Seoul、Incheon、Gyeonggi、Jeollabuk-do、Chungcheongnam-doの20才以上の男女	Jejuを除いた全国の20才以上の男女																																																																																				
資料収集	アンケートに基づいた直接訪問の面接調査	アンケートに基づいた直接訪問の面接調査	アンケートに基づいた直接訪問の面接調査																																																																																				
測定した健康状態	42個	17個	101個																																																																																				
主な価値評価方法	TTO	TTO	TTO																																																																																				
修正されたR ²	コメントなし (MAE=0.074)	0.4321 (MAE=0.028)	0.984 (MAE=0.029)																																																																																				
領域別の相対的な重要度 (最終モデルの基準[とりわけ、レベル3の回帰係数を参考])	<p>SC3 (自己管理)、UA3 (日常活動)、AD3 (不安・うつ病)、M3 (運動能力) & PD3 (痛み・不便性) の順</p> <table border="1"> <tr><td colspan="2">기본 LN 모형</td></tr> <tr><td>cons</td><td>-2.68</td></tr> <tr><td>M2</td><td>0.267</td></tr> <tr><td>M3</td><td>0.554</td></tr> <tr><td>SC2</td><td>0.471</td></tr> <tr><td>SC3</td><td>0.819</td></tr> <tr><td>UA2</td><td>0.374</td></tr> <tr><td>UA3</td><td>0.662</td></tr> <tr><td>PD2</td><td>0.318</td></tr> <tr><td>PD3</td><td>0.488</td></tr> <tr><td>AD2</td><td>0.313</td></tr> <tr><td>AD3</td><td>0.603</td></tr> <tr><td>N3</td><td>-</td></tr> <tr><td>I2sq</td><td>-</td></tr> </table>	기본 LN 모형		cons	-2.68	M2	0.267	M3	0.554	SC2	0.471	SC3	0.819	UA2	0.374	UA3	0.662	PD2	0.318	PD3	0.488	AD2	0.313	AD3	0.603	N3	-	I2sq	-	<p>M3 (運動能力)、UA3 (日常活動) & PD3 (痛み・迷惑)、AD3 (不安・うつ病)、SC3 (自己管理) の順</p> <table border="1"> <tr><td colspan="2">N3+I2sq</td></tr> <tr><td>cons</td><td>0.164</td></tr> <tr><td>M2</td><td>0.003</td></tr> <tr><td>M3</td><td>0.274</td></tr> <tr><td>SC2</td><td>0.058</td></tr> <tr><td>SC3</td><td>0.078</td></tr> <tr><td>UA2</td><td>0.045</td></tr> <tr><td>UA3</td><td>0.134</td></tr> <tr><td>PD2</td><td>0.049</td></tr> <tr><td>PD3</td><td>0.132</td></tr> <tr><td>AD2</td><td>0.044</td></tr> <tr><td>AD3</td><td>0.102</td></tr> <tr><td>N3</td><td>0.345</td></tr> <tr><td>I2sq</td><td>0.014</td></tr> </table>	N3+I2sq		cons	0.164	M2	0.003	M3	0.274	SC2	0.058	SC3	0.078	UA2	0.045	UA3	0.134	PD2	0.049	PD3	0.132	AD2	0.044	AD3	0.102	N3	0.345	I2sq	0.014	<p>M3 (運動能力)、UA3 (日常活動)、AD3 (不安・うつ病) & PD3 (痛み・不便性) & SC3 (自己管理) の順</p> <table border="1"> <tr><td colspan="2">N3 OLS</td></tr> <tr><td>cons</td><td>0.05</td></tr> <tr><td>M2</td><td>0.096</td></tr> <tr><td>M3</td><td>0.418</td></tr> <tr><td>SC2</td><td>0.046</td></tr> <tr><td>SC3</td><td>0.136</td></tr> <tr><td>UA2</td><td>0.051</td></tr> <tr><td>UA3</td><td>0.208</td></tr> <tr><td>PD2</td><td>0.037</td></tr> <tr><td>PD3</td><td>0.151</td></tr> <tr><td>AD2</td><td>0.043</td></tr> <tr><td>AD3</td><td>0.158</td></tr> <tr><td>N3</td><td>0.05</td></tr> <tr><td>I2sq</td><td>-</td></tr> </table>	N3 OLS		cons	0.05	M2	0.096	M3	0.418	SC2	0.046	SC3	0.136	UA2	0.051	UA3	0.208	PD2	0.037	PD3	0.151	AD2	0.043	AD3	0.158	N3	0.05	I2sq	-
기본 LN 모형																																																																																							
cons	-2.68																																																																																						
M2	0.267																																																																																						
M3	0.554																																																																																						
SC2	0.471																																																																																						
SC3	0.819																																																																																						
UA2	0.374																																																																																						
UA3	0.662																																																																																						
PD2	0.318																																																																																						
PD3	0.488																																																																																						
AD2	0.313																																																																																						
AD3	0.603																																																																																						
N3	-																																																																																						
I2sq	-																																																																																						
N3+I2sq																																																																																							
cons	0.164																																																																																						
M2	0.003																																																																																						
M3	0.274																																																																																						
SC2	0.058																																																																																						
SC3	0.078																																																																																						
UA2	0.045																																																																																						
UA3	0.134																																																																																						
PD2	0.049																																																																																						
PD3	0.132																																																																																						
AD2	0.044																																																																																						
AD3	0.102																																																																																						
N3	0.345																																																																																						
I2sq	0.014																																																																																						
N3 OLS																																																																																							
cons	0.05																																																																																						
M2	0.096																																																																																						
M3	0.418																																																																																						
SC2	0.046																																																																																						
SC3	0.136																																																																																						
UA2	0.051																																																																																						
UA3	0.208																																																																																						
PD2	0.037																																																																																						
PD3	0.151																																																																																						
AD2	0.043																																																																																						
AD3	0.158																																																																																						
N3	0.05																																																																																						
I2sq	-																																																																																						
点数計算及び範囲	<p>点数 = 1 - exp(-2.68NPH + 0.267M2 + 0.554M3 + 0.471SC2 + 0.819SC3 + 0.374UA2 + 0.662UA3 + 0.318PD2 + 0.488PD3 + 0.313AD2 + 0.603AD3)</p> <p>(NPH, 11111他の健康状態)</p> <p>範囲 -0.562(33333) ~ 1(11111)</p>	<p>点数 = 1 - (0.164NPH + 0.003M2 + 0.274M3 + 0.058SC2 + 0.078SC3 + 0.045UA2 + 0.134UA3 + 0.049PD2 + 0.132PD3 + 0.044AD2 + 0.102AD3 + 0.345N3 + 0.014I2sq)</p> <p>(NPH, 11111の外の健康状態; N3, level 3が最小1つ以上; I2sq、(level 2の個数 - 1)の二乗)</p>	<p>点数 = 1 - (0.05NPH + 0.096M2 + 0.418M3 + 0.046SC2 + 0.136SC3 + 0.051UA2 + 0.208UA3 + 0.037PD2 + 0.151PD3 + 0.043AD2 + 0.158AD3 + 0.05N3)</p> <p>(NPH, 11111他の健康状態; N3, level 3が最小1つ以上)</p>																																																																																				

		範圍 :- 0.243 (33333) ~ 1 (11111)	範圍 :- 0.171 (33333) ~ 1 (11111)
--	--	---------------------------------------	---------------------------------------

MAE: mean absolute error, OLS: ordinary least squares, TTO: time trade-off

□ 国内にて行ったEQ-5D-5Lに関する価値評価研究

EQ-5D-5Lは、健康の変化に関するEQ-5D-3Lの感度を改善し、全領域に問題がないと報告する天井効果（ceiling effect）を減らすべく考案された評価道具として、EQ-5D-3L同様な領域で構成されているが、領域別の回答レベルが3段階から5段階に拡張された。表 35にて提示されたように、現在、国内では二つのEQ-5D-5Lに関する価値評価研究がある。これらの研究で、EQ-5D-3Lと比較する際、天井効果減と信頼性と感度は有意義に増加したと報告された（Jo Minwoo, 2011; Kim et al., 2016a）。

まず、Jo Minwoo（2011）はSeoul・Incheon・Kyeongin地域に住む大人500人を対象に、SG方法をもってEQ-5D-5Lに関する価値評価の研究を行った。計70の健康状態を選定し、各回答者が7つの健康状態と追加的に55555及び無意識の状態を共に評価するようにした。interceptのない主効果モデルを最終モデルとして選定したが、この際、修正されたR2は0.75として相対的に高い説明力が表れた上、MAEも0.031で良好なレベルであった。最終モデルでは、移動性（運動能力）と自己管理が相対的に重要に評価され、質の加重値の範囲は-0.159（55555）から1（11111）であった。

Kim et al.（2016a）は、Jejuを除いた全国の19才以上の男女1,080人を対象に、composite TTO（cTTO）方法を用いてEQ-5D-5Lの価値評価に関する研究を行った。cTTOは死亡より良い健康状態に関しては、既存のTTO方法を、死亡より悪い健康状態に関しては、lead time TTO方法を用いた。同研究では、計86の健康状態を選定し、回答者が10の健康状態を評価するようにした。最終モデルでは、N4項をはじめとしたモデルが選定されたが、N4はレベル4以上が一つでもある場合、活性化される項である。同モデルのR2は0.538、MAEは0.027と、良好な数値が表れた。移動性（運動能力）と痛み・不便性の領域は相対的に重要に評価され、質の加重値の範囲は-0.066（55555）から1（11111）として、先行研究より有意義な差が表れた。

<表 35> 国内にて行ったEQ-5D-5Lに関する価値評価研究

項目	Jo Minwoo（2011）	Kimなど（2016a）
標本数	500人	1,080人
標本特性	Seoul、Incheon、Gyeonggi地域の満19才以上の男女	Jejuを除いた全国の19才以上の男女
資料収集	アンケートに基づいた直接訪問の面接調査	パソコンに基づいた一対一の面接調査
測定した健康状態	72個（55555 & 無意識を含む）	86個
主な価値調査方法	SG	Composite TTO

R ²	0.75 (MAE=0.031)	0.538 (MAE=0.027)																																																																																												
領域別の相対的な重要度 (最終モデルの基準 [とりわけ、レベル4、5の回帰係数を参考])	M5 (移動性)、SC5 (自己管理)、PD5 (痛み・不便性) & UA5 (日常生活) & AD5 (不安・うつ病) の順 <table border="1"> <thead> <tr><th colspan="2">1-5 (상수항 없음)</th></tr> </thead> <tbody> <tr><td>cons</td><td>-</td></tr> <tr><td>M2</td><td>0.068</td></tr> <tr><td>M3</td><td>0.114</td></tr> <tr><td>M4</td><td>0.238</td></tr> <tr><td>M5</td><td>0.396</td></tr> <tr><td>SC2</td><td>0.042</td></tr> <tr><td>SC3</td><td>0.081</td></tr> <tr><td>SC4</td><td>0.127</td></tr> <tr><td>SC5</td><td>0.232</td></tr> <tr><td>UA2</td><td>0.010</td></tr> <tr><td>UA3</td><td>0.034</td></tr> <tr><td>UA4</td><td>0.129</td></tr> <tr><td>UA5</td><td>0.170</td></tr> <tr><td>PD2</td><td>0.014</td></tr> <tr><td>PD3</td><td>0.036</td></tr> <tr><td>PD4</td><td>0.096</td></tr> <tr><td>PD5</td><td>0.198</td></tr> <tr><td>AD2</td><td>0.047</td></tr> <tr><td>AD3</td><td>0.054</td></tr> <tr><td>AD4</td><td>0.116</td></tr> <tr><td>AD5</td><td>0.164</td></tr> <tr><td>N4</td><td>-</td></tr> </tbody> </table>	1-5 (상수항 없음)		cons	-	M2	0.068	M3	0.114	M4	0.238	M5	0.396	SC2	0.042	SC3	0.081	SC4	0.127	SC5	0.232	UA2	0.010	UA3	0.034	UA4	0.129	UA5	0.170	PD2	0.014	PD3	0.036	PD4	0.096	PD5	0.198	AD2	0.047	AD3	0.054	AD4	0.116	AD5	0.164	N4	-	M3 (移動性)、PD3 (痛み・不便性)、UA3 (日常生活) & SC3 (自己管理) の順 <table border="1"> <thead> <tr><th colspan="2">N4 모형</th></tr> </thead> <tbody> <tr><td>cons</td><td>0.096</td></tr> <tr><td>M2</td><td>0.046</td></tr> <tr><td>M3</td><td>0.058</td></tr> <tr><td>M4</td><td>0.133</td></tr> <tr><td>M5</td><td>0.251</td></tr> <tr><td>SC2</td><td>0.032</td></tr> <tr><td>SC3</td><td>0.050</td></tr> <tr><td>SC4</td><td>0.078</td></tr> <tr><td>SC5</td><td>0.122</td></tr> <tr><td>UA2</td><td>0.021</td></tr> <tr><td>UA3</td><td>0.051</td></tr> <tr><td>UA4</td><td>0.100</td></tr> <tr><td>UA5</td><td>0.175</td></tr> <tr><td>PD2</td><td>0.042</td></tr> <tr><td>PD3</td><td>0.053</td></tr> <tr><td>PD4</td><td>0.166</td></tr> <tr><td>PD5</td><td>0.207</td></tr> <tr><td>AD2</td><td>0.033</td></tr> <tr><td>AD3</td><td>0.046</td></tr> <tr><td>AD4</td><td>0.102</td></tr> <tr><td>AD5</td><td>0.137</td></tr> <tr><td>N4</td><td>0.078</td></tr> </tbody> </table>	N4 모형		cons	0.096	M2	0.046	M3	0.058	M4	0.133	M5	0.251	SC2	0.032	SC3	0.050	SC4	0.078	SC5	0.122	UA2	0.021	UA3	0.051	UA4	0.100	UA5	0.175	PD2	0.042	PD3	0.053	PD4	0.166	PD5	0.207	AD2	0.033	AD3	0.046	AD4	0.102	AD5	0.137	N4	0.078
	1-5 (상수항 없음)																																																																																													
cons	-																																																																																													
M2	0.068																																																																																													
M3	0.114																																																																																													
M4	0.238																																																																																													
M5	0.396																																																																																													
SC2	0.042																																																																																													
SC3	0.081																																																																																													
SC4	0.127																																																																																													
SC5	0.232																																																																																													
UA2	0.010																																																																																													
UA3	0.034																																																																																													
UA4	0.129																																																																																													
UA5	0.170																																																																																													
PD2	0.014																																																																																													
PD3	0.036																																																																																													
PD4	0.096																																																																																													
PD5	0.198																																																																																													
AD2	0.047																																																																																													
AD3	0.054																																																																																													
AD4	0.116																																																																																													
AD5	0.164																																																																																													
N4	-																																																																																													
N4 모형																																																																																														
cons	0.096																																																																																													
M2	0.046																																																																																													
M3	0.058																																																																																													
M4	0.133																																																																																													
M5	0.251																																																																																													
SC2	0.032																																																																																													
SC3	0.050																																																																																													
SC4	0.078																																																																																													
SC5	0.122																																																																																													
UA2	0.021																																																																																													
UA3	0.051																																																																																													
UA4	0.100																																																																																													
UA5	0.175																																																																																													
PD2	0.042																																																																																													
PD3	0.053																																																																																													
PD4	0.166																																																																																													
PD5	0.207																																																																																													
AD2	0.033																																																																																													
AD3	0.046																																																																																													
AD4	0.102																																																																																													
AD5	0.137																																																																																													
N4	0.078																																																																																													
点数計算及び範囲	点数 = 1 - (0.068M2 + 0.114M3 + 0.238M4 + 0.396M5 + 0.042SC2 + 0.081SC3 + 0.127SC4 + 0.232SC5 + 0.010UA2 + 0.034UA3 + 0.129UA4 + 0.170UA5 + 0.014PD2 + 0.036PD3 + 0.096PD4 + 0.198PD5 + 0.047AD2 + 0.054AD3 + 0.116AD4 + 0.164AD5) 範囲 : -0.159(55555) ~ 1(11111) (*四捨五入により、点数表と計算値にて、僅差が発生。55555の計算時、0.160)	点数 = 1 - (0.096NPH + 0.046M2 + 0.058M3 + 0.133M4 + 0.251M5 + 0.032SC2 + 0.050SC3 + 0.078SC4 + 0.122SC5 + 0.021UA2 + 0.051UA3 + 0.100UA4 + 0.175UA5 + 0.042PD2 + 0.053PD3 + 0.166PD4 + 0.207PD5 + 0.033AD2 + 0.046AD3 + 0.102AD4 + 0.137AD5 + 0.078N4) 範囲 : -0.066(55555) ~ 1(11111)																																																																																												

MAE: mean absolute error, SG: standard gamble, TTO, time trade-off

□ その他、国内にて行われたMAUIに関する価値評価研究

- 7) EQ-5Dの外にも、国内にて一般人を対象に価値評価を行った暮らしの質に関する道具は、最近、国内にて開発されたHINT-8とCH-6Dがある。
- 8) HINT-8は、韓国人のHRQoLを測定し、今後、国民健康栄養調査にて用いるために開発された韓国型の暮らしの質を測定する道具である (Jo Minwoo, 2014)。身体的、社会的、精神的、肯定的の領域に基づいた8の項目 (身体的 - 階段上り、痛み、気力; 社会的 - 働き; 精神的 - うつ病、記憶、睡眠; 肯定的 - 幸福) と項目別に4つの回答レベルからなっている。
- 9) HINT-8に関する価値評価の研究 (Jo Minwoo, 2017) は、個人の面接にパソコンで補助する方

法で、全国の男女大人1,015人を対象に行われた（表 36）。計65,536の健康状態の中、111の

健康状態に関しSGを用いて評価し、回答者は完全な健康状態である11111111、44444444、死亡状態をはじめ、計9の健康状態を評価した。最終モデルであるR2は0.563、MAEは0.043と報告された。さらに、身体的な領域である痛みと階段上りが相対的に重要と評価され、最終モデルから算出された質の加重値の範囲は、0.132（44444444）から1（11111111）として、共に死亡（0）より良い状態と評価された。

CH-6Dは、国内では初めて児童の暮らしの質を測定するべく開発された選好に基づいた一般的な暮らしの質の測定道具である（Kang Eunjungなど、2014）（表 36）。児童・青少年を対象に、海外から開発された一般的なHRQoL道具はあるが（例えば、EQ-5D-Y、HUI-2、CHU-9D）、国内にて一般大衆を対象に価値評価が行われた児童向けの一般的な暮らしの質に関する道具は、現在CH-6Dが唯一である。CH-6Dは、6つの領域（勉強、運動・走り、力、気持ち、痛み・不便性、友達と遊び）と各領域別の3~4つの回答レベルからなっている。

Hankyoreh（新聞社） & Kang Eunjung（2015）は、Seoulに住む20歳以上の男女173人を対象に、TTOを用いて2,304の健康状態にて26の健康状態に関する調査を行った。最終モデルのR2は、0.554、MAEは0.0631であった。レベル4までの領域にて、レベル4の回帰係数サイズは友達と遊ぶ、勉強、運動・走り、力の順であり、レベル3まである領域でレベル3の回帰係数は気持ち、痛み・不便性の順であった。価値評価の方法に関する例やアルゴリズムは提示されなかった。両道具の妥当性と信頼性は、HINT-8の場合は一般大衆（Lee Hyunjeongなど、2016）及び手術後の乳がん患者（Kim et al., 2019）を対象に確認されたことがあり、CH-6Dの場合は一部地域の児童（n=112）と全国のパネルから抽出された高校生（n=229）を対象に確認されたことがあるが（Kang Eunjung & Hankyoreh, 2018）、より様々な集団を対象に追加的な検証が必要であると判断される。

<表 36> その他、国内における一般的な暮らしの質向けの道具を用いた価値評価の研究

項目	Jo Minwoo (2017)	Hankyoreh & Kang Eunjung (2015)
道具	HINT-8	CH-6D
道具の開発目的	健康に関する暮らしの質を測定する韓国型の道具であるHINT-8に、一般大衆の選好を反映したtariffの開発	児童に使用できる選好度に基づいた一般
健康状態の総計	65,536（一部の回答レベルが重なる）	2,304
領域及びレベル	4つの領域に基づいた8個の項目：身体的 - 階段上り (CL)、痛み (PA)、活力 (VI) 社会的 - 働き (WO) 精神的 - 鬱 (DE)、記憶 (ME)、睡眠 (SL) 肯定的 - 幸福 (HA) 各項目別、4つの回答レベル (tariff) を計算する際、一部のレベルが重なる	6つの領域（勉強、運動/走り、力、気持ち、痛み・不快感、友達と遊ぶ）。各領域別、3~4つの回答レベル

標本数	1,015人	173人
標本の特性	Jejuを除いた全国の19才以上の男女	Seoulに住む20才以上の男女
資料収集	パソコンで補助する個人面接の方法	アンケートに基づいた直接訪問調査
測定した健康状態	111	26
価値調査の方法	VAS、SG	TTO
基本モデル R ²	0.563 (MAE=0.043)	0.554 (MAE=0.0631)
領域別、相対的な重要度の傾向	<p>痛み > 階段上り、記憶 > その他</p> $= 1 - (0.073 + 0.018 \times CL2 + 0.072 \times CL3 + 0.122 \times CL4 + 0.055 \times PA2 + 0.116 \times PA3 + 0.188 \times PA4 + 0.019 \times VI23 + 0.070 \times VI4 + 0.004 \times WO2 + 0.028 \times WO3 + 0.036 \times WO4 + 0.012 \times DE2 + 0.044 \times DE3 + 0.098 \times DE4 + 0.014 \times ME2 + 0.058 \times ME3 + 0.109 \times ME4 + 0.020 \times SL3 + 0.090 \times SL4 + 0.014 \times HA2 + 0.068 \times HA3 + 0.082 \times HA4)$ <p>範囲 : 0.132 (44444444) ~ 1 (11111111)</p>	<p>L4を含んだ領域の回帰係数 サイズ : 友達と遊ぶ、勉強、運動・走り、力の順</p> <p>L4を含まない領域の回帰係数 サイズ : 気持ち、痛み/不快感の順</p> <p>(質の加重値に関する算出例やアルゴリズムの提示はなかった)</p>
特徴のサマリー	全国調査により、韓国を代表する標本。国民健康栄養調査にて勝たする目的で開発	一部地域に住む173人を対象。国内の一般児童を対象に価値評価が行われた唯一の選好基盤の一般的な暮らしの質を測定する道具

MAE: mean absolute error, SG: standard gamble, TTO: time trade-off, VAS: visual analogue scale

(3) 標準的なMAUI及びtariffの提示可否

最も広く用いられているMAUI (例 - EQ-5D、SF-6D、HUIなど) の中、現在、国内の一般大衆を対象に価値評価が行われた道具はEQ-5Dのみであるが、EQ-5D-3Lの対象3つ (Kang Eunjung、2006; Jo et al.、2008; Lee et al.、2009) とEQ-5D-5Lの対象2つ (Jo Minwoo、2011; Kim et al.、2016a) がある。これと共に、まだ検証が必要であるが、国内にて新たに開発されたHINT-8対象1つ (Jo Minwoo、2017) とCH-6D対象1つ (Hankyoreh & Kang Eunjung、2015) がある。しかし、上記のように、同じ道具であってもどのtariffを用いるかにより、質の加重値とQALYは大きく変わる。

道具の選択によるQALY値の差はICER値へ影響を与え、結果的には政策の決定に影響を与えることがある (Arnold et al.、2009; Richardson et al.、2015a; Richardson et al.、2015b; Brazier et al.、2017; Briggs et al.、2019) 。2009年に発表されたISPOR

Task Forceの合意研究 (consensus paper) では、公平なリソースの割り当てに向け、統一された道具選定の必要性を提示し (Drummond et al.、2009) 、一貫した政策を決定するためにNICE (2013) はEQ-5Dを標準道具として明示した。ただ、ISPORの合意研究では統一された道具選定の必要性を認知しているにもかかわらず、標準道具を提示することには潜在的なリスクがあり得るとコメントした。MAUI毎に、測定する健康の属性が少しずつ違いため、一つの道具が選定される場合、同道具が重点を置く属性に影響を与える仲裁 (intervention) を中心にリソースが割り当てられかねないと述べた。

現在の指針（2011年版）が改正された当時、価値評価が行われた道具はEQ-5D-3Lが唯一であった。指針では、国内にて妥当性が検証された道具を用い、MAUI別に複数のtariffが存在する場合には、なるべく直近で開発された代表性のある標本から算出されたtariff値を基本値として使用するようにした。ここ5年間で提出された製薬会社による資料分析の結果、直近で開発されたLee et al.

（2009）のtariffが最も多く用いられたことが分かり、指針に合致する結果を表した。2018年の経済性評価TFTのレポート（健康保険審査評価院、2018）でも標本の代表性に基づきLee et al.

（2009）のtariffがより選好され、EQ-5D-5Lの場合も同理由でKim et al.（2016a）のtariffがより選好されることが分かった。

しかし、代表性に基づいて選定された両tariffにおいても、健康領域別に選好度の差が存在することが分かった。例えば、Lee et al.（2009）のtariffでは、痛み・不快感の領域が相対的に重要ではないと評価されたが、Kim et al.（2016a）のtariffでは、同領域が相対的に重要と評価された。こうした差が、少なくとも一つの研究にて代表性が欠けたことを意味するか、それとも両価値評価の研究が違うタイミングに行われ、同タイミングによる選好に変化が発生したことを意味するかは分からない。こうした選好に差が発生する状況にて、代表性を根拠に道具別に一つのtariffを選定するのは論理の矛盾を招きかねない。だからといって、MAUI及びtariffを全く設定しなければ、評価の一貫性が損なわれるだけでなく、資料の提出者が有利な方向にこれを選定する恐れがある。ちなみに、現在、イギリスのNICEではEQ-5D-3Lのみを標準道具として認めており、EQ-5D-5Lの場合は価値評価の研究に関する質の問題が提起され、使用を保留している。NICEは、製薬会社にて臨床試験を行う際、EQ-5D-5Lの使用を控える必要はないが、受容可能なイギリスのtariffが開発されるまでは、EQ-5D-5L資料をEQ-5D-3L点数に変換して使用するよう勧めている。質の評価を担当したHernandez Alava et al.（2018）はEQ-5D-5LとEQ-5D-3Lで測定された結果は、両道具間の回答レベル及びtariffの差により、かなり違う傾向が表れるため（Wailoo et al., 2017a）、一貫した政策の決定が求められる場合は両道具を相互交換的に（interchangeably）使用できないとした。一部では、EQ-5D-3L tariffに関する質の問題も提起しているが、最も代表的な 이슈の一つが線型回帰分析使用の適切性である（Briggs et al., 2019）。Briggs et al.（2019）は、既存のtariffに関する批判を考慮し、NICEなどの該当機関にて、より適切なtariffを提供しなければならない、どのtariffを使用すべきかに関し、強力な指針を提供すべきだと主張した。ちなみに、国内ではLee et al.（2009）のtariffをはじめとした国内のMAUI価値評価の研究に関し、質の評価を公式的に行ったことはない。道具及びtariffの選定がリソースの割り当てに及ぼす影響を考慮すると、価値評価の研究に関し、徹底した評価が先行され、その結果により審評院若しくは関連機関にて別途の価値評価に関する研究を行い、より適切なtariffを提供する方法も考慮する必要がある。要するに、全ての疾病領域にて最も適切かつ完璧な道具はなく（Brazier et al., 2017）、どのtariffが現時点で大衆の選好を最も反映しているかも不確実である。従って、一貫性に重点を置き、一つの道具とtariffを選定するか、それとも、疾病に一番ふさわしい道具を用いることができるように、道具とtariffの選定を柔軟にしなければならないか、これに関する議論と合意が必要である。

オ) 利害関係者の意見

(1) 専門家の意見

- 効用に関する不確実性を減らすべく、方法を明示化する必要がある。審評院にて選好する道具、推定方法などを明示。
- 特定の道具を勧告するのは適切ではないという意見もある。
- 意思決定に及ぼす重要性を考慮し国で主導する事業として、代表性のある大規模な標本を用いた標準tariffの算出研究を行う必要がある。

(2) 製薬会社の意見

- 効用測定の道具に関しては、どの道具が優れているという標準（gold standard）はない。よって、EQ-5Dを選好するように制限を設けるよりは現在の指針通りに、研究者が選んだ道具を正当化して使用できるようにすることが望ましいと思う。EQ-5D、SG、TTOを選好するが、その中で最も正当な事由のある効用値を選択し、経済性の評価に適用することが望ましいと判断する。
- 効用値を推定する際、疾病の特性、測定道具、研究の形、国や社会的な特性により結果が違ってもかかわらず、韓国ではこうした特性を看過し一律的に特定の効用tariffのみ適用しているため、医薬品の実質的な特性を反映できていない。疾病の特性を考慮し、柔軟に効用のtariffを選択し適用できるようにしなければならない。
- 暮らしの質を測定する道具（EQ-5D）を国内向けの効用に換算する際、適用するtariffに関する不確実性が大きい。経済性の評価にて、効用を算出する際に適用できる3つのtariffにより、効用値の絶対値のみならず、健康状態による効用値の変化の差が大きい。費用効果的に評価された薬剤が費用効果的に評価できない可能性もある。実際、経済性評価を行う際、保守的なtariffのみを適用し医薬品の価値が低評価されている。健康状態による効用が十分に反映されたtariffの開発が必要である。
- 国内のtariffでは、どのような方法が最も適切かに関する勧告が不在である。

- 経済性評価の制度改善TFTのレポート（健康保険審査評価院、2018）：製薬会社側では、同分野の専門家が行ったtariffの研究に関し、プライオリティを提示し難いため、一つのtariffを定めるよりは全般的な参照を許す案を提案した（製薬会社の意見）。

カ) 指針の改正方向に関する提言

政策決定の一貫性のためには、一つの道具とtariffを選定することが必要である。このため、今まで国内及び国際的に最も多く用いられているEQ-5D-3Lを一つの道具として選定し、EQ-5D-3Lの価値評価研究の中で、標本サイズが最も大きく、直近で行われたLee et al. (2009) の研究にて開発されたtariffを標準tariffとして選定する案を第1案として検討した。

しかし、国内ではまだHRQoLの道具とtariffに関する比較研究や質の評価が真面に行われたことがないため、第1案に関する実証の根拠がかなり足りない状況である。評価の一貫性も重要であるが、十分な根拠に裏付けられない意思決定は、ともすれば誤った決定を招きかねないため、既存の指針（第2案）を維持することを提言する。

- 第1案（イギリスの案）.政策決定の一貫性と代表性を根拠とし、EQ-5D-3LとLee et al. (2009) が差出したEQ-5D-3L tariffの使用を選好すると明示し、その他のtariffは敏感度の分析にて用いることにする。評価の一貫性のため、EQ-5D-5Lが用いられた場合にも算式（mapping）を用いて、EQ-5D-3L点数に変換するようにする。

第2案（既存の指針を維持した案）。道具の妥当性（validity）、敏感性

（responsiveness）、信頼性（reliability）、実行可能性（feasibility）などを基に、選好に基づいた一般道具を選択するように明示する。選択した道具の国内人のtariffが複数存在する場合、最も代表性があり、直近で行われた研究にて算出されたtariffを使用することを推奨する。

ただ、tariffの選定が意思決定及びリソースの割り当てに及ぼす影響を考慮し、今後、審評院が中心となり、既存のtariffに関する質の評価を行い、必要に応じては代表性のある大規模な標本を用いて標準tariffを開発及び提供することを提言する。

4) Mappingに関する詳細な指針

ア) 現在の指針

(1) 主な指針

基本分析では、選好度に基づいた一般道具の使用が勧められ、疾病別の道具、選好度に基づいていない道具19) を用いて測定した結果は追加的に提出することができる。

19) 選好度に基づいた道具をもって変換 (transformation) させ、変換に用いた算式 (mapping function) の構築方法論、歪みの発生可能性に関し、詳しく記述した上、同方法論を正当化し、モデルの予測性及び適合度に関し記述する。

(2) 指針に関する追加説明

個人の選好度が反映されていないSF-36、Nottingham Health

Profile (NHP)、Sickness Impact Profile (SIP) などのような道具及び選好に基づいていない疾病別の道具は算式 (mapping) により、選好に基づいた値に変換 (transformation) する方法がある。しかし、こうした算式 (mapping) は大体、予測可能性 (predictability)、モデルの適合度 (model fit) において、かなりの変異があるという根本的な限界を持っていると報告され (Brazier et al, 2007)、Chancellorなど (1997) はそもそも算式 (mapping) を変換させることが不可能だと提起するなど、方法論的な問題が存在する。

現在の指針では、選好度に基づかない道具をもって、健康状態を測定した場合、算式 (mapping) を用いて、選好度に基づいた値に変換させた上、この結果を「追加的に」提出できると記載している。指針の追加説明には、算式 (mapping) の限界及び方法論的な問題を中心に記述されている。

イ) 既存の提出資料を検討した結果

表 37にて提示されたように、過去5年間 (2014-2018) で算式 (mapping) を用いて効用を算出したケースは計6件であった。6件は全て、既存の文献から提示した算式 (mapping) もしくはモデル設定 (model specification) を用いたが、算式 (mapping) の文献考察過程を報告したのは3件のみであった。算式 (mapping) の構築過程、モデルの適合性、予測力などに関し、詳しく報告した事例もほとんどなかった。

臨床試験にて用いられた疾病別の元の道具は、疾患により様々であったが目標道具 (target measure) は、全てEQ-5D-3Lと同様であった。EORTC QLQ-C30を元の道具として用いたケース (計2件) には、国内にて開発された算式 (mapping) を用いて韓国の一般大衆の選好が反映されたEQ-5D点数に変換した。この他、疾病別の道具が用いられた場合には、大体イギリスにて開発された算式 (mapping) もしくはモデルの設定 (model specification) を用いて、EQ-5D点数に変換したが、これは算式 (mapping) の可用性が反映された結果とみなし得る。イギリス・NICEの指針では、算式 (mapping) を用いる場合もイギリスの一般大衆の選好が反映されたEQ-5D-3L点数に変換することを勧めている。これにより、算式 (mapping) の多くがイギリスの一般大衆の選好を反映したEQ-5D-3L点数を目標変数にして開発された (Mukuria et al., 2019)。

しかし、算式 (mapping) を用いた場合であってもこれを全ての健康状態に適用したケースは3件のみ

であった。他の事例（3件）では、「維持状態」や「基本的な健康状態」のみに算式（mapping）を用いて算出された質の加重値を適用、「重症の健康状態」には標本数などの理由で既存の文献にて提示した値を用いた。なお、臨床試験では元の疾病別の道具を用いて測定された暮らしの質の改善が仲裁群の間で類似であると報告されたが、EQ-5D点数に変換する際、治療群の間で差が発生し、治療群のタイプ別に違う質の加重値を健康状態に付与した事例も1件あった。

(1) 今後、意思決定の過程にて算式（mapping）の利用に関する不確実性を減らすためには、少なくとも（1）算式（mapping）を用いる際、これに関する情報を詳細に提供し、評価者が算式（mapping）に関わる質の評価を十分に行えるようにしなければならない、（2）主な臨床試験に疾病別の道具もしくは選好に基づかない道具が含まれたら、これに関する資料も共に提出して変換された値を正当化するようにしなければならない。

<表 37> 算式（mapping）を用いて質の加重値を導き出した事例の分析

事例	対象疾患	元の道具	目標道具	選好度	算式の出典	算式の文献考察	報告する項目
1	転移性去勢抵抗性前立腺がん	FACT-P	EQ-5D-3L	イギリスの一般大衆	Spencer et al. (2011) (口頭発表の資料) (同一がんの対象)	はい	最終モデル：OLS、モデルの適合度及び予測力に関するコメントはなし
2	再発性慢性重症湿疹	PGA (臨床資料) をDLQIに換算	EQ-5D-3L	イギリスの一般大衆	既存の文献 PGA (臨床資料) * DLQIに換算 → EQ-5Dに換算	コメントなし	コメントなし
3	重症の喘息	AQLQ	EQ-5D-3L	不明確	Sheffield HE 回帰モデル (Tsuchiya et al., 2002 [対象者をコメント	はい	最終モデル：OLS、モデルの適合度及び予測力に関する正確な数値のコメントはなし

					なし]) を用いて直接回帰分析を実行		* 喘息が悪化した健康状態の場合、既存の文献値を利用
4	多発骨髄種	EORT CQLQ-C30	EQ-5D-3L	韓国の一般大衆 (Tariff : Lee et al, 2009)	Kim et al. (2012a) (様々な癌を対象)	はい	最終モデル : OLS、修正された $R^2 = 0.51$ * Post progression 健康状態の場合、既存の文献値を利用
5	好酸球性重症喘息	SGRQ	EQ-5D-3L	イギリスの一般大衆	Starkie et al. (2011) (COPD の対象)	コメントなし	最終モデル : OLSRMSE = 0.1723 * 喘息の悪化 健康状態の場合、既存の文献値を利用
6	転移性すい臓がん	EORT CQLQ-C30	EQ-5D-3L	韓国の一般大衆 (Tariff : Lee et al., 2009)	Kim et al. (2012b) (転移性乳がん)	コメントなし	最終モデル : OLS モデル適合度及び予測力に関するコメントなし

※ 2014-2018年に国内にて提出された製薬会社の資料分析 (最初に提出された資料の主な基本分析の基準)
 AQLQ: asthma quality of life questionnaire, COPD: chronic cbstruction pulmonary disease, DLQI: dermatologylife quality index, EORTC QLQ-C30: European Organization for Research and Treatment core quality of lifequestionnaire, FACT-P: functional assessment of cancer therapy-prostate, OLS: ordinary least squares, PGA:physicians global assessment, RMSE: root mean square Error, SGRQ: St. George's respiratory questionnaire

ウ) 外国の指針

カナダ (CADTH、2017)、オーストラリア (PBAC、2016)、イギリス (NICE、2013) の指針では、算式 (mapping) を制限的に許していることが分かった。オーストラリアやイギリスの場合は、MAUI (イギリスはEQ-5D-3L) の資料がない場合、算式 (mapping) を用いることが可能とし、カナダの指針では算式 (mapping) の利用を勧めないが、他の代案がない場合は制限的に用いることができるとした (表 38)。各指針では、算式 (mapping) を用いる際、これに関する正当化を行い、算式 (mapping) に関する詳しい情報を提供すると共に、関連した不確実性を十分に検討することを明示している。

(1) カナダ (CADTH、2017)

カナダの指針では、説明力 (explanatory power) を主な理由とし、算式 (mapping) の利用を推奨しないと明示している。元の道具 (source measure) 及び目標道具 (target measure) が共に選好に基づいた一般道具の場合には、両道具間で測定する健康領域が類似であるため、両道具間の相関 (correlation) が高い可能性があるが、測定する健康領域が相対的に違う疾病別の道具と一般道具を用いる場合は両道具間の相関が低いいため、算式 (mapping) の説明力が相対的に低いことがあるとコメントした。従って、元の道具と目標道具のタイプ、算式 (mapping) アルゴリズムのタイプ、含まれた健康状態の重症度などにより、予測値が大きく変わることがあるため算式 (mapping) の利用を推奨しないとした。ただ、算式 (mapping) によってのみ質の加重値の推定が可能な場合は制限的に算式

(mapping) を用いることができるが、この場合は検証された算式 (mapping) のアルゴリズムを用いて、これに関する実証資料 (empirical data) を提示するようにしている。なお、算式 (mapping) に関する不確実性を評価へ十分に反映することを明示している。

(2) オーストラリア (PBAC, 2016)

オーストラリアの指針では、臨床試験に含まれた患者の報告成果指標 (patient reported outcome, PRO) の道具がMAUIではない場合、これに関する正当化と共に、含まれたPRO道具に関する詳細な情報を提供し、算式 (mapping) を用いてこれを選好に基づいた値に変換することを明示している。算式 (mapping) を用いる場合は算式 (mapping) アルゴリズムの算出に用いた推定標本 (人口学的・臨床的な特性、標本サイズなど) に関して詳しく説明し、別途の検証 (validation) 標本があるかを記述するようにしている。また、元の道具 (source measure) と目標道具 (target measure) の指標 (index) 及び領域 (dimensional) 、算式 (mapping) のアルゴリズムを推定するために用いた統計モデル及び方法についても詳しく記述するようにしている。アルゴリズムの実行能力 (performance) と妥当性 (validity) を測定するために用いた方法と算式 (mapping) が提出された資料に相応しいか、とりわけ推定標本と関連付けて議論するようにしている。また、算式 (mapping) の不確実性と共に予測された質の加重値を提示することを明示している。

(3) イギリス (NICE, 2013)

イギリスのNICE指針では、EQ-5D-3L資料がない場合、算式 (mapping) を用いて他のHRQoL道具をもって測定された資料をEQ-5D-3L点数に変換できると明示している。指針では、EQ-5D-5Lも基本分析に用いられることができると明示しているが、現在、受容可能なイギリスのEQ-5D-5Lのtariffがない状態であるため、EQ-5D-5L資料もvan Hout et al. (2012) が開発した算式 (mapping) (“cross-walk”) を用いて、EQ-5D-3L点数に変換するようにしている。算式 (mapping) を選ぶ際は、EQ-5D-3Lと他のHRQoL道具を共に含めた資料から算式 (mapping) が算出されたか、統計的な属性が十分に記述され正当化されたか、算式 (mapping) が資料に相応しいかなどを考慮するようにしている。また、算式 (mapping) アルゴリズムを用いて算出された結果値を敏感度の分析にて提示するように明示している。

(4) その他

EUnetHTA (2015a) によると、多くのヨーロッパ諸国における指針では、算式 (mapping) に関する立場を提示していないが、他の代案がない場合は算式 (mapping) を認める国も多数あると把握されている (例えば、イギリス、チェコ、アイルランド、イタリア、ノルウェーなど) (EUnetHTA, 2019) 。

しかし、EUNETHTA（2015b）にて発行した効用に関わる相対的な効果の評価（relative effectiveness assessment）に関する指針によると、算式（mapping）は原則的に直接的な資料の収集方法より劣等（inferior）であるため、一般的に勧めないとしている。なお、算式（mapping）の必要性を減らすには、同HTA機関にて資料の提出者が臨床試験のプロトコルに、選好に基づいた道具を常に含めるように勧めなければならないとコメントした。

<表 38> 算式（mapping）に関する海外の主な国における指針内容

国	指針内容の原文	サマリー
カナダ CADT H 指針（4 版） (2017)	<p>Statistical mapping functions have been used to convert from onemeasure to another (e. g., HUI to EQ-5D) in an effort to maintain consistency, but mapping as a means of deriving health utilities is not recommended. (p47)</p> <p>When there is no alternative but mapping to obtain utility estimates, empirical data are required. The mapping should be undertaken using a validated algorithm. The full uncertainty in the mapped estimates, reflecting parameter uncertainty in the source algorithm, parameter uncertainty in the mapping algorithm, and uncertainty regarding which HRQoL states are represented in the clinical health states included in the model, should be incorporated into the evaluation. (p47)</p>	<p>算式（mapping）の利用は勧めない。他の代案がない場合、検証されたアルゴリズムを用いて、提出。算式に関する福確実性を十分に検討。</p>
オー ストラ リア PBA C 指針 version 5 (2016)	<p>Mapping of generic and disease-specific scales</p> <p>Nonpreference-based patient-reported outcome measures will require a mapping algorithm to be transformed into preference-based measures to estimate utilities. Where this occurs, detail the source of the mapping algorithm. Describe the estimation sample (population demographic and clinical characteristics, sample size etc) and whether there is an external validation sample. Provide details of the source and target measures (eg index, dimensional), and the statistical model and methods used to estimate the mapping algorithm. Detail the statistical association or operations that constitute the algorithm. Discuss methods used to measure the algorithm performance and validity. Present the resulting predicted utilities with associated uncertainty. Discuss the applicability to the submission data, particularly in relation to the sample in which the algorithm was developed. (p78)</p>	<p>算式を用いて選好に基づいた効用値推定可能。算式を用いる際、算式のアルゴリズム、標本、検証（validation）標本の可否、元の道具と目標道具、統計モデル、予測力などに関し、詳しく記述し不確実性を検討。</p>

イギリス NIC E 指針 (2013)	When EQ-5D data are not available, these data can be estimated by mapping other health-related quality of life measures or health-related benefits observed in the relevant clinical trial (s) to EQ-5D. The mapping function chosen should be based on data sets containing both health-related quality of life measures and its statistical properties should be fully described, its choice justified, and it should be adequately demonstrated how well the function fits the data. Sensitivity analyses to explore variation in the use of the mapping algorithms on the outputs should be presented. (p41-42)	EQ-5Dの資料がない場合、算式を用いてEQ-5D点数に変換可能。 選択した算式の統計的な特性に関して詳しく記述し正当化、モデルの適合度を証明。算式に関する敏感度の分析。
----------------------------------	---	--

工) その他、文献考察の結果

(1) 算式 (mapping) に関する研究現況

選好に基づいた一般道具の代わりに、疾病別の道具や選好に基づいていない一般道具が用いられた際は、一般大衆の選好を反映した質の加重値の算出が原則的に不可能である。この際、代案として用いられる一つの方法が、算式 (mapping) であるが算式 (mapping) を用いて同道具をもって測定された資料を選好に基づいた値に変換することができる。検証された算式 (mapping) がある場合、これを用いて質の加重値を相対的に容易かつ迅速に推定できるため、ここ10年間余りで算式 (mapping) の研究はかなり増えた (Mukuria et al., 2019)。とりわけ、イギリス・NICEの指針では、EQ-5D-3Lの代わりに他のHRQoL道具が用いられた場合、算式 (mapping) を用いてこれをEQ-5D-3L点数に変換することを明示しているため、EQ-5D道具を目標道具 (target measure) とする算式 (mapping) の開発研究がイギリスを中心に活発に行われている。ちなみに、Oxford HERC (Health Economics Research Centre) では、2013年に算式 (mapping) に関するデータベースを構築し、現在182の算式 (mapping) の研究とこれに含まれた算式 (mapping) アルゴリズムに関する情報を提供している。

先行研究からすると、既存 Brazier et al. (2010) が行った体系的な文献の考察では、1996年~2007年の初めに発表された計30の算式 (mapping) に関する研究を確認してこれを考察に含めた。最も一般的な統計分析方法は、最小二乗法 (ordinary least squares, OLS) を用いた回帰分析であり、他の分析方法は殆ど用いられなかった。

モデルの適合度と予測力に関する情報を提供するべく、殆どのR²、RMSE (root mean square error; 平均二乗偏差) もしくはMAE (mean absolute error; 平均絶対誤差) などを報告したが、含まれた研究の説明力は0.17~0.71、RMSE若しくは0.084~0.2、MAEは0.011~0.19レベルであった。Brazier et al. (2010) は疾患の重症度が高い場合は質の加重値が大きく推定され、低い場合は小さく推定される傾向があると報告したが、とりわけ、重症患者の質の加重値を推定する場合は過大推定 (即ち、健康状態の効用が高く推定) に関する注意が必要であると述べた。

最近、行われた体系的な文献の考察 (Mukuria et al., 2019) では、2007年~2018年に発表された一般道具を目標道具とする算式 (mapping) の研究を検討した。同期間中に発表された算式 (mapping) に関する論文は計180、算式 (mapping) のアルゴリズムは計233として、既存の Brazier et al. (2010) による文献考察からの結果と比べ、10年間で算式 (mapping) に関する研究が大幅に増えたことが分かった。算式 (mapping) のアルゴリズムにて、最も多く用いられた目標道具はEQ-5Dとして、計147のアルゴリズムにて用いられた。既存のBrazier et al. (2010) による報告とは違い、

様々な回帰分析の方法が算式（mapping）の推定に用いられた。効用資料の偏り、天井効果、範囲の制限などを考慮し、Tobitモデル、CLAD（censored least absolute deviation; 中途切断最小絶対偏差）モデル、Two-partモデルなど、様々な代案がOLSと共に用いられる事例が多かった。しかし、予測力の面では、OLSが他の代案並みの水準か、より良い評価を受け、基本分析では相変わらず最も多く用いられていると報告された。また、予測力に関しては、依然としてRMSEとMAEを最も多く報告していることが分かったが、以前とは違い、症状の重症度により予測力を別途に報告する研究が次第に増えていると報告された。Mukuria et al.（2019）は、算式（mapping）研究における質の向上に向け、算式（mapping）の開発研究者はISPOR Task Force（Wailoo et al., 2017b）、Ara et al.（2017）、Longworth & Rowen（2013）などが開発した指針を参考とし、算式（mapping）を開発、Petrou et al.（2015）が開発したMAPS（MApping onto Preference-based measures reporting Standards）の報告指針などに則り、報告することを推奨した。なお、算式（mapping）の研究が活発に行われることは別途に臨床試験や観察研究にて、一般道具の使用を回避しようとする目的で算式（mapping）を使用してはいけないと強調した。

(2) 国内における算式（mapping）に関する研究

(3) 表39に提示されたように、現在まで発表された国内における算式（mapping）の研究は、計7つと推定される。同研究では、癌、転移性乳がん、関節炎、COPD（chronic obstruction pulmonary disease; 慢性閉塞性肺疾患）患者のみならず、一般大衆を対象に測定した資料を用いて算式（mapping）を開発した。元の道具には、EORTC-QLQ-30などの疾病別の道具及びSF-36などの一般道具も含めたが、目標道具は全てEQ-5Dとして同様であった。基本分析では、OLS系統の分析モデルを利用、殆どの研究にてR²、RMSE、MAEなどを用いてモデルの適合度及び予測力を報告した。

(4) より詳しく同研究を調べると、まずKim et al.（2012a）は893人の癌患者（乳がん32.9%、大腸がん20.0%を含む）を対象に測定されたEORTC QLQ-C30資料を用いてEQ-5D-3L点数を予測した。EORTC QLQ-C30項目及び領域のタイプが異なるようにモデルを構成、分析方法としてはOLSを用いた。最終モデルの修正されたR²は0.516、RMSEは0.095、MAEは0.069として良好なレベルが表れたが、相対的に深刻な患者群では効用が過大推定される傾向があると報告した。重症度別に区分して分析する際、重症度が相対的に低い患者群ではMAEが0.056、高い群では0.078として後者にて予測力が少し低下することが分かった。モデルの検証は、別途の大腸がん患者123人を対象に行われたが、モデルの適合度及び予測力の面で、基本分析と類似な結果が表れた。

Kim et al.（2012b）は、199人の転移性乳がん患者を対象に測定されたEORTC QLQ-30とEORTC QLQ-BR23の資料を用いて、EQ-5D-3L点数を予測する算式（mapping）を開発した。199人の患者の中、149人の資料は算式（mapping）の推定に用い、他の50人の資料は算式（mapping）の検証に用いた。EORTC QLQ-30点数、EORTC QLQ-BR23点数、人口学的な特徴を含む可否により、計6つのOLSモデルを構成した。この中で、EORTC QLQ-30資料のみ使用したモデルにて説明力及び予測力が最も高かったが、修正されたR²は0.487、MAEは0.092であった。同結果は、Mukuria et al.（2019）が行った文献の考察（Mukuria et al., 2019）にて報告したと類似なレベルであったが、方向性の面では予測とは裏腹の回帰係数が一部報告された。Kim

et al. (2012b) は、EORTC C-30項目間で知られていない相互作用を可能な理由として提示、類似な既存の事例も述べたが、明確な理由に関してはより研究が必要とみられる。

最近、発表されたLim et

al. (2019) の研究では、299人のCOPD患者を対象にOLSのみならず、GLM (Generalized Linear Model; 一般化線形モデル)、Tobitモデル、Beta回帰分析などを用いて、COPD assessment test資料をEQ-5Dに変換する算式 (mapping) を開発した。患者の資料を50%ずつ分けて、それぞれ推定標本と検証標本として用いた。OLSを使用した際、RMSEが最も少なかったが、この際、RMSEが約0.11、MAEが約0.08であり、修正されたR²は別途に提示されなかった。研究では、標本サイズが相対的に小さく、重症患者が多く含まれなかったことを制限事項として提示した。

一方、他の研究 (Kim et al., 2016b) では、関節炎患者3,557人を対象に測定されたHAQ-DI (health assessment questionnaire disability index)、痛み-VAS、DAS28 (disease activity score-28) 資料を用いて、EQ-5D点数に変換するための算式 (mapping) の研究を行った。標本の80%は推定のために、使い残りの20%は検証のために用いた。HAS-DI、痛み-VAS、DAS28を含む可否と統計方法 (OLS、Tobit モデル、two-part モデル) により複数のモデルを構成したが、OLSを用いる際に修正されたR²が0.576と最も高く、RMSEは0.1654、MAEは0.1222と報告された。ちなみに、RMSEとMAEは検証のほうが少し高かった。他の3つの研究 (Kim et al., 2014; Park et al., 2017; Lee Jungeun & An Jeonghoon, 2019) では、一般大衆の一部若しくは全部を対象にし、算式 (mapping) の研究を行った。まず、Kim et al. (2014) は一般大衆、糖尿病患者、脳卒中患者2,211人を研究に含めたが、この中で1,660人の資料は推定のために使用し、551人の資料は検証のために用いた。この他にも、大腸がん患者123人の資料を別途に収集し、これを算式 (mapping) の検証に用いた。同研究では、一般道具SF-36の資料をEQ-5D点数に変換するために、OLS、two-partモデル、多項ロジスティック回帰モデルを用いて分析した。元の道具と目標道具を全て一般道具として使用し分析した結果、修正されたR²は0.6366~0.7476として、相対的に高い説明力を表し、この際、RMSEは0.133 ~ 0.160、MAEは0.087 ~ 0.109レベルであった。Kim et al. (2014) は、一般大衆から脳卒中患者まで様々なHRQoLレベルを含んだ資料を用いて算式 (mapping) を算出し、別途の患者資料を用いて追加的にこれを検証したため、様々な患者群にてこの算式 (mapping) を利用できる可能性もあると述べた。ただ、外部の検証資料に含まれた患者の重症度が相対的に低かったため、重症患者を対象に算式 (mapping) を見直す必要があると述べた。Lee Jungeun & An Jeonghoon (2019) の研究でも、一般大衆220人を対象に収集されたHINT-8 (HRQoL Instrument with 8 items) (一般道具) の資料を用いて、EQ-5D-3L点数とEQ-5D-5L点数を予測する分析を行った。OLSを用いて分析した結果、修正されたR²がEQ-5D-3Lでは0.53、EQ-5D-5Lでは0.66として、後者にてより高い説明力を表した。しかし、RMSE、MAEなどの誤差に関する情報や検証の可否に関するコメントが別途に提示されなかったため、算式 (mapping) の質を評価し難く、全ての領域にて「問題なし」と回答した事例がEQ-5D-3Lの場合51.4%、EQ-5D-5Lの場合は47.3%と報告され、重症度別に算式 (mapping) を追加検証する必要があると述べた。

最後に、Park et al. (2017) は癢痒症-VAS資料をEQ-5D-3L点数に変換する算式 (mapping) を開発するべく、268人の一般大衆を対象に断面研究を行った。仮想の健康状態3

つを各参加者へ提示し、これを瘡痒症-VASとEQ-5D-3Lを用いて評価するようにした。収集された資料の50%は推定に用い、残りは検証に用いた。回答者の反復測定（repeated measure）を考慮し、多レベル（multilevel）分析を行ったが、分析した結果、Log likelihoodが345.32、RMSEが0.157、MADが0.145として、誤差レベルが他の研究より少し高く報告された。

<表 39> 国内における算式（mapping）の研究

研究	元の道具	目標 道具	対象研究	従属変数	統計方法	モデル適合度及び予測力
Kim et al. (2012a)	EORTC CQLQ-C30	EQ-5D-3L (Lee et al.,2009)	様々な癌 n=893 (別途の検証標本[大腸癌] n=123)	EQ-5D 点数	OLS	adj.R2 : 0.516 RMSE : 0.095 MAE : 0.069
Kim et al. (2012b)	EORTC QLQ-C30 EORTC QLQ-BR23	EQ-5D-3L (Lee et al.,2009)	転移性乳がん n=199 (推定標本 n=149; 検証標本 n=50)	EQ-5D 点数	OLS	QLQ-C30 adj.R2 : 0.487 MAE : 0.092 QLQ-BR23 adj.R2 : 0.246 MAE : -
Kim et al. (2014)	SF-36	EQ-5D-3L (Lee et al.,2009)	一般の韓国人、 糖尿病の外来患者、 脳卒中患者	EQ-5D 点数 (EQ-5D 項目のレベル)	OLS (Two-part model) (Multinomial)	adj.R2 : 0.6366 - 0.7476 RMSE :

			n=2211 (推定標本 n=1660; 検証標本 n=551) (別途の検証 標本 大腸がん 患者 n=123)		logistic model)	0.133- 0.160 MAE : 0.087- 0.109
Kim et al. (2016b)	HAQ-DI VAS DAS28	EQ-5D-3L (Tariff明示せず)	関節炎 (n=3557) (推定標本 80%;検証標本 20%)	EQ-5D 点数	OLS (Tobit model) (Two-part model)	adj.R2 : 0.576 RMSE : 0.1654 MAE : 0.1222
Park et al. (2017)	搔痒症-VAS	EQ-5D-3L (Lee et al.,2009)	韓国の一般人 n=268 (推定標本 n=137; 検証標本 n=131)	EQ-5D 点数	Multilevel分析 (同一患者の複数の観察地を提供)	Log likelihood : 345.32 RMSE : 0.157 MAD : 0.145
Lim et al. (2019)	COPD assessment test (CAT)	EQ-5D-3L (Lee et al.,2009)	COPD n=299 (推定標本 50%;検証標本 50%)	EQ-5D 点数	OLS (GLM) (Tobitモデル) (Beta回帰) (Two-part モデル)	adj.R2 : NA RMSE : 約0.11 MAE : 約0.08
Lee Jungeun & Ahn Junghoon (2019)	HINT-8	EQ-5D-3L (Lee et al.,2009) EQ-5D-5L (Kim et al., 2016a)	韓国の一般人 (n=220) (検証に関するコメントなし)	EQ-5D-3L index EQ-5D-5L index	OLS	EQ-5D-3L adj.R2 : 0.53 EQ-5D-5L adj.R2 : 0.66

COPD: chronic obstruction pulmonary disease, DAS28: disease activity score-28, HAQ-DI: health assessment questionnaire disability index, GLM: generalized linear model, HINT-8: HRQOL instrument with 8 items, MAE: mean absolute error, OLS: ordinary least squares, RMSE: root mean square error, VAS: visual analogue scale

(5) 算式 (mapping) の手順及び報告に関する推奨事項

最近、ISPOR Task Forceにて開発した指針 (Wailoo et al., 2017b) をはじめ、算式 (mapping) の手順及び報告に関する指針が複数出版された (Longworth & Rowen, 2013; Wailoo et al., 2015)。こうした指針開発の影響で、算式 (mapping) の研究は量的にも質的にも改善していると報告された (Mukuria et al., 2019)。ISPOR Task Forceの指針 (Wailoo et al., 2017b) とこれを修正 - 変更したAra et al. (2017)の指針では、算式 (mapping) に関する研究の開始前から報告までの全てのプロセスを包括的に取り扱っている。これに

反し、Longworth & Rowen (2013) が開発した指針では、算式 (mapping) の開発そのものに重点を置き、推奨事項 (recommendations) を提示している。この指針では、EQ-5Dを目標道具として設定して説明したが、これは他の一般道具にも同様に適用できる。

Longworth & Rowen (2013) によると、**算式 (mapping) を導き出すプロセスは大きく①**

算式 (mapping) を導き出す資料の構成要素定義②モデル設定 (model specification) ③統計的な推定方法の選定④モデルの適合度 (model fit) と予測力 (predictability) の確認⑤導き出した算式 (mapping) を必要な対象の資料に適用する段階に区分することができる。これをより細分化し、推定標本 (estimation sample)、従属変数 (dependent variable)、独立変数 (independent variable)、モデルの選択及び設定 (model selection and specification)、モデルタイプ (model type)、性能 (performance)、検証 (validation)、不確実性 (uncertainty) の面にて考慮事項を提示したが、これに関する詳細は表40を参考とすることができる。

ちなみに、同指針では、算式 (mapping) の限界と注意事項なども共に提示したが、とりわけ、Ara et al. (2017) は①元の道具と目標道具において重複される概念が殆どない場合、②対象集団に元の道具や目標道具の使用が不適切な場合、③不適合なモデルとモデルの設定 (model specification) により算式 (mapping) が推定された場合、④算式 (mapping) を算出した標本がこれが適用される資料を真面に代表できない場合は同算式 (mapping) の利用は不適切 (inappropriate) であり、不正確な結果を提供するはずとした。

<表 40> 算式 (mapping) の手順と推奨事項

項目	算式(mapping) の手順及び推奨事項
推定標本 (Estimation sample)	<ul style="list-style-type: none"> • 推定標本と算式(mapping)アルゴリズムが適用される目標 (target)標本間の臨床的 - 人口学的な特徴が類似でなければならない。 • 推定標本と目標標本にて、算式(mapping)に含まれた共変量(covariates)の分布が重ならない。 • EQ-5D点数に影響を与えかねないと思われる目標標本内に含まれた変数は推定標本にも含まなければならない。 • EQ-5D資料と説明変数(explanatory variable)を共に含む資料がなければ、EQ-5D資料を収集し算式(mapping)を推定しなければならない。
従属変数 (Dependent variable)	<ul style="list-style-type: none"> • EQ-5D点数(index) • EQ-5D領域別のレベル(dimension levels)
独立変数 (Independent variable)	<ul style="list-style-type: none"> • 疾病別の道具：点数の総計、要約点数、項目のレベル点数、項目レベルのダミーレベル、相互作用項 (interaction terms)、二乗項 (squared terms)、三乗項(cubic terms)。 • 一般道具：点数の総計、要約点数、項目のレベル点数、項目レベルのダミー変数、相互作用項、二乗項、三乗項。 • 臨床装置(clinical measures)：総点、要約点数、範疇型ダミー変数、社会人工学的な変数。 • その他の適切な健康関連の資料。

モデルの選択及び設定(Model selection and specification)	<ul style="list-style-type: none">• 臨床的関係(clinical relationships)に関する事前知識(prior knowledge)の利用。• 算式(mapping)を持った推定前に、標準統計技法を用いて資料を検討 (例 - 頻度、相関分析)。• 回帰モデルの推定に向け用いられた資料に関し、十分に記述 (例 EQ-5D 範囲とEQ-5D分布を表すグラフなどを含む)。
---	--

	<ul style="list-style-type: none"> 費用-効果のモデルに用いられたEQ-5D予測値の範囲に関し十分に記述
モデルタイプ (Model type)	<ul style="list-style-type: none"> Linear ordinary least squares Tobitモデル Censored least absolute deviation Two-part model Generalized linear model Latent class mixture model Censored mixture model Multinomial logit model
モデルの適合度及び予測力 (Performance)	<ul style="list-style-type: none"> モデル適合度(goodness of fit) <ul style="list-style-type: none"> 統計的な有意性、回帰係数の方向(sign)とサイズ(size)、R²、修正されたR²、AIC、BIC。 モデル適合度の検定のために、追加でRamsey RESETを検定、Parkの検定、Jarque-Beraの検定などを実行。 モデルの仮定(model assumptions)が妥当かをグラフを用いて検討。 予測力(predictive ability) <ul style="list-style-type: none"> Root mean squared error(RMSE) and mean squared error(MSE) EQ-5D若しくは予測値の重症度の範囲 (severity range)により、RMSE, MSE, mean error, mean absolute error(MAE) 監査されたEQ-5D点数と予測されたEQ-5D点数を用いたグラフ
検証 (Validation)	<ul style="list-style-type: none"> 検証標本(validation sample)に適用する際、開発された算式(mapping)の適用(application)及び評価(assessment) 検証標本は推定標本と無関係の別途の患者資料である可能性もあり、算式(mapping)の推定に用いる資料を無作為で推定標本と検証標本にわけられる可能性もある。
不確実性 (Uncertainty)	<ul style="list-style-type: none"> 質の加重値に関する不確実性は、経済性評価に反映されなければならない。複数の利用可能な算式(mapping) アルゴリズムを用いて、質の加重値に関する敏感度の分析を行うことができる。

出典: Longworth & Rowen (2013)

※ EQ-5Dを目標道具として説明したが、この推奨事項は他の目標道具にも同様に適用できる。
 参考: Ramsey RESETの検定、model misspecification(誤ったモデルの設定) 検定; Park検定、分散不均一性(heteroscedasticity)の検定; Jarque-Beraの検定、正規性(normality)の検定
 AIC: Akaike's information criterion, BIC: Bayesian information criterion

オ) 利害関係者の意見

(1) 専門家の意見

- 現在は算式(mapping)の受容れが時によって変わる傾向がある。算式(mapping)が、いつ利用可能かに関する基準を明確にする必要がある。

(2) 製薬会社の意見

- 現在、国内ではEQ-5Dを基本分析として資料を要請しているが、臨床試験にて測定された資料をEQ-5D点数に変換した算式(mapping)も基本分析として受け容れる必要がある。

カ) 指針改正方向に関する提言

最近、算式（mapping）の手順及び報告に関し、複数の指針が開発され、こうした指針開発の影響を受け算式（mapping）に関する研究が量的にも質的にも改善されていることが分かった。これを反映し、指針の改正版で一般道具をもって測定された資料がない場合、算式（mapping）も受容可能であると明確にすることを提言する。ただ、この場合は評価者が算式（mapping）の質に関する評価を十分行えるよう、下記のいくつかの項目を報告する。ようにする。必要がある。

- 選好に基づいた一般道具を用いない場合、これに関する理由を十分に説明する。ように明示する。算式（mapping）をもって、選好に基づいた値に変換する。際、算式（mapping）に関する文献の考察を行い、算式（mapping）の構築方法及び限界に関し詳しく記述する。と共に、敏感度の分析により算式（mapping）に関する不確実性を十分に検討する。ように明示する。
- また、算式（mapping）を用いる場合は下記の項目に関する情報を追加的に提供する。ようにする。
 - ① 元の道具と目標道具
 - ② 目標道具に適用されたtariff（出典及び選好度の測定対象）
 - ③ 算式（mapping）の開発研究に含まれた標本サイズ
 - ④ 算式（mapping）の開発研究に含まれた回答者の人口学的 - 臨床的な特徴と算式（mapping）が適用された研究に含まれた回答者の人口学的 - 臨床的な特徴の類似性を検討
 - ⑤ モデル設定（model specification）の根拠
 - ⑥ 統計的な予測方法（model type）と選定理由
 - ⑦ モデルの適合度及び予測力、できれば重症度のレベルによってもモデルの適合度及び予測力に関する情報を提供
 - ⑧ 算式（mapping）の検証（validation）可否及び検証におけるモデルの適合度及び予測力
 - ⑨ 算式（mapping）の限界
 - ⑩ の使用による効用の不確実性を検討
 - ⑪ 算式（mapping）の結果と臨床試験から収集された元の道具をもって測定された資料を比較

5) 直接測定に関する詳細な指針

ア) 現在の指針

(1) 指針内の追加説明

(2) 直接測定時に報告すべき項目

資料の提出者が直接選好度を測定する場合、下記の項目に関する情報を提供すること。

- ・ 用いられた道具、選定の根拠を提示（信頼性、妥当性、実行の容易性）
- ・ 標本サイズ
- ・ 回答者の（該当の疾患に関する。）健康状態及び人口学的な特徴（必要最低限で年齢、性別）
- ・ インタビュー方法（郵便・対面・インターネット）
- ・ 標本地域
- ・ 回答率
- ・ 調査者間の差を評価
- ・ 完全な健康の定義
- ・ シナリオに基づき、選好度を導き出した場合、別添5に基づくこと

*別添5：シナリオに基づき、選好度を導き出した場合に考慮しなければならない項目

- 1) 各シナリオに健康状態を導き出した根拠を提示し、各治療法に関する。患者の経験を反映する。べく、どのような方法を用いたかを説明すること。（例えば、患者を対象にしたフォーカスグループを根拠にしたか、小規模なインタビューを行ったかなど）
- 2) シナリオが健康状態の選好度を妥当かつ敏感に測できるか、及び包括性があるかの可否を第3の(independent)検討者が評価したかを報告すること。
- 3) 一般大衆を対象にするため、表現の用語はなるべく分かり易くすること。
- 4) ex postの観点（即ち、回答者が現在患者だと想定）を取り、回答者に他の疾病がある可能性及び診断に関する。可能性などを排除する。
- 5) 表現及び言葉の選択により、回答者の反応が異なりかねないため、回答者に精一杯治療に関する。全体的な情報を提供し、どのような方法で提供したかを提示すること。これにより、回答者の理解程度を評価することを推奨する。
- 6) スポンサー（sponsor）による歪みを防ぐため、できればスポンサー（sponsor）名は明かさないようにすること。
- 7) 一般大衆の対象に質問する。場合、回答者の集中度向上と視覚的な補助器具（visual aid）などを用いることができるため、一般的に対面インタビューのほうが選好される。もし、対面インタビューを用いた場合は調査者の経験・教育程度などを明示する。

現在の指針では、シナリオに基づき健康状態の効用を直接測定する際、（SGとTTO方法に関する説明の外）報告すべき項目は提示しているが、シナリオの構成及び評価方法に関する詳細な指針は別途に提示していない。

イ) 既存の提出資料の検討結果

表41から提示されたように、去る5年間（2014-2018）¹¹⁾ 直接測定方法を用いて質の加重値を導き出したケースは計9件であった。直接測定方法を選んだ理由に関して明確に説明した事例は殆どなかったが、資料の分析結果、直接比較可能な臨床資料の不在6件、HRQoL資料の不在・制限2件、疾病別道具の使用1件が、その理由として把握された。

シナリオの構成に向け、初期に文献を考察した事例が計4件、文献考察と共にEQ-5D及びK-ALSFRS-R (Korean amyotrophic lateral sclerosis functional rating scale-revised) を基に、シナリオを構成した事例が1件、既存の研究から開発されたシナリオを用いた事例が2件、EQ-5DとSF-6Dを基に構成した事例が1件、コメントなしが1件あった。シナリオの草案に関する臨床専門家の検討は全ての事例にて行われた。

シナリオは、計4～18個からなり、大体、症状、治療反応、異常反応などが考慮された。回答者へ完全な健康状態の定義を提供した事例は計7件で、他の2件ではこれに関する別途の案内はなかった。直接測定は、韓国的一般大衆を対象に行った事例が8件、イギリスの一般大衆を対象にした事例が1件で、大体、ギャング・サーベイ (gang survey) の方法行われた。1件以外の全ての事例にて、回答者の特徴に関する情報が提供された。

質の加重値を測定する主な方法として、TTOを用いた事例が7件、SGを用いた事例は1件、SGとTTOを共に用いた事例は1件あった。この他にも、EQ-5D若しくはVASなどが併用された事例が6件あった。TTOを用いた場合の基準時間は10年3件、20年2件、30年1件、24か月1件とさまざまであったが、基準の設定に関する理由については大体、別途のコメントはなかった。

<表 41> 健康状態のシナリオに基づいた質の加重値に関する直接測定方法の事例を分析

事例、	対象の疾患	シナリオ文献の考察	シナリオの構成	完全な健康状態の定義	質の加重値の算出方法	標本	その直接測定理由
1	転移性去勢抵抗性前立腺がん	はい	n=8、症状、治療反応、異常反応	コメントなし	TTO (基準: 10年) EQ-5D-3L、EQ-VAS	n=280, 5個 (Seoul, Incheon, Busan, Gwangju, Daejeon) に住む一般男性	直接比較可能な臨床資料の不在
2	全身性未分化大細胞リンパ腫	既存の研究にて開発されたシナリオ	n=6、反応率と異常反応	はい	TTO (基準: 20年)	n=123, Seoulに住む一般大衆	直接比較可能な臨床資料の不在

5) 2014~2019年の5年間、審評院に健康保険登載の目的で提出された経済性評価の資料47件 (CUAのみ含む)。

6) .

3	骨髄線維症	既存の研究にて開発されたシナリオ	n=4、症状及び日常生活の障害	はい	TTO (基準: 20年)	n=128、Seoulに住む一般大衆	疾病別の道具を使用
4	多発性硬化症	コメントなし	n=18 (詳細な健康状態を含む)、日常生活の障害、異常反応治療法	はい	TTO (基準: 30年)、VAS	n=106、Seoulに住む一般大衆	直接比較可能な臨床資料の不在
5	分化型甲状腺癌	EQ-5DとSF-6Dを基に、構成	n=7、治療反応、異常反応	はい	TTO (基準: 10年)、VAS	n=100、イギリス一般大衆 (韓国人の選好度を反映するため、人口社会的な特性を補正)	直接比較可能な臨床資料の不在
6	慢性心不全	はい	n=5、症状、異常反応	コメントなし	TTO (10年の基準) EQ-5D-5L	n=200、Seoulに住む一般大衆	HRQoL 資料なし
7	非小細胞性肺癌	はい	n=5+7, 5個 (末期がんに反映し、死亡までの時間でシナリオを構成), 7個 (異常反応)	はい	TTO (基準: 20年 - 不明確)	n=205, Seoulに住む一般大衆	直接比較可能な臨床資料の不在 (比較代案3つ)
8	筋萎縮性側索硬化症	文献の考察, K-ALSFRS-RとEQ-5Dを基に構成	n=9、症状、日常生活の障害	はい	SG EQ-5D-3L	n=202、Seoulに住む一般大衆	直接比較可能な臨床資料の不在
9	固形癌の骨転移	はい	n=9、症状、治療	はい	SG TTO (基準: 24か月) EQ-5D-5L EQ-VAS	n=400、全国に住む一般大衆 (特徴に関する記述はなし)	HRQoL 資料なし (不確実)

※ (最初提出資料の) 主な基本分析を基準に、分類

HRQoL : health-related quality of life、K-ALSFRS-R : Korean amyotrophic lateral sclerosis functional rating scale-revised、SG : standard gamble、TTO : time trade-off

ウ) 外国の指針

オーストラリア (PBAC、2016) の指針では、直接測定方法に関し簡単な指針を提供しているが、カナダ (CADTH、2017) とイギリス (NICE、2013) の指針ではこれに関する。推奨事項や指針を提供していない (表 42)。

(1) カナダ (CADTH, 2017)

カナダの指針では、使いやすさ (ease of use)、比較可能性 (comparability)、解釈可能性 (interpretability) などの理由で、EQ-5DやHUIのような一般道具を用いた間接測定方法を勧めている。TTOやSGなどの方法を用いて、質の加重値を直接測定する。方法は設計及び実行が複雑で、健康状態をどのように記述するかにより質の加重値が大幅に変わりかねないと述べている。健康状態シナリオの構成方法やこれに関する。適切性の評価に関する。詳細な指針は別途に提示していない。

(2) オーストラリア (PBAC, 2016)

現在の指針では、健康状態のシナリオに基づいた直接測定方法に関し段階別に詳しく説明し、潜在的な偏向を最小化する。ように述べているが、これに関する。具体的な方案は提示していない。ただ、臨床試験の期間を外れる健康状態をシナリオに含める場合は、臨床試験にて評価した健康状態もシナリオと一緒に含め、検証すること述べている。

以前の指針 (例、- version 4.5 (PBAC, 2015)) では、直接測定方法に関しより具体的な指針を提示したが、主な内容は下記の通りである。

- 各シナリオにて記述した健康状態の算出根拠を提示し、治療法に関する臨床結論 (therapeutic conclusion) とシナリオに記述された内容間の関連性を説明すること。
- 各治療法に関する患者の経験をシナリオに反映するために、どのような措置を行ったかを説明すること (例えば、患者対象のフォーカスグループを根拠にしたか) なお、回答者は健康状態の構成属性 (attribute) が5~9個を超えると、認知的過負荷 (cognitive overload) により、ろくにその差を認識し回答し難いことがあるが、様々な患者の経験からこうした属性をどのように選択したかを説明すること。
- 妥当性 (validity)、信頼性 (reliability)、変化への敏感性 (responsiveness)、臨床的な重要性 (clinical importance) の面で、シナリオを評価すること。各シナリオにて反映している期間と選択に基づいた質の加重値の算出方法にて想定した期間を比較して評価及び報告すること。
- ex post観点 (即ち、十分な診断後、診断した通りに健康状態を経験すると想定。未来に発生しうる不確実なリスクに関する考慮はしない) を取ることを選好するため、関連性のない内容 (例えば、診断過程及び可能な様々な予後 (prognoses)) は排除すること。

表現及び言葉の選択により、回答者の反応はかなり異なり得るため、できれば回答者に治療に関する全体的な情報を提供すること。「癌」などの疾病ラベリング (labelling) もしくはフレーミング

(framing) 効果に注意すること。健康状態に関する。追加的な質問により回答者の理解程度を評価する方法も考慮すること。

- シナリオ数を正当化すること。
- 回答者の理解を高めるために、対面インタビューが選好される。電話インタビューや郵便調査などを用いる場合、これに関する正当化をすること。
- 論理的な順により情報を提供し、必要に応じては視覚的な資料を用いること。
- スポンサーによる歪みを最小にする。べく、インタビューを行う際はスポンサー名を明かさないこと。
- 資料を提出する際、回答者へ提供した情報も共に提出すること。

(3) イギリス (NICE, 2013)

イギリス・NICEの当指針では、健康状態のシナリオに基づいた直接測定方法に関しては述べていない。ただ、EQ-5Dを用いた間接測定方法の代案を説明する。NICE DSU TSD 11 (Brazier & Rowen, 2011) では、シナリオ (vignette) の妥当性を重要に述べている。シナリオ (vignette) の妥当性はシナリオがどれほど厳格 (rigour) に設計されたかによるが、深層面接などの広範囲な質的作業、HRQoL資料、第三者による独立的な検討などを基に、シナリオを記述すればシナリオの妥当性は改善できるとした。それにもかかわらず、健康状態シナリオは患者の経験を十分に反映できず、HRQoL資料に基づいていない直接測定はNICEの基準に相応しくないため、検証されたHRQoL道具をもって測定された資料がない場合は制限的に用いることもできると述べた。

<表 42> シナリオに基づいた効用の直接測定に関する。海外の主な国による指針

国	指針内容の原文	サマリー
カナダ・ADTHの指針 (4版) (2017)	(Direct methods of measurement refer to methods that enable the direct translation of preferences onto a utility scale, while indirect methods require a conversion scale to derive utilities. The most common direct methods include time-trade-off, standard gamble, and discrete choice experiments.) Direct approaches are complex to design and implement and, unless individuals own health states are being measured, likely to result in estimates that are highly dependent on the validity of the descriptions of the health states.76,77 (p45)	直接的な算出は設計や実行が複雑である。健康状態に関する。記述の妥当性により、推定値がかなり変わり得る。
オーストラリア・PBACの指針 version 5 (2016)	(Scenario-based methods use vignettes to describe the symptoms of a health state to a sample population, usually a representative general population sample, from which utility weights are elicited preferences include the standard gamble, time trade-off and discrete choice experiments, and other stated preference methods.) If using a scenario-based utility valuation to value health outcomes beyond the time horizon of the trial, include one or more health states captured and valued within the trial in the scenario-based study to validate the commonality of the trial-based and scenario-based utility weights. Present supporting evidence for any claim of increased sensitivity of a scenario-based approach to identify real differences in utility. Describe all stages of a scenario-based study in detail and explain efforts to minimise potential bias. It is difficult to minimise the many sources of analyst bias that are intrinsic to the scenario-based utility approach, including the nonblinded nature of the construction and presentation of the scenarios (eg incomplete inclusion and differential focus on alternative aspects of quality of life), the design of the methods to elicit values, and the analysis and interpretation of the results. (p78)	臨床試験の期間を過ぎる健康状態をシナリオに含める場合、臨床試験にて評価した健康状態もシナリオに含めて検証。シナリオに基づいた研究に関し、段階別に詳しく説明。潜在的なバイアスを最小化。

イギリス・ICEの指針 (2013)	<p>当指針にはコメントなし</p> <p>NICE DSU TSD 11 (Brazier & Rowen, 2011): Vignettes that have not been based on HRQL data do not meet the NICE Methods Guidance for alternatives to EQ-5D. However, vignettes may have a limited role where there are no data available using validated HRQL measures. (p23)</p>	<p>HRQoL資料に基づいていない健康状態のシナリオはEQ-5Dの代案としては不適合。しかし、検証されたHRQoL道具をもって測定した資料がない場合は制限的に用いることができる。</p>
--------------------	---	--

工) その他、文献の考察結果

健康状態のシナリオ (vignette) に基づいた直接測定は標準化したアンケートを用いる代わりに、文献の考察、臨床資料、患者の経験、専門家の意見などの様々な資料を基に、疾病と治療の影響 (impact) を記述し、これに関する。質の加重値をTTOやSGなどの選択に基づいた道具を用いて推定する。方法である。一般道具や疾病別の道具で測定した適切な資料がない場合、一つの代案としてしばしば用いられる (Tosh et al., 2011)。直接測定は、患者や一般大衆全てを対象に行えるが、機関への提出向けとして準備する。場合は当機関の選好を反映して対象を選定する。シナリオは叙述式もしくは箇条書きの形で記述されかねないが、これに関する。詳しい例、は既存の文献を参照することができる (例、 - 関節炎 (Iqbal et al., 2012)。調弦病 (Briggs et al., 2008)、糖尿病 (Matza et al., 2007)、喘息 (Lloyd et al., 2008)、ADHD (Lloyd et al., 2011; Matza et al., 2005)、うつ病 (Revicki & Wood, 1998)、慢性痛み (Eldabe et al., 2010)、癌 (Tolley et al., 2013))。

しかし、直接測定の場合は具体的な基準がないため、資料の提出者にとって有利に健康状態が記述されることができ、標準化した道具を用いないため、比較可能性 (comparability) が劣り得るという短所が存在する。(Blumenschein & Johannesson, 1998; Dolan et al., 1996; Brazier et al., 2017)。健康状態がどのように記述されるかにより、質の加重値が大きく変わり得るにもかかわらず、健康状態シナリオの構成や評価方法に関する。公式的な指針はまだ開発されていない (Brazier et al., 2017)。

ただ、最近、Brazier et al. (2017) は健康状態シナリオ (vignette) の構成 (construction) と価値評価 (valuation) に関し、下記のようないくつかの指針 (good practice suggestions) を提示した。

- 健康状態シナリオの個数は患者集団の重症度分布を反映すること。
- 健康状態シナリオはこれを価値評価しなければならない回答者が明確に理解すること。
- 健康状態シナリオは臨床専門家のインタビュー、患者のインタビュー、臨床試験資料、既存の文献など、様々な出典を通じて算出された最適の根拠 (best available evidence) を基に、作成することである。こうした目的のため、患者や臨床専門家を対象に質的研究を行う場合は合意 (consensus) を行い、これ以上の新たな情報が出ないまでインタビューを行うこと。
- 強度 (intensity) や頻度 (frequency) に関する。陳述を提示する場合は、なるべく患者レベルの量的資料を用いること。これに関する例は文献 (例-Lloyd et al., 2011; Matza et al.,

2014a) を参考にすること。

- 健康状態に関するラベリング (labelling) や属性 (attribution) は明確性 (clarity) のために必要かも知れないが (Matza et al., 2014a)、無関係の情報 (extraneous information) を取り出したり偏った (biased) 反応を招いたりすることもあるため、できるだけ用いないこと (Rowen et al., 2012)。
- 不確実な説明は評価を行う際、困難を招きかねないため、避けるべきである (例-「こういう症状がある可能性もある」)。
- もし、健康状態のシナリオが時間上の変化を記述しているなら、(即ち、経路状態 (path states))、ここで算出された質の加重値をモデルに適用する場合はシナリオにて想定した所要時間の総計 (例 - 価値評価の期間) も共に考慮しなければならない (Matza et al., 2014b)。
- 該当機関 (例、: NICE、SMC、CVZ、その他の期間) にて認める (recognized) 方法 (methods) と対象 (population) を用いて、健康状態シナリオ (vignette) の価値評価を行うべきである。

ちなみに、シナリオに関する根拠の客観性と質を上げるには、(1) 体系的な文献考察及び臨床資料を通じて草案作成、(2) 臨床専門家の検討後、草案の再修正 (3) 患者のフォーカスグループへのインタビューなどにより、患者の経験を反映 (4) 一般人を対象に事前調査を行い、シナリオを検討及び修正 (5) 当研究実行などの手順を経ることが必要である (例 : Matza et al., 2014b)。

オ) 利害関係者の意見

(1) 専門家の意見

— 提示された意見なし

(2) 製薬会社の意見

- 直接測定に関する指針を提示（シナリオの例など）

カ) 指針の改正方向に関する提言

質の加重値を直接測定する。場合、健康状態シナリオに関する。根拠の客観性と質を高めるため、下記のような詳細を提供することを提言する。

- 健康状態に関するシナリオ（vignette）を構成し、資料の提出者が直接質の加重値を測定する場合、これに関する正当化を行うようにする。健康状態に関するシナリオの構成過程、回答者の抽出と募集方法などを含んだ全体的な価値評価過程を詳しく記述するように明示する。敏感度の分析を通じ、直接測定に関する不確実性を十分に検討するように明示する。
- 健康状態のシナリオに関する根拠の客観性と質を高めるため、下記の注意事項を考慮してシナリオ（vignette）を作成し、提示された過程を基に直接測定を行い、資料の提出時は次の報告項目を考慮して報告するようにする。
 - **健康状態に関するシナリオ（vignette）を作成する際の注意事項**
 - 健康状態に関するシナリオの個数は、なるべく患者集団の重症度分布を反映して決めること。
 - 健康状態に関するシナリオは回答者が容易く明確に理解できるように記述すること。
 - 健康状態に関するシナリオは、臨床専門家のインタビュー、患者のインタビュー、臨床試験資料、既存の文献など、様々な出典から算出された最適の根拠（best available evidence）に基づいて作成すること。とりわけ、仲裁群における差は確実に明確な最適の根拠を基に記述すること。
 - 副作用や症状のような場合は（区別し難い差よりは）大きく影響を与えかねないものを中心にシナリオを構成すること。
 - 強度（intensity）や頻度（frequency）に関する説明を含める場合は、なるべく患者レベルの量的な資料から算出した根拠に基づいて作成すること。
 - 不確実な説明は回答者が価値評価を行う際、困難を招きかねないため、なるべく避けることである（例 - 「こうした症状が起きる可能性もある」）。
 - 疾病名（condition labeling）と疾病の属性（attribution）を記述することは、明確性のために必要な場合もあるが、偏った反応を誘導する可能性もあるため、注意して用いること。

- 直接測定のプロセス

- 文献調査及び臨床資料により主な症状、治療効果、副作用などを把握
- これに基づき、臨床専門家及び患者のインタビューに向けた体系的なインタビューのガイドを作成
- 臨床専門家とのインタビューを通じ、健康状態に関するシナリオの草案を作成、患者とのインタビューを通じては患者が経験した治療および症状などをシナリオに反映
- 一般大衆を対象に、事前調査（pilot study）を実行。SGやTTOなどのような選択に基づいた方法を用いて健康状態の質の加重値を測定。一般大衆が容易かつ明確に理解できるように健康状態が記述されているかを確認し、必要な場合は文言を修正
- 一般大衆を対象に、最終的な価値評価の研究を実行

- 直接測定時における報告項目

- シナリオの作成に向けた文献調査のプロセス及び最終的な考察に含まれた研究
- 臨床専門家とのインタビューに向けた臨床専門家の選定プロセス及び基本情報（臨床専門分野、当分野における経歴期間など）
- 患者へのインタビュー実行の可否及びインタビューにおける患者選定プロセス及び患者に関する。基本情報（年齢、性別、疾患の重症度など）
- 事前調査（pilot study）に関する情報：一般大衆の募集方法及び参加者数など
- 最終的な健康状態に関するシナリオと完全な健康の定義
- 最終的な価値評価研究に関する情報：一般大衆の募集方法、参加者数、回答率、回答者の人口学的な特徴（年齢、性別、居住地域、該当疾患に関する。健康状態など）、調査方法（対面インタビュー・オンライン調査・郵便調査など）、直接測定方法（SG、TTOなど）及びプロセス、各シナリオによる質の加重値の結果（平均、中位数、標準偏差、最大値、最小値など）など

6) 効用のMCID

ア) 現在の指針

コメントなし

臨床的に有意義な最小限の差を表すMCID（minimal clinically important difference）に関する。内容は現在の指針に含まれていない。

イ) 既存の提出資料を検討した結果

2014年～2018年の5年間で、審評院に健康保険へ登載する目的で提出された経済性評価向けのCUA資料47件を検討した結果、効用のMCIDに関する情報を提供した事例はなかった。ただ、資料評価の当時、MCID以下の効用値使用（例：異常反応に関する非効用（disutility））が問題になったことはある。さらに、臨床試験にて測定された治療群の暮らしの質の改善の程度を報告した事例も多くないため、暮らしの質の改善に関する臨床的な有用性の検討そのものが殆ど不可能だということが問題として指摘され続けたことがある。

ウ) 外国の指針

ナダ（CADTH、2017）とイギリス（NICE、2013）の指針には、MCIDに関するコメントが全くない。オーストラリアの指針（PBAC、2016）では、よく使われる一般MAUIのほか、他のMAUI若しくはPRO道具が用いられた場合、同道具に関する詳細を別途に添付することを明示しているが、その一つがMCIDである（表 43）。

指針では、MCIDの出典として下記を提示している。

- プロトコル（標本サイズの計算などの理由で、作成される場合が多い）
- PBACにて受容られたことのあるMCID（但し、臨床試験の対象と提示された適応症（proposed indication）に相応しくならなければならない）
- よく文献にて受け容れているMCID（但し、臨床試験の対象と提示された適応症に相応しくならなければならない）
- スポンサーにより行われた内部研究（準拠基盤（anchor-based）法、専門家の合意（expert consensus）、統計的方法（statistically based analysis））
- 提示された適応症と類似な適応症として、よく文献にて受け容れているMCID（但し、提示された適応症で一般化を期待できるべき）

MCIDを報告する際、MCIDの出典、算出方法（準拠基盤法、専門家の合意、統計的な方法など）、MCIDが算出された研究との比較、対象人口（population）、結果（outcome）の定義、道具のベースライン（baseline）値なども共に、表として提示するようにしている。

なお、指針ではスケール（scale）を用いて測定された場合は準拠基盤法を選好、生存資料や2進数資料の場合はMCIDの決定が容易ではないため、専門家の合意に基づいて算出することが一般的だと述べている。主な結果変数（primary outcome）が代理変数（surrogate）である場合は最終的な結果に有意義な変化をもたらしかねない代理変数上の変化に基づかなければならないと明示している。

<表 43> 海外の主な国におけるMCIDの指針内容

国	指針内容の原文	サマリー
オーストラリア PBA C 指針 version 5 (2016)	<p>Where a patient-reported outcome measure is used, or a MAUI that is not listed above, provide, in an attachment, a discussion of (or reference supporting) the (p37):</p> <ul style="list-style-type: none"> domains of quality of life, symptoms or function that are covered by the instrument scoring method of the instrument validity of the instrument reliability of the instrument responsiveness of the instrument to differences in health states between individuals and to changes in health states over time experienced by an individual clinical importance of any differences detected by the instrument (see Subsection 2.4.4 for guidance on minimal clinically important differences [MCIDs]). <p>Likely sources for an MCID may be (p38):</p> <ul style="list-style-type: none"> the protocol (often for the purposes of powering the study) a previously accepted MCID by the PBAC, as indicated in a public summary document, that is relevant to both the trial population and the proposed indication a commonly accepted MCID in the literature, relevant to the trial population and the proposed indication an internal study by the sponsor (anchor-based analysis, expert consensus, statistically based analysis) a commonly accepted MCID in the literature for a similar indication that can reasonably be expected to be generalisable to the proposed indication. 	<p>上記にて述べたMAUI以外のMAUIやPRO道具が用いられた場合、この道具のMCIDを含めた追加的な情報を添付。</p> <p>MCIDの出典は、プロトコル、PBACにて受け容れられた既存値、文献にてよく引用されるMCID値、提出者の内部分析。MCID値を引用する場合、臨床試験の対象人口と適応症に相応しいかを検討。</p>

※ カナダ・CADTHの指針（2017）とイギリス・NICEの指針（2013）では、MCIDに関するコメントなし。

工) その他の文献考察の結果

(1) MCIDの定義及び測定

患者中心の診療 (patient-centered care) が強調され、患者の主観的な暮らしの質若しくは経験が次第に重要となり、これにより複数の研究にて様々なHRQoLもしくはPRO道具が重要な測定道具として用いられている (Wyrwich, 2004; Wyrwich et al., 2013) 。しかし、PROは熟練した臨床専門家ではない患者の評価により算出されるため、これに関する解釈の困難が伴う。即ち、PRO道具を用いて測定された点数の変化が果たして有意義な変化であるが、もしくは患者の管理を他の方法に変更しなければならぬほどの変化であるかなどは判断し難い。統計的な有意性を根拠と判断する方法もあるが、統計的な有意性は標本サイズにより、小さな差にも有意に現れたり大きな差にも有意に現れなかったりする場合もある。従って、統計的な有意性だけでは、臨床的な有意性の判断には限界がある。これに関する代案として、患者に重要なPROの変化がどのレベルかを把握するための体系的な試みが行われたが、なかんずく、カナダのMcMaster大学の研究チームを中心に関連した研究が活発に行われた。同研究チームは、「最小限の臨床的に有意義な差」 (minimal clinically important difference、MCID) に関する基準を提供するために努力したが、Jaeschke et al. (1989) はMCIDを下記のように公式的に定義した。

「the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management.」(p.408)

同定義によると、MCIDは患者が利益 (beneficial) として認識し、(煩わしい副作用や過度な費用が発生しなくても) 患者の管理に変化が必要になり得る関心領域における最も小さな点数の差を意味する。しかし、「臨床的 (clinical) 」という用語が患者の日常経験よりは臨床領域 (clinical arena) に、より焦点を当てる傾向があるという主張が提起 (Schünemann et al., 2005a) 、以来、「臨床的」という用語を除いた「最小限の有意義な差」 (minimal important difference、MID) を用い始めた (Juniper et al., 1994; Schünemann et al., 2005b) 。Schünemann et al.2005b) はMIDを下記のように公式的に再定義した。

「the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and that would lead the patient or clinician to consider a change in the management.」(p.82)

即ち、MIDは患者 (informed patients) もしくは代理者 (informed proxies) が重要とする、即ち、利益 (beneficial) になったり、危害 (harmful) を与えたりすると認識し、患者や臨床医が管理に変化を考慮できる関心の結果 (outcome of interest) における最も小さな点数の差を意味する。この他にも、感知できる最小限の変化 (minimum detectable change、MDC) 、感知できる最も小さな差 (smallest detectable difference) など、様々な用語が混用されている。ここでは、オーストラリア・PBACの指針同様に、MCIDという用語として通称し、用語の定義及び変遷史はKing et al. (2011) などによる他の研究を参照とすることができる。

HRQoL道具及びその他のPRO道具のMCIDを推定する。方法としては、大きく（1）準拠（anchor）になり得る測定道具を同時に用いて、変動の様相を比較する。準拠基盤法（anchor-based methods）と（2）測定結果の分布変動にて統計的な差を調べる分布基盤法（distribution-based methods）がある（Wyrwich et al., 2013）。

この他にも、専門家の合意（consensus）、MAUIを対象に開発された道具の定義アプローチ法（instrument-defined approach）（Luo et al., 2010）、アメリカ・FDA（2009）の指針で提示した反応累積分布関数（cumulative distribution of response）などがある。

□ 準拠基盤法（anchor-based method）

準拠基盤法は準拠（anchor）に狩りかねない他の道具を用いて、最小限の有意義な差を推定する。方法である。準拠は測定道具と相関関係がないと準拠として用いられないが、既存の研究ではSF-36質問項2（即ち、1年前の比べた健康状態）、患者の変化グレード（global ratings of change）、Eastern Cooperative Oncology Group（ECOG） performance statusなどの様々な道具を準拠として用いた。

準拠基盤法は、分析方法によりさらに細分化されるが、Coretti et al.（2014）はこれを①回帰方法②平均変化アプローチ法（average change approach）③最小に感知できる変化アプローチ法（minimum detectable change approach）④変化差のアプローチ法（change difference approach）、⑤ROC（receiver operating characteristic、受信者操作特性）曲線アプローチ法と分類・整理した（表 44）。一、回帰方法は関心道具をもって測定された変化値を従属変数、

準拠道具をもって測定した独立変数とし、OLSもしくはGLMを行った上、準拠道具の回帰係数値を関心道具のMCIDで推定する方法である。準拠道具の値はダミー変数に変更し、分析に含めることもできるが、この際、ダミーの基準値（threshold value）は患者の選好や分析者の判断に基づいて算出できる。二、平均変化アプローチ法は準拠道具を基準に反応子（responders）を分類し、これらの点数変化の平均値をMCIDとして用いる方法である。三、最小感知できる変化のアプローチ法は測定の誤差（measurement error）以上の最小の差を測定する方法である。準拠道具を基準に、非反応子（non-responders）を分類し、これらの点数変化の平均値に関する信頼区間の上限値をMCIDと用いる。四、変化差に関するアプローチ法は反応子と非反応子との平均点数の変化差をMCIDとして用いる方法である。五、ROC曲線のアプローチ法では、陽性反応（positive response）に関する敏感度（sensitivity）若しくは（and/or）特異性（specificity）を最大にする点数の変化値をMCIDと定義する。検査Xの敏感度は疾患のある患者にて、検査Xにより陽性と診断された患者の比重で計算できるが、準拠基盤法では準拠道具により、「改善（improvement）」と分類された患者を前者と見なし、関心道具にて基準値以上と分類された患者を後者と成すことができる。同様に、

特異度は準拠道具を用いて、「改善（improvement）」と分類されなかった患者の中で、関心道具により基準値以下に分類された患者の比重と見なすことができる。分類正確度（accuracy）は、ROC曲線のAUC（area under the curve）として提示できるが、0.7～0.8レベルであれば適切、0.8～0.9レベルであれば良いレベルと言える（Parker et al., 2011）。

準拠基盤法の主な限界としては、（1）想起に基づいた準拠を用いる場合、想起バイアス（recall bias）の発生可能性（2）どの準拠道具を用いたかにより結果偏差の発生可能性（3）相対的に困難な標準化（4）準拠道具の併用による回答の過程における一致化（concordance）発生の可能性などである（Wyrwich et al., 2013; Parker et al., 2011; Bae Jongmyun, 2015）。

<表 44> 準拠基盤法（anchor-based method）方法の比較

方法	手順	算出物（MCID）
回帰方法（Regression methods）	関心道具をもって測定された変化値を従属変数準拠道具で測定された変化値を主な独立変数とし、OLSもしくはGLM実行。準拠道具で測定された値はダミー変数に変更し、分析に含め得る。この際、ダミー基準値（threshold value）は患者の選好もしくは分析子から算出。他の共変量が含まれる可能性あり。	回帰係数
平均変化アプローチ法（Average change approach）	◆ 準拠道具を基準に分類された反応子（responders）の点数変化の平均値	平均値
最小感知変化アプローチ法（Minimum detectable change approach）	信頼区間が与えられた状態で、測定誤差（measurement error）以上の最小の差。準拠道具を基準に分類された非反応子（non-responders）の点数変化の平均を利用	非反応子（non-responders）にて感知された平均変化の信頼区間の上限値
変化差のアプローチ法（Change difference approach）	◆ 反応子と非反応子との平均点数の変化の差	差
受信者操作特性曲線（ROC curves）	感度をx軸、（1-特異度）をy軸にしたROCを利用。感度は準拠道具により、「改善（improvement）」と分類された患者の中、関心道具により基準値以上と分類された患者の比重。基準値として中央値がよく用いられる。特異度は準拠道具により、「改善（improvement）」と分類されなかった患者の中で、関心道具により基準値以下と分類された患者の比重。	陽性反応に関し、最大の感度若しくは（and/or）特異度を提供すること。AUC（area under the curve）は分類の正確度を提示。

出典：Coretti et al. (2014)

参照 MCID : Minimum clinically important difference、MDC : Minimum detectable change、ROC : Receiver operating characteristics curve、OLS : Ordinary least squares、GLM : generalized linear models

□ 分布に基づく方法 (distribution-based method)

分布に基づく方法は統計学的な分布に基づき、MCIDを推定する。方法として様々な測定法が存在する。が、共通的な要因は点数変化の差と変動性 (variability) という概念を用いるということである (Wyrwich et al., 2013)。

最も代表的な方法では、効果量 (Effect size, ES) と測定の標準誤差 (standard error of measurement, SEM) がある。ESはグループ間における平均点数変化の差を標準偏差で割って計算できるが、標準偏差のタイプにより、様々なバージョンが存在する。(Norman et al., 2007)。分布に基づいたES測定方法のメリットは単位がないため、様々なグループ間、結果間の比較ができるということである。よくESが0.2であれば差が小さく、0.5であれば中間であり、0.8以上なら大きいと解釈する (Cohen, 1988)。(Cohen, 1988)。Norman et al. (2003) は38のPROに関する。研究を分析した結果、PRO道具のMCIDは標準偏差の0.5倍若しくは0.5 ESに近いと主張した。SEMは誤差の変動性 (即ち、測定誤差の標準偏差) を表す概念として道具の標準偏差 $\times \sqrt{1 - \text{道具の信頼係数}}$ で計算できるが、重要な変化を推定する。場合、よく使われる (Nunnally et al., 1994)。1 SEMは準拠基盤法をもって測定した様々なPRO道具のMCIDとかなり密な関連性を表す傾向があり、SEMを用いてMCIDを推定する。ときの重要な基準値としてよく用いられている (Wyrwich et al., 2002; Wyrwich, 2004; Wyrwich et al., 2013)。

こうした分布に基づく方法は、然るべき準拠がない状況にて、PROの点数変化を解釈する。際、容易く使用できるが、臨床的な妥当性と反応度を考慮しないという短所を持っている。このため、オーストラリアPBAC (2016) の指針では準拠に基づく方法をより好むと述べ、アメリカのFDA (2009) から分布に基づく方法は準拠に基づく方法を支える程度の役割を果たしうると明示した。

□ その他

この他にも、専門家の合意 (consensus)、道具定義アプローチ法 (instrument-defined approach)、累積分布関数 (cumulative distribution of response) などがある。道具定義アプローチ法は比較的最近に開発された方法として、MAUIのMCID推定に用いられた (Luo et al., 2010、McClure et al., 2017)。同方法は、患者の資料なしにシミュレーションを用いて通じMCIDを推定する。方法として、MAUIにより定義される健康状態が1ステップ変わる際、これに当たる質の加重値の変化をMCIDと想定する。例えば、EQ-5D-5Lの場合は33333と定義された健康状態にて23333、32333、33334のように一つの健康領域にて一つレベルのみ変わる場合、最小限の有意義な臨床的变化が発生する。と想定し、この際の質の加重値の変化をMCIDと定義する。EQ-5D-5Lと定義され得る健康状態は計3,125個であるが、それぞれの健康状態にて1ステップのみ変化して生成され得る健康状態は計25,000個である (McClure et al., 2017)。道具定義アプローチ法では、適切なtariffを適用し25,000個の健康状態の変化による質の加重値の変化を計算し、これの平均をMCIDをもって推定する。反応子の累積分布関数は米FDA (2009) の指針にて提示した方法であるが、臨床的に有意義な変化を、一つの値ではなく分布をもって情報を提供する。同累積分布関数は道具点数の変化をX軸として、この点数の変化を切断点 (cut-off point) と使用する。場合、累積反応子 (responders) の比率をY軸にして描くことができるが、MCID推定の不確実性に関する一つの代案として提示された。

(2) MAUIのMCIDに関する研究検討

7) EQ-5DやSF-6DなどのようなMAUIに関する。MCID研究も、過去10年間活発に行われた。Coretti et al. (2014) はイギリス・一般大衆の選好を反映したEQ-5D点数のMCIDを推定するため、文献の考察を行い、計18の文献を最終的に選定し、これらの研究結果を検討した。結果は、表45に提示された通りであるが、臨床試験から断面研究まで様々な資料とMCIDの算出方法が用いられ、MCIDも0.03 ~ 0.54とさまざまに報告された。同研究においても、使用した分析方法と準拠などにより、様々な値が算出された。例えば、Parker et al. (2011) は二つの準拠と平均変化アプローチ法 (average change approach) 、最小感知変化アプローチ法 (minimum detectable change approach) 、変化差のアプローチ法 (change difference approach) 、ROC曲線を用いてMCIDを推定し、0.15~0.54の範囲を報告した。研究チームは同方法の中で、最小感知変化アプローチ法が最も適切だと述べたが、その理由としては測定における誤り以上の値であり、状態が改善された患者らが報告した平均の変化値に最も近く、準拠の影響を最も受けないためと述べた。最小感知変化アプローチ法をもって測定したMCIDは0.46であった。他の研究 (McClure et al., 2017) では、道具定義アプローチ法を用いて、5か国のEQ-5D-5LのMCIDを比較した。カナダ、中国、スペイン、日本、イギリス、ウルグアイにて開発されたEQ-5D-5L tariffを用いてMCIDを推定した結果、日本のtariffを用いる場合は0.048と最も低く表れ、中国のtariffを用いる場合は0.069と最も高く表れた。EQ-5Dの領域別の最大値を除いて分析した場合は0.037 (カナダ、イギリス) ~ 0.058 (中国) と推定された。韓国では、MAUIを用いたMCID研究が殆ど行われなかった。Kim et al. (2015) による研究が唯一であるが、同研究では脳卒中患者487人を対象に、10か月間のEQ-5DとSF-6Dの点数変化を用いて各道具のMIDを計算した。EQ-5Dの場合、国内のtariff (Lee et al., 2009) を適用したが、SF-6Dの場合は国内にて開発されたtariffがないため、イギリスのtariffを適用した。修正されたRankin scaleとBarthel indexを準拠とし、最小限の改善と最小限の悪化を定義し、この範囲に含まれた患者らの平均点数の変化をMIDと定義した。研究では、EQ-5DのMIDを0.08 ~ 0.12、SF-6DのMIDを0.04 ~ 0.14と報告した。研究結果をまとめると、対象人口、分析方法、準拠タイプ、tariffなどにより、MCIDにかなりの差が生じ、研究間の差のみならず研究内における差も無視できないことが確認できた。

<表 45> イギリス・一般大衆の選好が反映されたEQ-5D点数に関するMCIDの研究

研究	研究設計	対象疾患領域	測定方法	準拠の測定道具	用語	結果
Larsen et al. (2008)	非無作為化臨床試験	筋骨格系 (n=98)	断面基準分布 (Reg)	-	CRD	0.08
Marra et al. (2005)	横断研究	筋骨格系 (n=313)	断面基準分布 (SD)	-	MID	0.05
Solberg et al. (2013)	前向きコホート研究	筋骨格系 (n=692)	縦断基準準拠 (ROC)	- Global Perceived Scale of Change	CVS	0.30

Soer et al. (2012)	前向きコホート研究	筋骨格系 (n=151)	縦断基準準拠 (ROC)	- Roland Morris Disability Questionnaire - Numeric rating scale - Pain Disability Index	MCIC	0.03
Parker et al. (2013)	前向きコホート研究	筋骨格系 (n=69)	縦断基準準拠 (AC, MDC, CD, ROC)	- North American Spine Society Patient Satisfaction Scale	MCID	0.24
Parker et al. (2012a)	後ろ向きコホート研究	筋骨格系 (n=47)	縦断基準準拠 (AC, MDC, CD, ROC)	- Health transition index of SF-36 - Patient's satisfaction after the surgery	MCID	0.14-0.24
Impellizzeri et al. (2012)	前向きコホート研究	筋骨格系 (n=99)	縦断基準準拠 (ROC)	- Symptoms specific well-being - Global Treatment Outcome	MCIC	0.16
Parker et al. (2012b)	後ろ向きコホート研究	筋骨格系 (n=53)	縦断基準準拠 (AC, MDC, CD, ROC)	- Health transition index of SF-36 - Patient's satisfaction after the treatment	MCID	0.29-0.52
Parker et al. (2011)	後ろ向きコホート研究	筋骨格系 (n=45)	縦断基準準拠 (AC, MDC, CD, ROC)	- Health transition index of SF-36 - Patient's satisfaction after the treatment	MCID	0.15-0.54
McDonough et al. (2011)	前向きコホート研究	筋骨格系 (n=1000)	縦断基準準拠 (AC)	- Symptom satisfaction - Self-perceived health - Oswestry Disability Index - Progress rating	MID	0.12-0.15
Boonen et al. (2007)	横断研究	筋骨格系 (n=120)	断面基準分布 (LA)	-	SDD	0.36
Staerke et al. (2011)	横断研究	筋骨格系 (n=51)	縦断基準分布 (SEM)	-	MDC	0.36
Kvam et al. (2011)	前向きコホート研究	癌 (n=239)	縦断基準準拠 (AC) 分布 (SD)	Patient's perceived improvement	MD	0.08-0.10

Pickard et al. (2007)	後ろ向きコホート研究	癌 (n=534)	断面基準準拠 (MDS) 分布 (SEM, SD)	- Eastern Cooperative Oncology Group performance status ratings - Functional Assessment of Cancer Therapy-General score	MID	0.07-0.12
Le et al. (2013)	ランダム化比較試験	外傷後のストレス障害 (n=200)	縦断基準準拠 (Reg, ROC) 分布 (SD)	- Clinical Global Impression-Severity - Clinical Global Impression-Index - Treatment response status	MCID	0.04-0.10
Stark et al. (2010)	横断研究	炎症性腸疾患 (n=502)	縦断基準準拠 (Reg)	- Patient's perceived improvement after the treatment	MD	改善 : 0.076 悪化 : 0.109
Shikier et al. (2006)	ランダム化比較試験	乾癬 (n=147)	断面基準準拠 (AC) 分布 (SEM)	- Psoriasis Area and Severity Index - Physician's Global Assessment	MID/MCID	0.09-0.22
Walters & Brazier (2005)	縦断研究	混合 (n=283)	縦断基準準拠 (CD) 分布 (ES, SRM, SD)	- Global Rating of Change	MID	0.07

出典 : Coretti et al. (2014) の文献考察内容の再整理

AC: Average change, CD: Change difference, CRD: Clinically relevant difference, CVS: Cut-off values of success, ES: Effect size, LA: Limits of agreement, MCIC: Minimum clinically important change, MCID: Minimum clinically important difference, MD: Meaningful differences, MDC: Minimum detectable change, MDS: Mean difference score, MID: Minimum important difference, REG: Regression analysis, ROC: Receiver operating characteristics curve, SD: 0.2 or 0.5 standard deviation, SDD: Smallest detectable difference, SEM: Standard error of measurement, SRM: Standardized response mean

オ) 利害関係者の意見

(1) 専門家の意見 (諮問会議)

- 質の加重値を適用する際、時々臨床的に有意義な値なのかに関し悩むことがある。MCIDという言葉が与える負担があるため、これを緩和して表現する。方法への検討が必要。

(2) 製薬会社の意見

- なし

カ) 指針の改正方向に関する提言

暮らしの質の改善が、はたして臨床的に有意義なのかを考えてみるきっかけを与えるため、下記のような

内容を指針の説明に含める案を検討した。

質の加重値の変化や差が有意義であることを証明するため、資料の提出者は臨床試験にて用いられた MAUIのMCID (minimal clinically important difference) の報告を考慮することができる。MAUIが用いられなかった場合は臨床試験にて用いられた他の詳しい質の道具のMCIDを報告することができる。報告の際は、MCIDの算出対象及び方法なども共に記述する。

しかし、質の加重値の場合は様々な資料元（例、既存値の引用、直接測定）にて算出されることもあり、対象群の暮らしの質の改善が必ずしもモデルにて用いる治療群間の質の加重値の差として表れるのではないため、MCIDの対象道具が明確でない可能性がある。従って、質の加重値を取り扱う効用パートにて、これを述べるよりは分析技法にてCUAの対象可否を議論する際、上記の内容、即ち、暮らしの質の改善における臨床的な有用性の検討に関する内容を簡単に述べることを提言する。

ケ. 経済性評価のモデル構築における統計的な考慮事項

1) 現在の指針

現在の指針は、統計的な考慮事項を別途の指針として提示していない。しかし、統計的な争点事項に関し、主な利害関係者らから継続的な要求事項があったところ、主な争点事項らの除外国の指針を検討し指針の改正方法を提言した。

2) 効果の推定

経済性評価のモデル及び適切な資料元を選定し、資料元から推定された申請薬の効果をモデル内のパラメータ推定値として用いる。経済性評価モデルの構造及び目的、信頼性、一致性に合致する。適切な資料元の選択（複数の資料元が可能）及び適切な効果の推定値を算出することが重要である。

ア) 除外国の指針

(1) カナダ・CADTH

ナダの指針では、指針の全般にわたり、統計的な事項を含めている。まず、効果の推定における統計的な考慮事項は下記の通りである。

- 統制された最適の環境（臨床試験では、一般的にランダム化比較試験、randomized controlled trial、RCT）にて行った臨床試験における処置結果を効果（efficacy）と言い、実際の環境、即ち、日常の診療環境にて評価された処置の結果を実際の効果（effectiveness）という。
- 意思決定者は一般的に実際の環境にて、処置が適用された患者らから算出された効果に関心がある。これは、基本分析にて生存期間と生存の質に関わる臨床的に有意義な結果が必要であることを意味する。
- 重要なことは、RCT資料のどこまでが実際の環境における実際効果を反映するかを評価することである。これは、臨床試験の外部妥当性に当たる経済性評価では、臨床試験における統制された効果よりは実際の効果を反映することを意思決定者はより有意義と考えるためである。
- 資料にて、実際の効果を推定する場合、臨床試験の環境と実際の環境にて差が生じかねない要因を選別しなければならない（例：順応度）。実際の環境における要因を分析にて考慮する場合、重要に判断すべきことは潜在的な仲裁効果の修正要因を患者の健康結果に繋げられる資料の利用可能性を確認しなければならない。処置に関する潜在的な異常反応の評価が必要である。比較しようとする処置に関する異常反応は費用と結果へ影響を与えかねないため、これを経済性評価にて考慮することは重要となる。とりわけ、異常反応の特性、頻度、期間、重症度などにより異常反応は患者の順応度、死亡、病気の状態、効用、リソースの使用などへ影響を与えることがある。臨床的に有意義な異常反応に重点を置かなければならず、これは実際の環境にて比較

する処置の期待費用と効果へ影響を与えかねない。

□ 実際の効果と安全性の評価における母数

- 効果と安全性を評価するための母数は、評価変数により違う形になりかねない。例えば、効果もしくは安全性の変数が連続型である場合は、絶対リスク差（absolute risk difference）により、2進数である場合は相対危険度（relative risk）もしくはオッズ比（odds ratio）を用いて、対照群比、処置群の効果サイズを測定することができる。ことができ、評価母数として用いることができる。
- もし、効果もしくは安全性の変数がprogression free survival（PFS）もしくはoverall survival（OS）のように、事件が発生するまでの時間である場合は母数的、准母数的、もしくは非母数的な生存分析の結果であるハザード比（hazard ratio）を評価母数として用いることができる。経済性評価にて、母数の推定値としてハザード比を用いる場合、相対的な効果は分析期間中に一定であると仮定する。
- 母数の推定値として資料元から推定されたハザード比を用いる場合、比例リスクの仮定が適切かを全体的な分析の流れにて評価しなければならない。こうした過程の合理性を評価する際、時間によって変わる他の変数がハザード比へ及ぼす影響を考慮しなければならない。実際、環境と相反する程度により、結果から発生し得る潜在的なバイアスを考慮することが重要である。
- 比較効果の推定が妥当でなかったり、治療の選好により比較研究が非倫理的だったりする場合、相対的な条件よりは絶対的な条件、即ち、各仲裁に関し別途に推定値を提供する資料元を利用する必要があると想定される。これは、異常反応のリスクに関する情報を必要としたり、迅速かつ致命的な疾病の進行があったりする場合、処置の効果を評価する際に発生し得る。
- これと同様に、MRI装置やcomputed tomography scannerのような新たな技術を評価する場合、比較できるような代案がないことがある。こうした場合、処置の効果は確認された件数などの絶対的な値を利用することができる。
- なお、絶対的な値は比率として提示されることもある。確率と比率は、一般的に相互交換的に用いられる意思決定のモデルにて、確率を推定しようとする場合、瞬間的な事件の比率（instantaneous event rate）は特定期間中の確率に変換しなければならない。
- 変換のためには、比率が特定の期間中に一定するという仮定が必要である。例えば、個人の健康状態における転移を説明したり、診断検査の正確性を評価したりするための感度や特異度などは確率を評価母数として設定することができる。

□ 資料元と妥当性の評価

- 実際の効果、異常反応などに関する母数の推定値を求めるための資料の出典としては、RCT、観察研究、行政データベース、単一群の研究、専門家集団の意見などがある。経済性評価のための目標の母集団、比較群、給与基準などを反映した関連資料元を定義しなければならない。意思決定の問題に定義されたように、臨床的な実際の効果と異常反応に関する情報を検索する。こうした検索は包括的であり第三者により再現可能でなければならない。
- 資料元を評価する際は、経済性評価の全般を考慮しなければならない。経済性評価モデルの母数に関する情報を持っている資料元を利用する必要があり、究極的には評価の目的に合致しなけれ

- ばならない。これに関する評価は資料元と意思決定の問題における差を調査することである。特別な処置を受けるような患者のタイプ、特別な異常反応の発生が見込まれる患者のタイプ、処置の適用方法もしくは容量の差、一般的な処置との差などのような様々な事案を考慮しなければならない。
- もし、以前の検索もしくはレビューがある場合は、経済性評価の目的に合致する程度の範囲、包含・除外の基準を評価することができる。同様に、メタもしくはネットワークメタ分析の経済性評価に関する目的の合致性は範囲、包含・除外の基準及びこうした分析が意思決定の問題とどれほど関連するかを評価する。
 - 母数の推定値の信頼性も考慮しなければならない。これは、推定値のバイアスと関連している。信頼性を評価する際、資料元の方法論的な厳格性、即ち、実際の値に関し、信頼できるほどの推定値を提供するかに関する評価が必要である。ベースライン特性の処置群間の非均衡性と performance bias と detection bias のような潜在的なバイアスの発生要素を考慮しなければならない。
 - 資料元を評価する際、経済性評価モデルの他の母数を推定するために用いる他の資料元との一致性を評価しなければならない。資料元における資料の収集過程と経済性評価モデルの他の複数の母数との関連性があるかを考慮しなければならない。

○ 母数の推定

- 経済性評価のため、実際の効果及び異常反応に関する母数を推定する際、究極的には様々な資料元における目的の合致性、信頼性、一致性の潜在的な差を考慮した方法を用いて、全ての可能な資料元から習得した資料を統合する必要がある。なお、統合した資料はなるべく、情報の全体性を反映しなければならない。
- 実際の効果及び異常反応（安全性）の評価変数に関する確率分布を仮定しなければならず、不確実性を評価する。
- 現在も全ての可能な資料元において、根拠を合体させる方法論の開発が勧められている。全ての可能な資料元を合体させる適切な方法がない場合、メタ分析やネットワークメタ分析を用いることができる。経済性評価における意思決定の問題にて、比較群が2つ以上か直接的に両群を比較できない場合が存在する可能性がある。こうした場合、一般的に両群を比較するメタ分析より、複数の仲裁群を比較することができ、間接比較を可能にするネットワークメタ分析を愛好する。
- 経済性評価に関し、既に行われたレビューもしくは追加したいものがある場合、考慮している資料元を追加的に分析し更新することができる。このためには、追加的な目的の合致性、信頼性、一致性を考慮する必要がある。こうした追加的な分析は、以前のサマリーの推定値を現在の分析に結合して更新することができる。
- ただ、一つの資料元がある場合や資料元間の差があり、資料の合体が不適切な場合、各資料元はモデルの他の部分に用いられた資料元との目的合致性、信頼性、一致性の観点から評価しなければならない。複数の資料元から一つの資料元を選択する場合、どのような資料元がこうした基準に最適かを判断する必要がある。例えば、母集団のレベルにて意思決定を行う場合、非無作為割り当ての観察研究もしくは行政データベースと理想的な環境、truncated horizon、制限的な選定・除外の基準の下で行われたRCTと比較することは相対的な目的の合致性とそれぞれの資料元の信頼性にて選択を必要とする。

- 同様に、比較単一群の研究を用いて、効果と異常反応に関する母数推定値の資料元として用いる場合、その資料元の患者の性格がモデルにて他の母数を評価することに用いられた資料元と似て
- いても母数の推定値の信頼性に関する懸念がある。
- もし、意思決定が患者の報告による異常反応の推定値に基づく場合、こうした単一群に関する研究と方法論的に適切な専門家の意見に基づいた資料元（この資料元は、信用度は高いが目的の合致性には合わない可能性がある）間のトレードオフが生じる可能性がある。目的の合致性、信頼性、一致性において、最も適切と判断した資料元の選択過程及び母数の推定値方法に関しては経済性評価レポートに詳しく記述しなければならない。
- 基本確率的な分析の前後内容やシナリオの分析を用いて、こうしたトレードオフの潜在的な意味を考慮する必要がある。即ち、信頼できるほどの推定値であるが、評価の目的に合致するかに関する懸念（例：効果 推定値）は、効果と区間を適切に変更して確率的な分析を行い、評価する。母数の推定値に関する効果のサイズは専門家（expert elicitation）もしくは実証的な資料をもって説明する。
- 推定された区間は、効果のサイズに関する不確実性の程度を反映して、より拡大する必要がある。代案として、基準間における最も良いトレードオフと判断される資料元は、初期の基本分析にて用いられることができ、他の潜在的な資料元に関するその結果の敏感度は続くシナリオの分析にて評価することができる。
- 母数を推定するための情報が足りない場合、不確実性の分析は重要な役割を果たす。少ない資料とそれによる不確実性に関する影響はreference case probabilistic analysisにて用いられる分布の広さと形に反映されなければならない。推定値の他の構造的な仮定や資料元を含むシナリオの分析を通じて評価しなければならない。少ない資料をまとめる方法（即ち、事件のない研究群を取り扱う場合）に関する議論は Cooperを参照することができる。

(2) オーストラリア・PBAC

- それぞれの臨床試験（資料元）に関し、全ての試験母集団に関する一次的な評価変数及び関連した評価変数の結果を提示する。
 - 一般的に、下記のような細部項目を提示する。
 - リスクレベル（at risk）もしくは結果に当たる対象者の数
 - （もし、適切であれば）事件が発生した対象者数
 - 事件が発生した対象者の比率、群別の平均（標準偏差）もしくは中央値（四分位数）
 - 各群における結果に関する信頼区間
 - 結果の解釈
 - 候補として設定されたMCIDの観点からの結果に関する議論
 - 結果表にて、結果評価のタイミングの提示は重要である。もし、資料元毎に、結果の測定時期と追跡観察の時期が異なる場合、各結果表の下に差を提示する。結果の解釈において、臨床試験にて中間分析を行ったか、これによる早期終了はあったかについて説明する。

□ 2進数の資料

- 各群別の事件発生の頻度、相対頻度、相対リスクの推定値及び信頼区間、リスク差の推定値及び信頼区間などと提示する。

□ 連続型データ

- 殆どの臨床試験にて、連続型の変数はベースラインと臨床試験の進行において、前もって決めたタイミングに再測定する。処置効果は、統計的な方法で算出する。最も多く用いる方法は共分散分析（ANCOVA）であるが、他の方法も使用可能である。ANCOVAの結果は、ベースラインを補正した平均変化量の差である。統計的な補正が適用された場合、補正した共変数を提示し、用いた統計的方法の仮定が適切かを検討しなければならず、推定された処置効果における共変数補正の影響の程度を説明する。
- もし、結果変数が一回以上測定されたことにより、一つのタイミングが選択されたら、選択したタイミングに関する根拠を提示する。選択していない他のタイミングにおける処置効果があるかを検討し提示する。
- 経済性評価にて連続型変数を範疇化して分析した場合、範疇化するためのカットポイントに関する根拠を提示する。もちろん、同カットポイントは臨床的に関連がなければならない。なお、他のカットポイントを考慮した敏感度の分析を行わなければならないか、群別に連続型変数及び累積分布関数を提示する。

□ 事件までの時間資料（生存資料）

- 各資料元別に、 Kaplan-Meier 曲線を提示する。（一般的に、比較群である場合）もし、資料へアプローチできなかったり、Kaplan-Meier 曲線が論文として掲載されていなかったりする場合、これを明確に記述しなければならない。
- 生存資料の分析方法に関し記述する。分析の仮定及び過程の適切性に関する検定を記述する。例えば、生存資料の分析をもってCoxの比例ハザードモデルを行った場合、ハザード比及び95%の信頼区間を提示し、比例ハザードの仮定が十分であるかを検定し結果を提示する。
- もし、比例ハザードに関する仮定が不十分である場合、比較効果を推定するための代案を提示する。限られた平均生存時間が用いられた場合、限られた平均の推定値及び群間の差を臨床試験の期間中における様々なタイミングにて計算する。

□ 順序データもしくはカテゴリカルデータ

- もし、臨床試験の資料が順序データもしくはカテゴリカルデータを用いて測定される場合は、連続型データ資料にて提示したような方法を用いる。こうした資料を分析する場合、専門的な臨床統計学者の諮問が役立つ（とりわけ、こうした資料をメタ分析する場合）。
- 群別に、各タイミングにて95%の信頼区間と共に、MAUIの結果を提示する。各タイミング別に、アンケート可能な対象者の数及び回答した対象者の数を提示する。提示できない場合、その理由を説明し、試験計画書に明示された結果を提示する。

- 試験の中間（もしくは適切な）タイミングまで求めた平均効用加重値間の積分としての群間の差及び95%の信頼区間を提示する。もし、MAUIを比較する他の方法を用いる場合はその方法を詳しく記述する。
- スコアリングアルゴリズムが一般の母集団から算出できない場合、可能であれば他のスコアリングアルゴリズムを用いた敏感度の分析を行う。含まれた研究にて、二つ以上のMAUIが行われた場合はその結果を比較した結果を提示する。
- 臨床試験にて測定する他の評価変数に関する結果を提示する。

□ MCIDの観点における実際の効果

- MCIDに関する1次評価変数及び主な評価変数の結果を説明する。対照群と比較する場合、処置群で計画通りのMCIDが表れたかを記述する。信頼区間にて、計画通りのMCID比、小さな差値を含めたかを記述する。累積分布関数を提示する。
- 適切な統計的検定を用いて、各群にて計画したMCIDより大きな回答を達成した患者数を比較する。可能であれば、他のMCIDを用いて評価する。

□ 非劣性限界の適用

- 単測95% 信頼区間の下限（もしくは上限）値と非劣性限界値値を比較し、非劣性限界が含まれるかを評価する。非劣性の検定はITT群とPP群と共に行い、差を記述する。臨床試験にて、比較は一般的に相対的な測定値を用いるが、重要な差は一般的に絶対差を提示する。この場合、非劣性の限界として他の結果測定値を用いる場合、変換する方法に関して記述する。もし、非劣性が十分ではない場合、その理由を説明する。非劣性の結論を追加的に裏付けられるような考慮事項を記述する。（例：同じ種類の申請薬剤と対照薬にて、申請薬剤の点推定値がより良い場合、申請薬剤の安全性がいいか、毒性が低い場合）。

イ) 指針の改正方法

- 現在、統計的な考慮事項を別途に切り分け、独立的に記述する指針書はない。統計的な事項は全般的な経済性評価過程にて、考慮されなければならないため、別途に切り分け、記述することよりは指針の随所にて、必要に応じ記述する必要がある。とりわけ、モデルの構築、間接比較などを用いて適切な統計及び結果の提示方法に関し、記述しなければならない。さらに、変化を続ける臨床試験の設計などを考慮し多補正方法などを詳しく記述する必要はないが注意すべきこと、もし、補正方法を用いた場合は根拠と制限、敏感度の分析などに関する内容を含めなければならない。

3) 長期的な効果を推定するための外挿

ア) 現在の指針

現在の指針では、外挿の妥当性及び仮定に関する根拠を提示し、敏感度分析の実行に関して下記のように勧告している。

□ 外挿の正当性

外挿のために用いられる仮定は、必ず臨床的な妥当性を持たなければならず、透明に報告する必要があり、仮定の使用を正当化しなければならない。外挿に関する様々な仮定の影響を比較するには、複数のシナリオを考慮しなければならないが、例えば、治療効果の期間中における治療の利益をいつにするかに関するシナリオは①なし、②治療期間同様で同じレベルと維持、③ long-termにて減少などと考えることができる。モデリング過程にて伸びた期間に関し、費用と健康利益を外挿するケースがある可能性がある。治療効果を外挿するために用いられる仮定は、必ず臨床的な妥当性を持たなければならず、透明に報告されなければならず、明確に正当化されなければならない。

- カナダのCADTH及びオーストラリアのPBACは、外挿に関するより詳しいガイドラインを提示している。
- イギリスのNICEは、別途のTechnical Support Documentを通じて、詳しく説明している。

イ) 諸外国の指針

(1) カナダのCADTH

- 生存資料の分析にて、資料が観察された期間を過ぎた長期的な効果を推定するためには、外挿法が必要な場合がある。長期的な効果における外挿は、状態や疾病の自然史 (natural history) と利用可能な資料の期間を過ぎた期間における仲裁に関する実際の効果の推定が必要な場合に行うことがある。
- 短期的な効果の母数推定値を用いて長期的な効果の外挿に向け、一般的に母数的なモデルを利用した生存分析を使用することができる。個人単位の生存資料を分析する生存分析法はたくさん開発されている。
外挿モデルの適切性の考慮における重要な議論事項は、観察期間が過ぎた処置効果に関する仮定に関わるものである。まず、観察期間を過ぎて発生した効果推定値の比率を報告し、正当化しなければならない。この比率は、単にモデルにて考慮する全体的な分析期間にて推定された incremental QALYに関し、資料が観察された期間から推定された incremental QALYの比で測定する。
- なお、処置が中止された後に蓄積もしくは増加する利益推定値の比率も報告しなければならない。モデルにて考慮する全体的な分析期間中に、推定された incremental QALYに関し、処置が適用された期間中に推定された incremental QALYの比をもって測定する。この二つの測定値が臨床的

に可能かを評価することは外挿法の適切性を評価することに役立つ。

- どんな状況でも、効果の母数推定値の外挿に関する根拠を提示しなければならない。個人単位の資料を利用する場合、Survival Model Selection Process Algorithmに定義された適切な方法を用いるべきである。確率的な分析を利用し、母数の不確実性に関する全ての考慮が必要である。
- 構造的な不確実性を探索するためのシナリオの分析は、母数的な分析モデル（分布を含む）が観測された資料にどれほどふさわしいかを評価するため、比例ハザードを仮定した分析と比例ハザードを仮定していない分析をそれぞれ行い、その結果を比較しなければならず、様々な母数的な分布を用いるべきである。
- 資料の観察期間を過ぎ、臨床的な効果が続く期間とそのサイズは外挿に関する重要な判断根拠である。外挿の適切性を正当化するには、外部の資料元を参照する必要がある。適切な根拠なしに、相対的な実際の効果（effectiveness）が処置中に続くという仮定は不適切である。
- 処置の実際の効果は、単に資料が観察された期間中のみ持続すると仮定したり、時間の経過により減ると仮定したり、処置中は続くとは仮定したりすることができる。この三つの仮定の中、二番目の仮定に関しては、事件発生リスクの相対的な減少は、特定のタイミング後に減少が始まって以来、安定的な減少効果が表れたり、処置の効果が消えるまでの時間により、減少を続けたりすると仮定することができる。
- 全体的な分析期間において、早いタイミングに外挿を行う場合は注意が必要である。これは、全体的な分析期間に関し、平均生存時間を推定するために、潜在的にかなりの期間を外挿しなければならないためである。

(2) オーストラリアのPBAC

- 観察された資料にて、分析期間に関する費用と効果推定に関する処置と比較群間の全ての重要な差を推定できない場合、外挿は正当化されやすい。外挿を行う場合、外挿方法に関する詳細を記述する必要がある。

外挿方法を用いる場合、少数の試験対象者からこれ以上の事件が発生しないことで、観測された資料を信頼し難いタイミングまでは、外挿のモデルを用いるより、観測された生存資料を用いなければならない。

- どのタイミング後から外挿された転移確率を用いるかに関して記述し、根拠を提示しなければならない。例えば、臨床試験の資料に相応しい生存曲線が、より長い追跡観察期間中に、より多くの同程度の患者のコホート資料にて、相応しい生存曲線から外れるタイミングを利用することができる。敏感度の分析では、様々なタイミングを選んで評価する必要がある。
- 上記にて決めたタイミング後の転移確率を外挿するために、なるべく観察された個人資料に基づいて、適切に推定された母数的な生存曲線を利用する。

○ 細部項目は下記のようなものである。

- 観察期間後も、比例ハザードの仮定が適切かを評価する。
- 観察された資料に、様々な分布を仮定した母数的な生存分析モデルを適用する。（例：指

数、ログ-正規、ログ-ロジスティック、ガンマ、ゴンパーツ) 生存曲線屈折の様々なポイント (point) を考慮した比較的柔軟な外挿法を含めることができる (例: piecewise spline models)。

- 視覚的な検査、Akaike's information criterion (AIC)、Bayesian information criterion (BIC) を用いて、モデルの適合性を評価する。基本分析のための最も相応しい母数的なモデルを選び、敏感度の分析により他の母数的なモデルを評価する。
 - 観察期間後の予測値に関する妥当性を提示する。(例: 進むハザード比、処置効果、交差点、各群における残りの生存など)
 - 外挿方法をもって推定した処置効果を用いて、分析期間に関する生存曲線を生成する。もし、処置効果が維持されるか増えるが、こうした現象が臨床的に可能でなければ、適切なタイミングにて両群の効果は同様であると仮定できる。
 - 外挿した処置効果を考慮する場合、処置の中止もしくは継続における臨床的な決定に間する明確な考慮を提示しなければならない。これに関わる全ての仮定を記述し、正当化する必要があり、それぞれの処置費用をモデル化する場合、一致性があるように適用しなければならない。
 - 経済性評価において、外挿の方法を参照できる論文は複数ある。

□ 個人患者の外挿に関する他のイシュー

- 多重の中間もしくは結果事件の経験を表現する範疇型資料の場合、どの事件に関する時間をモデル化し、合わせられた事件がそれぞれの競争的な事件に関する確率を定義するには、多項ロジスティックモデルを用いる2段階の過程を利用することができる。

できれば、時間変化の確率を反映するために、多項ロジスティック回帰分析に時間変化の共変量を含める。他の代案としては、各事件に関する独立的な競争のリスクモデルを考慮することができる。しかし、同方法は多重事件が独立という仮定により、母数の不確実性を過大に推定する傾向がある。
- 連続型変数に関しては、連続型変数を範疇化するか、一般化推定方程式の方法を用いることができる。

□ 出版された生存結果を用いた外挿法

- もし、個人患者の生存資料を用いられない場合、グラフデジタイザーソフトウェアを用いて、出版された Kaplan-Meier 曲線から生存確率を外挿することもできる。代案として観測された資料が、不安定なタイミングの最後の屈折される地点後に抽出された生存資料に関し、一定のハザード関数 (指数) や単調増加するか減少するハザード関数 (ワイブルもしくはゴンパーツ) を適用することもできる。代案として、Kaplan-Meier 曲線から無増悪として残っている患者数が少なく、観察された資料が不安定となるタイミング前の生存資料を抽出してリスクが一定のハザード関数 (指数関数) やリスクが単調増加するか減少するハザード関数 (ワイブルもしくはゴンパーツ) を適用することができる。
- 他の生存曲線を適用後、適合度の検定を行い、基本分析にて最も適合度の高い曲線を用いなければならない。

ウ) 指針の改正方向に関する提言

- 外挿に関する方法と注意事項を指針書に含めなければならない。主に、生存分析にてOSを推定するための外挿がおおく発生するため、生存分析をはじめ外挿部分を記述する必要がある。

4) 処置転換の補正

倫理的な理由などにより、処置転換を考慮した臨床試験の設計は、比較的最近に行われている。こうした臨床試験においては、一般的に対照群にてprogressionが発生すると、その後、新たな処置への転換を許して追跡観察を行う。この場合、intention to treat分析群を用いて、処置群と対照群の間のOSを比較すると、対照群のOSが過大評価され、処置群と対照群の間のOS差は過少評価される可能性がある。統計的な因果関係推論の観点から、こうした処置転換を補正した方法論が開発されているが、現在の指針ではこれを反映していない。

ア) 現在の指針

現在の指針では、処置転換に関する統計的な考慮事項を別途の指針として提示していない。

イ) 諸外国の指針

(1) オーストラリア・PBACの処置転換の補正

- 基本的に、処置転換が行われない資料や根拠を優先する。
- モデルに用いた資料元の中で、処置を転換した対象者がいる場合、転換のパターンが比較群における現在の臨床慣行と処置に関する未来の臨床の実行が一致するかを評価しなければならない。もし、そうでなければ、観測された比較処置の効果が、期待される処置効果を反映しない可能性がある。こうした時は、補正が適切なケースもある。
- 処置の転換を補正する方法の適切性は、同方法にて用いた仮定に頼る。しかし、こうした仮定の妥当性を評価することはかなり難しい。補正方法は、実際の比較処置効果を過大推定するリスクを下げる方法でなければならない。

○ 選好する方法

- 臨床試験の各群における処置転換のメカニズムを説明する必要がある。各群に関し、下記を記述する。

- 転換が行われた薬物
- 転換の程度
- 比較群から処置群への転換が、現在の臨床環境を反映するかの可否（もしくは差）
- もし、処置に転換した薬物リストがあれば、処置群における転換が現在の臨床環境を反映するかの可否

- もし、転換（見込まれる転換の比率）が現在（比較群）もしくは未来（処置群）の臨床環境と似ているなら、こうした群における処置転換の効果を補正することは不適切である。
- もし、転換（もしくは転換の程度）が現在の臨床環境を反映しなければ、差を記述し、次の 이슈ーを取り扱わなければならない。
- 処置転換を補正するための分析方法が、臨床試験の計画書にあらかじめ明示されたかを記述する。処置転換の補正を説明した臨床試験計画書のセクションを参照して記述する。
- 処置転換を行う前の参加者の特性と共に、変換群と非変換群のベースラインの特性を提示する。もし、試験対象者が病気や状態の悪化により、処置転換が行われたが、実際に悪化され処置を転換した対象者群と、悪化されたが処置を転換しなかった対象者の特性を提示する。
- 転換の理由（例：症状の悪化、異常反応など）及び理由別の患者数を提示する。
- 処置転換を補正して、生存曲線を推定するいくつかの方法が開発、用いられている。もし、比較群の推定値が明確に無効と推定されたら、単なる補正方法は受容可能である。もっと複雑な補正方法は、妥当性に関する検定がかなり難しい仮定に頼る。（例：inverse probability of censoring weights, a rank-preserving structural failure time model, 2-stage methods）
- 適用した方法の仮定と、どのようにその仮定を検定したかに関する詳細を報告し、各方法の仮定の適切性をはじめ、選択した方法の正当性を詳しく記述しなければならない。
- 処置転換を補正した処置効果の推定値に関する不確実性を減らし得る追加的な根拠と議論事項を提示する。もし、複雑だったり新たな方法論が用いられたりしたら、多く採用している他の補正方法の結果も提示し、なぜこのような方法を用いたかに関する正当性を提示すべきである。
- より複雑な方法が用いられた場合、比較のために、より単純な方法の結果も提示しなければならない。
- 方法に関する細部的な議論が必要な場合、技術的な追加ドキュメントとして提示する。

□ 処置転換に関する補正の結果

- 処置転換を補正した処置の効果を推定するための各方法に関し、補正された推定値と95%の信頼区間を提示する。適用した補正方法間の処置効果の異質性を記述する。処置転換のない場合に関する処置効果の推定値及び95%の信頼区間を提示する。

- 可能であれば、処置転換を補正した後、各群におけるカプラン・マイヤーの生存曲線（95%信頼バンドを含んで）を提示する。図と共に、臨床試験の追跡期間の間、定期的なタイミング別に、中途脱落した対象者及びまだ死亡が発生していない対象者の数を提示する各群別のリスク頻度表を作成する。
- 複雑な統計的方法を用いた場合、処置転換がなかった処置群（例：過去の対照群）を研究報告した論文を十分に検索する。検索された論文における結果を現在の試験に適用できるかを考慮する。処置転換がなかった研究におけるカプラン・マイヤー曲線を現在補正されたカプラン・マイヤー曲線と比較し、差があったらその理由を説明する。
- 処置転換が発生する前の結果を用いた推定値（例：PFS）がある場合、その結果が臨床的に関連性のある結果（例：OS）における代理指標になり得るかを説明する。例えば、もしPFSがOSの適切な代理指標であれば、PFSを変換して推定したOS推定値と統計的な方法をもって、処置転換を補正し推定したOSを比較する。
- 可能であれば、処置転換の補正に向け、様々な統計的なアプローチ法を利用する。様々な方法を適用した結果が同水準であれば、これは推定値に関する不確実性を減らし、結果の信頼性を高めることができる。過去の対照群との比較もしくは代理指標をもって変換して推定したOSは統計的なアプローチ法における信頼性を高めることができる。

□ 副集団もしくは間接比較による処置転換の補正

- 副集団もしくは間接比較を用いて処置転換を補正する場合、推定値にバイアスが発生する恐れがある。副集団もしくは間接比較を用いることは低いレベルの根拠と言いかねないため、こうした方法は並行しないほうが良い。

(2) イギリス・NICE

- 臨床試験の対照群に含まれた患者が追跡観察の期間中、どのタイミングに試験処置群への転換を許した場合、処置転換が発生する。処置転換は癌治療剤の臨床試験にて、よく発生する。
- 処置転換が発生した場合、ランダムに割り当てられた通りに群を維持して分析するintention-to-treat分析を用いてOSを分析すると、結果にバイアスが生じかねない。もし、対照群の患者が処置群へ処置転換し、新たな処置からの利益が発生したら、新たな治療のOSに関する処置効果は過少推定される。

生存に影響を与える処置に関し、NICEは分析期間の全体を考慮する経済性評価を必要とする。この場合、ITT分析は適切でないため、処置転換によりバイアスが生じかねない。

- 処置転換が行われた場合、生存の推定値を補正するための様々な統計的な方法がある。しかし、こうした方法は強い仮定を基にする上、仮定にかなり頼っている。処置転換のタイミングに、途切れたと見なしたり、処置転換された対象者を分析から除いたりする単なる補正方法は処置転換が予後に関わる可能性が高いため、選択のバイアスが発生する可能性がかなり高い。
- もし、仮定に満たされる場合、理論的にバイアスのない結果を与えるより複雑な補正方法を用いるこ

とができる。Rank Preserving Structural Failure Time Models (RPSFTM) と Iterative Parameter Estimation (IPE) アルゴリズムは、反事実的な生存時間（処置転換された患者において、もし、処置転換を行わなかったら観測できたはずの生存時間）を推定するためのランダムな割り当てに基づいた分析方法である。

- Inverse Probability of Censoring Weights (IPCW) 方法は、観測された資料に基づいた分析として、処置転換が行われた場合、そのタイミングに途切れたものと考え、処置転換が行われなかった対象者は加重値を用いて補正する方法である。
- こうした方法別の仮定、メリット、デメリットをはじめとする方法論をNICEのTSDに詳述する。

□ NICE TSD: Rank preserving Structural failure time model (RPSFTM) and iterative parameter estimation (IPE) algorithm method

- RPSFTMとIPE方法は、ランダムな割り当てに基づき、処置転換を補正する方法である。RPSFTMは、モデルの仮定を満たす効果を推定するため、g-推定値方法を用いる。一方、IPEアルゴリズムは効果を推定するため、反復的に母数的な過程を用いる。よって、IPE方法は生存時間に関し、母数的な分布によるという追加的な仮定が必要である。

○ RPSFTM

- 一般的な単一母数RPSFTMは、各患者から観測された生存時間（Ti）に関し、処置を受けない期間における生存時間（TAi）と処置を受ける期間における生存時間（TBi）という二つに分ける。処置群としてランダムに割り当てられた群と、対照群として処置転換が行われない群のTAiは0であり、TBiは0である。しかし、処置転換が行われていない群（即ち、順応が不完全な群）では、TAiとTBiは共に、0より大きい。
- 同方法にて、潜在的な結果変数であるUiは下記のような因果的なモデルと設定する。

$$U_i = T_{A_i} + e^{\psi_0} T_{B_i} \quad (1)$$

- ここでは、仲裁に関するacceleration factor (AF) である。対象者が仲裁群である場合は1、対照群である場合は0と、2進数の過程を定義すると、因果モデル (1) は下記のように改めて定義することができる。

$$U_i = \int_0^{T_i} \exp[\psi X_i(t)] dt$$

- 同モデルにて、母数はg-推定法を用いて推定することができる。ψの潜在的な値らの範囲を考慮する。モデル (1) を用いて推定された U_i で、処置効果であるψに関し、可能な値を含む範囲に関する格子探索方法を行い、母数を推定する。ψの実際値は、 $U(\psi)$ はランダムに割り当てられた群と独立している。MarkとRobins (1993) に記述された通り、ベースライン (反事實的) 生存曲線は、両群において同様であるというRPSFTMのg-検定は非母数的にログ順の検定やWilcoxonの検定を用いることができ、母数的にはWald検定を行うことができる。RPSFTM 方法をもって、推定した反事實的な生存時間が与えられたら、Coxの比例ハザードモデルを用いて 補正されたハザード比 (hazard ratio : HR) を推定することができる。
- RPSFTMは、intention to treat分析におけるp-valueを維持する。補正された信頼区間は、これに基づいて計算することができる。即ち、仮にITTが統計的に有意な処置効果を表せなかったら、RPSFTMをもって補正した分析も有意な処置効果を表せられない。
- BransonとWhiteheadは、IPE過程による母数的な方法を用いて、RPSFTMを拡張した。RPSFTMのようなACFモデルを用いるが、初期のψを推定するべく、補正されていないITT資料に母数的なfailure time model (例：ワイブル、ログ-ロジスティック、ログ-正規分布) を適用する。
- 処置転換した対象者から観測されたfailure timeは、 e^{ψ} と数式 (1) の反事實的な生存時間のモデルを用いて変換され、処置群と対照群を再び母数的なモデルを用いて比較する。同過程は、ψを更新し変換群の反事實的な生存時間を再推定する過程を繰り返す。同過程は、 e^{ψ} が受け入れられるまで続く。
- IPEとRPSFTM方法は、同じ仮定に基づいている。処置が行われた時間に関する処置効果は、全ての対象者に同様である (common treatment effect assumption) 。処置の生存時間の不在はランダムな割り当て群と独立している。BransonとWhitehead方法は、追加的に生存時間が母数的な分布に従うと仮定する。
処置を受けた群の反事實的な中途切断時間と予後との関連性のため、IPEとRPSFTM方法にて中途切断は問題がある。Administrativeの中途切断の時間である C_i と $C_i \exp(\psi)$ において、最小値から $U_i(\psi)$ を再中途切断することにより、中途切断と受けた処置間の関連性を断絶し、バイアスを減らすことができる。もし、 $D_i^*(\psi) < U_i(\psi)$ である場合、 $U_i(\psi)$ を反事實的な生存時間である $D_i^*(\psi)$ の中途切断時間に代替する。

- IPEとRPSFTM方法の主な制限事項は、対象者は常に「on」もしくは「off」状態の単純な二つの処置状態のみを考慮することである。（即ち、処置を受けている時「on」もしくは受けていない時「off」）の比較群が活性の場合、これは分析にて問題が発生する可能性がある。多重の状態を考慮した多重母数モデルを考えることができるが、こうした分析は検定力が足りず、因果影響を有意義に説明するための点推定値を得られないというデメリットがある。
- 二つの状態を考慮したモデルに制限し、「on treatment」に基づいたIPEとRPSFTM分析方法は、「on treatment」ではない期間は「off treatment」と見なす。そして、同期間中に受けたどの処置も効果が似ていると考慮する。比較群が偽薬群であったり、姑息的治療であったりする場合は、両状態のモデルは合理的である。しかし、比較群が活性処置であるケース、事後治療が非活性であるケースの両状態の「on treatment」モデルは恐らく臨床的には適切ではないはずである。
- 活性対照群と非活性の事後治療を一つの「off」軍に括ることは、比較群に関する効果を薄れさせ、処置の効果を過大推定する恐れがある。
- こうした場合、対照群が受けたシーケンスと比較し、処置群が受けた全ての処置シーケンスの処置効果を推定するために、処置群の分析（treatment group）を行うことが合理的な可能性がある（処置転換を考慮するために、補正しなかった場合）。対照群から処置群へ転換する場合、この対象者は処置群の処置シーケンスに転換すると仮定する。
- 転換した対象者が受ける処置シーケンスは、初期段階にランダムな割り当てにより、処置群に割り当てられた対象者が受ける処置シーケンスと同様であり、そのシーケンスの相対的な効果は処置群における相対的な効果と同様であるという仮定である。従って、「common treatment effect」の仮定は処置群の単独を意味するものではなく、転換後に受ける処置の集合を意味する。
- 「on-treatment」分析にて、単なるtwo-exposure model群は相変わらず「on」状態にて処置を受けた時間は、今後の処置中止を無視し、試験群の処置開始後の全ての時間と処置の「off」状態における時間は、試験の処置を開始する前の全ての時間、もし、転換が行われなかった場合は対照群の処置を中止した時間を含めて合算する。
- 従って、このような分析は試験の処置及び活性対照処置に関する実際の効果を薄れさせる恐れがある。両処置群に関し、処置中止後の時間は、他の処置に従うからである。しかし、新たな処置の因果効果を提供しない反面、同方法は対照群よりはランダムに割り当てられた試験処置群の処置効果を妥当に推定することができる。

「on treatment」分析にて、単に比較群の効果を薄れさせるより「treatment group」のアプローチはpost-treatmentが似ており、処置転換した対象者と最初から処置群に割り当てられた対象者との効果が似ている場合と、特別に試験の処置に関する利益が処置の中止後も維持される場合、両群共に効果を薄れさせる。従って、比較群が活性であり、post-treatmentと違う場合に選好される。

- 「treatment group」のアプローチは、ランダムに割り当てられた群にて、post-study treatmentsを受けた群が実際的な処置の経路を代表するという仮定が必要である。この場合、もし、以前の処置によるpost-study therapy後に発生した利益が異なる場合は、実際的な医療技術の評価もしくは経済性評価の観点からこのような利益は試験処置の結果として見なす

のが適切な可能性がある。

○ Inverse Probability of Censoring Weights (IPCW)

- IPCWは、informativeが途切れた場合、処置の効果を補正するための周辺構造モデル (marginal structural model) にて用いられる方法である。臨床試験の資料は、対象者の観測値が死亡もしくは中途切断までの時間と記録される。対照群にて、試験処置をもって処置転換を行う場合、この対象者は転換タイミングにて中途切断される。
- 対照群に残っている他の対象者ら、即ち、処置転換が発生していない対象者らの生存時間はベースラインと時間による予後共変量の値に基づいて計算された加重値を用いて補正する。処置転換のタイミングにて、人為的に中途切断された対象者は、もし処置転換がベースラインと時間依存的な予後要因と関連する場合は、選択におけるバイアスが生じかねない。
- IPCW方法は、こうした選択のバイアスを最小にするべく、安定した加重値を利用する。処置転換した対象者と似ている特性を持った未転換対象者に大きな加重値を与える。根本的に、IPCWは処置転換を行わない対照群の対象者に加重値を与えることにより、対照群に関する擬似個体群 (pseudo-population) を構成する。従って、相対的に少数の対照群の対象者が転換しない場合は誤差が生じかねない。
- 安定化した加重値を求め、これを用いて補正された処置効果を推定するために、周辺構造モデル (MSM) を利用する。MSMは、加重値、ベースラインの予後共変数を含める時間依存Coxモデルを用いることができる。同モデルの結果は、IPCWをもって補正されたハザード比 (hazard ratio) として提示できる。ブラン・マイヤーの推定値やログ-順位の検定もIPCWをもって補正することができる。
- I IPCW方法は、「no unmeasured confounders」の仮定に依存する。これは、独立的な処置転換により発生するinformativeの中途切断を予測できる死亡に関する全ての時間依存的な予後要因に関する資料がないと、IPCWがバイアスのない結果を提供できないという仮定である。従って、こうした補正方法を用いるには、臨床試験にて補正に用いる変数及び資料を前もって十分に収集しなければならない。
- 同仮定の妥当性に関し、観測した資料では検定できない。処置転換を考慮した生存モデルは正しく設定しなければならない。もし、処置転換を常に招く共変数があれば、この方法は適切ではない。なお、「no unmeasured confounder」の仮定が成立しても対照群の標本数が少ない場合はIPCWの結果はバイアスが存在する可能性がある。

ウ) 指針の改正方向に関する提言

変わりつつある臨床試験の設計及び環境を反映し、処置転換の補正を考慮する内容を指針書に含めることができる。イギリスのTSDのように、詳しく記述する必要はないが、オーストラリアのように補正において考慮及び注意すべき事項を含めて、指針に記述する必要がある。

5) 新たな試験の設計

ア) 現在の指針

現在の指針は、新たな試験の設計に関する統計的な考慮事項を別途の指針として提示していない。

イ) 諸外国の指針

(1) カナダ・CADTH

- 適応型臨床試験の設計は、新たな設計において最も多く用いる設計である。同設計は、臨床試験にて中間分析を行い統計的な方法を用いて、第1種の誤りを適切に調整し標本数、割り当て比、処置群の数などを変更して、臨床試験が継続できるような設計である。
- 適切な資料元を選ぶ際、資料の解釈と使用に影響を与えかねない特性を考慮して臨床試験の設計を評価する。新たな臨床試験に関する設計方法が開発・利用される傾向であるため、他の資料元がある場合、基本分析ではこうした資料元を利用し、新たな設計からの情報の影響はシナリオの分析にて考慮する。

ウ) 指針の改正方向に関する提言

- 新たな臨床試験に関する内容は、カナダ・CADTHにて簡単に記述している。改正される指針書には、同内容を簡単に含めることができる。

6) 非劣性臨床試験及び非劣性の限界

ア) 現在の指針

現在の指針では、新たな試験の設計に関する統計的な考慮事項を別途の指針として提示していない。

イ) 諸外国の指針

(1) オーストラリア・PBAC

- 非劣性というのは、処置群が活性対照群より、効果と安全性において劣等でないという意味である。非同等性の検定にて、統計的な有意性を確保できないことが適切に非劣性を証明するのではない。非劣性検定方法として、最も多く用いられる方法は、効果群間の差の信頼区間が、予め決めた臨床的に重要な差の程度を含めないことである。

□ 非劣性を立証する変数の選択

- 第1次評価変数もしくは患者に関わる最も重要な結果変数に関し、非劣性の限界を設定する。もし、処置群が二つの違う適応症に影響を与えたり、二つの適応症に影響を与える二つの成分からなったりする場合、二つの適応症に第1次評価変数と患者に関わる最も重要な結果変数に関する非劣性の限界を設定する。

□ 非劣性の限界設定における根拠

- 処置群が活性対照群より、重要な差だけに比劣等でないことを確認できる非劣性の限界の限界を設定する。
- 結果変数にて、非劣性の限界として用いることができる臨床的に重要ではない差の程度を提示する。非劣性の限界の設定方法に関する根拠を提示する。これは、統計的な方法だけでは足りず、様々な資料元における根拠及び専門家の同意が必要である。臨床的に重要な最小限の差より小さな値として設定することが一般的である。

□ 前もって、決めた非劣性の限界

- 行われた臨床試験にて、予め非劣性の限界を既に設定した場合、その限界を提示し根拠を記述する。この限界の設定根拠を記述することにおいて、行われた臨床試験の計画書に記述された根拠を提示する。もし、行われた臨床試験の計画書に記述された根拠が不十分であれば、裏付けられる追加的な根拠を提示する。
- ある非劣性の臨床試験は、新薬が偽薬より優れていることを共に証明できるように設計された場合がある。しかし、こうした場合にも非劣性の限界は活性対照群より、処置効果の重要な減少を

許す。この場合、限界を再設定し非劣性を再検定する。

□ 前もって、決めていない非劣性の限界

- もし、前もって非劣性の限界を設定していない場合、後で限界を設定したり、正当化したりすることは容易でもなければ選好もされない。こうした場合は、保守的に限界を設定する必要がある。後で、非劣性の限界を設定する場合は下記の項目を含めることができる。

- 新薬比、比較薬の優越性の検定を行ったが、これを証明できない場合
- 共通対照群を用いた申請群と比較群の間接的な比較
- 行った非劣性臨床試験におけるプライマリー エンドポイントと経済性評価効果の結果変数が異なる場合

- 事後的に限界を設定する場合、できれば複数の資料元を用いて、限界設定の根拠を提示する。この際、複数の資料元における非劣性の限界の設定は同様なレベルでなければならない。非劣性の限界として一つの値を設定した根拠を提示する。

㉔) 指針の改正方向に関する提言

- オーストラリアのように、簡単に非劣性に関する臨床試験と限界を説明し、経済性評価にて非劣性を立証する場合、記述しなければならない内容を指針に含めることができる。

7) 生存分析及び生存分析モデルの選択

ア) 現在の指針

現在の指針では、新たな生存分析に関する統計的な考慮事項を別途の指針として提示していない

イ) 諸外国の指針

イギリスの場合、Technical Support Documentにて生存分析及び外挿に関する内容を詳述しており、カナダとオーストラリアは外挿の部分にて生存分析に関して記述している。

ロ) 指針の改正方向に関する提言

下記の内容を指針に含めることができるはずである。もし、下記の内容を指針に含める場合、詳細の程度に関しては必要に応じてイギリスのように詳細な指針書や記述ドキュメントに分けて作成する方法も可能である。

- 開始タイミングから事件が発生するまでの時間を効果の評価変数として用いる場合、生存分析の方法を適用することができる。
 - 臨床試験では、主に処置開始から病気が進むタイミングまでの時間（無増悪生存）もしくは死亡が発生するまでの時間（全体的な生存）を評価する。
 - 生存資料の特性としては、時間は0より大きく一般的に正規分布の仮定が適切ではなく、何よりも臨床試験の期間中に全ての患者から事件が発生するのではない。さらに、追跡観察中は臨床試験から中途切断が発生し得、こうした対象者の生存時間は途切れたという特性を持つ。
 - こうした特性を考慮した生存資料の分析法は母数的、準母数的、非母数的な方法がある。一般的に、生存分析の方法では、こうした中途切断はランダムで発生したと仮定する。（即ち、non-informative censoring）。この仮定は評価しなければならない。
 - 生存資料を用いた処置群の効果は、対象者を追跡観察した期間中に限って評価する。ランダムな割り当て臨床試験における生存の比較は、非母数的な方法であるカプラン・マイヤー曲線をもって、生存曲線を推定し、ログ-順位の検定により群間の差を検定することができる。比較効果の推定は、準母数的な方法であるCoxの比例ハザードモデルを用いて、生存資料を適用後、群間効果の比較測定値をハザード比（hazard ratio; HR）として提示することができる。
- しかし、医薬品の経済性評価においては、モデルの分析期間全体の効果を推定するために、追跡観察期間後の効果を外挿しなければならない場合がある。外挿を行うには、分布仮定をしない非母数的もしくは準母数的な方法よりは母数的な方法を用いる。
- 母数的な生存分析の方法は、仮定した分布により外挿の結果が敏感になるかねないため、最も適切な分布を選ぶことが重要である。
 - 適切な分布の選択方法は、カプラン・マイヤーの生存曲線の上に、様々な分布を考慮した分析の結果で推定した生存曲線を重ね書きし、視覚的にどれほど適合しているかを評価することができる。
 - なお、様々なモデルを考慮する場合、AICとBICなどの基準値を用いる方法がある。こうした基準

値は、可能度の適合値に生存モデルに含んだ母数及び資料の数を考慮した関数を用いた定義により、こうした基準値が小さいほどモデルがより適合であると判断する。

- ログ-累積のハザード図を視覚的に判断し、モデルの適合性を判断することができる。ログ-累積のハザード図は、 $\log\{-\log(t \text{ タイミングにて推定された生存曲線})\}$ をy軸に、 $\log(t \text{ タイミング})$ をx軸に描いた図として、指数の分布とワイブル分布の場合、仮定が適切であれば同図は直線に近くなる。他の分布を評価するには、修正されたログ-累積のハザード図を用いることができる。
- さらに、カプラン・マイヤーを適用したログ-累積のハザード図は、ハザードがどのタイミング後から変わるか、比例ハザードモデルが可能か、どの分布を適用するか、piecewise生存曲線を適用するかに関する情報を提供することができる。
- 上記にて、記述した方法は資料がある区間にて、処置の効果を評価するのに適切な方法である。観察期間後は、資料がないため、上記のような方法をもって適切だと判断したモデルが実際は、適切でない可能性がかなり高く存在する。
- 一つのモデルにて、群の変数を含めて、両群の効果を推定したり、比較したりすることが一般的である。これは、比例ハザードなどのモデルの仮定が満たされる場合に適切である。しかし、モデルにて比例ハザードなどの仮定が適切でない場合は、各群別に適切な分布の仮定が満たされる生存モデルを適用し、生存曲線及び効果を推定することが必要な場合もある。
- さらに、観察期間中に適用したモデルにおける群間の差を用いて、追跡観察後を外挿するため、両群の差が継続的に反映され、全体的な効果の差が過大推定される可能性が高い。これは、保守的なモデルを選択しなければならない根拠である。
- 従って、適切な外挿を選ぶ際は、様々な方法で評価しなければならず、その結果は臨床的な推論に適合しなければならない。
- 様々なモデルを適用後、その中で最も適切で、相対的に保守的なモデルを選択し、基本分析として用いなければならない。
- 不確実性を評価するために、複数の違う仮定における結果も敏感度の分析を用いて提示しなければならない。

もし、処置群と対照群の生存曲線が追跡観察の期間中に、逆交差したり合致したりする場合は、そのタイミング後の群間の効果は同様であると仮定することが合理的である。

- この場合、一つの問題は臨床試験の観察期間が長いため、両群の生存曲線が交差することを確認する場合には、そのタイミング後の効果差がないと外挿するが、もし、臨床試験の観察期間が短い場合は、その期間中の効果の差が観察期間の長い臨床試験と同様であり、その後合致する可能性が高いが、資料の不在により、長期的な効果があるもので外挿することができる。この場合は、短期間で観察した臨床試験を資料元として用いる場合の効果が過大推定されかねないため、注意が必要である。
- 場合によっては、観察期間中は臨床試験の資料を用いて、観察期間後は外部の過去（historical）資料元を用いる場合がある。この場合は、資料元の妥当性、即ち、外部の資料元がどれだけ現実を反映するか、臨床試験の資料元とは差がないかなどに関し、詳しく評価する必要がある。

4. 割引率

1) 現在の指針

ア) 主な指針

未来に発生する費用と結果を現在の価値で評価するため、費用と結果共に、年間5%の比率で割引する。併せて、費用と結果共に、割引しないケース(0%)と3%、7.5%の割引率を適用したケースを敏感度の分析を用いて提示する。

ナ) 指針における追加説明

... 割引率(discount rate)の水準に関しては、国別に少し差があるが、これは各国における経済的な状況、そして公共投資部門における経済性分析に用いる告知された割引率値が存在するか、その値がいくらかにより変わる。保健医療技術の評価に関する複数のガイドラインでは、5%を基本的な割引率として提示している。割引が結果へ与える影響を調べるため、そして、後で他の割引率を適用した場合と比較するときに備えて割引を行っていない場合と3%の割引率を適用した場合を敏感度分析するようにしている。

韓国の場合、公共投資部門における妥当性の分析には7~13%の割引率が適用されてきた。しかし、限界生産性と消費利息率の下落により、最近では適正割引率が下がっているところ、Lee Gyubangなど(2001)の研究によると、1999年と2000年の社会的な平均割引率は7.3%レベルであり、今後数年間は7~8%のレベルを適用することが望ましいと述べた。しかし、このような割引率は一般的な公共投資事業の妥当性の評価に適用する割引率として、水資源の開発事業、環境施設の事業、停滞地域の開発事業など、間接的な便益が大きい、政策的な配慮が必要な事業に関しては、他の割引率を適用することができる。Lee Gyubangなど(2001)は、こうした場合、一般公共投資事業の妥当性の評価に適用する割引率より2%程度低い割引率を適用することが望ましいと指摘した。

2006年、経済性評価指針の初版を発行して以来、10年間で5%の同一な割引率を適用してきたが、国内の経済状況が変化し、社会的な割引率が継続的に下方修正され、経済性評価に関する指針の割引率も引き下げの必要性があるという主張が提起され続けてきた。これを受け、割引率をどの程度に下げするか、また、過去のように保健医療分野に関する政策的な配慮を受け、社会的な割引率より低い割引率を適用するかに関する検討が必要である。

2) 既存の提出資料に関する検討結果

既存、審評院に提出された50件の経済性評価資料を調べると、割引を行った場合が45件（90%）、割引を行わなかった場合が5件（10%）であったが、後者の場合は分析期間が1年内であると把握される。

割引を行った45件にて、割引率に関する敏感度分析の実行可否を調べると、1件を除いた44件にて敏感度の分析を行った。

割引率に関する敏感度分析の際、用いられた割引率の種類はかなりさまざまであったが、まず経済性評価の指針による割引率（0.3、7.5%）を適用したケースが29件（65.9%）で最も多く、次は0%、7.5%を適用したケースが5件（11.4%）、その他に指針に提示されていない割引率を適用したケースも10件（22.7%）あった。

一方、一般的な予想のように低い割引率を適用するほどICERが下がる傾向が表れたが給与可否の決定に影響を与えるほどの大きな変化ではないと把握された。低い割引率を適用した場合に、ICERが逆に高まる場合も1件あった。

<表 46> 既に、提出された経済性評価の資料における割引適用の現況

区分		頻度	区分		頻度
割引	実行	45	敏感度の分析	頻度	44
	未実行	5		頻度	1
	計	50		頻度	45

<表 47> 既に、提出された経済性評価の資料における敏感度の分析に適用された割引率の現況

区分	頻度(%)	区分	頻度(%)
0%, 3%, 7.5%	29 (65.9)	0%, 7%	2 (4.5)
0%, 7.5%	5 (11.4)	2.5%, 7.5%	1 (2.3)
3.5%, 7%	4 (9.1)	3.5%, 6.5%	1 (2.3)
3%, 7%	2 (4.5)	計	44 (100)

3) 外国の指針

ア) カナダ・CADTH

2017年に改正されたカナダの指針（4版）は、2006年の指針（3版）以来、11年ぶりに改正されたものとして、ここではまず、社会的な割引率に関する定義を下しているが、これは政府の投資に割り当てられたリソースの限界、社会的な機会費用（marginal social opportunity cost）を反映しなければならず、政府の債権の実質的な利率同様だという。

これを根拠として経済性評価の際、費用は州政府の債権の実質的な利息率を用いて割引し、結果は予算の制約が存在するという仮定の下で、実質的な利息率から費用-効果の閾値（cost-effectiveness threshold）の増加率を引いた割引率を適用することが理論的に妥当であるが、費用-効果の閾値に関する不確実性及び見込まれる変化に関する不確実性により、実際は費用と結果共に同一な割引率を適用することを勧めている。

具体的に、基本分析では1年後から発生する費用と結果共に、年間で1.5%の割引率を適用し、現在の価値に換算するが、この際、割引率1.5%は州政府の債権の名目利息率にてカナダ中央銀行の目標物価上昇率（現在、年間2%）を補正した値である。

不確実性の反映と割引率の変化による影響を把握するため、0%と3%の割引率を用いた敏感度の分析を行う必要があるとし、指針改正（4版）前の分析と比較する目的の場合、必要に応じて既存の割引率（5%）を適用した敏感度の分析を追加的に行いかねないとする。一方、時間により可変的な割引率（non-constant discount rate）を適用する場合に及ぼす影響を調べる場合、敏感度の分析に含める可能性もあるとする。

ナ) オーストラリア・PBAC

オーストラリアの指針では、基本分析として費用と健康の結果共に、年間5%の割引率を適用しなければならない。敏感度分析の場合、費用と結果共に3.5%及び0%の割引率を適用して提示するようにしている。

必要に応じては、基本分析の割引率ではない他の割引方法（例：他の固定割引率、期間による差等割引率の適用など）を用いて、追加分析の結果を提示することができ、この際用いた方法に関する根拠を提示するようにしている。

ダ) イギリス・NICE

費用の効果性は、数年間にわたり発生する費用と健康結果の現在の価値を反映しなければならないが、費用及び健康結果に関し、3.5%の割引率を同一に適用しなければならないという。

併せて、基本分析と共に、費用と結果共に1.5%の割引率を適用した敏感度分析の結果を提出することもあるとするという。とりわけ、該当の治療がなかったら、死亡したり深刻な健康喪失を経験したりしたはずの患者の健康が回復され、同健康結果が少なくとも30年以上続く場合は、費用と結果に1.5%の割引率を適用した敏感度分析の結果が考慮されることがあるとする。

ラ) その他の指針

ドイツの場合、費用に関しては3%、便益に関して1.5%の割引率を適用し、割引率に関する敏感度分析のために、他のシナリオも提示しなければならない。

フランスの場合は2005年後に定まった4%の社会的な割引率を費用と効果に関して同一に適用、分析期間が30年以上である場合は2%の割引率を適用する。割引率に関する評価結果の強健性の確認のために、敏感度分析が必要である。

デンマークの場合は、財務部にて定めた社会的な割引率である4%を費用と効果に関し、同様に適用する。

ノルウェーの場合も、財務部にて定めた社会的な割引率4%を費用と健康結果に同様に適用する。併せて、同割引率は実質的な利息率であるため、医薬品の経済性評価に物価上昇率を適用してはならないとされている。

スウェーデンの場合は、費用と効果に関し、3%の割引率を同様に適用しなければならない。敏感度分析では、費用と効果に同様に0%と5%の割引率を適用した場合及び費用は3%、効果は0%の割引

率を適用した場合に関しても提示するようになっている。

4) その他の文献考察の結果

□ 民間投資事業の収益率に関する研究 (Lee Gyubangなど、2001)

Lee Gyubangなど (2001) は、SOC施設に関する経済的な妥当性の分析及び投資効果の評価に必要な社会的な割引率を算出するための研究を行った結果、今後、数年間は7~8%の割引率を適用することが望ましいと判断し、道路、鉄道、港湾などの一般的な公共投資事業には7.5%の割引率を適用し、水資源の開発事業、環境施設事業、停滞地域の開発事業など、間接的な便益が大きい政策的な配慮が必要な事業に関しては2%p程度低い割引率を適用することが望ましいとした。

□ KDI公共投資管理センター (2018) にて発行した「公企業・準政府機関による事業の予備妥当性調査を実行するための一般指針の修正研究 (第2版)」のレポートによると、社会的な割引率は公共投資事業の経済的な妥当性の評価に影響を与える重要なパラメータの一つと規定している。1999年、政府の予備妥当性調査制度の導入により、資本の潜在価格に基づき、社会的な割引率を定める方法を選んだが、実質的には7.5%の割引率を適用し始めた (水資源部門の場合、長期的に考慮し 6%)。以来、経済環境の変化により、2004年の社会的な割引率は6.5%と下がり (水資源部門は30年まで6.5%、以後の20年は5%と差等適用)、2007年に再び5.5%と下げた。

□ 適正な社会的な割引率の推定のために、基準利息率、社会的な時間選好率 (social time preference rate)、財務的な割引率などを考慮するが、社会的な時間選好率を推定式により計算した結果、5.0~5.5%が適正な範囲であることが分かった。基準理想率を調べても同範囲に収まり、財務性分析の実質的な割引率も4.5%が適正であると分析された。

□ 2017年8月、企画財政部 (企財部) で予備妥当性調査制度を改変した際、市場金利及び経済成長率の下方傾向など、経済環境の変化を反映し、社会的な割引率を4.5%と下方修正して提示し、これにより公共機関の予備妥当性調査にて、社会的な割引率4.5%を同様に適用するようになっている。

□ 公共投資事業に関する経済性の分析に向けた社会的な割引率の推定 (Choi Jieunなど、2015)

2008年、リーマンショック以来、市場利息率の継続的な下落により、公共投資事業における調達金利も下落したということに着目し、Ramsey方法を適用し時間選好率、消費の限界効用弾力性、1人当たりの実質消費の年間増加率を推する方法を用いて、公共投資事業の社会的な割引率を推定した結果、複数の仮定により、3.3~4.5%の値が算出され、既存の社会的な割引率5.5%の追加的な値下げへの必要性を提案した。

□ Discounting and the Evaluation of Health Care Programs (Paulden M., 2017)

カナダにて、指針の改正に向け、割引率に関し検討したレポート (Paulden, 2017) では、経済性評価指針 (4版) を改正する際、割引率を5%から1.5%に下げることとした。これは、州政府の債権

(provincial government bonds) の実質的な利息率を反映したものと、1.5%は州政府の債権の名目利息率からBank of Canadaが提示した物価上昇率の目標値である2%を引いた値である。敏感度を分析するため、0%、3%の割引率を用い、改正前の割引率である5%は前の研究と比較する目的として使用可能であるとし、時間により可変的な割引率 (time-varying discount rate) の適用は敏感度分析で提示可能とする。

5) 利害関係者の意見

□ 経済性評価制度改善のTF

(製薬会社の意見) 韓国における割引率も、社会的な割引率の下方修正と諸外国における保健医療部門の割引率の下方修正を考慮し、現在より低く調整しなければならないと判断する。外国の割引率を参考する場合、1.5~5%まで可能であり、社会的な割引率のレベルを参考して低く設定する場合、2.5%前後の割引率も可能と判断される。

(審評院の意見) 韓国の場合、2006年に初めて「医薬品の経済性評価 指針」を制定して以来、約10年間同一な割引率を適用してきたが、その間、国内の経済状況が変化し、社会的な割引率が継続的に下方修正された。現行、経済性評価指針の割引率は国内の社会的な割引率より高く、調整の必要性はある。ただ、既存に政策的な配慮として社会的な割引率比2%程度低い割引率を経済性評価の割引率に適用したが、社会的な割引率が下がっている状況にて同様な比率を適用するには不確実性がある。なお、諸外国の事例を参考する場合、複数の国にて社会的な割引率比経済性評価の割引率を低く設定しているが、一部では同様に設定しており、政策的に配慮する方法においても差があり、国内に適用可能な最も妥当な方法に関し十分な議論が必要である。

□ Scoping諮問会議の結果

割引率のレベルに関し、専門家の中に異見が存在した。

- KDI (韓国開発院 : Korea Development Institute) の割引率 (4.5%) と同様に設定すべきかに関する検討必要
- KDIと割引率を同様にすべきかに関し疑問がある
- 米国の事例を見ると、保健医療分野の割引率が、他の分野の割引率より高い。KDIが4.5であれば、我々は5%を維持することが適正かもしれない。

費用と結果に同様な割引率を適用することに関しても、提案された意見は下記のようなものである。

- 費用と効果の割引率を違って設定するのは、考慮しないかが気になる。
- 現在のmajorityは同様であるべきだということ。

□ Scopingの意見調査の結果

製薬会社の回答者 (9人) において、殆ど (7人) が割引率の引き下げを提案、殆どが3%台の引き下げを提案する中、3.5%もしくは3%未満を提示した場合もあった。3%に引き下げの場合、敏感

度分析のための割引率は0%、1.5%、5%を提案した。

研究者（3人共に経済性評価を実行した経験あり）の全員が割引率の引き下げを提案したが、予備妥当性調査の割引率（4.5%）と同レベルもしくは3%台へと引き下げを提案した場合と現実化の必要性のみ提案し、具体的な数値は提示しない場合もあった。

□ 専門家による家諮問会の結果

割引率の引き下げは必要であり、企財部の予備妥当性調査に関する指針のような割引率（4.5%）を適用するが、分野別の差等適用はしないことを提案した。

- 実質金利が下がる傾向であるため、割引率も下方修正が必要であり、割引率が高い場合は、未来の世代に得になる仲裁らの価値が小さく評価されかねないため、割引率の下方修正が必要である。

- 分野別に投資収益率、未来へのリスクが異なるため、分野別の割引率を違って設定することも可能であるが、KDIにて分野別の割引率を認めないことを考慮する場合、企財部（企画財政部）の予備妥当性の調査指針に合わせる事が実務における混乱を避けるはずである。

○ **割引を行う際、費用と結果に同様な割引率を適用しなければならない。**

- 費用と効果に割引率を違って適用するのが妥当だという外国の研究があるが、韓国は全般的に企財部の予備妥当性調査の指針（便益と費用に関し、違う割引率を認めない）に従うことにしたため、違って適用すると、論理的な矛盾が発生する。

○ **分析期間により、割引率に差等を設けるかに関しては、企財部の予備妥当性調査の指針に従い、30年以後から3.5%を適用することも可能とみられる。**

- 一般的に長期利息率が短期利息率より低いため、長・短期の割引率を違って適用してもよさそうである。
- 分析期間が30年以上の鉄道と水資源事業は、運営30年間は4.5%を適用し、以後は3.5%の割引率を適用する企財部の指針の連動することがよさそうだ。

6) 指針の改正方向に関する提言

社会経済的な環境の変化により、社会的な割引率が継続的に下がりつつある状況にて、経済性評価のための割引率も引き下げる必要がある。ただ、割引率をどのレベルまで下げるかも保健医療分野の特殊性を考慮するかに関する合意が必要である。

今度の指針改正では、経済性評価に適用する適正なレベルの割引率を予備妥当性調査に適用される割引率と同様な4.5%レベルにすることを提案する。初めて、経済性評価指針が制定された当時は、社会的な割引率が7.5%とかなり高かった状況にて、保健医療分野に関する政策的な配慮及び外国の割引率のレベルなどを考慮し、5%の割引率の適用を提案したことがある。

これまで、社会的な割引率が下がり続け4.5%まで下がった状況にて、保健医療分野に関する政策的な配慮の観点で追加的に低い割引率を適用することより予備妥当性調査にて適用する割引率同様にすることが政策の一貫性の面でより妥当だと判断される。併せて、外国の割引率の範囲を考慮する場合も適正レベルと判断される。

費用と結果に同様な割引率を適用することに関しては、以前と同じように同様に適用し、敏感度分析のためには割引しない場合（0%）と3%割引率を適用することを提案する。分析期間により、割引率を差等適用することに関しては、二つのケースがあり得る。一つは予備妥当性調査の指針に従い、30年以後から3.5%を適用することであり、もう一つは分析期間と関係なく4.5%を適用することである。前者は結果が長期的に現れる場合、とりわけ、世代間の選択問題に関する場合に未来世代の便益をより反映しなければならないという立場であるが、後者は経済性評価の主な適用対象が医薬品であり、その便益が殆ど当事者に発生するということを挙げ、期間による差等割引率の適用が妥当では二という立場である。

保健医療分野の経済性評価対象の分析期間が30年以上の長期である場合、基本分析では全期間で4.5%の割引率を適用することにし、子供に関する治療のようなケースに必要なだと判断する場合、30年以後から3.5%の割引率を適用することを敏感度分析にて提示できるはずだ。

カ. モデルの構築

1) 改正が必要な領域

経済性評価制度改善のTF、関連した団体を対象にした意見調査の結果、特別な意見はなく、専門家による諮問会の結果で妥当性の評価を補う意見があった。

2) 現在の指針

各治療代案の最終結果を直接観察できない場合、モデルを構築してこれを推定する。モデルを用いる場合、モデルが必要な理由に関して記述し、モデルの構造に関する下記事項に関し具体的に記述するようにする。

- モデルが適用される人口集団
- モデルの種類
- モデルに用いられた仮定
- モデルの構造（疾病状態、分析期間、周期期間）
- 移転確率などのモデルの母数（parameter）
- 費用・結果
- さらに、検討射干モデルの構造と分析過程を一目で把握できるように、表もしくはグラフをもってモデルの構築結果を提示し、各ステップ別に、該当疾病状態にある人々の数（コホートシミュレーションの場合）、費用などを記述するようにする。
- モデルの構築過程にて用いた資料元に関しては、透明に記述するようにし、外国にて開発されたモデルを用いる場合は同モデルを国内の状況においても適用できるか検討し、これをレポートに記述するようにする。

以前の指針にて、モデルの構造（structure）、資料（data）及び透明性（transparency）に関する勧告事項はあったが、妥当性の部分は修正が必要だと思われる。

3) 既存の提出資料の検討結果

既存、審評院に提出された資料50件の中で、費用の最小化分析を行った2件の他、全体的に提出されたレポートの中で、80%以上がマルコフモデルを用い、dynamic modelingもしくはDESなどはなかった。また、全体の68.7%が外的な妥当性など、validationを行わなかった。妥当性はモデルの信頼度及び質においてかなり重要（Eddy et al, 2012; Afzali & Karnon, 2011; Sculpher et al, 2001; Phillips et al, 2001）にもかかわらず、大体行われなかった。国内にて経済性評価の目的で、提出されたモデルにおいて、既存の経済性評価に関する文献考察をお粉なかったのは約73%であり、外的な妥当性及びcross-validationは殆ど行われず、改正が必要である。

<表 48> 提出されたモデルのタイプ

区分	件数	%
マルコフ	40	83.3
決定木	3	6.3
Microsimulation (1st order monte carlo simulation)	2	4.2
その他	2	4.2
コメントなし	1	2.1
総	48	100
	*	

*費用最小化分析の2件を除く; その他= partitioned survival model,

UKPDS

<表 49> 既存のモデルを考察

考察方法	件数	%
既存文献の選択的な言及	11	22.9
体系的な文献考察	2	4.2
コメントなし	35	72.9
総	48*	100

*費用最小化分析の2件を除く

<表 50> 外的妥当性の評価

区分	件数	%
O*	15	31.
X	33	68.7
総	48	100

*専門家の諮問、他の資料元の比較; 資料元の妥当性、独立性は大体ノーコメント.

<表 51> cross-validation

考察方法	件数	%
O	5	10.4
X	43	89.6
総	48	100

4) 外国の指針

ア) カナダ (CADTH 2017)

モデルの構造は臨床的に適切でなければならず、必要に応じては臨床専門家の意見を受けなければならない。また、現在決めなければならない質問に適切であるべきであり、モデルの構造は現在使用した資料により、モデルの方向が決まっていけない。しかし、使用した資料はモデルの開発過程の制限により、モデルの構造が修正されなければならないこともある。妥当性が検証され、上手く構成されたモデルが既にある場合は同モデルを考慮してモデルを構築しなければならない。即ち、既存のモデル開発者との接触及び協力 (collaborate) もしくは単純に既存のモデルと似ている構造的な仮定、もしくは類似な natural history 資料 (関連があれば) を用いることができる。既存のモデルを認知し、モデルを開発するのはモデルの妥当性などに役立つ。(既存のモデルに関する文献考察の必要性を提示)

モデルをどこまで詳しく構成することが望ましいかを決めるのはかなり難しい (Roberts et al., 2012)。臨床 pathway に関する全ての詳細な内容を含む必要はないが、治療群と対照群にて、費用と効果における差が算出されかねないレベルは含める必要がある。しかし、必要以上に複雑な必要はない。様々なモデル構造があり、コホートモデル (マルコフ)、患者単位の以前のモデル (マイクロシミュレーションもしくは 1st-order)、dynamic model、discrete event simulation モデルなどがある。モデル構造の選択に関する根拠が提示される必要があり、選択したモデルは方法論的に妥当かつ透明でなければならず、good modeling practice guideline に従う必要がある。(Roberts et al., 2012; Eddy et al., 2012; Siebert et al., 2012; Caro et al., 2012; Briggs et al., 2012)。資料の提出者は、潜在的な健康状態の異質性及び dynamic evolution を共に考慮しなければならない。

□ 母数値を含む (incorporating parameter estimate)

Natural history: baseline characteristics 과 효과값이 같은 자료원이라면, 사용된 자료원이 Natural history : baseline characteristics と効果値が同様な資料元であれば、用いられた資料元に代表性があるかを考慮すべきである (Dias et al., 2013)。とりわけ、baseline の資料はカナダの

セッティングに当たる値でなければならない。もし、baseline 値が臨床値と違う値であり、カナダの値を代表することより適切であれば (例えば、請求資料もしくはコホートの観察研究)、baseline 値として別途の値を用いることも合理的な可能性がある。もし、一つ以上の資料元が適切な場合はシナリオ分析などを行うことができる。もし、国内 (カナダ) の値がなく、外国にて観察された値とかなり違うと思われる根拠がある場合、シナリオの分析を通じてその影響を評価することができる。資料元 (例えば、専門家の意見など) に関する説明とその根拠を提示すべきである。natural history 資料が不完全な場合、シナリオ分析などを通じ、後続的な不確実性 (structural uncertainty) を探索する。効果値を含む: baseline 比、相対値 (Dias et al., 2013) もしくは絶対値 (Cooper et al., 2015) をもって考慮することができる。もし、効果値が時間により変わる場合はそれに関し補正する必要がある。

副作用を含む: モデルに含まれた副作用値がどのように算出され、含める場合はどのような方法を用いたかを説明すべきである。副作用が健康結果や医療利用・費用に些細な影響を与える場合、副作

用は含めないほうが望ましい可能性がある。もし、治療に関する副作用の費用、有病率、disutility値を提示すると透明性は高まる。

治療に無関係の健康結果を含むことに関する可否：まだ議論している。病気が根治され、長生きする中で生じる副作用や健康結果を含むことの可否はまだ議論中である。無関係の健康結果は、別途のシナリオ分析により、提出できる（may）。

□ 専門家意見の算出

専門家の意見算出は、1) 資料が足りない部分に関する把握、2) 代表性のある専門家集団の把握、3) 算出法の適用、4) 専門家から算出された資料の合成という段階に分けることができる（Bojke et al., 2010; Bojke et al., 2014）。

研究者は、適切な資料元を下がるため、継続的に努力する必要がある。

□ モデルcalibration (Karnon et al., 2011; Stout et al., 2009)

モデルに用いられていない資料元とモデルの結果値を比較し、未知の母数を測定する方法と定義される。12) 現在、用いた資料にて足りない部分に関し、calibrationを通じて測定することができる。calibration方法は、定期的に用いられないため、研究者は適切な資料元を探すために努力する必要がある。

□ モデルの妥当性検証 (model validation)

face validityは、モデルの構造、仮定、母数などが臨床現実をよく反映するかに関する定性的な評価である。大体、モデルの開発初期に行われ、iterativeに継続的に評価される。

内的な妥当性（internal validityもしくはinternal verification）は数学的な計算と母数の算出に関する質の評価である。モデルの開発に直接関与しない独立的な人が数学的なlogicの合理性とエラーに関する確認により、各equationとコーディングの正確度、母数が元の資料をよく反映していることを含む。コーディングの正確度はdouble programming（二人の独立的なプログラマーがモデルの一部分を構築）、0もしくは極値を入力しモデルの検査、結果値が知られた通りのシナ

7) the estimation of unknown model parameters by achieving agreement between the model outputs and other sources of data that are external to the model and not used to parameterize the model

リオで検査などを含む。直接観察と反対の検査結果は内的な妥当性に問題がある可能性を示し得る。類似な他のモデルがあれば、研究者は先行モデルと比較する必要がある（cross-validation）。しかし、モデル間の類似な構造、方法、資料元を用いたものでなければcross-validationは意味がない。モデルの予測値が実際値（信頼でき、なるべくモデルに用いられていない独立的な資料元）と一貫しているかを考察する外的妥当性（external validity）の検証を行われなければならない（Eddy et al., 2012）。もし、資料元とモデルの予測値において、かなりの差があると、信頼度、一貫性などの面で説明しなければならない。

妥当性の検証は、具体的にどの方法を用いたかを資料で詳しく提示し、同モデルを検討したり用いたりする人が信頼できるようにする。

4) オーストラリア（PBAC 2016）

□ 母数のモデルにを含むことに関する事項

治療群別に、効果が異なる転移確率の場合、一般的に臨床資料から算出される。それぞれの転移確率値は、点推定値及び範囲（95%の信頼区間）をISPOR TF 2012（Briggs et al., 2012）を参考して算出することを勧告する。

□ 外挿

患者数が少ないことで、観察資料の数値に関する信頼度が下がる地点の前（before）までは、実際の観察値がモデルにて推定された値より選好される。この際、外挿された値が適用されたタイミングに関し、外部の資料を用いて正当化し、patient-level資料を活用して適切なparametric survival curveを算出する方法が勧告される。proportional hazardの仮定が満たされたか、様々な生存モデルを適用してモデルのfitは視覚的及びAIC/BICで評価する。Bagust et al.（2014）、Grieve et al.（2013）、Karnon et al.（2011）、Latimer et al.（2013）の論文を参考する。

□ モデルのoperational validation（主に、妥当性の検証に間し提示）

□ モデルにて治療群・比較群のtracesを提示するようにし、外的な妥当性を評価する。traceを用いてモデルの論理が適切かを評価し、モデルに用いることができない既存の研究値と比較する。既存の他のモデルとのcross-validationを行うことにする。

□ AdvISHE（Verner et al., 2016）を用いて妥当性を評価する。

㉔) イギリス (NICE 2013)

□ モデル方法

モデルに関する仮定と根拠を明確に提示し、仮定に関する現実的な代案があれば、敏感度分析を用いて、その効果を評価する。モデルが必要な事例としては、1) 一つの臨床試験から全ての根拠資料を得られない場合 2) 臨床試験の患者群が全体の患者群を表せない場合 3) 中間指標が用いられた場合 4) 適切でない比較群が選定された場合 5) 観察期間を過ぎる場合などと提示する。治療効果値は、体系的な文献考察に基づかなければならず、構造的な仮定は完全に (fully) 正当化する必要があり、input資料は明確に根拠を提示しなければならない。

モデルの開発に用いられた方法とモデルの妥当性検証に関する結果値を提示する必要がある。結果値は、各部分別に分けて (disaggregated) 提示し、生命延長 (Life year gained)、死亡率、臨床結果指標の頻度などは表として提示しなければならない (should)。

最終指標が中間指標より有意義であり、最終指標を使えない場合、中間指標を用いるが、最終指標との関係を明示する必要がある。モデル構築のための資料元を探し、批評的に評価する (critically appraise) 過程は、分析対象の人口集団と比較し、その適切性を説明しなければならない (justified)。

観察期間を過ぎて外挿する場合、内的・外的な妥当性を検証しなければならない。外的妥当性は、既存の力学資料と外挿された資料の比較と、臨床上・生物学的な現実性 (plausibility) を考慮することなどを含む。観察された値と外挿された値の相対的なfitを比較し、内的妥当性を評価する場合、統計的な方法の限界も共に記録しなければならない。治療効果は、外挿期間中、1) なかったり2) 続いたり3) 減ったりするなどのシナリオも一緒に考慮すべきである。

臨床試験にて、交差投薬 (cross-over) され、ITT分析がこれ以上、適切でない場合、switchingに関する統計的な補正値を提示することができる。交差投薬された患者のみ除き、分析する単純補正は避けて、選択された方法の相対的な長・短所に関しては、その選択の根拠に関して提示する必要がある。考慮しなければならない要因としては、臨床試験にて交差投薬が行われた機転、baselineにおける分析可能な資料の現況、時間依存的な要因、患者がずっと残っている際、期待できる治療効果などに基づくことがある。

㉕) EUnetHTA (2015/2019)

□ モデルのタイプ (表 52) を参考

受容可能なモデルのタイプに制限はなく (“all kinds of models accepted”)、大体decision treeとマルコフが取り上げられた。ベルギーの場合、モデルが複雑だと用いた資料が不足する可能性が高いため、モデルはなるべく簡単に作るように勧告した。

ソフトウェアのタイプとしては、オランダ（TreeAgeとMicrosoft Excelのみ、受容可能）、フィンランド（Microsoft Excel）、イギリス（Microsoft Excel、TreeAge、Winbugs or R）などを述べた。

□ **外挿**

大体、様々なシナリオをもって、敏感度を分析するようにした。ポーランドとフランスはoptimistic、pessimistic、neutralの三つの仮定をもって、外挿するように勧告、スコットランドとイギリスは効果が、1) 観察期間後なし 2) 観察期間後も持続 3) 観察期間後、継続的に減少というシナリオをもって分析するように勧告した。

□ **妥当性**

2015年に内的・外的な妥当性に関し述べ、2019年のレポートの場合はAdviSHE（Verner et al.2016）を用いた妥当性の検証を具体的に提示した。cross-validationの際、モデルがdependentな場合、その価値が下がり、モデル間の値が異なる場合はそれに関する説明を提示する必要がある。外的な妥当性の場合、independentな資料の利用、比較の可能性、妥当な結果値である必要がある。EUnetHTA（2019）：model verification（技術的な正確度）& validation（妥当性）に関し、下記のように提示した。

- 1)Face validity：モデルの構造、仮定、母数、分布、結果値が合理的であるか？
- 2)Internal validity（内的妥当性、technical verification）：モデルが適切に構築されたか？
- 3)Cross model validation（モデル間の妥当性）：独立的に開発されたが、類似な結果を予測する他のモデルと類似な値を提示するか？
- 4)External validation（外的妥当性）：モデルに用いられていない外部の資料とモデルの結果値を比較した際の結果は？

各妥当性の項目に関し、AdviSHE（Verner et al., 2016）をもって、妥当性をチェックすることを提案した。

<表 52> EUnetHTA（2015）国別のモデルに関する指針の整理

国	受容可能なモデルと求められる具体的な事項 (Acceptance of modeling and specific requirements)
Austria	Modelling accepted All kinds of models Internal and external validation should be analysed
Belgium	Modelling accepted Markov and decision trees are mentioned as major categories Extrapolation should be presented scenarios

	Internal and external validation requirements
Croatia	Modelling accepted; No further information available
Czech Republic	Modelling accepted; All kinds of models
Denmark	Modelling accepted; No further information available
England	Modelling accepted Microsoft Excel, DATA, Winbugs or R is recommended(process guide). If extrapolation, compare several alternative scenarios The methods of quality assurance used in the development of the model, and model validation should be detailed
Estonia and Latvia	Modelling accepted No further information available
Finland	Modelling accepted; No further information available
France	Modelling accepted; Many types of model can be used Effect of extrapolation tested in scenarios The ability of a model to produce results that are consistent and suited to the reality of the decision-making process is tested
Germany	Modelling accepted;All kinds of models Requirements on validation
Hungary	Modelling accepted All types of models are accepted Modelling software is not specified
Ireland	Modelling accepted Available modelling techniques including decision-tree analysis, state transition or Markov models, and discrete-event simulation. No requirement on methods for extrapolation Validation requirements exist
Italy	Modelling accepted All types of models are accepted, requirements on documentation and motivation No specific program of modelling methods recommended Models should be carefully validated
The Netherlands	Modelling accepted In general: Markov, decision trees, discrete-event simulations (template for applicants) Only models in MS Excel or TreeAge are accepted (template for applicants) Internal and external validation is required
Norway	Modelling accepted The choice of approach should be justified. Requirements on documentation of extrapolation Models should be carefully validation
Poland	Modelling accepted Markov model is accepted. Other models are not mentioned.If data in the model are extrapolated over time horizon of the primary trials, the following scenarios should be analyzed: optimistic, pessimistic and neutral. Internal and external validation are required

Portugal	Modelling accepted The following types of models are accepted: Decision trees, Markov's Model, Extended revision of literature relevant to the clinical and economic analysis of the problem. There is no specific program recommended. External validation is recommended.
Russia	Modelling accepted; No further information available
Scotland	Modelling accepted; There is no requirement on methods for extrapolation. Requirements on internal or external validation are not stated
Slovakia	Modelling accepted No further information available
Slovenia	Modelling accepted No further information available
Spain	Modelling accepted (Spanish recommendation, Osteba, AETSA, CatSalut) All types of models (only decision trees and Markov models addressed in Osteba) No additional information concerning programs or methods for extrapolation or internal/ external validation.
Sweden	Modelling accepted No further information available
Switzerland	Modelling accepted No further information available

자료원: EunetHTA 2019

Box 27: The AdViSHE tool

"Assessment of the Validation Status of Health-Economic decision models (AdViSHE) is a validation-assessment tool in which model developers report in a systematic way both on validation efforts performed and on their outcomes. Subsequently, model users can establish whether confidence in the model is justified or whether additional validation efforts should be undertaken. In this way, AdViSHE enhances transparency of the validation status of HE models and supports efficient model validation."[179]

The AdViSHE tool consists of 13 questions, divided into five parts, each covering an aspect of validation:[179]

Part A: Validation of the conceptual model (2 questions based on face validity and cross validity of conceptual model)

Part B: Input data validation (2 questions based on face validity of input data and model fit testing when input parameters are based on regression models)

Part C: Validation of the computerized model (4 questions based on: the external review of the computerized model by modelling experts; extreme value testing; testing of traces; individual sub-modules testing).

Part D: Operational validation (4 questions based on: face validity and cross validity testing of model outcome; validation against outcomes using alternative input data; validation against empirical data).

Part E: Other validation techniques (1 question to describe any other validation techniques performed).

자료원: EunetHTA 2019

<表 53> AdvISHEの評価道具 (Verner et al., 2016)

		説明	基準
A. Validation of conceptual model			
A1	Face validity	専門家にモデル概念の適切性に関し、 諮問したか?	Which experts? Why? どこまで適切である同意したか?
A2	cross validity	他のモデルと比較したか?	行わなかったら、なぜ行わなかったか?
B. Input data validation			
B1	Face validity ¹³⁾	専門家にinput dataの適切性に関し、 諮問したか?	Which experts? Why? どこまで適切だと同意したか?
B2	Model Fit testing ¹⁴⁾	もし、input dataが回帰分析に基づいたら、 統計検証を行ったか?	行ったら説明、行わなかったら、なぜ行わなかったか?
C. Validation of computerized model			
C1	外部の検討	モデル専門家に検討してもらったか?	Which experts? Why? 専門家は独立的か?行わなかったら、 なぜ行わなかったか?
C2	Extrem evaluate testing ¹⁵⁾	コーディングエラーを見つけるために、 極値を入れて行ったか?	行わなかったら、なぜ行わなかったか?
C3	Testing Traces	患者を追跡し、モデルのロジックが 合うか把握したか?	行わなかったら、なぜ行わなかったか?
C4	Unit testing ¹⁶⁾	モデルのsub module別の検査を行ったか?	検査、基準、acceptance normなどは 事前に提示されたか?
D. Operational validation			
D1	Face validity testing	モデル結果値の適切性に関し、 専門家に諮問を求めたか?	Which experts? Why? どこまで適切である同意したか?
D2	cross validation testing	類似な問題を取り扱う他のモデルの 結果と比較したか?	- 発表された結果のみ比較したか、 それとも他のモデルに直接アプローチしたか? - もし、差があったら差に関し、 説明できるか?
D3	Validation against outcomes ¹⁷⁾	他の資料元を用いた際に出た結果に 関し、妥当性を検証したか?	行わなかったら、なぜ行わなかったか?
D4	Validation against empirical data	empirical dataと比べたか? dependent/independent validation (モデル構築の使用可否により)	行わなかったら、なぜ行わなかったか?
E. Other validation technique			
E1	Other validation technique ¹⁸⁾	他の妥当性方法が用いられたか?	簡単にサマリーを提示

叶) 海外の指針に関する小結論

- 大体、モデルは大きな方向において、既存のISPOR & SMDM TF (2012) などから提示されたことと類似である。
- 臨床現実に適合し、効果値はSRを通じて提示;イギリスは唯一にcross-overに関し手述べた。モデルは透明に提示され、再現できなければならないということが一貫して述べられた (Roberts et al., 2012; Siebert et al., 2012; Eddy et al., 2012; Karnon et al., 2012; Pitman et al., 2012)。カナダは、最も広範囲にモデルに関して提示し、イギリス、オーストラリアは外挿に関しては具体的に述べた。
- 最近、妥当性に関する指針が補われたが、その具体性 (例えば、AdViSHEの適用程度; オーストラリア、EUnetHTAは勧告) に差がある。

- **AdviSHE** (Verner et al, 2016) : 文献の内容は、既存に比べて新たな内容はないが、かなり具体的に提示した。
- 指針の目的が一貫した政策の決定、現在の科学レベルを反映して現実性も考慮するとした場合、AdviSHEは政策決定者には役立つが、資料提出者 (製薬会社) にとっては負担になりかねない。妥当性の重要性 (モデルの信頼度 & 質) を考慮する場合、妥当性は重要であるため、より柔軟に提示する必要がある。
- AdviSHEのA1、B1、C1、D1に明示された「専門家から諮問を受けたか」にて、専門家はC1のモデル専門家を除いては全てが明確でなく、(例えば、臨医師? 統計学者? 経済学者?) C1、C3、D3、E1の場合は実現可能性が低い。

1) 指針の改正方向に関する提言

-
- 8) potential for bias; generalizability to the target population; availability of alternative data sources;
 - 9) Examples of tests include but are not limited to: comparing model fit parameters (R², Akaike information criterion(AIC), Bayesian information criterion (BIC)); comparing alternative model specifications (covariates, distributional assumptions); comparing alternative distributions for survival curves (Weibull, lognormal., logit); testing the numerical stability of the outcomes (sufficient number of iterations); testing the convergence of the regression model; visually testing model fit and/or regression residuals
 - 10) Examples include but are not limited to: zero and extremely high (background) mortality; extremely beneficial., extremely detrimental., or no treatment effect; zero or extremely high treatment or healthcare costs.
 - 11) turning sub-modules of the program on and off; altering global parameters;
 - 12) Alternative input data can be obtained by using different literature sources or datasets, but can also be constructed by splitting the original data set in two parts, and using one part to calculate the model outcomes and the other part to validate against.
 - 13) double programming (two model developers program components independently and/or the model is programmed in two different software packages to determine if the same results are obtained).

<表54> モデルの構築に関する改正方向

項目	現在の指針内容	改正方向
モデルの構造 (structure)	<ul style="list-style-type: none"> - 疾病の状態若しくは治癒（もしくは悪化）過程は、疾病の生物学的な特徴をよく反映しなければならず、モデル構造の根拠となった資料元及び説明を全て提示。なるべく、簡単であり透明であるべき。 - モデル構造の適切性を分析の目的に基づいて正当化すること。 - 全ての関連した仮定は根拠を提示する必要があり、十分な領域にて敏感度の分析が行われなければならない。臨床資料より長い観察期間を外挿する場合、様々なシナリオに基づき、不確実性を検討することを勧告。 	そのまま
資料 (data) の選定、分析及びを含む。	<ul style="list-style-type: none"> - モデルの目的に相応しい資料が選定されたかを提示することを勧告。 - 資料は、体系的に選定し、排除された文献の場合は根拠を提示することが望ましい。 - モデルに含まれた資料元は十分に詳しく説明し、検討者が評価できるようにすること。 - 点推定値の場合、敏感度の分析に用いられる領域を提示し、確率分布として用いる場合は分布を明示し、根拠を提示することが望ましい。 	そのまま
妥当性	- 含まれた資料元は互いに相反していないか、内的・外的な妥当性に関し、議論することが勧められる。モデルの分析対象となる人口集団を国内の給与対象人口集団に一般化できるかに関し、議論することを勧告。	内的妥当性、外的妥当性、cross-validation calibrationなどを具体的にコメント

- 現在の指針では、モデルの妥当性部分を追加し大きな方向は海外の指針同様にすることでしているが、A AdviSHEをどこまで適用するか（即ち、基本分析にするか、それとも選好するとすべきか）議論が必要である。

- validationに関し、AdviSHEなどの道具導入に関しては議論が必要（extensive）

指針の目的が現在の科学レベルを反映し、現実性も考慮するとした際、AdviSHEは資料提出者（製薬会社）に大きな負担になるが、妥当性の重要性（モデルの信頼度 & 質）を考慮する場合、AdviSHEは標準化された客観的な評価道具であるため、考慮が必要である。

- 第1案.妥当性の重要性強調（カナダの指針及びISPOR TF; Eddyなど 2012、AdviSHEなどの提出を選好（勧告ではない）

製薬会社の現実性を考慮し、今度はEddy/EUnetHTAの論文レベルに勧告し、AdviSHEは選好するとし、柔軟に適用することを提案する。下記は、EUnetHTA（2019）の原文である。

- 1) Face validity : モデルの構造、仮定、母数、分布、結果値は合理的か?
- 2) Internal validity (内的妥当性、technical verification) : モデルは適切に構築されたか?
- 3) Cross model validation (モデル間の妥当性) : 独立的に開発されたが類似な結果を予測する他のモデルと類似な値を提示するか?
- 4) External validation (外的妥当性) : モデルに用いられていない外部の資料と、モデルの結果値を比較した場合の結果は?差があれば、説明しなさい。

□ **第2案.AdviSHEの勧告**

妥当性評価の標準と基準を提示するというのは、有意義であるが、製薬会社にとっては負担になりかねない。製薬会社など、利害当事者の意見の受け容れが必要である。この際、C1、C3、D3、E1の場合、実現可能性が低いため、除いて進めることを提案する。

エ. 診断検査に伴われる薬物

1) 現在の指針

現在の指針では、診断検査に伴われる薬物に関する別途の指針は提示されていない。

2) 既存の提出資料に関する検討結果

診断検査に伴われる薬物は、診断検査を通じて選別された患者を対象に、投薬が行われる薬物を意味し、既存に経済性評価の資料を提出した50件の中では6件がこれに該当した。

最も早く検討された件は、局所進行性もしくは転移性非小細胞性肺癌患者の第2次治療剤として、FISH (fluorescence in situ hybridization; 蛍光 in situ ハイブリダイゼーション) の検査結果、未分化リンパ腫キナーゼ (ALK) の養成診断を受ける患者には申請薬剤を、陰性診断を受ける患者には、既存の治療剤を投薬する代案とFISH検査なしに全ての患者に既存の治療剤を投薬する既存の治療法を比較するケースであった (「test and treat」vs「treat all」)。経済性評価小委員会による議論の結果、検査後の治療戦略と冒頭の治療戦略を比較する基本モデルの受け容れを決めたが、FISH検査の偽陽性、偽陰性の診断による結果も経済性評価のモデルに反映することが適切だという意見が提示されたことがある。続いて、議論された件は「BRAFV600E変異が確認された手術が不可であったり転移性であったりする悪性黒色腫 (第1次治療)」を対象にしたもので、典型的な患者合わせ型モデル (診断検査過程及び遺伝子変異発現率を反映した) を構築せず、BRAF陽性患者のみを対象にしたモデルを構築後、有病率の逆数で診断検査費用を追加した。当時、製薬会社側では、患者合わせ型モデルと提示されたモデルが結果的に同様な結果をもたらすことを主張したが、経済評価小委員会 (経評小委) では、期待値は同様であっても、BRAF陽性患者のみを対象にして構築されたモデルにおいては、診断検査の敏感度、特異度などを巡る不確実性を確認できないことを指摘した。

以後も、T790M検査、PD-1検査などの薬物使用の過程にて、診断検査を伴う複数の件が提出されたが、診断検査の敏感度、特異度が反映された典型的な患者合わせ型モデルではなく、有病率の逆数で検査費用のみを追加したモデルであった。これに、経評小委では既審議事例及び検査法の敏感度、特異度に関する情報が制限的であることを考慮し、検査費用を考慮する範囲にて、提出モデルを受け容れたことがある。

3) 外国の指針

多くの国にて、診断法に関する評価委員会を別途に運営し、評価指針も別途に運営している。しかし、診断検査と薬物が互いに結合された場合に関する評価指針を提供するケースは多くなかった。当研究では、診断検査に伴う薬物に関する評価指針を述べたオーストラリアとイギリスの事例を検討した。

ア) オーストラリア

オーストラリアでは、診断記述はMSACが、医薬品はPBACが評価する。しかし、診断検査と医薬品が互いに連携された相互依存技術 (codependent technologies) は、両委員会にて一緒に検討する

方向に進んでおり、PBACのガイドラインでは総合依存技術（codependent technologies）に関し、別途の項目をもって、詳細な指針を提示している（product type 4）。

指針では、「他の技術（例、病理学的な検査もしくは映像技術）の使用により、ある医療技術（例、医薬品）の健康成果が改善される場合、両技術が総合依存的だとする」と定義し、下記のような情報を求めている。

- 直接的な根拠（Direct evidence）：「直接的な根拠」は、現在用いられる診断検査/検査戦略を受けるグループと提案された診断検査/検査戦略を受けるグループを直接的に比較する研究として、診断方法により患者の健康成果に与える影響を直接的に測定できるという長所がある。
- 連携根拠（Linked evidence）：提案された検査/検査戦略の臨床的な効果を図るために提案された検査/検査戦略と現在、用いられる検査/検査戦略間の正確度を比較した結果と別途の資料元を通じて確認した治療効果に関する根拠を連携する。
- バイオマーカー検査と医薬品間の関係を明確にすること。とりわけ、バイオマーカーのレベルにより、治療効果が変わるのか（治療効果のmodification）、予後効果のためか、それとも両方共によるものかを明確にしなければならない。総合依存技術を統合して提出した申請書（An integrated codependent submission）は医薬品が検査により確認したバイオマーカーと相互作用して患者の健康成果を改善するといことを立証しなければならない。もし、総合依存的な技術が背景的な予後効果と治療効果の修正共にに関する根拠を持っているなら、背景的な予後効果を統制した純治療効果を提示するようにする。
- 総合依存技術に関し、PBACとMSACが求める根拠を満足するために求められる情報は、全部で62項目として、各項目は、（T）、（M）、（O）と表示されている：（T）/検査、（M）/医薬品、（O）/両方ともに重複。以下、PBACが提示する具体的な指針は付録で提示する。

イ) イギリス

イギリスは、コンパニオン診断（Companion diagnostics）という表現を用いて、指針を提示している。指針によると、特定のバイオマーカーの存在可否より、技術の使用が変わることがあるため、ただ、特定の技術の治療決定をサポートする目的で、診断検査が行われる場合、診断検査による関連した費用は臨床効果、費用効果の評価に含めなければならない。そして、診断検査の費用を含めない結果を敏感度分析で提示する必要がある。必要な場合、診断検査の正確度を検討し、経済性評価に含ませるようにする。

現実では、該当の技術が臨床試験にて用いられたものとは違う代案的な診断手順も可能な可能性がある。必要であれば、代案的な検査を使用する可能性（臨床試験に用いることと診断の正確度が異なる場合がある）が対象患者群の選択と該当治療法の費用効果性に影響を及ぼす可能性についても検討する。

<表55> 診断検査に伴う薬物に関する海外の指針内容

国	指針内容の原文
---	---------

<p>オーストラリア・PBACの指針 version 5 (2016)</p>	<p>Product type 4 – Codependent technologiesでsectionの全体を割く (p144~)</p>
<p>イギリス・NICEの指針 (2013)</p>	<p>The use of a technology may be conditional on the presence or absence of a particular biomarker (for example a gene or a protein). If a diagnostic test to establish the presence or absence of this biomarker is carried out solely to support the treatment decision for the specific technology, the associated costs of the diagnostic test should be incorporated into the assessments of clinical and cost effectiveness. A sensitivity analysis should be provided without the cost of the diagnostic test. When appropriate, the diagnostic accuracy of the test for the particular biomarker of treatment efficacy should be examined and, when appropriate, incorporated in the economic evaluation.</p> <p>5.92 The appraisal will take account of any requirements of the marketing authorisation, including tests to be completed and the definition of a positive test. In clinical practice in the NHS, it may be possible that an alternative diagnostic test procedure to that used in the clinical trials of the technology is used. When appropriate, the possibility that using an alternative test (which may differ in diagnostic accuracy from that used in the clinical trials) may affect selection of the patient population for treatment and the cost effectiveness of the treatment will be highlighted in the appraisal guidance.</p> <p>5.93 It is expected that assessments of multiple companion diagnostic test options will generally be undertaken in the NICE diagnostics assessment programme. (p93-94)</p>

4) 他の文献考察の結果

診断検査を伴う薬物は時に、個人合わせ型医薬品という用語として呼ばれることもある。従って、個人合わせ型医薬品の経済性評価を巡る様々な議論は、診断検査を伴う薬物の評価にもそのまま適用できる。

大体、多く取り上げられる内容は、比較対象、診断検査の敏感度・特異度に関する情報をどのように考慮すべきか、バイオマーカーの有病率が結果に与える影響、互いに違う資料元の連携による不確実性といったものがある。

まず、ISPOR個人合わせ型医薬品SIGにて提出したレポート（Fulkner et al., 2012）では、個人合わせ型医薬品の経済性評価モデルは現実の意思決定を反映しなければならず、この際に考慮される選択は新薬Aと既存薬Bの中ではなく、「全て治療」戦略と「検査後治療」戦略における選択であることを強調した。この場合、バイオマーカーの有病率が重要となる。レポートでは、偽陽性、偽陰性率と偽陽性、偽陰性の結果もモデルにて考慮しなければならないと強調する。この他にも、著者らは、検査結果と治療戦略、そして、患者間の成果の関係が互いに違う文献から算出される場合、同過程にて不確実性が膨らみかねないため、モデルの額面妥当性（結果がmake senseなのか）、外的妥当性（結果が一般化可能で、現実の資料と一貫しているか）に関する評価が必要であり、個人合わせ型医薬品の価値を立証するためには、直接的な根拠の提示が求められるということも述べた。

Annemans et al.（2013）は、個人合わせ型医薬品の経済性評価を巡る方法論的な 이슈を 検討し、下記のように勧告した。i) 敏感度、特異度、偽陽性、偽陰性、予測度を明瞭に考慮する必要がある、ii) バイオマーカーの有病率が重要である、iii) 検査の特性は臨床的な効用に繋がらなければならない（translated）、iv) 個人合わせ型医薬品の二つの要素、即ち、診断要素と治療的な要素が連携した表れる価値は臨床試験にて検討しなければならない、v) 分析のモデルが複雑になる場合、現実の診療pathwayをより上手く反映することはできるが、資料不足、そして長期的な外挿による不確実性などにより、それほどに不確実性が膨らむ問題がある。

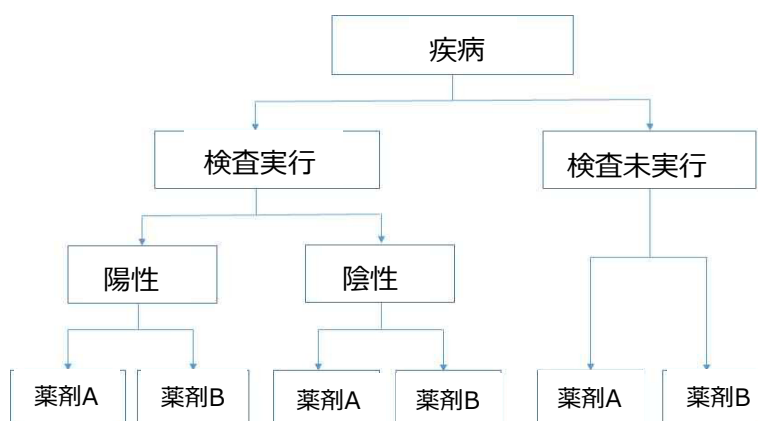
そのほか、Annemans et al.（2013）はQALYの利得が個人合わせ型医薬品の価値を記述する最上の方法なのかに関し、疑問を提起し過程がもたらす効用（process utility）、力量（capability）などに関する考慮も必要だという立場を披歴した。

Seo Hyesun・Park Mihyeなど（2014）は、既存の文献とこれまで審評院にて検討した個人合わせ型医薬品の事例を基に、個人合わせ型医薬品の経済性評価の方法論に関し、下記のようにまとめたが、同レポートから提案した内容は経済性評価小委員会の審議過程にて、一部が考慮された。

- 比較代案としては、選別検査なしに全ての患者に用いられる標準治療剤の選定
- 戦略対戦略を比較する特性上、検査の性能と費用が評価に含まれるが、該当遺伝子変異の有病率サイズ及び診断検査の価格、性能により個人合わせ型医薬品の経済評価の結果は変わることがある。なかんずく、有病率は評価の結果にかなり影響を与えかねないため、適切な有病率がモデルに反映されなければならない。
- 検査の敏感度（sensitivity）と特異度（specificity）及びこれに関する偽陽性（false positive）、偽陰性（false negative）のレベルを経済性評価のモデルにて考慮する必要がある。

- 同伴診断検査により算出された経済性評価の結果は、他の診断検査の経済性評価の結果として転用することはできない。
- 互いに違う資料元を用いて、複数の根拠を連携する場合、同資料元の内的妥当性と外的妥当性を共に考慮し、受容れの可否を判断しなければならない。
- 同伴診断検査から算出された経済性評価の結果は、他の診断検査の経済性評価の結果として転用できないということも強調した。

続いて、Seo Hyesun・Park Mihyeなど（2014）は、診断検査-個人合わせ型医薬品の望ましい経済性評価モデルとして、図8を提示した。これは、診断検査の敏感度と特異度を共に考慮するものである。



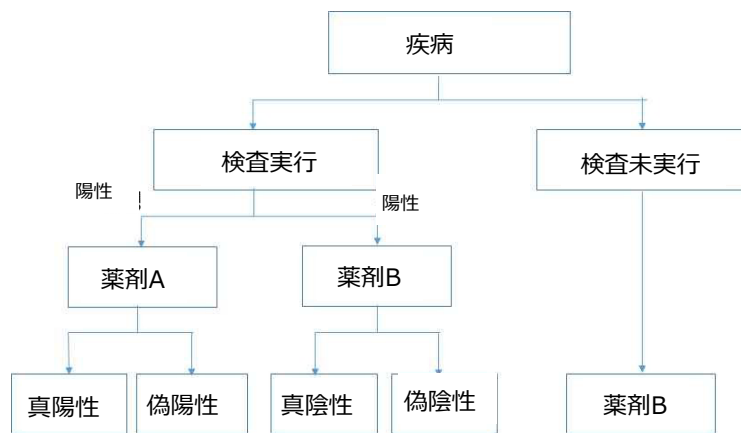
【図8】 診断-患者合わせ型医薬品の経済性評価に関するモデル

しかし、著者らは現実にて用いた資料の不在や偽陽性、偽陰性のレベルがかなり低く、複雑なモデルが不要な場合もあるため、同モデルをより単純化したものをいくつか提示した。図9の（a）は「検査しない」代案と「検査後の治療」の代案を比較したもので、現実的に偽陽性、偽陰性のレベル及びこれによる健康結果に関する情報がなければ、図8を適用することは唯ではないという判断で、図8を単純化したものである。

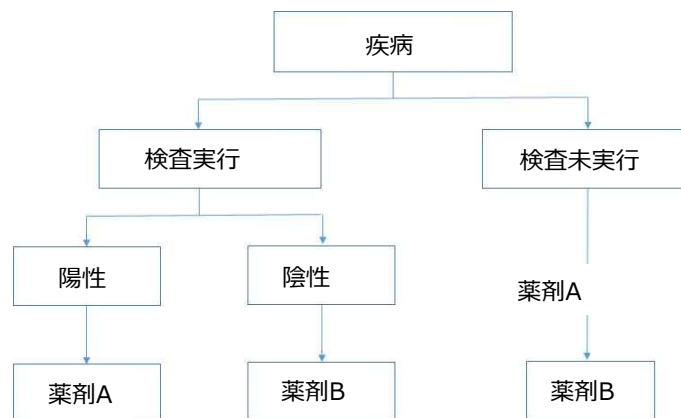
一方、（b）は偽陽性と偽陰性の比率がかなり低く、これによる健康結果及び費用への影響が無視できるような場合に用いる単純化した評価方法として提示したものである。

最後に、（c）は制限的な評価方法として、遺伝子の変異が確認された患者に個人合わせ型医薬品と標準治療剤を投薬するダインを比較する方法である。これは、同検査の臨床的な有用性と費用効果が既に立証されたということ为前提し、臨床試験は遺伝子の変異が確認された患者を対象に行われた場合である。この場合、診断検査は効果には全く影響を与えず（与えないと仮定）、費用だけに影響を与えるようになる。

即ち、モデルの個人合わせ型医薬品の代案に有病率の逆数をかけ、検査費用だけを適用する。著者らは、偽陽性、偽陰性のレベルが高いため、これによる影響を考慮する必要があるれば、同モデルを適用することは相応しくないと提案した。



(a)



(b)



(c)

[図9] 診断検査-個人合わせ型医薬品の経済性評価に関するモデル (単純化)

出典 : Seo Hyesun・Park Mihyeなど (2014)

その他、Kip et al. (2018) は、診断検査とバイオマーカーの経済性評価を行う際、含めなければならない項目を表として提示し(付録を参照)、Yang et al. (2018) も診断検査のモデル基盤の経済性評価にて確認する項目をチェックリストの形で提示した(付録を参照)。

既存の文献における勧告項目をまとめると、下記のようなものである。

-診断検査を伴う薬物の場合、比較代案は薬物対薬物ではなく治療戦略対治療戦略である。バイオマーカーの有病率を考慮しなければならない。

-診断検査の敏感度、特異度を考慮しなければならない。

- 互いに違う根拠を連携する場合、同資料元の内的な妥当性、外的な妥当性を評価しなければならず、不確実性に関する評価が必要である。

5) 利害関係者の意見

- なし

6) 指針の改正方向に関する提言

診断検査を伴うオーダーメイド型医薬品の場合、同医薬品が導入されなかったら、選別検査なしに全体の患者群に標準治療剤を用いたはずであるため、この場合は比較代案は選別前の全体患者群となる。

診断検査後、治療対象を選別する場合であれば、検査費用も含めて、申請薬剤の費用比効果を評価することが望ましい。ただ、治療対象を選別する検査に当たるとしても申請薬剤の投薬可否に関係なく、行われる検査であれば、費用から除き、申請薬剤の投薬のために、追加的に行われる検査であれば費用に含める。

バイオマーカーの有病率により、選別検査の人員比、投薬人員の比が変わるため、バイオマーカーの有病率を考慮する必要がある。

検査の感度 (sensitivity) と特異度 (specificity) 及びこれに関する偽陽性 (false positive)、偽陰性 (false negative) のレベルを経済性評価のモデルにて考慮しなければならない。ただ、偽陽性率と偽陰性率が無視できるほど、十分に低いという根拠があれば、省いて簡素化したモデルを用いることができる。

検査、治療、健康結果間の関連性に関する根拠を提示しなければならない。直接的な根拠があるのが最も望ましいが、互いに違う資料元から算出された根拠を連携するアプローチも可能である。この場合、各根拠の内的な妥当性、外的な妥当性を適切に評価しなければならない。

構造、母数を巡る不確実性を十分に検討しなければならない。

ス 不確実性

1) 改正が必要な領域

経済性評価制度改善TF、学会及び関連した団体に関する意見調査の結果や専門家による諮問会議の結果共に、これといった意見はなかった。

2) 現在の指針

経済性評価は、用いられる資料の不足と単一な方法論の不在などにより、ある程度の不確実性を持っている。従って、経済性評価を行い際は、不確実性の問題を明確に検討することにより、意思決定者の評価結果がどこまで安定的であるかを確認できるようにしなければならない。

不確実性を検討する最も基本的な手段は敏感度の分析である。経済性評価過程にて用いた全ての仮定とその他の不確実性を招く源泉に関し、単変量(一元)敏感度の分析を行うことにする。単変量敏感度の分析結果、結果指標に分館な影響を与える変数に関しては、多変量(多元)敏感度の分析を行うことを推奨する。そして、敏感度分析を行った全ての変数と敏感度分析の結果を表やグラフでまとめて提示することにより、意思決定者が不確実性の程度と様相を検討できるようにしなければならない。

最近では、費用比効果を推定するタイミング推定値のみならず、95%信頼区間を共に提示しており、模型にて用いられた母数の不確実性を検討する手段として確率的な敏感度の分析を行っているため、こうしたアプローチ方法も試みることを推奨する。さらに、費用比効果分析の結果を費用比効果の受け入れ曲線(cost-effectiveness acceptability curve)として提示することを推奨する。

3) 既存の提出資料に関する検討結果

審評院に提出された50件中、単変量分析は全ての件にて行われたが、2つ以上の変数を一緒に変化させる多変量分析は1件のみ行われ、確率的な敏感度分析(PSA)は全体にて約40%が行われた。PSAは、必ずしも行わなくても構わないにもかかわらず40%行われた。

<表 56> 不確実性評価におけるタイプ別の分類

分析の種類	件数	%
単変量分析 ¹⁹⁾	50	100
多変量分析 ²⁰⁾	1	2
確率的な敏感度の分析(PSA)	20	40
計	50	100

* 重複可能

大体、治療費用（80%）、効用（88%）にて、単変量の敏感度分析が行われたが、他の文献値（78%、治療費用；61%、効用）が最も多く用いられ+-10%などのような任意値も、約22%程度が用いられた。

<表 57> 母数の不確実性の評価（DSA）

区分*	件数	%
薬価	10	34
薬価以外の医療費用	40	80
Relative effectiveness	11	22
効用	44	88
Baseline risk	18	36
割引率	5	10
その他	29	58
計	50	100

* 重複可能

<表 58> 母数の不確実性の評価（DSAの詳細）

区分	治療費用		Relative effectiveness		効用		baseline risk	
	件数	%	件数	%	件数	%	件数	%
95%信頼区間	0	0	6	54	7	16	1	5
互いに違う文献	6	78	1	10	27	61	8	45
他の任意値*	34	22	4	36	10	23	9	50
計	40	100	11	100	44	100	18	100

* +-10%など

7) univariate analysis

8) multivariate analysis

PSAの場合、変数の特徴により大体、適切な分布（費用はgamma、効用は beta）が用いられ、使い易いが勧告されないtriangular、uniformなどは用いられたが33%未満と観察された。

<表 59> PSAに用いられた分布

区分	費用		効用		Relative effectiveness	
	件数	%	件数	%	件数	%
Lognormal	0	0	0	0	2	22
Beta	0	0	9	56	1	12
Gamma	14	82	2	13	0	0
Normal	2	12	3	19	3	33
その他*	1	6	2	12	3	33
計	15	100	16	100	9	100

* triangular, uniform分布など

国内の現況を分析した結果、計50件の中で一元敏感度の分析は大体で行われていたが、任意の範囲（+-10%）で行われたケースが多く、2つ以上の変数を変異させた多元敏感度分析は1件しかなかった（2%）。PSAは、全体の50の中で20（40%）にて行われ、大体適切な分布が用いられたが、そうではない場合（例：費用に関して正規分布、全体の12%）も観察された。構造の不確実性に関しては観察期間を違って設定した場合が最も多かった。

<表 60> 構造の不確実性に関する評価

区分*	件数	%
分析対象の人口集団	19	38
time horizonの変化	38	76
Extrapolationモデル	16	32
その他の過程を違って設計し	17	34
計	50	100

* 重複可能

4) 外国の指針

ア) ギリス（NICE、2013）

- 構造の不確実性（structural uncertainty）：シナリオ分析（scenario analysis）

モデル構造の仮定は明確に説明しなければならず、その根拠もしくは仮定を裏付けられる根拠を提示する必要がある。モデル構造の仮定に関する不確実性は現実性のあるシナリオに基づいた別途の分析を通じて、探索しなければならない。観察期間を過ぎて、外挿などの仮定に関する不確実性（他の仮定）、費用及び効果資料元の選択による不確実性（他の資料元）などの場合、シナリオ分析をもって不確実性を探索する。

□ **資料元の選択による不確実性：他の資料元をもって敏感度分析**

費用の資料元が各医療機関にて変異が大きいか、適切な効用資料元が複数あったりする場合などは、様々な資料元を用いて敏感度分析を行うことを勧告した。

□ **母数の不確実性（Parameter uncertainty）：確率的な敏感度の分析（PSA）**

適切な分布（distribution）及び母数間のcorrelationなども考慮しなければならないことを提示した。母数の不確実性の場合、cost-effectiveness plane、cost-effectiveness acceptability curve、cost-effectiveness acceptability frontierなどで、結果を表せることを提示した。一元敏感度分析、best/worst caseの敏感度分析は結果へ敏感に影響を与える変数を把握するのに役立つが、母数の数字が大きくなることにつれ、複数の資料元のcombined効果を表すのに限界が生じるため、母数の不確実性はPSAと表すように勧告した。

イ) カナダ（CADTH、2017）

□ **構造的な不確実性（structural uncertainty）**

の場合、方法論に関わる不確実性（methodological uncertainty）は scenario analysisを勧告。

他の方法を用いたnon-reference caseとreference caseを比較することを勧告し、他の仮定、費用及び効果資料元の選択による不確実性（他の資料元）などを用いて不確実性を評価するように勧告した。

□ **母数の不確実性（parameter uncertainty）：母数のtrue**

valueの不確実性に関する評価は、PSAをもって行うべきである（用語をPSAの代わりに、probabilistic analysisと変更）。

母数の不確実性は、確率的な分布で表示され不確実性はPSAをもって、評価するように勧告した。この際、費用と効果の過少/過大評価を防ぐため、母数間のcorrelationを考慮することを勧告した。

決定論的な敏感度分析（単元、多元、閾値分析など）の場合、モデルが非線型であったり母数間でcorrelationされたりする場合、歪んだ結果をもたらしかねないため、勧告しないことを明示した。不確実性はCEAC（cost-effectiveness acceptability curve）とCEAF（cost-effectiveness acceptability frontier）で表すようにした。

ウ) オーストラリア（PBAC、2016）

□ **母数の不確実性**

PSAの場合、適切な分布を用いることを提示した。即ち、以前の確率はbeta、RR/HRの場合は

log-normal、費用はgamma、効用はlog-normal distributionなどの分布を用い、他の分布を用いる場合は根拠を提示するようにした。変数間のcorrelationを考慮し、とりわけ、治療群と対照群の場合、相対的な治療効果の適用にjoint ertaintyを考慮するように勧告した。

単元敏感度の分析は、tornado diagramを提示し、敏感な変数は多元敏感度分析を行うように勧告した。多元敏感度分析の際、DSAに追加しPSAを提出（すなわち、基本はDSA）とし、イギリス、カナダと少し違う方向である。

□ **構造的な不確実性 (structural uncertainty) : scenario analysis**

複数のモデル構造が定義された場合、モデルの結果値を比較し他の構造への影響を比較するようにした。ちなみに、オーストラリアは translational uncertaintyを述べているが、仮定の変異という面では構造的な不確実性の範囲内と見なしかねない。

<表61>「translational uncertainty」の事例と不確実性分析を勧告した例

Translations Incorporated into base-case analysis	Suggested uncertainty analysis
連続型資料を2進数に変形	閾値 (thresholds) を変えて分析
処置転換の補正	他の法制方法、補正なしの治療効果
サブグループに基づいた治療効果	ITTに基づいた治療効果
治療効果の資料元を選択	他の資料元の選択若しくはメタ分析資料・資料元を選択
中間指標を最終指標に変換	現実的な値の範囲
観察期間後、資料の外挿	他の外挿地点、他の外挿モデル、今後治療交換に関する他の過程
Externally sourced utility values	Alternative values or sources (Subsection 3A.5.1)

資料元：オーストラリア・PBAC Guidelines (2016)

I) EUnetha (2019, 2015)

不確実性の概念に関して整理したISPOR/SMDMのBriggs et al. (2012) の研究のような概念である。

Stochastic uncertaintyは1st order, random variability between otherwise identical patientsで十分なiteration/simulationを用いて安定した値を算出するようにした。parameter uncertainty (2nd order)はPSA、structural uncertaintyは過程を違って居設定した際、結果値へ与える影響を評価することを整理した。なお、モデルの構築過程にて結果へ影響を与える変数を把握し、これに関して敏感度の分析を行うようにした。

<表62> 不確実性の用語及び概念の整理 (Briggs, 2012)

用語	概念	Other Terms Sometimes Employed	Analogous Concept in Regression

Stochastic uncertainty	Random variability in outcomes between identical patients 同じ患者間の結果値の変異	Variability: Monte Carlo error; first-order uncertainty	Error term
Parameter uncertainty 母数の不確実性	The uncertainty in estimation of the parameter of interest 母数の測定における不確実性	Second-order uncertainty	Standard error of the estimate
Heterogeneity 異質性	The variability between patients that can be attributed to characteristics of those patients 患者間の異質性	Variability: observed or explained heterogeneity 観察若しくは説明された異質性	Beta coefficients
Structural uncertainty 構造の不確実性	The assumptions inherent in the decision model モデルの仮定	Model uncertainty	レア分析モデルの形 (例: linear、Logなど)

□ **points for consideration (EUnetHTA 2019)**

平均値のみならず、upper/lower bounds. 値を把握することが重要であり、CI区間が狭過ぎるのは懸念事項になり得ることを提示した。適切な分布を用いて、合理的な範囲内に入るようにした。例えば、確率やutilitiesはbeta distribution、費用はgamma distribution、RR/HRの場合はlognormal distributionをコメントした。

異常値を無視せず、結果の不確実性を全体を反映できるように包括的に提示（変数や仮定を含む可否及び範囲）することを勧告し、変数間のcorrelationを考慮したか、使用した資料と仮定にてcriticalな不足分に関して議論したかの考慮が必要である。

□ **DSA vs. PSA (2015, 2019)**

殆どの国にて、DSA/PSAを共に行うようにした。EUnetHTA（2015）は、敏感度分析に関してはDSAとPSAを共に行わなければならないと提示した。2015年のレポートでは、one-way sensitivity analysis²¹⁾、multiple sensitivity analysis²²⁾ が多くの国にて勧告された。

9) Austria, Belgium, England, Finland, France, Germany, Hungary, Ireland, the Netherlands, Norway, Scotland, and Spain

10) Belgium, France, Germany, Hungary, Ireland, Norway, Scotland, and Spain

parameter uncertainty)に関しては、多くの国にてPSAを選好すると提示された²³⁾。
deterministicモデルの場合、変数間の関係を考慮せずに平均値を入れる場合、変数間に
correlationあるか、分布が偏ったケースを無視する懸念があるというイギリスの事例が
取り上げられた。

オ) 海外の指針に関する小結論

- **構造の不確実性に関してはシナリオ分析、母数の不確実性に関してはPSAがDSAより選好される傾向が観察されるが、オーストラリアはDSA（多元/一元）を選好する。**

カナダの場合、DSAに関する懸念を明示し、PSAを基本的に行うことを勧告する。イギリスは、以前のバージョンからPSAが選好されることを明示した。オーストラリアもPSAを述べたが、多元敏感度分析においては、DSAがPSAよりも優先される。一元敏感度の分析として提示されたのは、構造的な不確実性によるシナリオの分析のためと考えられる。

- **構造の不確実性 (structural uncertainty) : シナリオ分析 (scenario analysis)**

大体、モデル構造の仮定（観察期間を過ぎて外挿などの仮定に関する不確実性）、費用及び効果資料元の選択による不確実性（他の資料元）などの場合、シナリオ分析によりオーストラリア、イギリス、カナダ及びEUnetHTAにて提示され、一貫されている。

- **母数の不確実性 (Parameter uncertainty) : 確率的な敏感度分析 (PSAイギリス、カナダ) vs. DSA (オーストラリア、EUnetHTA)**

一元敏感度の分析、best/worst caseの敏感度の分析は結果に敏感に影響を与える変数を把握することに役立つが、母数の数が上がることにより複数の資料元のcombined効果を表すには限界があり、母数の不確実性はPSAと表すようにする。（イギリス）； 結論的な敏感度分析（単元、多元、閾値分析など）の場合、モデルが非線型であるか、母数間でcorrelationされた場合、歪んだ結果をもたらしかねないため、勧告されなかった（カナダ）。適切な分布（distribution）及び母数間のcorrelationなども考慮されるべきであることを提示する。

オーストラリアの場合、一元敏感度分析及びtornado diagramを提示し、敏感な変数は多元敏感度分析を行うようにする。多元敏感度分析の際はDSAに追加し、PSAを提出（即ち、基本はDSA）； PSAの考慮事項を提示した。

EUnetHTAの殆どの国にて、DSA/PSAを一緒に行うようにした。敏感度分析に関してはDSAとPSAが共に行われなければならないとEUnetHTA（2015）にて提示した。

14) Austria, Belgium, Croatia, England, Finland, France, Germany, Hungary, Ireland, Italy, The Netherlands, Norway, Poland, Scotland, Slovakia, and Spain

5) 他の文献考察の結果

<表 63> 決定論的な敏感度分析のメリット・デメリット

メリット	デメリット
<ul style="list-style-type: none"> 一元敏感度の分析は用い易く、方法が論理的であるため、CEA構造を理解し易い上、追加的な研究用データが必要かに関し判断できるような基準を提示する。(Walker and Fox-Rushby, 2001) 高質のサンプルデータがない場合、単純敏感度の分析は特定な母数に関する不確実性を説明するのに、最も効果的である。(Briggs など, 1994; NICE, 2008) 費用効果分析のkey driver (中核的な要素) は単変量の分析を用いて、明かすことができる (Walker and Fox-Rushby, 2001; NICE, 2008) 	<ul style="list-style-type: none"> 多変量分感度分析の場合、母数が増えると、解釈が困難 (NICE, 2008 : Critchfield and Willard, 1986 : Briggs など, 1994 : Walker and Fox-Rushby, 2001 : Ades など, 2006) 分析に含まれる母数は、度々任意的に選ばれるため、selection bias問題がある) O'Brien et al., 1994) ある敏感度分析の結果に関し、変動性のレベルに関する基準がなく) どの程度がrobustかに関するガイドがなく) 主観的な解釈が可能 (O'Brien など, 1994) 決定論的な敏感度の分析は、複数の母数間の相互作用と相関関係を反映し難く、母数効果を個別的に考慮するのは、全体における不確実性のレベルを過少評価し易い (Briggs and Gray, 1999; Briggs, 2000)

2012. 保険医薬品の経済性評価に関する改正方法の研究にて、再引用及び再構成

<表 64> 確率的な敏感度分析のメリット・デメリット

メリット	デメリット
<ul style="list-style-type: none"> 費用効果分析にて、母数の不確実性を把握するべく、PSAの使用を推奨 (L Andronis et al., 2009) 確率分布を用いることにより、全ての母数の不確実性を特性化することができ、平均的な費用効果の結果にて不確実性に関する解釈も可能。(Briggs et al., 1994; NICE, 2004) 母数間の相関関係がモデルにて正確に具体化され、非線型モデルにおいても、平均費用と効果に関する正確な推定値を提供。(Ades et al. 2006; NICE, 2008) 	<ul style="list-style-type: none"> 母数間の独立性の過程が一般的に成立されなければならず (Claxton et al., 2005) 母数分布の選択が妥当ではない場合、懸念があり(Claxton et al., 2005) モデル構造、方法、仮定に関する不確実性はPSAを用いて調べることが困難。(Briggs et al., 1994; Lord et al., 1999)

2012. 保険医薬品の経済性評価に関する改正方法の研究にて、再引用及び再構成

6) 針の改正方法に関する提言

□ 母数の不確実性 (Parameter uncertainty) : PSA vs. DSA

1) DSAに反対:

- 一元敏感度の分析は、全体的な不確実性を過少評価する母数の分布及び様々な変数を共に変えさせるPSAがモデルの不確実性より現実性を反映する。(Claxton et al., 2005)。²⁴⁾
- モデルが非線型か、母数がcorrelationされた場合、DSAは歪んだ結果をもたらしかねないため、勧告しない (CADTH 2016)。DSAの不確実性の範囲 (極値) は変数間のcorrelationがある場合、歪みかねない (Ades et al., 2006)。一方、PSAはjoint posterior probability distributionからMonte Carlo samplingを通じ、適切な不確実性の分析を行うことができる (Ades et al., 2006)。

2) DSAに賛成:

- 敏感度を評価する。において、一元敏感度の分析で十分であり、(Meltzer et al., 2001) 一元敏感度の分析は母数一つの結果値に関する。影響を表せるため、考慮が必要である (EunetHTA, PBAC) ; 敏感度の分析に関しては、DSAとPSA共に行わなければならないと提示した複数の国があり²⁵⁾ 母数の不確実性に関する。DSAとPSAで提示されることがある (Weinstein et al, 2003)。

3) 提言 : 母数の不確実性も、まだ決定論的な敏感度の分析をなくすには少し負担である。(EUの諸国、オーストラリア)

- 一元敏感度の分析により、変数一つの変異を表すが、任意的に +-50%などの値より有意義な (例えば、十分な範囲の (95%)) 値を利用することを提言する。
- 審評院に提出された資料の中、全体の40%にPSAが提出され、大体母数の特徴に合致する分布を用いた。多元敏感度の分析は殆ど行われず、PSAを勧告することを提案する。多元敏感度の分析ではDSAの代わりにPSAを勧告し変数間のcorrelationと分布をまともに考慮し、triangularやuniformなどの分布は除外する。ように (Briggs et al., 2012) 明示することを提案する。一元敏感度の分析に関してはトルネード・フラット、PSAに関してはCEAC / cost-effectiveness planeと提示する。ことを提案する。

□第1案 : 構造的な不確実性は、一元敏感度の分析。母数の不確実性に関しては、PSAと一元敏感度の両方が考慮される。ただ、一元敏感度の分析は臨床的かつ統計的に十分に広範囲な範囲 (例、 : 95%CI) にて行い、任意的な変異 (例えば、+-20%) は避けるように²⁶⁾、PSAも共に実行する。が、母数に適切な分布及びCorrelationを考慮して行うことを勧告する。

15) 多元敏感度分析の場合、PSAより時間がかかり、computationally intensiveである。

16) “Nevertheless, univariate sensitivity analyses are stated to be particularly informative to identify parameters which may have substantial impact on the results of economic evaluations and both the use of DSA and PSA is recommended.”, EunetHTA 2015

- 第2案：母数の不確実性に関する。評価として、PSAを基本分析と提案する。（PSAよりDSAを
選好）DSAも考慮されかねないが、臨床的かつ統計的に有意義な範囲にて行う。

17)completely arbitrary 任意的に選定された領域（+-50%）を用いて、敏感度分析を行うのは無意味（Briggs、
2012,ISPOR Consensus）

セ. 付録：専門家意見の算出方法

1) 現在の指針

【別添1】専門家の意見収集

専門家の意見に基づく場合、本指針に提示された様式(表 1)に従うことを推奨する。専門家意見の収集は、臨床試験の研究(無作為臨床試験など)結果を臨床現実に適用させる場合、外国にて行われた研究を国内の臨床現実に一般化させる場合、モデルの仮定設定、資料不足や現存資料の信頼度が低くて仮定が必要な場合に、根拠として用いられることがある。

専門家の意見は、様々な形の資料元の中、最も根拠レベルが低いと報告された。専門家の意見は、妥当な方法論による科学的な根拠に大体できず、主な臨床効果や単位費用などを算出する場合の専門家の意見だけに基づいて算出してはいけない。

専門家の意見収集が用いられた場合、使用できる資料元がないことを確認し、算出された結果値は臨床的かつ統計的に妥当な範囲内にて敏感度の分析を行い、結果の固さを観察し、用いられた領域の根拠を提示する。

<表 65> 専門家の意見収集における提出様式

提供すべき情報	説明
専門家の選定基準	当専門家若しくは専門家集団で選定した基準を提示することにする(ランダム・特定の学会・特定の地域・病院)
接触した専門家数	明示
参加した専門家数	欠席した専門家の特徴及び欠席の程度が意見の代表性に影響を与えるかを評価すること。
意見の受け容れのための方法	専門家が全員参加した会議から意見を受け容れたか、個別的に接触したか、アンケートを用いて意見を受け容れたかなど、意見受容れの方法を明示すること。
意見受容れの手段	郵便・電話・対面・インターネット
質問内容	アンケート作成の根拠を提示し、もし事前テスト (pilot test) を行ったら、その結果も一緒に提示すること。 アンケートは、申請薬剤に中立的かを評価し、根拠を提示すること。 アンケートを含め、open・closed questionなのかを提示すること。 専門家が回答した原文も提示すること。
合意を導き出した方法	Delphi・多数決・平均・中間値の根拠を提示すること。

変異の評価	一つに評価された値に関する変異を評価するため、定量的な資料の場合は標準偏差及び極値を提示し、定性的な資料の場合は外れ値（outlier）を提示することを推奨。
利害相反の可否	意見の収集に参加した専門家が申請薬剤と利害関係があるかの可否を専門家本人の虚名を含めて、記載するようにすること。（例えば、申請薬剤の開発過程に参加した経歴の有無など）これは、意見の収集に参加した専門家数が少数の場合に、より推奨。
専門家に提示された background 情報	意見の収集に参加した専門家に、申請薬剤及び比較薬剤に関し、どのような情報が提供されたか、コピー本を提出しなければならない。 もし、専門家が専門的（technical）な意見を提供したら、専門家が使った用語に関する定義を提出しなければならない（どのような意味で、使ったかを明示する必要がある）。
各質問別の回答率	特定集団の意見を代表するかの可否を評価するために、非回答者と回答者の特徴を比較する。

注：関連した質問が申請薬剤もしくは疾患にし、不適切だったり作成が困難であったりする場合は、その根拠を提示することとする。

現在の指針では、専門家に関する意見調査時に提出様式をはじめ、比較的詳細な指針が提示されている。実際、提出された資料を分析した結果に基づき、提出様式の改正必要性を検討しようとする。

2) 既存の提出資料に関する検討結果

- 既存の提出資料の中、専門家への意見調査を含めた件を検討した結果
 - 参加した専門家数は、2~22人に及んだ（K薬は1人）
 - 調査結果をまとめるにおいて、多数が平均値、もしくは受け入れの意見を用いた。semi delphi を行ったと述べた件は3件、中央値を考慮した場合は1件であった。
 - 利害相反の可否に関しては、明かした場合もあったが、全くコメントのない場合もあった。
 - 大体、専門家の意見調査に用いた調査紙そのものは提出したとみられる。

3) 外国の指針

ガ) オーストラリア

- PBACガイドラインは、付録を通じ専門家意見の算出方法に関する指針を提示している。その内容にて、関連性の高いいくつかを調べてみると、
 - 専門家の意見が他の観察資料より選好されるのではないが、観察資料を意思決定の状況に適用できるかに関する。判断には役立ち、この場合、専門家の意見は不確実性を減らしようとコメント。
 - 専門家の意見を用いる場合は調査紙や専門家らへ提示した情報のコピー本を提出すること。
 - 専門家の意見と専門家らにおける変異をまとめる際、結果の解釈と共に、選択した（算出）方法の限界とバイアスに関し議論すること。
 - 複数の処方者が存在する疾病に関し、専門家の意見を求める場合は少数の処方者を対象にした調査に頼らないこと。これは、不確実性を高める結果をもたらすはずである。専門家への調査により算出された最終的な推定値は、PBACの効果や費用-効果性の過大推定か、財政への影響を過少推定するリスクを最小化する方向に使用すること。専門家の意見に関する不確実性を減らすために算出された値を中心にした敏感度の分析結果を提示し、互いに違う推定値に分析の結果が敏感ではないことを明確に説明すること。
- 下表は、専門家の意見を収集する方法を提示した表である。

<表66> Table A1.1 Methods to collect and collate expert opinion

Information to be provided	Notes
Criteria for selecting experts	Prefer a random or comprehensive set of prescribers likely to prescribe the proposed medicine, or the appropriate medical specialty group. In general, an advisory board created by the sponsor, or for advising on the drug development program or marketing may not be representative of experts in Australian clinical practice. The generalisability of expert opinion derived from such boards is difficult to assess
Number of experts approached ^a	Where the likely number of prescribers is large, it is less acceptable to provide expert opinion derived from a small number of prescribers
Number of experts who participated ^a	Assess whether the extent and characteristics of the <u>nonresponders</u> are likely to diminish the representativeness of the opinions provided, compared with the intended sample approached
Declaration of potential conflicts of interest from each expert or medical specialty group whose opinion was sought	Provide a signed statement from each expert and specialty group specifying <u>any potential conflict of interest and stating the nature of any contractual arrangement, including how much payment was offered and accepted</u> . Where the collection of expert opinion has been <u>contracted out</u> , the contractor should provide this statement, reporting on <u>both the arrangements made between the sponsor and the contractor, and the arrangements made between the contractor and those whose opinions were sought</u>
Background information provided and its consistency with the	Include a <u>copy of any background information</u> provided in the technical document or attachment. If background information has been provided, ask the experts to define the comparative clinical place of the proposed

totality of the evidence provided in the submission	medicine and the main comparator based on this background information. Including the experts' definitions in the technical document or attachment allows an <u>assessment of the consistency of the background information with the evidence provided in the submission</u>
Method used to collect opinions	For example, were the experts <u>approached individually or was a meeting convened?</u> Was any incentive used to maximise responses?
Medium used to collect opinions	For example, was information gathered by <u>direct interview, telephone interview or self-administered questionnaire?</u>
Questions asked ^b	Explain the <u>design of the tool</u> (quantitative or qualitative). Describe its development. Indicate <u>whether it was pilot tested</u> and, if so, provide the results of that testing and explain how the results were used to improve the questions. <u>On a question-by-question basis, assess the extent to which each question is neutral or biased, and the extent to which each question is open or closed.</u> To allow an independent assessment, include the questionnaire or an outline of the interview questions in the technical document (or attach a copy)
Whether iteration was used in the collation of opinions and, if so, how it was used	The Delphi technique, for example, uses an iterative approach
Number of responses received for each question ^a	Assess whether the <u>extent of any nonresponse</u> is likely to diminish the representativeness of the opinions provided to particular questions, compared with the intended sample approached
Whether all experts agreed with each response	If not, specify (i) the <u>approach used to finalise the estimates</u> (eg the majority opinion or a Delphi technique could be applied; for quantitative results, point estimates [such as the mean, median or mode] could be presented), and (ii) the <u>approach used to present the variability</u> in the opinions (eg present the range of opinions expressed, including common and outlying views; for quantitative results, measures of variance [such as confidence intervals, range, centiles] could be presented)

a Tabulate these information items. b The way the questions are asked is an important source of potential bias in obtaining expert opinion. A particularly influential extension question extends the respondent beyond 'what' the opinion is (eg what would be done, what extent of benefit would be clinically important) to ask 'why' (eg explain why would you do this, explain why this is important). Conveying these reasons alongside expert opinion-based estimates might help improve their acceptability. Including these explanations in the technical document or attachment would allow the opinions to be assessed on the basis of the underlying reasoning rather than only depending on the authority of the experts

4) イギリス & カナダ

- イギリス & カナダ

4) 他の文献考察の結果

- 文献の考察

Iglesias et al. (2016) は、専門家による判断を専門家の意見 (expert opinion) と専門家の算出 (expert elicitation) と区分した。前者は質的な判断を引き出すことで、フォーカスグループ、名目集団技法、デルファイ調査などを例とした。後者は統計的な形で量的に表現された判断を引き出すことで、定量的な方法 (例、roulette、tertile、probabilityなど) を用いる。著者などは専門家らにおける合意過程を通じ、専門家の調査結果を報告する。基準を提示した。複数の基準を提示したが参考できるものは下記のようなものがある。

- 専門家集団に関する定義 - 彼らがどのような専門性を持っているか、なぜ調査対象になったかを提示する。一人の専門家では不十分である。該当専門分野を代表うできる人々の中で調査対象者を選定したほうがいい。
 - 専門家の数：専門家を選択する。基準と最終的に参加した専門家の数。この際、選択の基準は詳細に提示し、参加率のようなものがあれば、これも提示するようにすること。
 - 専門家に関する調査計画書を提出
 - 資料の収集方法を詳細に記述すること。
 - 調査紙と調査方法を提示すること。
 - 専門家へ与えられた背景乗法があるか、あるならこれに関しても提示すること。
 - アンケート紙など、一切を提示すること。
 - 結論の算出方法
 - 個別回答者の結果、合算結果、分布を共に提示すること。
 - 不確実性に関する考慮
 - 倫理的な面
 - Delphi調査を行う場合：反復回数、調査方法に関する詳細を記述、前のラウンドにおける結果がどのように専門家らへフィードバックされたか、毎ラウンド別の質問紙を提示、毎ラウンド別の回答率を提示、Delphi調査の場合はパネルの構成が結果へ影響を与えることがある。
- 最近、Bojke et al. (2019) も、保健医療の意思決定に用いる専門家の意見算出に参考する。プロトコルを開発した。専門家の意見算出における主な原則とHTA向けのプロトコルは下表のようである。いくつかの内容を調べると、
- 専門家の構成においては、該当分野の専門家らが持つ様々な見解を包括できるように構成すること。利害相反を最小化するか、利害相反に関する記録が必要。該当研究の内部者を含めないようにし、少なくとも5人の専門家を含めるように勧告する。
 - 定量的な値を算出する場合は、なるべく単純かつ観察可能な値を算出し、ワーディングは明確に提示されなければならないと勧告する。
 - 専門家らの見解は個別的に算出することを勧告する。そして、専門家らにおける変異を探求するように、とも勧告。
 - 妥当性の評価に向け、内外部的な検討が行われることがある。

- 意見算出の過程にて対面アプローチを勧告。そして、専門家に算出意見の分布をフィードバックし、フィードバック後、専門家らへ見解を修正する機会を与えることを強調。
- 専門家らを対象にした（意見算出に関する。）訓練のセッションを含む。訓練の内容はバイアスを避け、不確実性を表現することに焦点を当てること。
- テスト調査を実行。
- 意見算出後、専門家らがどのように彼らの判断を形成したかに関するrationaleを収集。
- 意見の算出に関する全ての方法的な選択を記録し正当化すること。

<表 67> Key principles of SEE in HCDM

Principle	Key message
1. Transparency	SEE should be transparent and reproducible.
2. Fitness-for-purpose	Elicited information should be fit-for-purpose to be used as an input to further analysis.
3. Consistency, but respecting constraints of the decision making context.	The SEE needs to adapt to the practical and logistic constraints faced by different contexts/decision-making bodies, but maintain a level of consistency in methods used across evaluations.
4. Reflecting uncertainty at the individual expert level	SEE must seek to elicit uncertainty in experts judgements.
5. Recognising and acting on biases	SEE must recognise common expert biases and employ strategies to minimise these.
6. Suitability for substantive experts who are less likely to be normative	SEE must utilise methods that are appropriate for experts with lower levels of normative skills.
7. Recognising where adaptive skills are required	When required SEE must employ methods that incorporate or promote the adaptive skills of experts.
8. Recognising between-expert variation	SEE must attempt to capture any between-expert variation, understand the reasons why it exists, and explore its potential impact on the decision.
9. Promoting high performance	SEE must motivate experts to best express their beliefs about a quantity of interest.

<表 68> A reference protocol for HTA

Element	Reference methods suggested
Experts	<ol style="list-style-type: none"> 1. Recruitment will be driven by the context, however the structured expert elicitation (SEE) should pursue diversity, representing the full range of valid expert beliefs. Experts should be willing to participate. 2. Focus on gathering substantive expertise or experience. Normative skills can

	<p>be developed during the training session as part of the SEE.</p> <p>3. Minimize and record conflicts of interest among the experts. Include expert external to the SEE task, i.e. not those involved in developing the task.</p> <p>4. At least 5 experts should be included in the SEE.</p>
Quantities elicited	<p>1. Simple observable quantities should be elicited where possible; ratios or complex parameters such as regression coefficients should not be elicited directly.</p> <p>2. Dependence between variables should be captured in SEE. Expressing dependent variables in terms of independent variables is preferable when experts do not have strong normative skills.</p> <p>3. Wording should be clear and quantities should be decomposed where this means a better fit with experts mental models.</p>
Approach to elicitation	<p>1. Beliefs should be elicited from experts individually, even if a group interaction follows.</p> <p>2. Although interaction between experts can be structured through face-to-face sessions, constraints in healthcare decision-making (HCDM), such as a lack of experienced facilitators, will usually mean that this will take place via a Delphi style remote process.</p> <p>2. Between-expert variation should be explored explicitly.</p>
Method	<p>Both VIM(variable interval method) or FIM(fixed interval method) work well, however decision makers should aim for consistency across applications.</p>
Aggregation	<p>1. Statistical distributions should be fitted to experts individually-elicited judgements.</p> <p>2. Following fitting, a summary of the individual distributions should be obtained using linear pooling with equal weighting of experts.</p> <p>3. Any adjustments applied should be to improve coherence and consistency not reduce variability. Internal and external review can be used to assess validity.</p>
Delivery	<p>1. Face-to-face where possible to allow a facilitator to deliver training to the expert.</p> <p>2. Feedback to experts should be given during the SEE. Following feedback, experts should be given an opportunity to revise their distributions, either during or after a SEE session.</p>
Training & piloting	<p>1. Training is crucial and should focus on avoiding bias and expressing uncertainty.</p> <p>2. Piloting should be undertaken.</p>
Rationales & documentation	<p>1. Rationales for how the experts made their judgements should be collected post SEE.</p> <p>2. All methodological choices for the SEE must be documented and justified.</p>

5) 利害関係者の意見

- なし

6) 指針の改正方向に関する提言

□ 検討方向

- 1第1案は、既存の表と説明の中で不要と判断される部分は除き、重要な部分を浮き彫りにする方向に修正
 - 専門家の選択基準、専門家数、利害の相反
 - 専門家に提供した情報とアンケート紙のコピーを提出
 - 複雑な指標よりは経験により容易に導き出せる単純な値を質問。ワーディングは明瞭に
 - 専門家別の調査結果を提示
 - 専門家の意見をまとめる際は、平均のみならず中央値、変異を提示。代表値としては中央値を喜好
 - 敏感度分析
 - 調査計画書の提出
 - 意見の受容れに関する全ての過程を記述
 - 意見の受容れに関する全ての過程を記述
- 第2案は、既存の表と説明が十分に詳細であるため、これを維持する案である。

4. 結論

これまで、現在の医薬品経済性評価指針にて、改正が必要な項目を確認し、項目別の改正方向を提案した。このため、国内にて行われた経済性評価の方法論を巡る議論内容と外国の指針改正の現況を考察し暫定的な改正の範囲を決定し、利害関係者に関する。アンケートと専門家による諮問会議を通じ、改正が必要な範囲に関する。意見を求めた。こうした過程を経て確認された改正必要の項目に関しては項目別の改正作業を行った。製薬会社で審評院へ提出した資料を分析し、現在の指針に関する遵守の可否と程度を確認し、経済性評価の小委員会の審議結果により、主に争点となった内容は何か、より具体的な指針が必要な部分は何かを確認した。さらに、諸外国の指針を密に分析し、現在求められる根拠のレベルを把握、追加的に関連した文献の考察を通じ、各項目別に理論的・実的に争点になっている内容が何かを把握した。最後は、利害関係者が提出した意見も一緒に考慮し、指針の改正方向を提言した。

改正の方向を提案するにおいて、研究チームは下記の項目を考慮した。一、指針の具体性、明瞭性の確保である。現在の指針も経済性評価を行う際に基準となる方向を提示しているが、利害関係者に関する調査や専門家による諮問会議の結果を振り返ってみると、使用者らは同様な指針を置いても、互いに違う解釈をするケースが多かった。例えば、指針にて明瞭に禁じたり否定的に述べたりすることがなくても、指針に当方法論が具体的に提示されていないため、審評院で当方法論を受け入れる意思がないと考える傾向もあった。なお、同一な指針に関し、互いに違う解釈を行う場合もあった。このため、経済性評価の実行者と検討者が各々の業務を行うことにおいて、実質的な基準となる明瞭な指針を提供したかった。

二、これまで行われた方法論上の進歩と合意のレベルを十分に反映する。ために努力した。このため、国内外の指針のみならず、関連した文献を包括的に検討した。三、経済性評価の資料を準備しなければならない国内の企業と研究者の受け入れ態勢も考慮した。いくら良い方向であっても、実際にこれを行わなければならない主体が従えない指針であれば有名無実な指針になってしまうためである。

四、一貫した意思決定を裏付けられなければならないということも考慮した。このため、基本分析は指針にて提示した方法論通りに行い、他の妥当なアプローチは敏感度の分析を行うか、別途の分析結果を提示する。ようにした。最後に、費用-効果性を巡る不確実性を重要に考慮した。

以上の原則を基に、各パート別に、まず現在の指針から改正が必要な部分を検討し、既存に提出された申請件らを分析し、国内の経済性評価における実行現況を調べた。続いて、カナダ、オーストラリア、イギリス、そして他の国における指針とその他文献の検討を通じ、現在、世界的に通用かつ受け入れられている方法論、資料要求レベルを確認し、最後に利害関係者が提出した意見を考慮し、改正案を設けた。改正方向に関する。提案の内容は研究チームで準備した改正案に関する。専門家による諮問会議を経て最終的に設けられた。

当研究にて提案した内容、そして当レポートと共に提示する。指針の草案は、今後様々な利害関係者の意見受容れ過程を経て、最終改正ガイドラインとして確定される予定である。同レポートは、今後、発表される指針の背景乗法を提供する。役割を果たす。

併せて、指針の改正案を設ける過程にて、一部の追加的な研究が必要な部分が確認された。今後、これらの研究が進んだら、これに合わせて追加的な指針の改正が必要となる可能性がある。外国の指針の改正過程と韓国の指針改正過程を比較して見ると、外国は全面的な改正前であっても、必要な部分に関する

局所的な改正を行っているが、韓国は現在まで審評院で公開するドキュメントをもって、審議の方向を一部明かしたが、そのほか、全面的な改正までの指針の改正は行わなかった。変わりつつある状況に、より柔軟に対応するには全面的な改正のみならず、随時的な改正も行う必要がある。さらに、経済性評価をすることに置いて、当指針だけで十分ではない。外国のケースを見ても、医薬品の経済性評価指針のほか、費用マニュアル、様々な統計方法論に関するマニュアルもしくはドキュメントがウェブサイトに掲載され、資料の提出を準備する側としても活用できるようにしている。韓国の場合、間接的な比較の指針程度が提示されており、費用マニュアルは韓国保険医療研究院にて2013年に発表した冊子が活用されている。同指針の改正と共に、関連した指針も共にアップデートするか、追加的に提供する必要がある。

付録 1.総合依存医療技術に関する詳細な指針（オーストラリア）

□ Section 1 – Context

- 下記の情報は、PBACに提出する申請書のPart A、Section 1に関するもの

○ 生物学的標識子、検査、医薬品に関する詳細

- 追加的な要請資料
- 1 (O) 検査と医薬品の現在の給与状況の記述
- 2 (T) 検査のsponsor確認
- 3 (M) 医薬品のsponsor確認
- 4 (O) 生物学的標識子に関する記述
- 5 (T) 提案された検査に関する記述
- 6 (O) 適応症に関する記述
- 7 (O) 適切な臨床管理経路（clinical management pathways）に関する記述

○ 相互依存性に関するrationale

- 追加要請情報
- 8 (O) 生物学的な標識子を定義しなさい。
- 9 (O) 申請薬剤が提案されたバイオマーカーを目標とするという生物学的なrationaleを提供しなさい。
- 10 (O) 医薬品の相対的な治療効果を修正（modify）する他のバイオマーカーに関して定義しなさい。
- 11 (O) 検査を受ける可能性のある人々の中で、目標とする状態の有病率を定義しなさい。

○ 相互依存技術が現在の臨床診療へ与える影響

- 追加要請情報
- 12 (T) 提案された検査は疾病や状態の経路に関するものを含み、時間の流れに沿って一貫すると期待されるかを陳述しなさい。
- 13 (T) 提案された検査は疾病や状態の経路に関するものを含み、時間の流れに沿って一貫すると期待されるかを陳述しなさい。
- 14 (T) 提案された検査は現在用いられている他の検査に付加的な検査なのか、代替的な検査なのか、それとも両方か（検査結果により、他の検査に付加的に若しくは一部の検査を代替し）を陳述しなさい。

- 15 (T) 提案された検査はオーストラリアにて、どのように提供されるか (will be offered) を記述しなさい。
- 16 (T) 検査のために必要なbiospecimenもしくはsampleを確認しなさい。そして、こうしたspecimenがその検査だけのために、特異的に収集される必要があるのか、それとも他の目的で既に収集されたものかを確認しなさい。
- 17 (T) 新たなsomatic mutationsの発生をモニタリングするため、そして・もしくは相互依存的な医薬品の容量若しくは治療中止を決めるために、後続 (subsequent) 検査が必要かを記述しなさい (適切であれば)
- 18 (O) 提案された医薬品が提案された検査のほか、該当バイオマーカーを確認するための他の特異的な検査と共に用いられるかを確認しなさい。バイオマーカーの検査にて用いた方法を記述しなさい。

□ Section 2 – Clinical evaluation

○ **直接根拠に基づいたアプローチ (Approach based on direct evidence)**

- Section 2a – バイオマーカーの予後効果
- Section 2d – 相互依存技術の臨床的な評価 (結合使用の根拠)
- そして・もしくは
- 連携根拠に基づいたアプローチ (Approach based on linked evidence)
- Section 2a – バイオマーカーの予後効果
- Section 2b – 提案された検査の成果 (performance) と正確度
- Section 2c – clinical managementの変化
- Section 2d – 相互依存技術の臨床的な評価 (別途)
- 各Section 2は、このガイドラインのPart Aに提示されたステップに従う。

○ **直接根拠へのアプローチ (direct evidence approach)**

- 追加要請情報：直接根拠
- 19 (O) バイオマーカーの検査が提供されたclinical managementとは無関係に、患者の健康成果における差を予測できるかを決定しなさい。
- 20 (O) 直接根拠の検索が包括的であり選定過程が偏って居合ことを確認しなさい。
- 21 (O) 歪み、confounding、直接根拠に提示された結果へ与える偶然 (chance) の影響を評価しなさい。

○ **連携根拠へのアプローチ (Linked-evidence approach)**

- 追加要請情報：連携根拠

- 22 (T) 提案された検査の分析的な成果 (analytical performance) を記述しなさい。
- 23 (T) 提案された検査の成果 (performance) を測定する際に用いる参考値もしくはgold standardを定義しなさい。
- 24 (T) 提案された検査における診断の正確度や予測測正確度に関する根拠の検索は包括的であったか、そして根拠の選定過程は偏らなかったかを確認しなさい。
- 25 (T) 提案された検査の診断正確度もしくは予測正確度を報告する根拠が (i) 良質のものか、そして (ii) 要請されたMBS目標集団に適用科野なものかを確認しなさい。
- 26 (T) 診断の正確度若しくは予測正確度の面にて提案された検査のperformanceを報告しなさい。もし、複数の検査が提案されたり若しくは特定の検査を特定しなかったりすると、どの検査のperformanceが最もいいかを確認しなさい、もし、検査の正確度を決定できなかったら検査における一致度やconcordanceを計算しなさい。
- 27 (T) どの検査が最もアプローチ性が良いか・利用可能か・用いられるかを確認しなさい。(いくつかの検査が提案されたが、特定の検査が特定されない場合のみ)

□ section 2c Change in clinical management

- 追加要請資料：連携の根拠
- 28 (O) 検査に関する知識が治療を担当する臨床医の患者管理に変化をもたらすかを立証しなさい。検査にバイオマーカーが存在するということを確認したにもかかわらず、managementが変わらない例を確認しなさい。

□ Section 2d相互依存技術に関する臨床的な評価（別途separate）

- 追加要請情報：連携根拠
- 29 (T) 全体的な検査過程に影響を与えかねない安全性に関する (any safety consideration) 考慮を確認。
- 30 (M) 提案された医薬品の治療効果に関する根拠の検索が包括的であり、根拠の選定過程が偏らなかったかを確認しなさい。(indicate)
- 31 (M) 提案された医薬品の治療的な効果を報告する根拠が良質であるかを確認しなさい。
- 32 (O) 適切であれば、バイオマーカーの状態による治療効果の変更（即ち、相互作用）の根拠を提供しなさい。
- 33 (O) 適切であれば、検査を行うことにより、医薬品に最もよく反応する患者をより上手くターゲティングするという根拠を提供しなさい（即ち、疾病や状態悪化の規定リスクを決めるために、バイオマーカーの予後効果を用いることにより）。
- 34 (O) 検査やバイオマーカーの結果により、conditionedされた医薬品の効果が患者に関する健康成果の面で、臨床的に重要であり統計的に有意な効果があったか確認しなさい（安全性と効果共に）。

○ 相互依存技術効果の適用可能性

- 追加要請情報
- 35 (O) 相互依存技術の臨床的な効果を支える根拠がオーストラリアの国民へ、そして各技術を用いる状況にて、適用可能かを確認しなさい。

□ Section 3 – 経済性評価

○ モデルの構造

- 追加要請情報
- 36 (O) モデルの構造が似ている状況にて行われた（検査や治療を決める前に始める）他の経済性評価と一貫しているかを確認しなさい。
- 37 (O) モデル構造が、項目7により提供された臨床的な経路と一貫したものかを確認しなさい。
- 38 (O) 適切なら、提案された検査の利用において、非健康に関する影響への補足的な分析を提供しなさい。

○ 検査結果に関する転移確率

- 検査結果に関する転移確率
- もし、Section 2にて、連携根拠アプローチが用いられたら、モデルに下記を計算し含みなさい。
- 39 (O) 39 (O) 提案された検査の養成予測値 (PPV)
- 40 (O) 提案された検査の1-PPV
- 41 (O) 提案された検査の陰性予測値 (NPV)
- 42 (O) 提案された検査の1-NPV
- 43 (O) モデルにて、(i) 真陽性患者と偽陽性患者に提案された医薬品に関する (ii) 真陰性と偽陰性の患者において、比較薬に関する;もしくはiii) 直接的な根拠にて報告された（検査の二重若しくは単一のRCTを使用した状況にて）副作用の発生率を提供しなさい。
- 44 (O) モデルにて、全ての検査を受ける人において、検査に関する副作用の発生率を含みしなさい。
- 45 (O) 治療効果の変更に加え、予後効果が働く場合、絶対的な治療効果を提示する際、モデルがこうした要素を補正したことを保障しなさい。

○ モデルに含まれたリソースの項目と費用

- 追加要請情報

- モデルに下記の費用を含めなさい。
- 46 (O) 単位検査費用
- 47 (O) (適切なら) サンプルング費用
- 48 (O) 検査administration費用
- 49 (O) 検査結果と治療計画に関する医療チームの患者相談費用
- 50 (O) 再検査と評価可能な結果の費用
- 51 (O) 検査に関する副作用の費用
- 52 (O) 提案された検査の結果として、追加的に行わなければならない検査の費用
- 53 (O) 検査結果が偽陽性である場合を含めて、医薬品に関する副作用の費用
- 54 (O) 適切な他の保健医療リソースの費用 (例、診断、医療、病院、allied health)

○ モデルにおける不確実性

- 追加要請情報
- 55 (O) 医薬品の治療効果を巡る不確実性を評価しなさい。
- 56 (O) Section 2にて連携根拠アプローチが用いられた際、検査の正確度を巡る不確実性を評価しなさい。
- 57 (O) Section 2にて連結根拠アプローチが用いられた際、バイオマーカーの有病率を巡る不確実性を評価しなさい。
- 58 (O) 適切なら、先行要件として提案された検査を行わず、医薬品をPBSに登録するオプションに関し、scenario analysisを提供しなさい。

□ Section 4 – 臨床における医薬品の使用

- 追加要請情報
- 59 (O) MBSとPBSを併せて、財政に関する影響分析をしなさい。そして、結果を部門別に分けなさい (公共、民間、患者、その他)。
- 60 (O) 検査結果により決まるバイオマーカーの有病率に基づき、疾病負担の力学的な推定値を計算しなさい。
- 61 (O) 検査対象である全ての患者の検査費用と評価不可若しくはindeterminate結果が出た際の再検査費用を推定しなさい。もし、治療が始まってからも検査が求められたら、(即ち、治療をモニタリングしたり、治療中止を決めたりするため)、この費用も含めなければならない。適切なら、提案された検査を先行要求事項とせず、医薬品をPBSに登録させることに関するscenario analysisも提供しなさい。
- 62 (O) 検査と医薬品が掲載される場合に発生する他のMBS費用も全て推定しなさい。

付録2.診断検査の経済性評価に関するチェックリスト

□ Kip et al. (2018) が提示したチェックリスト

<表 69> Reporting checklist to indicate which items were included in the health economic evaluation of diagnostic tests and biomarkers

Items of the Evaluation of Diagnostic Tests and Biomarkers		Included in evaluation	
		Yes	No
Time to presentation			
Onset of complaints or onset of suspicion by physician and start of diagnostic trajectory			
1.1	The study should contain a description of the individuals who enter the diagnostic pathway (i.e., a patient develops a [new] condition or disease, which may or may not result in symptoms or complaints, or undergoes diagnostic testing as part of screening or genetic testing).		
1.2	Consider the time to start of the diagnostic trajectory or the time until a monitoring or screening test is (repeatedly) performed (initiated by symptoms/complaints or initiated as part of regular screening or monitoring). (The time between 1.1 and 1.2 is the time during which individuals are at risk of complications from disease and progression, in the absence of a diagnosis and thus also in the absence of treatment.)		
Use of diagnostic tests			
Decision regarding which diagnostic test(s) is/are performed, in which patients, and in what order			
2	Specify for which purpose(s) the test is used (e.g., screening, diagnosing, monitoring, guide dosage, commencement or cessation of therapy, triaging, staging, prognostic) and define the entire diagnostic and clinical pathway.		
2.1	Consider whether more than 2 (possible) diagnostic strategies can be compared, each involving a single test or combination of tests.		
2.2	Consider whether the evaluated diagnostic strategies include multiple tests, which can be performed in parallel or in sequence.		
2.2.1	Consider whether some tests of the diagnostic workup are performed conditional on previous test outcomes, leading to a selection of patients undergoing specific tests.		
2.3	Consider whether subgroups can be defined based on explicit criteria or patient characteristics, in which different tests would be performed (not solely dependent on previous test outcomes).		
2.4	Consider whether different tests are applied based on implicit (shared) decision making (e.g., perceived condition or risk, or symptom presentation).		
Test performance and characteristics			
Diagnostic test performance and items related to the sampling and testing			
3.1	Specify the costs of the diagnostic test		
3.2	Specify test performance, in terms of sensitivity, specificity, negative predictive value (and its complement), and/or positive predictive value (and its complement), either or not combined with a decision rule or algorithm.		
3.2.1	Describe the evidence base for the estimated test performance.		

3.2.2	Describe the positivity criterion (i.e., cutoff value) applied to the test or testing strategy		
3.2.3	Consider whether the estimated test performance may be biased, for example, due to lack of evidence on conditional dependence or independence, lack of a (perfect) gold standard (i.e., classification bias), verification bias, analytic bias, spectrum bias, diagnostic review bias, and incorporation bias.		
3.2.3.1	Consider how likely/to what extent bias in the available/applied evidence affects the estimated test performance		
3.2.4	Describe how uncertainty/variation in the test performance (receiver operating characteristic [ROC] curve) was handled or explained, for example, due to interrater and intrarater reliability, or experience of the clinician.		
3.2.5	Describe the logic, or analysis, applied to choose the cutoff value (i.e., the point on the ROC curve) for the test, for example, depending on whether the test is used as a single test or part of a sequence of tests.		
3.2.6	Describe whether different test performances and cutoff values were considered for different subgroups of patients and/or environmental characteristics. For example: based on specific subgroup(s) of patients, timing of the test in the diagnostic trajectory, or selection of patients based on previous test outcomes (if any).		
3.2.7	Consider whether test performance is dependent on disease prevalence (which also includes the impact of spectrum bias on disease prevalence and, as a consequence, on test performance) or affected by other patient characteristics or conditions.		
3.2.8	Consider whether test performance is based on a combination of tests (and on a combination of areas under the ROC curves for each test).		
3.3	Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.		
3.4	Consider the occurrence of test failures or indeterminate/not assessable results.		
3.5	Consider costs of retesting (after obtaining insufficient/unusable sample or after test failure or indeterminate/not assessable result).		
3.6	Consider complications, risks, or other negative/positive aspects directly related to obtaining the sample and/or performing the diagnostic test (either in the intervention or in the control strategy).		
3.7	Consider the time taken to perform the test (including waiting time) until the test result is available or until a management decision or treatment is initiated based on this test result (either in the intervention or in the control strategy).		
3.8	Consider the impact of additional knowledge gained by performing the diagnostic test (i.e., for a genetic test) or the occurrence and impact of incidental findings (i.e., the unintentional discovery of a previously undiagnosed condition during the evaluation of another condition).		
3.8.1	The impact of incidental findings on performing additional tests is		

	addressed.		
	Patient management decisions Impact of a test on the diagnosis and/or patient management strategy (based on this diagnosis)		
4.1	Clearly specify the impact of the test in selecting the patient management strategy.		
4.2	Consider whether other aspects besides test results themselves are part of the decision algorithm (and included in the evaluation). These may involve a		

	shared decision-making process of the physician with patients/relatives or aspects including coverage or physician adherence to treatment guidelines.		
4.3	Consider whether the impact of the test result on resulting/selected diagnosis or management strategy varies across subgroups (this difference should not only be caused by differences in diagnostic performance of the test and does not need to include the impact on costs and/or health outcomes within this subgroup).c		
4.4	Consider the consistency of test results over time (e.g., genetic mutations may be affected by treatment prescribed after the initial diagnosis).		
4.5	Consider the impact of performing the test and providing and interpreting the result on the time spent/capacity of the health care professional(s) or the patient.		
Impact on health outcomes and costs Impact of the patient management strategy on diseased and nondiseased individuals, in terms of health outcomes and costs			
5	Evaluate the direct impact of the chosen patient management strategy on the number of (in)correctly diagnosed individuals, health outcomes, and/or costs.		
5.1	Consider the direct impact of the chosen patient management strategy on health outcomes and/or costs. This concerns the entire period in which patient management may affect a patient's health and/or costs and does not only involve the testing strategy itself.		
5.2	Consider whether the direct impact of the chosen patient management strategy on health outcomes and/or costs varies across subgroups. (This does not include only varying the incidence of a certain condition in a sensitivity analysis. The subgroups should be clearly defined and preferably be identifiable based on patient characteristics. c)		
5.3	Consider patient's adherence to treatment (which includes aspects that may indicate (partial) nonadherence (e.g., following only some of the treatment recommendations, as well as aspects that affect the degree of administration of treatment).		
5.4	Consider the occurrence (and consequences) of treatment-related adverse events.		
5.5	Describe the probability or time it takes to observe that the patient management strategy proves to be effective over time or that the patient cures spontaneously (regardless of whether the patient received a correct or an incorrect diagnosis).		
5.6	Describe the probability of or time it takes to repeat or extend the diagnostic workup when the patient management strategy proves to be ineffective, either directly or over time (regardless of whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which the patient receives no treatment or unnecessary treatment.		

5.7	Describe the impact of ineffective or unnecessary treatment or management on health outcomes and/or costs (including both side effects and costs and regardless of whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which incorrectly no treatment is provided or in which the treatment is delayed.		
5.7.1	Consider the impact of delay in treatment initiation on health outcomes and/or costs.		
Societal impact Wider (societal) impact of the chosen diagnostics and management strategy			

6.1	Consider the psychological impact of the diagnostic outcome and management strategy on patients, including the value of knowing (in terms of reassurance or anxiety), patient preferences regarding undergoing diagnostic tests, the (accompanying) impact on caregivers or relatives, and so on.		
6.1.1	Consider the impact of test outcomes on relatives themselves, regarding the value of knowing (spillover knowledge) and regarding subsequent testing and/or treatment in this group (in case of heritable genetic conditions or contagious diseases).		
6.2	Consider the additional impact of diagnostic outcome and management strategy on the health system or health care professionals.		
6.3	Consider the additional impact of diagnostic outcome and management strategy for society.		

a If an item is included in the quantitative analysis, indicate the corresponding model parameter(s) and evidence source(s).

b If an item is excluded from the quantitative analysis, please explain why the exclusion was necessary.

c Existing guidelines indicate that subgroup analyses are relevant when different strategies are likely to be (sub)optimal in different subgroups. Subgroup specific analyses can then be performed to address multiple decision problems. Here we consider scenarios where different tests may be used in different subgroups, depending on patient characteristics or previous test outcomes.

□ Yang Y et al (2019) が提示したチェックリスト

<表 70> Checklist for model-based economic evaluations of diagnostic tests

Domain	Question
1. Decision problem and scope specified	1. Is there a clear statement of the decision problem?
	2. Is the perspective of the model stated clearly?
	3. Has the target population been identified?
	4. Are the model inputs consistent with the stated perspective?
	5. Are the primary outcomes of the model consistent with the perspective, scope and overall objective of the model?
2. Identification and description of comparators	6. Have all feasible and practical options been identified?
	7. Have the comparators being evaluated been clearly described?
	8. If comparators have been excluded from the evaluation, have these exclusions been justified?
3. Appropriate data systematic and identification	9. Are the data identification methods transparent, appropriate given the objectives of the model?
4. Sufficient detail for data incorporation	10. Have all data incorporated into the model been described and referenced in sufficient detail?
	11. Where choices have been made between data sources, are these justified appropriately?

Domain	Question
	12. Are transition probabilities calculated appropriately?
	13. Has discounting been conducted?
5. Quality and incorporation of test accuracy data	14. Has the quality of the test accuracy data been assessed?
	15. Have diagnostic accuracy data been derived from high quality datasources (hierarchy of evidence)?
	16. Are tests in sequence treated dependently, where appropriate?
6. Quality and incorporation of treatment data	17. Has the quality of the treatment effect data been assessed?
	18. Have relative treatment effects been derived from high quality datasources (hierarchy of evidence)?
7. Source and incorporation of cost data	19. Has the source of cost data been presented clearly?
	20. Have costs been inflated to a specific year, where appropriate?
8. Source and incorporation of utility data	21. Is the source for the utility weights referenced and justified?
	22. Are the utilities incorporated into the model appropriately?
9. Model structure	23. Have the reasons behind the type of decision analytic model chosen been fully described and justified?
	24. Has a systematic review of existing economic evaluations been carried out?
	25. Is the structure of the model consistent with a coherent theory of the health condition under evaluation?
	26. Are the structural assumptions underpinning the model transparent and justified?
	27. Have the methods used to extrapolate short-term results to final outcomes been documented and justified?
	28. Has the time horizon been stated and justified?
	29. Has cycle length of Markov models been justified?
	30. Has parameter uncertainty been addressed via sensitivity analysis?
	31. Has probabilistic sensitivity analysis been carried out? If not, has this omission been justified?
	32. If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?
10. Uncertainty	33. If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?
	34. Have structural uncertainties been addressed via sensitivity

analysis?

Domain	Question
11. Validity	35. Have alternative assumptions related to final outcomes been explored through sensitivity analysis?
	36. Has value of information analysis been done?
	37. Has the face validity been reviewed by someone external to the model developers?
	38. Has the mathematical logic of the model been assessed? (e.g. using null and extreme values)
	39. Have the model and its results been compared to the findings of other models and studies, and any disagreements or inconsistencies been explained (cross-validity)?

<表 71> Checklist questions with poor performances n (%)

Questions	Met	Partially met	Not met	N (Excluding n/a)
If comparators have been excluded from the evaluation, have these exclusions been justified?	19 (50)	15 (39)	4 (11)	38
Has the quality of the treatment effect data been assessed?	10 (20)	15 (29)	26 (51)	51
Are transition probabilities calculated appropriately?	19 (66)	6 (21)	4 (14)	29
Are tests in sequence treated dependently, where appropriate?	11 (32)	8 (24)	15 (44)	34
If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	12 (26)	30 (64)	5 (11)	47
Have structural uncertainties been addressed via sensitivity analysis?	0 (0)	12 (22)	43 (78)	55
Has the face validity been reviewed by someone external to the model developers?	16 (29)	15 (27)	24 (44)	55
Have the model and its results been compared to the findings of other models and studies, and any disagreements or inconsistencies been explained (cross-validity)?	27 (50)	8 (15)	19 (35)	54

The percentages might not add up to 100% due to rounding. Percentages were calculated by n/N (i.e. n/a are excluded)

付録 3. 利害関係者へのアンケート

医薬品経済性評価指針の改正に関する意見調査

- 製薬会社向け -

お世話になっております。当研究は、健康保険審査評価院から配布する医薬品経済性評価指針の改正に関する研究です。当調査は、指針の改正に先立ち、指針改正の範囲に関する利害関係者の意見を収集するために計画されました。当調査は、自主的に参加への意思を明かした方に限り、行うつもりであり必要であれば担当研究員（ ）に質問することができます。あなたが回答する全ての内容は、個人情報なしに保管され、研究のための基礎資料としてのみ活用されます。当調査の回答期限は8月16日までであり、回答後は担当研究員までメールでアンケート紙を送ってください。ありがとうございます。

2019年 7月

研究責任者Lee Tae-Jin (Seoul National University・保健大学院)・Bae Eun-Young
(Gyeongsang National University・薬学大学)

1. 現在、所属はどちらですか。()
① 国内の製薬会社 ② 外資系の製薬会社
2. 経済性評価資料を提出した経験はありますか。()
① なし ② あり
※ 「なし」を選んだ場合は4番、「あり」を選んだ場合は3番に移動してください。
3. 経済性評価資料の提出時、資料はだれが作成しましたが。()
① 自社の社員 ② (開始芸会社の場合) 本社 ③ 外部の研究者 ④ その他
4. 下記は、現在の医薬品経済性評価指針案の内容を要約したものです。改正が必要だと思う項目を全てチェック (V) してください。

項目番号	項目	内容	改正必要
評価の一般事項			
1	観点	制限された社会的な観点	
2	分析期間	主な臨床結果を確認できるほど、十分に長くすることを推奨	

項目番号	項目	内容	改正必要
3	分析対象人口 集団	<input type="checkbox"/> 費用と結果が算出された人口集団を明示 <input type="checkbox"/> 給与対象人口集団に一般化できるかの根拠を提示	
4	分析方法	CEA若しくはCUAの最終指標の使用を推奨	
5	分析結果を提示	ICER,の総費用と総効果を別途に提示	
6	報告	段階別の結果と分析に含まれた資料を具体的に提示	

評価方法

7	比較対象の選定	<ul style="list-style-type: none"> 登載医薬品がある場合は、その中で最も多く使われているもの。同一成分の医薬品が複数ある場合は、シェアを用いた加重平均価格を比較薬の単位費用として利用。歪みの可能性が低い資料元（例、RCT）を優先的に選択。 	
8	資料元	<ul style="list-style-type: none"> 申請薬と比較薬を直接比較した資料元を选好。 <input type="checkbox"/>直接比較の資料元がない場合、共通対照群を用いた間接比較も許容包括的な検索。検索戦略を詳しく提示。 	
9	資料検索	資料選定作業を始める前に、包含・排除の基準を定立。レポートに詳述。排除文献の目録及び排除した根拠を提示。	
10	資料選定	最終選定資料に関し、各資料の特徴及び結果を要約・提示。質の評価結果も提示。	
11	資料の抽出及び 評価	異質性を考慮。	
12	メタ分析	費用項目の列挙 → リソース消耗量を測定 → 単位費用を適用 → 最終費用を算出。 <input type="checkbox"/> 制限的かつ社会的な観点を取ることで、直接保健医療費用、患者及び家族（若しくは介護人）の費用、その他、社会の他の部門に発生する費用を含む。	
13	費用	<input type="checkbox"/> 生産性損失費用及び同疾病の治療に関係していない未来への医療費用は基本分析に含まない。	
14	費用項目	<input type="checkbox"/> 保険者の観点から分析する場合、非給与項目は基本分析から除外。 各項目の使用量測定と単位費用の算出過程を詳述。単位費用とリソース使用量の資料元は国内の資料を利用。	
15	費用算出		

結果

16	結果指標	最終結果（final outcome）の使用を推奨	
17	効用・健康に関する暮らしの質	<ul style="list-style-type: none"> QALYを使用。質の加重値は千古度を反映したもの。外国にて開発された道具を用いる際は、国内にて妥当性が検証されたものを用いること。選択度具を指針の別添3、表4に基づいて正当化し、敏感度分析により、質の加重値の不確実性が結果へ与える影響を検討。 基本分析では、選好度に基づいた一般道具の使用を推奨。健康状態の選好度測定対象は、国内にある代表性のある一般大衆を选好。既存の文献にて、質の加重値に関する値を引用する場合、資料元の包含・排除の基準を詳しく説明し、根拠を提示するなど、選定の過程を透明に報告。さらに、該当の健康状態に適用可能かの可否を議論し不確実性を評価。 	
18	割引率	5%基本、0%、3%、7.5%の分感度分析	

項目番号	項目	内容	改正必要
19	モデル構築	<ul style="list-style-type: none"> モデルに関する詳述。 モデル分析の段階ごとに、該当疾病状態にある人（コホートシミュレーションの場合）の数、費用などを記述。 外国にて開発されたモデルの場合、国内の状況にもこれを適用できるかを検討。 	
20	不確実性	<ul style="list-style-type: none"> 経済性評価過程にて、用いた全ての仮定と他の不確実性を招く源泉に関し、臨床的・統計的に妥当な範囲にわたり、一元分感度の分析を実行。母数に関する不確実性はPSAを推奨。 給与を制限して申請する場合、食役処の許可事項にまで対象が拡大される場合に関する敏感度分析を追加的に提示。 	
21	一般化	<ul style="list-style-type: none"> 単位費用及びリソース使用量の資料は国内の資料を利用。 国内の有病率や診療パターンが外国と大きく違い、結果の国内への適用が問題になりかねない場合は、国内の資料に基づいて分析を実行、これを臨床試験の資料に基づいた分析結果と共に提示。 <p>複数国における臨床試験の場合、国内人被験者数が統計的に有意な結論を出せるほどに十分であれば、国内人被験者の試験結果を基本分析、全体被験者の試験結果をもって敏感度分析を実行。そうでない場合は、全体の被験者の臨床試験結果を基本分析で実行。</p>	
22	衡平性	衡平性に関し、どのような過程を行ったかを記述。	
23	財政への影響	給与決定による恩恵の集団を記述。	

5. 上記にてチェックした項目の改正が必要な理由、考えている改正方向、根拠（文献を基準）は何ですか。（別途のページに作成可能）

項目番号	改正が必要な理由	改正方向	根拠

項目 番号	改正が必要な理由	改正方向	根拠

6. その他、医薬品に関する経済性評価指針の改正において考慮する事項があれば、ご記入願います。

()

完了したアンケート紙は下記のメールアドレスへ送ってください。

ありがとうございます。

医薬品経済性評価指針の改正に関する意見調査

- 患者・市民団体向け -

お世話になっております。当研究は、健康保険審査評価院から配布する医薬品経済性評価指針の改正に関する研究です。当調査は、指針の改正に先立ち、指針改正の範囲に関する利害関係者の意見を収集するために計画されました。当調査は、自主的に参加への意思を明かした方に限り、行うつもりであり必要であれば担当研究員（ ）に質問することができます。あなたが回答する全ての内容は、個人情報なしに保管され、研究のための基礎資料としてのみ活用されます。当調査の回答期限は8月23日までであり、回答後は担当研究員までメールでアンケート紙を送ってください。ありがとうございます。

2019年8月

研究責任者Lee Tae-Jin (Seoul National University・保健大学院)・Bae Eun-Young
(Gyeongsang National University・薬学大学)

1. 現在、所属はどちらですか。 ()
① 患者団体 ② 加入者（消費者）団体 ③ その他
2. 薬剤給与決定関連の意思決定に委員として参加した経験はありますか ()
① なし ② あり
3. 下記は、現在の医薬品経済性評価指針案の内容を要約したものです。改正が必要だと思う項目を全てチェック (V) してください。

項目番号	項目	内容	改正必要
評価の一般事項			
1	観点	制限的・社会的な観点	
2	分析期間	主な臨床結果を確認できるほど、十分に長く設定することを推奨。	
3	分析対象人口 集団	<input type="checkbox"/> 費用と結果が算出された人口集団を明示 <input type="checkbox"/> 給与対象人口集団に一般化可能かの根拠を提示	
4	分析方法	CEAもしくはCUAの最終指標の使用を推奨	
5	分析結果の提示	ICERの総費用と総効果を別途に提示	
6	報告	段階別の結果と分析に含まれた資料を具体的に提示	
評価方法			

項目番号	項目	内容	改正必要
7	比較対象の選定	登載医薬品のある場合は、これらの中で最も多く用いられるもの。同一成分の医薬品が多数ある場合は、シェアを用いた加重平均価格を比較薬の単位費用として利用。	
8	資料元	<ul style="list-style-type: none"> 歪みの可能性が低い資料元（例、RCT）を優先的に選択。 申請薬と比較薬を直接比較した資料元を選好。 直接比較可能な資料元がない場合、共通対照群を用いた間接比較も許容。 	
9	資料検索	包括的な検索。検索の戦略を詳しく提示。	
10	資料選定	資料選定作業の開始前、包含・排除の基準を定立。レポートに詳述。排除した文献目録及び根拠を提示。	
11	資料の抽出及び評価	最終的な選定資料に関し、各資料の特徴及び結果を要約・提示。質の評価に関する結果も提示。	
12	メタ分析	異質性を考慮	
13	費用	費用項目の列挙→リソース消耗量の測定→単位費用の適用→最終費用を算出	
14	費用項目	<ul style="list-style-type: none"> 制限的・社会的な観点を取ることで、直接保健医療費用、患者及び家族（もしくは介護人）費用、その他の社会の他の部門に発生する費用を含む。 生産性損失費用及び該当疾病の資料に関わらない未来の医療費用は、基本分析に含まない。 保険者の観点から分析する場合、非給与項目は基本分析から除く。 	
15	費用算出	各項目の使用量測定と単位費用の算出過程を詳しく記述。単位費用とリソースの使用量に関する資料元は国内の資料を用いる。	
結果			
16	結果指標	最終結果（final outcome）の使用を推奨	
17	効用・健康に関する暮らしの質	<ul style="list-style-type: none"> QALYを使用。質の加重値は、選好度を反映したもの。外国にて開発された道具を用いる場合は、国内にて妥当性検証を行ったものを使用。選択道具は指針の別添3、表4に基づいて正当化し、敏感度分析を行い、質の加重値が結果へ与える影響を検討。 基本分析では、選好度に基づいた一般道具の使用を推奨。健康状態の選好度測定の対象は、国内の代表性のある一般大衆を選好。既存の文献から質の加重値に関する値を引用する場合、資料元の包含・排除の基準を詳しく説明し、根拠を提示するなど、選定の過程を透明に報告。さらに、該当健康状態に適用可能科の可否を議論し、不確実性を評価。 	
18	割引率	5%の基本、0%、3%、7.5%の敏感度を分析	
19	モデル構築	<ul style="list-style-type: none"> モデルに関する詳述 モデル分析の段階ごとに、該当の疾病状態にある人（コホートシミュレーションの場合）の数、費用などを記述。 外国にて開発されたモデルの場合、国内の状況にもこれを適用できるかを検討 	

項目番号	項目	内容	改正必要
20	不確実性	<ul style="list-style-type: none"> 経済性評価過程にて用いた全ての仮定と他の不確実性を招く源泉に関し、臨床的・統計的に妥当な範囲にわたり、一元敏感度分析を事項。母数に関する不確実性はPSAを推奨。 給与を制限して申請する際、食役処の許可事項までに対象が拡大される場合に関する敏感度分析を追加的に提示。 	
21	一般化	<ul style="list-style-type: none"> 単位費用及びリソースの使用量に関する資料は国内の資料を利用。 国内の有病率や診療パターンが外国と大きく違い、結果の国内への適用が問題になりかねない場合は、国内の資料に基づいて分析を実行。これを臨床試験の資料に基づいた分析結果と共に提示。 複数国における臨床試験の場合、国内人被験者数が統計的に有意な結論を出せるほどに十分であれば、国内人被験者の試験結果を基本分析、全体の被験者の試験結果で敏感度分析を実行。 そうでない場合は、全体の被験者の臨床試験結果をもって基本分析として実行。 	
22	衡平性	衡平性に関し、どのような過程を行ったかを記述。給与決定における恩恵集団を記述。	
23	財政への影響	導入後、3-5年間にわたり、保険の財政に与える影響を分析	

4. 上記にてチェックした項目の改正が必要な理由、考えている改正の方向、根拠（文献の根拠）は何ですか。（別途のページに作成可能）

項目番号	改正が必要な理由	改正方向	根拠

項目 番号	改正が必要な理由	改正方向	根拠

5. その他、医薬品に関する経済性評価指針の改正において考慮する事項があれば、ご記入願います。

()

完了したアンケート紙は下記のメールアドレスへ送ってください。

ありがとうございます。

医薬品経済性評価指針の改正に関する意見調査

- 研究者向け -

お世話になっております。当研究は、健康保険審査評価院から配布する医薬品経済性評価指針の改正に関する研究です。当調査は、指針の改正に先立ち、指針改正の範囲に関する利害関係者の意見を収集するために計画されました。当調査は、自主的に参加への意思を明かした方に限り、行うつもりであり、必要であれば担当研究員（ ）に質問することができます。あなたが回答する全ての内容は、個人情報なしに保管され、研究のための基礎資料としてのみ活用されます。当調査の回答期限は8月23日までであり、回答後は担当研究員までメールでアンケート紙を送ってください。ありがとうございます。

2019年8月

研究責任者Lee Tae-Jin・Bae Eun-Young

1. 現在、所属はどちらですか。()
① 大学 ② 研究所 ③ 企業 ④ その他
2. 医薬品経済性評価の指針に基づいた経済性評価の経験はありますか。()
① なし ② あり
3. 薬剤給与の決定過程に、委員として参加した経験はありますか。()
① なし ② あり
4. 下記は、現在の医薬品経済性評価指針案の内容を要約したものです。改正が必要だと思う項目を全てチェック (V) してください。

項目番号	項目	内容	改正必要
評価の一般事項			
1	観点	制限的・社会的な観点	
2	分析期間	主な臨床結果を確認できるほど、十分に長く設定することを推奨。	
3	分析対象人口 集団	<input type="checkbox"/> 費用と結果が算出された人口集団を明示 <input type="checkbox"/> 給与対象人口集団に一般化可能かの根拠を提示	
4	分析方法	CEAもしくはCUA.の最終指標の使用を推奨	
5	分析結果の提示	ICER.の総費用と総効果を別途に提示	

項目番号	項目	内容	改正必要
6	報告	段階別の結果と分析に含まれた資料を具体的に提示	
評価方法			
7	比較対象の選定	登載医薬品のある場合は、これらの中で最も多く用いられるもの。同一成分の医薬品が多数ある場合は、シェアを用いた加重平均価格を比較薬の単位費用として利用。	
8	資料元	<ul style="list-style-type: none"> 歪みの可能性が低い資料元（例、RCT）を優先的に選択。 申請薬と比較薬を直接比較した資料元を 선호。 直接比較可能な資料元がない場合、共通対照群を用いた間接比較も許容。 	
9	資料検索	包括的な検索。検索の戦略を詳しく提示。	
10	資料選定	資料選定作業の開始前、包含・排除の基準を定立。レポートに詳述。排除した文献目録及び根拠を提示。	
11	資料の抽出及び評価	最終的な選定資料に関し、各資料の特徴及び結果を要約・提示。質の評価に関する結果も提示。	
12	メタ分析	異質性を考慮	
13	費用	費用項目の列挙→リソース消費量の測定→単位費用の適用→最終費用を算出	
14	費用項目	<ul style="list-style-type: none"> 制限的・社会的な観点を取ることで、直接保健医療費用、患者及び家族（もしくは介護人）費用、その他の社会の他の部門に発生する費用を含む。 生産性損失費用及び該当疾病の資料に関わらない未来の医療費用は、基本分析に含まない。 保険者の観点から分析する場合、非給与項目は基本分析から除く。 各項目の使用量測定と単位費用の算出過程を詳しく記述。単位費用と 	
15	費用算出	<ul style="list-style-type: none"> リソースの使用量に関する資料元は国内の資料を用いる。 	
結果			
16	結果指標	最終結果（final outcome）の使用を推奨	
17	効用・健康に関する暮らしの質	<ul style="list-style-type: none"> QALYを使用。質の加重値は、選好度を反映したもの。外国にて開発された道具を用いる場合は、国内にて妥当性検証を行ったものを使用。選択道具は指針の別添3、表4に基づいて正当化し、敏感度分析を行い、質の加重値が結果へ与える影響を検討。 基本分析では、選好度に基づいた一般道具の使用を推奨。健康状態の選好度測定の対象は、国内の代表性のある一般大衆を 선호。既存の文献から質の加重値に関する値を引用する場合、資料元の包含・排除の基準を詳しく説明し、根拠を提示するなど、選定の過程を透明に報告。さらに、該当健康状態に適用可能科の可否を議論し、不確実性を評価。 	
18	割引率	5%の基本、0%、3%、7.5%の敏感度を分析	
19	モデル構築	<ul style="list-style-type: none"> モデルに関する詳述 モデル分析の段階ごとに、該当の疾病状態にある人（コホートシミュレーションの場合）の数、費用などを記述。 外国にて開発されたモデルの場合、国内の状況にもこれを適用できるかを検討 	

項目番号	項目	内容	改正必要
20	不確実性	<ul style="list-style-type: none"> 経済性評価の過程にて用いた全ての仮定と他の不確実性を招く源泉に関し、臨床的・統計的に妥当な範囲にわたり、一元敏感度分析を実行。母数に関する不確実性はPSAを推奨。 給与を制限して申請する場合、食薬処の許可事項にまで拡大される場合に関する敏感度分析を追加的に提示 	
21	一般化	<ul style="list-style-type: none"> 国内の有病率や診療パターンが外国と大きく違い、結果の国内への適用が問題 になりかねない場合は、国内の資料に基づいて分析を行い、これを臨床試験の資料に基づいた分析の結果と共に提示 <p>□ 複数国における臨床試験の場合、国内人の被験者数が統計的に有意な結論を出せるほどに十分であれば、国内人被験者の試験結果を基本分析、全体の被験者の試験結果をもって敏感度分析を実行。そうでない場合は、全体の被験者の臨床試験結果を基本分析で実行</p>	
22	衡平性	衡平性に関し、どのような過程を行ったか、記述。給与決定における恩恵集団を記述	
23	財政への影響	導入後、3-5年間で保険の財政に与える影響を分析	

5. 上記にてチェックした項目の改正が必要な理由、考えている改正の方向、根拠（文献の根拠）は何ですか。（別途のページに作成可能）

項目番号	改正が必要な理由	改正方向	根拠

項目 番号	改正が必要な理由	改正方向	根拠

6. その他、医薬品に関する経済性評価指針の改正において考慮する事項があれば、ご記入願います。

()

完了したアンケート紙は下記のメールアドレスへ送ってください。

ありがとうございます。

5.参考文献

- Kang Eunjung, Shin Ho-Sung, Park Hye-Ja, Jo,Min-Woo, Kim Na-Yeon 。 (2006) 。EQ-5Dを用いた健康水準の価値評価。保健経済と政策研究 (旧、保険経済研究) 、12 (2) 、 pp.19-43。
- Kang Eun-Jeong, Hankyoreh, Kim Hyo-Jin。 (2014) 。経済性評価のための児童の一般的な健康技術大家の開発。保健経済と政策研究 (旧、保険経済研究) 、20 (4) 、 pp.145-63。
- Kang Eun-Jeong, Hankyoreh. (2018) 。一般的な選好基盤の健康関連暮らしの質道具であるChild Health-6 Dimension (CH-6D) の妥当性研究。韓国母子保健学会誌、22 (1) 、 pp.53-61。
- 健康保険審査評価院。 (2018) 。2018年度、経済性評価制度改善TFTレポート。原注：
健康保険審査評価院。健康保険審査評価院。経済性評価小委員会ワークショップのレポート (1~5次) 。健康保険審査評価院の内部資料。健康保険審査評価院。 (2011) 。
医薬品経済性評価指針。 Seoul: 健康保険審査評価院。 Retrieved from:
<https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA030016000000&brdScnBltno=4&brdBltno=2058&pageIndex=2>
- 健康保険審査評価院。 (2014) 。臨床的有用性評価のための間接比較遂行段階別の資料提出指針。Seoul 健康保険審査評価院。 Retrieved from:
<https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA030016000000&brdScnBltno=4&brdBltno=2069&pageIndex=1#none>
- Kim yun-hee, Shin Sang-Jin, Park Ju-Yeon, Jeong Ye-Ji, im Ji-Min, ee Tae-Jin, Bae Eun-Young, Song Hyun-Jin。 (2013) 。保健医療分野にて費用算出方法。 Seoul : 韓国保険医療研究院。
- Bae Seung-Jin, Lee Soo-OK, Jang Sun-Mee, Choi Sang-Eun。 (2011) 。保健医薬品経済性評価ガイドライン改正方案研究。健康保険審査評価院。
- Bae Jong-Myon。 2015。患者報告成果の反応及び解釈指標。Korean J Fam Pract, 5(3), pp.161-6. Suh Hae-Sun, Park Mi-Hai。 (2014) 。個人合わせ医薬品の経済性評価方法論研究。Seoul: 健康保険審査評価院。Lee Gyu-bangなど。 (2001) 。民間投資事業の収益率に関する研究。国土研究院。
- Lee Jung-Eun, Ahn Jeong-Hoon。 (2019) 。HINT-8道具からEQ-5D道具らへのMappingを用いた転換式算出の研究。保健経済と政策研究 (旧、保険経済研究) 、25 (1) 、 77-103。
- Lee Hyun-Jung, Jo,Min-Woo, Choi Seon-Hye, Kim Youn-Jung, Oh Kyung-Won。 (2016) 。韓国型健康関連暮らしの質測定道具の開発と妥当性及び信頼度の評価。週刊健康と疾病、9 (24) 、 pp.447-54。
- Jo,Min-Woo, Yoon Sung-chul, Kim Seon-ha, Son Woo-seung. (2011). EQ-5D-5L健康状態の効用推定。University of Ulsan & 韓国健康増進開発院。
- Jo,Min-Woo. (2014)。国民健康栄養調査 健康関連暮らしの質測定道具の開発。University of Ulsan & 疾病管理本部。Jo,Min-Woo。 (2017)。韓国型健康関連暮らしの質測定道具 (HINT-8) の価値評価研究。University of Ulsan & 疾病管理本部。
- Choi Ji-Eun & Park Tong-Kyu. (2015).公共投資事業の経済性分析のための社会的な割引率の推定。社会科学研究,41(3), pp.145-167。

- Hankyoreh, Kang Eun-jung. (2015). Time Trade-Offを用いた児童の一般的な健康状態記述体系であるCH-6Dの価値評価。保健経済と政策研究(旧、保険経済研究), 21(4), pp.97-116. KDI・公共投資管理センター。(2018) 公企業・準政府機関の事業に関する予備妥当性調査の実行に向けた一般指針の修正・補完の研究(第2版)
- Ades, A. E., Claxton, K., & Sculpher, M. (2006). Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health economics*, 15(4), pp.373-381.
- Afzali, H. H. A., & Karnon, J. (2011). Addressing the challenge for well informed and consistent reimbursement decisions.
- Andronis, L., Barton, P., & Bryan, S. (2009). Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Perth: PrepressProjects Limited.
- Annemans, L., Redekop, K., Payne, K. (2013). Current methodological issues in the economic assessment of personalized medicine. *Value in Health*, 16(6), pp.S20-S26.
- Ara R, Rowen D, Mukuria C. (2017). The Use of Mapping to Estimate Health State Utility Values. *Pharmacoeconomics*, 35(Suppl 1), pp.57-66.
- Arnold D, Girling A, Stevens A, Lilford R. (2009). Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. *The BMJ*. 339, b2688.
- Bagust, A., & Beale, S. (2014). Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Medical Decision Making*, 34(3), pp.343-351.
- Baio, G., & Dawid, A. P. (2015). Probabilistic sensitivity analysis in health economics. *Statistical methods in medical research*, 24(6), pp.615-634.
- Blumenschein K, Johannesson M. (1998). An experimental test of question framing in health state utility assessment. *Health Policy*, 45(3), pp.187-93.
- Bojke, L., Claxton, K., Bravo-Vergel, Y., Sculpher, M., Palmer, S., & Abrams, K. (2010). Eliciting distributions to populate decision analytic models. *Value in Health*, 13(5), pp.557-564.
- Bojke L, Soares M. Decision analysis: eliciting experts' beliefs to characterize uncertainties. In: Culyer AJ, editor. *Encyclopedia of health economics*. Amsterdam: Elsevier; 2014.
- Bojke, L., Soares, M., Claxton, K., Colson, A., Fox, A., Jackson, C., Jankovic, D., Morton, A., Sharples, L., Taylor, A. (2019). Developing a reference protocol for expert elicitation in healthcare decision making. Retrieved from: eprints.whiterose.ac.uk/151174
- Boonen A, van der Heijde D, Landewe R, van Tubergen A, Mielants H, Dougados M, et al. (2007). How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis? *Ann Rheum Dis*, 66(6), pp.771-7.
- Branson, M., & Whitehead, J. (2002). Estimating a treatment effect in survival studies in which

- patients switch treatment. *Statistics in medicine*, 21(17), pp.2449-2463.
- Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. (2007). *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press.
- Brazier, J., Ratcliffe, J., Salomon, J.A., Tsuchiya, A. (2017). *Measuring and valuing health benefits foreconomic evaluation*. Oxford: Oxford University Press.
- Brazier J, Rowen D. (2011). NICE Decision Support Unit(DSU) Technical Support Documents 11: Alternatives to EQ-5D for Generating Health State Utility Values. London: National Institute for Health and Care Excellence (NICE).
- Brazier JE, Yang Y, Tsuchiya A, Rowen DL. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J HealthEcon*, 11(2), 215-25.
- Briggs, A. H., & Gray, A. M. (1999). Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess*, 3(2), pp.1-134.
- Briggs, A.H., & O'Brien, B.J. (2001). The death of cost-minimization analysis? *Health Economics*, 10(2), pp.179-184.
- Briggs A, Wild D, Lees M, Reaney M, Dursun S, Parry D, et al. (2008). Impact of schizophrenia andschizophrenia treatment-related adverse events on quality of life: direct utility elicitation. *Health Qual Life Outcomes*, 6, pp.105.
- Briggs, A.H., Weinstein, M.C., Fenwick, E.A.L., Karnon, J., Sculpher, M.J., and Paltiel, A.D. (2012) Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM modeling good research practices task force working group - 6. *Medical Decision Making*, 32(5). pp. 722-732.
- Briggs AH, Belozeroff V, Feeny D. (2019). Comparison of health state utility estimates from instrument-based and vignette-based methods: a case study in kidney disease. *BMC Res Notes*,12(1), pp.385.
- Brooks R. (1996). EuroQol: the current state of play. *Health Policy*, 37(1), pp.53-72.
- Brouwer, W.B.F. (2019). The Inclusion of spillover effects in economic evaluations: Not an optional extra. *PharmacoEconomics*, 37(4), pp.451-456. DOI:10.1007/s40273-018-0730-6
- Caro JJ1, Ishak KJ. (2010). No head-to-head trial? simulate the missing arms. *Pharmacoeconomics*, 28(10). pp. 957-67
- Caro, J. J., Eddy, D. M., Kan, H., Kaltz, C., Patel, B., Eldessouki, R., & Briggs, A. H. (2014). Questionnaire to assess relevance and credibility of modeling studies for informing health caredecision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value in health*, 17(2), pp.174-182.
- Chancellor JV, Coyle D, Drummond MF. (1997). Constructing health state preference values from descriptive quality of life outcomes: mission impossible? *Qual Life Res*, 6(2), pp.159-68.

- Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., & O'Hagan, T. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics*, 14(4), pp.339-347.
- Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooper NJ, Sutton AJ, Achana F, Welton NJ. (2015) Use of network meta-analysis to inform clinical parameters in economic evaluations. Ottawa: CADTH. Available from: <https://www.cadth.ca/sites/default/files/pdf/RFP%20Topic-%20Use%20of%20Network%20Meta-analy%20to%20Inform%20Clinical%20Parameters%20in%20Economic%20Evaluations.pdf>
- Coretti S, Ruggeri M, McNamee P. (2014). The minimum clinically important difference for EQ-5D index: a critical review. *Expert Rev Pharmacoecon Outcomes Res*, 14(2), pp.221-33.
- Critchfield, G. C., Willard, K. E., & Connelly, D. P. (1986). Probabilistic sensitivity analysis methods for general decision models. *Computers and biomedical research*, 19(3), pp.254-265.
- Dias, S., Welton, N. J., Sutton, A. J., & Ades, A. E. (2013). Evidence synthesis for decision making 5: the baseline natural history model. *Medical Decision Making*, 33(5), pp.657-670.
- de Vries, L.M., van Baal., P.H., Brouwer, W.B.F. (2019). Future costs in cost-effectiveness analyses: Past, present, future. *PharmacoEconomics*, 37(2), pp.119-130. DOI:10.1007/s40273-018-0749-8
- Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. (2018). Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ*, 27(1), pp.7-22.
- Dolan P, Gudex C, Kind P, Williams A. (1996). Valuing health states: a comparison of methods. *J Health Econ*, 15(2), pp.209-31.
- Dolan P. (1997). Modeling valuations for EuroQol health states. *Med Care*, 35(11), pp.1095-108.
- Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E, et al. (2009). Toward a consensus on the QALY. *Value Health*, 12(Suppl 1), pp.S31-5.
- Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., Torrance, G.W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press. pp.5-6.
- Eddy, D. M., Hollingworth, W., Caro, J. J., Tsevat, J., McDonald, K. M., & Wong, J. B. (2012). Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Medical Decision Making*, 32(5), pp.733-743.
- Eldabe S, Lloyd A, Verdian L, Meguro M, Maclaine G, Dewilde S. (2010). Eliciting health state utilities from the general public for severe chronic pain. *Eur J Health Econ*, 11(3), pp.323-30.
- European network for Health Technology Assessment (EUnetHTA) (2015). *Methods for health economic evaluation - a guideline based on current practices in Europe*. EUnetHTA.27)
- European network for Health Technology Assessment (EUnetHTA) (2015). Endpoints used for

RelativeEffectiveness Assessment: Health-related quality of life and utility measures
(adapted version)

18) 効用パートのEUnetHTA 2015a

).28)

- European network for Health Technology Assessment (EUnetHTA) (2019). Methodological guidelines: Practical considerations when critically assessing economic evaluations. ver 0.3. EUnetHTA.
- Faulkner, D., Annemans, L., Garrison, L., Helfand, M., Holtorf, A-P., Hornberger, J., Hughes, D., Li, T., Malone, D., Payne, K., Siebert, U., Towse, A., Veenstra, D., Watkins, J. (2012). Challenges in the development and reimbursement of personalized medicine-payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR personalized medicine special interest group. *Value in Health*, 15, pp.1162-1171.
- (U.S.) Food and Drug Administration (FDA). (2009). Guidance for industry on patient-reported outcome measures: Use in medical product development to support labeling claims. Food and Drug Administration.
- Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care*, 40(2), pp.113-28.
- Gold ME, Siegel JE, Russell LB, Weinstein MC. (1996). *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press.
- Grieve R, Hawkins N, Pennington M. (2013). Extrapolation of survival data in cost-effectiveness analyses: improving the current state of play. *Medical Decision Making*, 33(6), pp.740-742
- Griffin, S., Claxton, K., Hawkins, N., & Sculpher, M. (2006). Probabilistic analysis and computationally expensive models: necessary and required?. *Value in Health*, 9(4), pp.244-252.
- Hawthorne G, Richardson J, Osborne R. (1999). The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of Health-Related Quality of Life. *Quality of Life Research*, 8(3), pp.209-24.
- Hernandez AM, Pudney S, Wailoo A. (2018), Quality review of a proposed EQ-5D-5L value set for England. EEPRU report.
- Hutton B, Salanti G, Caldwell DM et al. (2015) The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. *Annals of Internal Medicine*, 162. pp. 777-84.
- Iglesias, C.P., Thompson, A., Rogowski, W.H. (2016). Reporting guidelines for the use of expert judgement in model-based evaluations. *Pharmacoeconomics*, 34, pp.1161-1172. DOI 10.1007/s40273-016-0425-9
- Impellizzeri FM, Mannion AF, Naal FD, Hersche O, Leunig M. (2012). The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. *Osteoarthritis and cartilage*, 20(7), pp.638-45.
- Iqbal I, Dasgupta B, Taylor P, Heron L, Pilling C. (2012). Elicitation of health state utilities

associated

18) 効用パートのEUnetHTA 2015b

- with differing durations of morning stiffness in rheumatoid arthritis. *J Med Econ*, 15(6), pp.1192-200.
- ISPOR Task Force. (2016). Response to Editorial: Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health*, 19(6), pp.702-3.
- Jaeschke R, Singer J, Guyatt GH. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*, 10(4), pp.407-15.
- Jo MW, Yun SC, Lee SI. (2008). Estimating quality weights for EQ-5D health states with the time trade-off method in South Korea. *Value Health*, 11(7), pp.1186-9.
- Juniper EF, Guyatt GH, Willan A, Griffith LE. (1994). Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol*, 47(1), pp.81-7.
- Kaplan RM, Anderson JP. (1988). A general health policy model: update and applications. *Health Serv Res*, 23(2), pp.203-35.
- Karnon J, Vanni T. (2011). Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. *Pharmacoeconomics*, 29(1), pp.51-62.
- Kim SH, Jo MW, Kim HJ, Ahn JH. (2012a). Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. *Health Qual Life Outcomes*, 10, pp.151.
- Kim EJ, Ko SK, Kang HY. (2012b). Mapping the cancer-specific EORTC QLQ-C30 and EORTC QLQ-BR23 to the generic EQ-5D in metastatic breast cancer patients. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 21(7), pp.1193-203.
- Kim SH, Kim SO, Lee SI, Jo MW. (2014). Deriving a mapping algorithm for converting SF-36 scores to EQ-5D utility score in a Korean population. *Health Qual Life Outcomes*, 12, pp.145.
- Kim SK, Kim SH, Jo MW, Lee SI. (2015). Estimation of minimally important differences in the EQ-5D and SF-6D indices and their utility in stroke. *Health Qual Life Outcomes*, 13, pp.32.
- Kim SH, Ahn J, Ock M, Shin S, Park J, Luo N, et al. (2016a). The EQ-5D-5L valuation study in Korea. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 25(7), pp.1845-52.
- Kim HL, Kim D, Jang EJ, Lee MY, Song HJ, Park SY, et al. (2016b). Mapping health assessment questionnaire disability index (HAQ-DI) score, pain visual analog scale (VAS), and disease activity score in 28 joints (DAS28) onto the EuroQol-5D (EQ-5D) utility score with the Korean Observational study Network for Arthritis (KORONA) registry data. *Rheumatol Int*, 36(4), 505-13.
- Kim, D.D., Wilkinson, C.L., Pope, E.F., Chambers, J.D., Cohen, J.T., Neumann, P.J. (2017). The influence of time horizon on results of cost-effectiveness analyses. *Expert Review of*

- Pharmacoeconomics & Outcomes Research, 17(6), pp.615-623.
DOI:10.1080/14737167.2017.1331432
- Kim J, Jo MW, Lee HJ. (2019). PCN153 VALIDITY AND RELIABILITY OF HINT-8 IN PATIENTS WITH BREAST CANCER. *Value in Health*, 22, pp.S85.
- King MT. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*, 11(2), 171-84.
- Kip, M.M.A., IJerman, M.J., Henriksson, M., Merlin, T., Weinstein, M.C., Phelps, C.E., Kusters, R., Koffijberg, H. (2018). Toward alignment in the reporting of economic evaluation of diagnostic tests and biomarkers: the AGREEDT checklist. *Medical Decision Making*, 38(7), pp.778-788.
- Kvam AK, Fayers PM, Wisloff F. (2011). Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. *European journal of haematology*, 87(4), pp.330-7.
- Larsen K, Hansen TB, Soballe K. (2008). Hip arthroplasty patients benefit from accelerated perioperative care and rehabilitation: a quasi-experimental study of 98 patients. *Acta orthopaedica*, 79(5), 624-30.
- Latimer NR. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011; available from <http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.pdf>
- Latimer, N. R. (2013). Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Medical Decision Making*, 33(6), pp.743-754.
- Latimer, N.R., Keith R. A. NICE DSU Technical Support Document 14: Adjusting survival time estimates in the presence of treatment switching. 2014
- Le QA, Doctor JN, Zoellner LA, Feeny NC. (2013). Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT). *Health Qual Life Outcomes*, 11, 59.
- Lee YK, Nam HS, Chuang LH, Kim KY, Yang HK, Kwon IS, et al. (2009). South Korean time trade-off values for EQ-5D health states: modeling with observed values for 101 health states. *Value Health*. 12(8), pp.1187-93.
- Lim J, Choi SE, Bae E, Kang D, Lim EA, Shin GS. (2019). Mapping analysis to estimate EQ-5D utility values using the COPD assessment test in Korea. *Health Qual Life Outcomes*, 17(1), pp.97.
- Lloyd A, Doyle S, Dewilde S, Turk F. (2008). Preferences and utilities for the symptoms of moderate to severe allergic asthma. *Eur J Health Econ*, 9(3), pp.275-84.
- Lloyd A, Hodgkins P, Sasane R, Akehurst R, Sonuga-Barke EJ, Fitzgerald P, et al. (2011). Estimation

- of utilities in attention-deficit hyperactivity disorder for economic evaluations. *The patient*, 4(4), pp.247-57.
- Longworth L, Rowen D. (2013). Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health*, 16(1), pp.202-10.
- Luo N, Johnson J, Coons SJ. (2010). Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Med Care*, 48(4), pp.365-71.
- Meltzer, D. (2001). Addressing uncertainty in medical cost-effectiveness analysis: implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *Journal of health economics*, 20(1), pp.109-129.
- Mark, S. D., & Robins, J. M. (1993). A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. *Controlled clinical trials*, 14(2), pp.79-97.
- Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med*, 60(7), pp.1571-82.
- Matza LS, Secnik K, Rentz AM, Mannix S, Sallee FR, Gilbert D, et al. (2005). Assessment of health state utilities for attention-deficit/hyperactivity disorder in children using parent proxy report. *Qual Life Res*, 14(3), pp.735-47.
- Matza LS, Boye KS, Yurgin N, Brewster-Jordan J, Mannix S, Shorr JM, et al. (2007). Utilities and disutilities for type 2 diabetes treatment-related attributes. *Qual Life Res*. 16(7), pp.1251-65.
- Matza LS, Devine MK, Haynes VS, Davies EW, Kostelec JM, Televantou F, et al. (2014a). Health state utilities associated with adult attention-deficit/hyperactivity disorder. *Patient preference and adherence*, 8, pp.997-1006.
- Matza LS, Chung K, Van Brunt K, Brazier JE, Braun A, Currie B, et al. (2014b). Health state utilities for skeletal-related events secondary to bone metastases. *Eur J Health Econ*, 15(1), pp.7-18.
- McCabe, C. (2019). Expanding the scope of costs and benefits for economic evaluations in health: Some words of caution. *Pharmacoeconomics*, 37(4), pp.457-460. DOI:10.1007/s40273-018-0729-z
- McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. (2017). Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores. *Value Health*, 20(4), pp.644-50.
- McDonough CM, Tosteson TD, Tosteson AN, Jette AM, Grove MR, Weinstein JN. (2011). A longitudinal comparison of 5 preference-weighted health state classification systems in persons with intervertebral disk herniation. *Medical decision making : an international journal of the Society for Medical Decision Making*, 31(2), pp.270-80.
- Michelle, T., Philip, C., Karin, T., Kim, D. (2019). Incorporating future medical costs: Impact on

- cost-effectiveness analysis in cancer patients. *Pharmacoeconomics*, 37(7), pp.931-941. DOI:10.1007/s40273-019-00790-9
- Mukuria C, Rowen D, Harnan S, Rawdin A, Wong R, Ara R, et al. (2019). An Updated Systematic Review of Studies Mapping (or Cross-Walking) Measures of Health-Related Quality of Life to Generic Preference-Based Measures to Generate Utility Values. *Appl Health Econ Health Policy*, 17(3), pp.295-313.
- Neumann, P.J. (2009). Costing and perspective in published cost-effectiveness analysis. *Medical Care*, 47(7 suppl 1), pp.S28-S32.
- Neumann, P.J., Sanders, G.D., Russel, L.B., Siegel, J.E., Ganiats, T.G. (2017). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.
- Neyt, M., Brabandt, H.V. (2011). The importance of the comparator in economic evaluations: Working on the efficiency frontier. *Pharmacoeconomics*, 29(11), pp.913-916.
- Norman GR, Sloan JA, Wyrwich KW. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*, 41(5), pp.582-92.
- Norman GR, Wyrwich KW, Patrick DL. (2007). The mathematical relationship among different forms of responsiveness coefficients. *Qual Life Res*, 16(5), pp.815-22.
- Nunnally JC, Bernstein IH. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- O'Mahony, J.F., Newall, A.T., van Rosmalen, J. (2015). Dealing with time in health economic evaluation: methodological issues and recommendations for practice. *Pharmacoeconomics*, 33, pp.1255-1268.
- Park SY, Park EJ, Suh HS, Ha D, Lee EK. (2017). Development of a transformation model to derive general population-based utility: Mapping the pruritus-visual analog scale (VAS) to the EQ-5D utility. *J Eval Clin Pract*, 23(4), pp.755-61.
- Parker SL, Adogwa O, Paul AR, Anderson WN, Aaronson O, Cheng JS, et al. (2011). Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. *Journal of neurosurgery Spine*, 14(5), pp.598-604.
- Parker SL, Adogwa O, Mendenhall SK, Shau DN, Anderson WN, Cheng JS, et al. (2012a). Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. *The spine journal : official journal of the North American Spine Society*, 12(12), pp.1122-8.
- Parker SL, Mendenhall SK, Shau DN, Adogwa O, Anderson WN, Devin CJ, et al. (2012b). Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. *Journal of neurosurgery Spine*, 16(5), pp.471-8.
- Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. (2013). Assessment of the minimum

- clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article. *Journal of neurosurgery Spine*, 18(2), pp.154-60.
- Paulden, M. et al. (2017). Discounting and the evaluation of health care programs. Canadian Agency for Drugs and Technologies in Health.
- Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. (2015). Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. *Pharmacoeconomics*, 33(10), pp.985-91.
- Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. (2016). NICE DSU TECHNICAL SUPPORT DOCUMENT 18: METHODS FOR POPULATION-ADJUSTED INDIRECT COMPARISONS IN SUBMISSIONS TO NICE. Available from <http://www.nicedsu.org.uk>
- Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. (2018). Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Med Decis Making*, 38(2). pp. 200-211.
- Philips, Z., Bojke, L., Sculpher, M., Claxton, K., & Golder, S. (2006). Good practice guidelines for decision-analytic modelling in health technology assessment. *Pharmacoeconomics*, 24(4), pp.355-371.
- Pickard AS, Neary MP, Cella D. (2007). Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes*, 5, pp.70.
- Pitman, R., Fisman, D., Zaric, G. S., Postma, M., Kretzschmar, M., Edmunds, J., ... & ISPOR-SMDM Modeling Good Research Practices Task Force. (2012). Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5. *Value in health*, 15(6), pp.828-834.
- Prosser, L.A., Wittenberg, E. (2019). Advances in methods and novel applications for measuring family spillover effects of illness. *Pharmacoeconomics*, 37(4), pp.447-450. DOI:10.1007/s40273-019-00794-5
- Ramos, M. C. P., Barton, P., Jowett, S., & Sutton, A. J. (2015). A systematic review of research guidelines in decision-analytic modeling. *Value in Health*, 18(4), pp.512-529.
- Revicki DA, Wood M. (1998). Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. *J Affect Disord*, 48(1), pp.25-36.
- Richardson J, Iezzi A, Khan MA, Maxwell A. (2014). Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. *The patient*, 7(1), pp.85-96.
- Richardson J, Khan MA, Iezzi A, Maxwell A. (2015a). Comparing and Explaining Differences in the Magnitude, Content, and Sensitivity of Utilities Predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D Multiattribute Utility Instruments. *Medical Decision Making*, 35(3), pp.276-91.

- Richardson J, Iezzi A, Khan MA. (2015b). Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 24(8), pp.2045-53.
- Roberts, M., Russell, L. B., Paltiel, A. D., Chambers, M., McEwan, P., & Krahn, M. (2012). Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices taskforce-2. *Medical Decision Making*, 32(5), pp.678-689.
- Rowen D, Brazier J, Tsuchiya A, Alava MH. (2012). Valuing states from multiple measures on the same visual analogue scale: a feasibility study. *Health Econ*, 21(6), pp.715-29.
- Schünemann HJ, Guyatt GH. (2005a). Commentary--goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res*, 40(2), pp.593-7.
- Schünemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. (2005b). Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). *COPD*. 2(1), pp.81-9.
- Sculpher, M. Fenwick, E., & Claxton, K. (2000). Assessing quality in decision analytic cost-effectiveness models. *Pharmacoeconomics*, 17(5), pp.461-477.
- Seiber WJ, Groessl EJ, David KM, Ganiats TG, Kaplan RM. (2008). *Quality of Well Being Self-Administered (QWB-SA) Scale. User's Manual* Health Services Research Center, San Diego:University of California.
- Shikar R, Willian MK, Okun MM, Thompson CS, Revicki DA. (2006). The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. *Health Qual Life Outcomes*, 4, pp.71.
- Shiroiwa T, Fukuda T, Ikeda S, Takura T, Moriwaki K. (2017). Development of an Official Guideline for the Economic Evaluation of Drugs/Medical Devices in Japan. *Value Health*, 20(3), pp.372-8.
- Siebert, U., Alagoz, O., Bayoumi, A. M., Jahn, B., Owens, D. K., Cohen, D. J., & Kuntz, K. M. (2012). State-transition modeling: a report of the ISPOR-SMDM modeling good research practices taskforce-3. *Medical Decision Making*, 32(5), pp.690-700.
- Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, Gupta SR, Mulani PM. (2010). Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *Pharmacoeconomics*, 28(10). pp. 935-45.
- Sintonen H, Pekurinen M. (1993). A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In: Walker SR, Rosser RM, editors. *Quality of Life Assessment: Key Issues in the 1990s*. Dordrecht: Springer Netherlands. pp. 185-95.
- Soer R, Reneman MF, Speijer BL, Coppes MH, Vroomen PC. (2012). Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *The spine journal : official journal of the North American Spine Society*, 12(11), pp.1035-9.

- Solberg T, Johnsen LG, Nygaard OP, Grotle M. (2013). Can we define success criteria for lumbar disc surgery? : estimates for a substantial amount of improvement in core outcome measures. *Acta orthopaedica*, 84(2), pp.196-201.
- Spencer M, Diels J. (2011). MA2 Mapping Fact-P to Country Specific Patient Health Status Measured by EQ-5D in Metastatic Castrate Resistant Prostate Cancer Patients. *Value in Health*, 14(7), A242.
- Staerkle RF, Villiger P. (2011). Simple questionnaire for assessing core outcomes in inguinal hernia repair. *The British journal of surgery*, 98(1), pp.148-55.
- Stark RG, Reitmeir P, Leidl R, König HH. (2010). Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. *Inflamm Bowel Dis*, 16(1), pp.42-51.
- Starkie HJ, Briggs AH, Chambers MG, Jones P. (2011). Predicting EQ-5D values using the SGRQ. *Value Health*, 14(2), pp.354-60.
- Stevens K. (2009). Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 18(8), pp.1105-13.
- Stout, N. K., Knudsen, A. B., Kong, C. Y., McMahon, P. M., & Gazelle, G. S. (2009). Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics*, 27(7), pp.533-545.
- Tolley K, Goad C, Yi Y, Maroudas P, Haiderali A, Thompson G. (2013). Utility elicitation study in the UK general public for late-stage chronic lymphocytic leukaemia. *Eur J Health Econ*, 14(5), pp.749-59.
- Tosh JC, Longworth LJ, George E. (2011). Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value Health*, 14(1), pp.102-9.
- Tsuchiya A, Brazier J, McColl E, Parkin D. (2002). Deriving preference-based single indices from non-preference based condition-specific instruments: converting AQLQ into EQ5D indices. Discussion Paper Series. School of Health and Related Research (SchARR), University of Sheffield.
- van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. (2012). Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*, 15(5), pp.708-15.
- Vemer, P., Ramos, I. C., Van Voorn, G. A. K., Al, M. J., & Feenstra, T. L. (2016). AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics*, 34(4), pp.349-361.
- Wailoo A, Hernandez Alava M, Grimm S, Pudney S, Gomes M, Sadique Z, et al. (2017a). Comparing the EQ-5D-3L and 5L versions. What are the implications for cost effectiveness estimates?. Sheffield: Decision Support Unit.

- Wailoo A, Hernandez Alava M, Manca A, Mejia A, Ray J, Crawford B, et al. (2017b). Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR GoodPractices for Outcomes Research Task Force Report. *Value Health*, 20(1), pp.18-27.
- Walker, D., & Fox-Rushby, J. A. (2001). Allowing for uncertainty in economic evaluations: qualitative sensitivity analysis. *Health Policy and Planning*, 16(4), pp.435-443.
- Walters SJ, Brazier JE. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 14(6), pp.1523-32.
- Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, et al. (2016). Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health*, 19(6), pp.704-19.
- Wyrwich KW, Tierney WM, Wolinsky FD. (2002). Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 11(1), pp.1-7.
- Wyrwich KW. (2004). Minimal important difference thresholds and the standard error of measurement: is there a connection? *J Biopharm Stat*, 14(1), pp.97-110.
- Wyrwich KW, Norquist JM, Lenderking WR, Acaster S, Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). (2013). Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res*, 22(3), pp.475-83.
- Yang, Y., Abel, L., Buchanan, J., Fanshawe, T., Shinkins, B. (2019). Use of decision modelling in economic evaluations of diagnostic tests: an appraisal and review of health technology assessments in the UK. *PharmacoEconomics*, 3(3), pp.281-291.
- Zhao, Y., Feng, H-M., Luo, J.Q.X., Ma, W-J., Tain, J-H. (2018). A systematic review of pharmacoeconomic guidelines. *Journal of Medical Economics*, 21(1). pp.85-96.
- Ziouani, S., Granados, D., Borget, I. (2016). How to select the best comparator? an international economic evaluation guidelines comparison. *Value in Health*, 19(7), pp.A471-A472. DOI: <https://doi.org/10.1016/j.jval.2016.09.726>

ウェブサイト

- Australian government: Department of health (PBAC). (2008). Guidelines for preparing a submissions to the Pharmaceutical Benefits Advisory Committee (Ver 4.3). Retrieved from: <https://pbac.pbs.gov.au/content/information/archived-versions/pbac-guidelines-v4-3-2008.pdf>
- Australian government: Department of health (PBAC). (2008). Report of the Indirect Comparisons Working Group to the Pharmaceutical Benefits Advisory Committee: assessing indirect comparisons. Retrieved from:

<http://www.pbs.gov.au/industry/useful-resources/pbac-technical-working-groups-archive/indirect-comparisons-working-group-report-2008.pdf>

Australian government: Department of health (PBAC). (2015). Guidelines for preparing a submissions to the Pharmaceutical Benefits Advisory Committee (Ver 4.5). Retrieved from: <https://pbac.pbs.gov.au/content/information/archived-versions/pbac-guidelines-v4-5.pdf>

Australian government: Department of health (PBAC). (2016). Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Ver 5.0). Commonwealth of Australia. Retrieved from: <https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf>

Australian government: Department of health (PBAC). (2019). Procedure guidance for listing medicines on the Pharmaceutical Benefits Scheme. Retrieved from: <http://www.pbs.gov.au/industry/listing/procedure-guidance/files/Procedure-Guidance-Listing-Medicines-on-the-PBS-v1-6.pdf>

The Canadian Agency for Drugs and Technologies in Health (CADTH). (2016). Guidance Document for the Costing of Health Care Resources in the Canadian Setting: Second Edition. Retrieved from: <https://www.cadth.ca/dv/guidance-document-costing-health-care-resources-canadian-setting>

The Canadian Agency for Drugs and Technologies in Health (CADTH). (2017). Guidelines for the economic evaluation of health technologies: Canada. 4th ed. Ottawa: CADTH. Retrieved from: <https://www.cadth.ca/dv/guidelines-economic-evaluation-health-technologies-canada-4th-edition>

The Canadian Agency for Drugs and Technologies in Health (CADTH). (2019). Procedure and Submission Guidelines for the CADTH Common Drug Review. Retrieved from: https://cadth.ca/sites/default/files/cdr/process/Procedure_and_Guidelines_for_CADTH_CDR.pdf

The National Institute for Health and Care Excellence (NICE). (2013). Guide to the methods of technology appraisal 2013. NICE. Retrieved from: <https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781>

The National Institute for Health and Care Excellence (NICE). (2014). Developing NICE guidelines: the manual - appendices. Retrieved from: <https://www.nice.org.uk/process/pmg20/resources>

The National Institute for Health and Care Excellence (NICE). (2019). Position statement on use of the EQ-5D-5L value set for England (updated October 2019). Retrieved from: <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l>