

非選好型アウトカム尺度から推計した 健康状態効用値へのマッピング —ISPORアウトカム研究班報告実施基準—

Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report

Allan J. Wailoo, MA, PhD¹, Monica Hernandez-Alava, MSc, PhD¹, Andrea Manca, MSc, PhD², Aurelio Mejia, MSc³, Joshua Ray, MSc⁴, Bruce Crawford, MA, MPH⁵, Marc Botteman, MS, MA⁶, Jan Busschbach, PhD⁷

訳 兼安貴子^{1,2)} 白岩健³⁾ 船越大⁴⁾ 村澤秀樹⁵⁾ 斎藤信也⁶⁾ 福田敬²⁾ 下妻晃二郎^{2,7)}

要 旨

質調整生存年 (QALY) あたりの費用の観点で行われる経済評価は、意思決定者にとって有益な情報を世界各地で提供している。理想的には、医療技術の有効性を評価する臨床研究が、QALYを計算するための健康効用に直接結びついたアウトカム尺度を含むことである。しかし実際にそのようなことはほとんどなく、たとえ行われたとしても、臨床研究が費用効用評価には十分ではないかもしれない。マッピングはこの問題を解決できる。マッピングは、臨床研究で測定されたアウトカムと健康効用の関係を推計するための追加のデータセットを使用し、推計する。これは、一つの測定単位から得られる、ある医療技術による効果の利用可能なエビデンスと、意思決定者がそれを別の測定単位 (QALY) で表すことを求めるというエビデンスギャップの橋渡しを

行う。2014年に、ISPORは、マッピング研究のためのアウトカム研究班実施基準を開発した。本班報告は、マッピング研究の結果を費用効用分析で用いたり、これらの研究を批判的にレビューしたりする、マッピング研究の分析者を対象とした推奨を提供する。本推奨は、あらゆる領域におけるマッピングの実践、すなわち、マッピング推計のためのデータセットの選択、モデル選定、性能評価、報告基準、そしてばらつきと不確実性を適切に反映した結果の利用をカバーする。本報告は、国際的な視点で行われており、マッピングの実践的側面を包括的にカバーし、そして現時点での最高水準の技術を反映している点で独創的である。

Copyright © 2017, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

キーワード：経済評価, 健康効用, マッピング, QOL

¹ シェフィールド大学健康および関連研究大学院 (英国 シェフィールド), ² ヨーク大学健康経済学センター (英国 ヨーク), ³ 医療技術評価研究所 (コロンビア ボゴタ), ⁴ F. Hoffmann-La Roche (スイス バーゼル), ⁵ アイ・エム・エス・ジャパン株式会社 (東京), ⁶ Pharmerit International (米国 メリーランド州ベセスダ), ⁷ エラスムス医療センター医療心理学・心理療法部門精神科 (オランダ ロッテルダム)

¹⁾ 立命館大学 生命科学部 生命医科学科 ²⁾ 立命館大学 総合科学技術研究機構 医療経済評価・意思決定支援ユニット ³⁾ 国立保健医療科学院 保健医療経済評価研究センター ⁴⁾ 立命館大学大学院 生命科学研究科 ⁵⁾ 盛岡大学栄養科学部 栄養科学科 ⁶⁾ 岡山大学大学院 保健学研究科 ⁷⁾ 立命館大学 生命科学部

緒 言

健康関連 QOL の評価は、医療技術とサービスの評価、規制と償還に関する決定において重要である。選好型尺度 (Preference-based measures: PBM) (選好に基づく尺度、も同義) は、これらの評価において中心的な役割を果たす。PBM は、患者が悪い健康状態の影響を記述するのに役立つ、各健康状態の記述に関係する「効用」値 (またはタリフ) を得ることができる。PBM では、1 の値は完全な健康、0 は死亡、負の値 (PBM によって定義される場合は) は死よりも悪い状態を意味する。そしてこれらの効用値は、多くの医療経済評価で使用されている、健康便益に関するアウトカム測定単位である質調整生存年 (Quality adjusted life years; QALY) を計算する際に使用できる¹⁾。

もっとも広く使われている PBM は包括的尺度である (すなわち、広い領域の疾患、患者、そして介入に適用可能)。たとえば、EuroQol five-dimensional questionnaire (EQ-5D)²⁾、six-dimensional health-state short form³⁾ (SF-36 由来)、そして Health Utilities Index⁴⁾ がある。イングランドおよびウェールズ⁵⁾、スペイン⁶⁾、フランス⁷⁾、タイ、フィンランド、スウェーデン、ポーランド、ニュージーランド⁸⁾、カナダ、コロンビアおよびオランダなど多くの国々の経済評価ガイドラインが、これらの包括的尺度を採用するように提言または要求している。いくつかの国は、特定の尺度を推奨しているが、その多くは EQ-5D である⁹⁾。

多くの場合、臨床試験に PBM は含まれない。臨床試験ではしばしば患者報告アウトカム尺度 (Patient-reported outcome measures: PROM) が 1 つ以上含まれるが、それらは関連する選好に基づくスコアリング方法を持たないため、完全な PBM ではない。このため、PROM では QALY の測定ができない。典型的な臨床試験では、「客観的に」測定された、すなわち、患者による解釈や報告ではない、身体的アウトカムも含む。PBM のアウトカムがない状態で、研究者はこれらの試験から QALY を推計するために「欠損している」PBM を導き出す必要がある。こうした状況下で、問題は、PBM が把握できるであろう価値を予測することが可能かどうか、

もし可能だとしても、どのようにして予測するか、である。ここでは、観察された臨床アウトカムをわれわれが知っていて、被験者個々の背景因子の媒介効果が得られることが前提である。「マッピング」は、この問題に答えようとする試みであり、それによって、一つの測定単位から得られる医療技術の効果に関する利用可能なエビデンスと、意思決定者がそれを別の測定単位 (QALY) で表すことを求めるというギャップに橋渡しをする。また、ある PBM から別の PBM へと、アウトカムを変換する際にも利用可能である。

マッピングは実験的というよりも観察的な別のデータセットを活用する。このデータセットは、関連する臨床研究で測定されたものと同じアウトカムと、標準的な PBM 尺度に対する患者の回答が含まなければならない。この外的なデータセットは、2 種類のアウトカム尺度間の統計学的関係を推計するためによく使用される。推計された統計学的関係を試験からのアウトカムデータと組み合わせることで、健康効用における治療効果の推計を可能とし、さらに QALY の計算に使用できる。統計モデルを健康効用値にあてはめる手法は、「マッピング」「クロスウォーキング」そして「効用値への転換」とさまざまな呼ばれてきた¹⁰⁾。「マッピング」が、一般的に利用されるようになったため、本報告を通して使用する。

経済評価の文脈で、マッピングの必要性を生じさせるエビデンスギャップは一般的に見られる。たとえば、Kearns ら¹¹⁾ は、英国国立医療技術評価機構 (NICE) による最近の 79 件の技術評価を精査し、ほぼ 4 分の 1 の事例でマッピングモデルが使用されていたことを明らかにした。これらには、乾癬患者の Psoriasis Area Severity Index (PASI)、子宮頸癌患者の Functional Assessment of Cancer Therapy-General (FACT-G)、および特に慢性の便秘を伴う女性の Patient Assessment of Constipation-Symptoms (PAC-SYM) と Patient Assessment of Constipation-Quality of Life (PAC-QOL) からのマッピングが含まれる。関連する臨床研究に (前述のように) PBM が含まれない、あるいはこれらの臨床研究だけでは、費用対効果を推計するための効用値情報を提供するのに不十分であるために、マッピングの必

要性が生じる。臨床試験において観察される健康状態の範囲を超えた外挿の要求、または複数の臨床試験からのエビデンスを統合する要求がありうるが、こうした臨床試験のすべてがPBMに関するエビデンスを含むわけではない。このようにマッピングは、PBMのない試験データに基づく経済評価と多くの経済モデル研究の両方において課題となる。そして、これまでに実施されてきた研究は、新技術の評価における対照として、基礎的エビデンスの一部であり続けることから、効用値推計の実施基準が現在の臨床試験に引き続いて使用されている場合でも、しばらくの間は必要とされるだろう¹²⁾。

現在のマッピングの方法には相当なばらつきが含まれ、それが費用対効果の推計における相違につながる事が知られている^{13,14)}。本班報告の目的は、QALYに基づくあらゆる種類の経済評価で使用されるマッピング研究の実施に適したアウトカム研究の

実施基準を設定することであった。本推奨は、たとえば医療提供者の能力の尺度として使用される効用値など、分析者が選好型アウトカムを他の変数の関数として推計したいと考える、すべての状況に幅広く関連するものである¹⁵⁾。本推奨は、マッピング方法の全領域に及ぶ。それは、マッピング推計のためのデータセットの選定、モデル選定と性能評価、報告基準、ばらつきと不確実性の適切な反映を含む結果の利用である。各節の最後にまとめの表を置いた。これらの推奨は、一貫性のない現在の手法、アプローチ間の結果における相当なばらつき、そして複数の方法におけるバイアスのリスクに鑑みて重要である。他の推奨も作成されているが^{16,17)}、本報告は、国際的な視点に立ち、マッピングの実践を包括的に扱っており、現在の最新技術を反映している点で他にないものである。

ISPOR 研究班の背景

2014年6月、ISPOR 医療科学政策会議は、ISPOR 評議委員会に対して、ISPOR 費用対QALY 経済分析のための非選好型尺度から健康効用値を推計するためのマッピング推計実施基準班の設置を推奨した。同月、評議委員会は班の設置を承認した。班のメンバーと主たるレビュアーは、基礎または応用的な効用値研究における専門性、経済モデル、あるいは医療技術評価の効用値測定の質評価に関する理解の度合いに基づいて選定された。班のメンバーと主たるレビュアーの選定では、各国の医療制度を代表するよう、相当の努力が払われた。中心的メンバーの一覧は班のホームページ：<http://www.ispor.org/Mapping-Estimate-Health-State-Utility-Non-Preference-Outcome-Measures-guidelines.asp> で参照できる。班は、「マッピング」研究の結果をどのように実施、評価、使用するかに関する指針を作成することで、マッピング分析の正確性、質および有用性を改善する必要性を同定した。

班は概要、レビューされた節の案、そして報告書の案を、電子メール、電話会議、およびアムステルダムのISPOR European Congress における対面での会議を経て

作成した。主なレビュアーだけでなくすべての班のメンバーは、口頭または書面でのコメントとしてフィードバックを行った。班報告書の案は、数回にわたってレビューされた。専門家である主たるレビュアーのグループによって1回、そして、費用対QALY 経済分析のための非選好型尺度から健康効用値を推計するためのマッピング推計(班)のレビューグループによって2回レビューされた。コメントはまた、2つのフォーラム(2014年および2015年のISPOR European Congresses)における発表を通して受領された。レビューの過程と発表で受け付けたすべてのコメントが検討され、議論され、報告書の改訂案で適切に対応された。すべての書面でのコメントは、taskforce@ispor.org へ電子メールで依頼すれば入手できる。班報告書およびホームページはISPOR ホームページ(<http://www.ispor.org>)の紫色のリサーチツールメニューから、ISPOR Good Practices for Outcomes Research を選び、Preference-Based Methods; Health-State Utilities – Mapping for Cost per QALY Economic Analysis というヘディングからアクセスできる。

訳者注：2023年1月現在、班のホームページは閉鎖されており、Good Practices Reports (報告実施基準/班報告書)は、ISPOR ホームページ (<https://www.ispor.org/heor-resources/good-practices/methodological-statistical-research>) にて確認できる。

モデリング前の検討事項

マッピングを目的とする統計学的解析に取り組む前に、分析者は、マッピングの用途案や潜在的な用途に関連する多くの要素を考慮しなければならない(表1参照)。これらの用途が統計学的解析および試験する対象となるデータセットの要件を生み出す(表2および表3)。

エビデンスギャップの理解

マッピングはほぼ常に所定の目的を持って行われ、多くの場合、特定の費用効果分析(Cost-effectiveness analysis: CEA)に情報提供するためのものである。取り組むべきエビデンスギャップを明確に理解するには、関連する既存効用値エビデンスの評価、分析結果を評価する意思決定機関の要件、および結果が使用されるCEA——たとえば、意思決定者は特定のPBMに基づくべき効用値を特定しているか——が必要である。これらの要因が分析上の選択肢を知るために役立ち、費用対効果研究におけるバイアスのない推計が可能になる。不確実性や、さらに状況によっては推計のばらつきを適切に反映することが要件となるだろう(例：費用対効果モデルで個々の患者をシミュレーションする場合)。

どの健康状態あるいはプロファイルが効用値のエビデンスを必要とするかに関しては、CEAのニーズが、これらの特定のニーズに対して適切な性能を期待できる方法とデータセットを分析者が選定する際の助けとなるだろう。決定分析モデルを用いて研究が行われる時、そのモデルにどの健康状態が反映されているか——それらがどのように定義され、どのようにその定義が臨床アウトカム尺度または関連した尺度およびターゲットとなるPBMの双方に関連付けられているか——を検討する必要があるだろう。臨床アウトカムとPBMにほとんど一致する部分がなければ、マッピングはおそらく失敗に終わるだろう。推奨されたPBMを含め、さまざまなアウトカム尺度の記述的な内容の比較は出発点として有用である。これにより、それぞれの測定尺度の健康における具体的側面が明らかになる。マッピングが適切なアプローチであるために、PBMや臨床アウトカムが同じ症状や機能的な能力/機能的障害を扱う

ことは必須ではないが、同じ基本的概念を測定する必要がある。

状態推移モデルなどの多くのモデルは、一般に比較的少数の離散的な健康状態を定義する。他の状況では、ある部分はマッピング研究に、そしてまた別の部分は他のエビデンスに由来する健康状態の組み合わせが必要となるかもしれない。たとえば、あるモデルは、疾患のアウトカム尺度と患者が受けている治療、または患者の経験した副作用、あるいは併存疾患に基づいて、健康状態を識別するかもしれない。状況によっては、マッピング研究は疾患特異的なアウトカム尺度と健康効用との関係のみを示すのに対し、別の臨床研究は副作用による効用値への影響についての情報を与えるのに向いているかもしれない。マッピングや他の既存のエビデンスは、これらのエビデンスギャップに取り組む際にさまざまな選択肢を提供できる。

PBMの1つである効用値へのアウトカムのマッピングは、通常、回帰分析を用いて行われる。変動域の一端では、関連する健康の基準を持つ患者の効用値の平均や分散を単に用いるだけで、回帰モデルが完全に回避されうるというまれな状況もある。重要だと考えられる新たな共変量がなく、各カテゴリー内の患者が十分に観察されて、効用値を説明できる単一の要約尺度が存在する場合、この簡単なアプローチは完全に合理的である。しかし、これらの条件を備えていない場合、他のCEAへのマッピングの一般化可能性は制限されうること留意しておくべきである。

測定データの範囲外の追加の共変量および/または外挿が必要となると、大抵の場合、回帰型の分析の実施が必要となる。これはおそらく、健康効用の推計にまとめて使用される、疾患の異なる次元を反映する複数の疾患特異的なアウトカム尺度が存在するからであろう。もしくは、健康効用に社会人口学的変数を組み込みたいと分析者が考えるからかもしれない。たとえば、年齢は多くの状況に関係する変数となりそうだが、それは健康およびQOLと関連することになるからである。マッピングに回帰分析を検討する別の理由としては、データで観察された疾患重症度の範囲を超えた外挿が必要な可能性があげられる。データ範囲を超える外挿はいかなる状況

表 1 モデリング前の検討事項の概要

1. マッピングの用途または潜在的用途を検討する。
 - a. コホート決定モデル, 患者レベルのモデル, または試験に基づく費用効果分析に使用するのか?
 - b. マッピングからの効用値の推計を必要とする健康状態は何か, そして, どのようにそれをPBMへ関連付けるか?
 - c. 効用値が必要とされる疾患重症度の範囲は何か?
2. 臨床上の説明変数, 従属変数のPBM, およびそれらが重複する範囲の説明を提供する。
3. 回帰によるマッピングが必要であれば評価する。
 - a. 効用値の推計が必要な健康状態がいくつあるか?
 - b. 重要な追加的な共変量はあるか?
 - c. 各カテゴリー内に十分な観察記録があるか?
4. 2つ以上のデータセットが推計に利用可能ならば特定する。候補データセットの特性を比較する。
5. サンプルデータセット内の患者データの分布は, 費用効果分析の対象をどの程度, 反映しているか? 特に, 疾患重症度の両端は代表できているか?
6. 患者が受ける治療の種類は健康効用と臨床上のアウトカム尺度の関係に影響を与える可能性があるか?

PBM, 選好にもとづく尺度

でも避けるのがベストだが, 常に避けられるとは限らない。特にデータセットが併存疾患や他の重症度に関し除外基準のあるランダム化比較試験に由来する場合, 患者の疾患重症度のすべての範囲を含まないデータセットに基づくマッピング研究がよくある。これは特に, 慢性疾患患者の決定モデルにおける必要性とは対照的で, 患者の寿命をモデル化して疾患のすべての可能な範囲を網羅する。

これらの回帰分析のいくつかの手法はバイアスを示すことがよく知られており, その程度は対象となる効用値尺度に依存する部分がある。詳細はモデリングとデータ分析の節で後述するが, 現時点で留意すべき点は, 極端な疾患重症度でバイアスが最大化しやすいことである。つまり, 重症疾患を伴う患者に対して, これらのアプローチは健康効用を過大評価し, 健康な人々の健康効用を過小評価する¹⁸⁾。この点に注意して, 分析者はCEAの要件を評価する必要がある。たとえば, 決定モデルが対応する疾患の範囲はどの程度か? この判断は対象となる技術が評価される患者経過のある時点(モデルベースライン)で対象患者の特性に対して行われるだけでなく, モデルによって扱われる将来の健康状態の範囲にも基づくべきである。(疾患の進行, 治療への反応, そして疾患の寛解を含む)患者が経験する多様な経過を取り巻く, 長期の外挿を扱う可能性があるため, 場合によっては非常に広範囲の疾患重症度をカバーできる。

マッピングのためのデータ

マッピング関数を推計するデータセットの要件にも, 同様の考慮が必要である。追加要件として明らかかなことは, 候補のデータセットが, 関連する臨床アウトカム尺度とターゲットのPBMを同時に備えた, 個人を対象とする研究に由来しなければならぬことである。マッピング研究にランダム化試験の方が望ましいという理由はない。上述しているように, 確かにランダム化試験は厳格な組入基準および除外基準と限られたフォローアップゆえに, 疾患重症度に関して他の種類の研究に比べて患者が多様化していないことが多い。観察研究は, 代表的な患者集団を対象とすることが多く, サンプルサイズが大きく, そして比較的 low コストで実施できる。2つ以上の候補データセットが存在する場合, 別の研究に含まれる追加データフィールドを設けるかどうかを検討すべきであり, それによって, 患者集団のサンプルサイズや一般化可能性, そして研究デザインによる潜在的なバイアスと同様に, 対象となるPBMがより正確に推計できる可能性がある。しかし, 引き続き行うCEAにおけるこれらの値の使用とバランスをとる必要がある。たとえば, 回答者の年齢等に関する情報の入手可能性は, おそらくモデルの適合性を改善するため, CEAに組み込まれるべきである。一般的な共変量が存在し, 患者と研究デザインの違いが共変量とPBMの関係に影響しないことが期待される場合に, データセットを組み合わせることが可能である。

推計における不確実性は最小化されるべきである。これは、多くの観察データセットを使用し、可能な場合においてはデータの範囲外での外挿を避けることで、部分的に促進される。データセットにおいて疾患重症度の範囲をCEAの集団と一致させることは重要であるが、マッピングモデルにおいて共変量として使用される他の患者特性の範囲もまた重要である。

最後に、分析者はデータセットのあらゆる潜在的なバイアスに注意を払う必要がある。この状況でのバイアスとは、説明変数として使用される臨床アウトカムの尺度の影響以外に、患者報告の健康効用値に影響する要因を意味する。たとえば、患者が受けている治療の種類が、マッピングのデータセットで測定されている臨床アウトカムとは無関係の有害事象と関連付けられる場合など、状況によっては何らのバイアスを生じる可能性がある。

モデリングとデータ分析

統計モデルの選定

効用値尺度は、多くの非正規分布特性を示す傾向がある。これらの尺度は、範囲の上下両方において一種の制限従属変数と考えることができる。定義では、1は得られる最大の値で、「完全な健康」に相当する。測定法または国別のタリフによって異なる下限があり、「ピット」状態と呼ばれることもある。効用値におけるこれらの限界は、「打ちきり」と同じではないことに留意して欲しい。

統計モデルの選定に影響を与える効用値の分布の追加的側面は、分布の大きなスパイク（完全な健康の上限で一般に見られる）、歪度、多峰性と取り得る値の範囲におけるギャップである。図1は、さまざまな疾患領域のEQ-5D-3L（UKタリフ）の分布例を示している。これらの特徴が現れる範囲は、マッピング研究の対象であるPBMの測定法とスコアリングアルゴリズム、そして患者集団の性質による。これらの特徴のいずれかの存在が、単純な統計的回帰法を難しいものにし、同時に存在する場合に複雑化する。

マッピングの文献でもっともよく研究されている患者報告アウトカムであるEQ-5D-3Lの分析に線形

回帰が使用される場合、これらの分布の特徴が系統的なバイアスの結果であるという確かなエビデンスが存在する¹⁹⁻²¹⁾。同様の所見が、(制限従属変数を扱うために設計された)Tobit²⁰⁾、Two-partモデル²²⁾(完全な健康に見られる観察量に対応)、打ちきりのある最小絶対偏差モデル^{23,24)}等に当てはまることが示されている。これらの報告に共通する知見は、中程度の健康状態に関連付けられる期待健康効用が過小評価される一方、より重症な健康状態では過大評価されることである。これらのバイアスを伴うマッピング研究が経済評価に使用される場合、臨床効果のある治療の費用対効果が実際よりも低く表される。複数の研究で、このバイアスの規模が小さくないことが示されている^{12,13)}。

最近の研究で、マッピングのさまざまな統計手法の性能が比較されている。一連の手法により、要約した効用値を直接推計できる。これらの直接的な手法の中で、2つの手法の性能を裏付ける経験的根拠が存在する。それらは、制限従属変数混合モデル法^{13,20)}とベータ分布に基づく回帰アプローチ^{18,25)}である。両者とも、あらゆる効用値に特有の制限的な性質を反映し、前者は前述した効用値分布の他の鍵となる特性も反映している。

別の方法として、間接的な方法は、二段階の手順の一部として効用値を推計する²⁶⁾。これらの方法は、ある設定では標準的な方法よりも改善されていることが実証されている^{13,27-29)}。第一段階の、いわゆる応答マッピングモデルは一連の(従属または独立な)個別的回帰関数を使用して、対象となるPBMの記述システムにおける各個別領域上のレベルを推計する。順序カテゴリデータに適したモデルは、この第一段階と組み入れられた次元応答間の相関に使用されるべきである²⁸⁾。その後、手順の第二段階として、記述システムにおける各健康状態に割り当てられた確率とそれら状態に関連する効用値に基づき、期待効用値が単純に計算される。この段階分けによって、分析者は、その必要性に従い、第一段階で、効用値のタリフを推計したモデルに適用することができる。しかし、モデルの適切性と適合性は、検証されたタリフに特異的であることに注意すべきである。さらに、応答マッピングモデルは、記述システムの各レベルで十分な観察を必要とする。これ

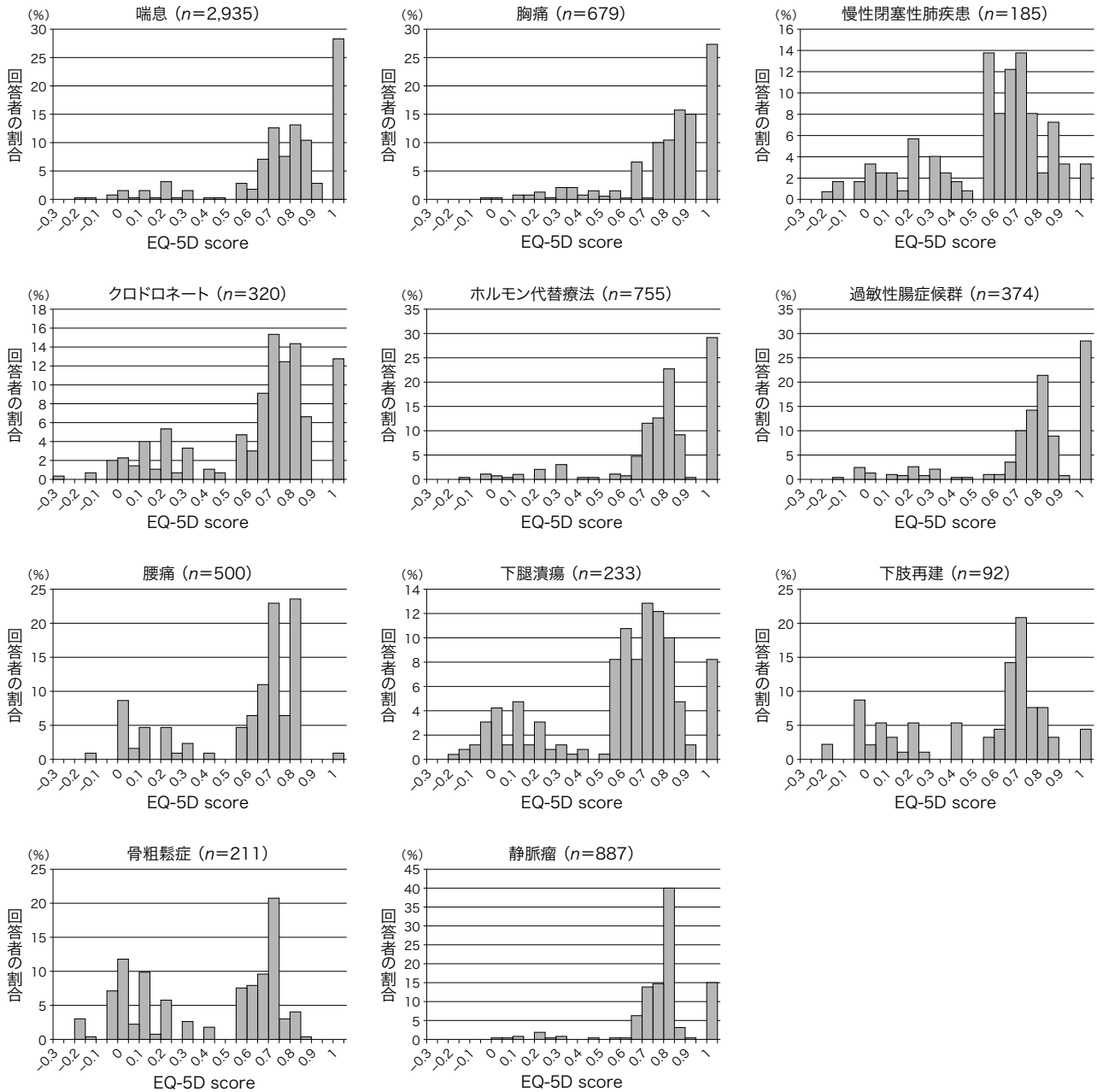


図1 一連の研究におけるEQ-5D-3L 英国タリフ・スコアの分布
EQ-5D-3L, three-level Euro Qol five-dimensional questionnaire

がなければ、モデルを推計することはできない。
異なる方法の性能は、対象の効用値尺度、問題とする疾患や患者集団、説明的な臨床変数、およびCEAで使用しようとする形式によって変化するため、われわれは具体的な方法を推奨しない。そのため、性能に優れるとの既存の経験的エビデンスがあり、対象の効用値尺度（特に費用対効果モデルでの実装結果において、問題を回避できる限られた範囲

の効用値)の主要な特徴を損なわない種類のモデルを使用することが賢明であることを示唆しておく。
言うまでもなく、マッピングでは、一般的な統計解析のための実施基準を分析者が遵守しなければならない。ここで、特にマッピングに関連する実施基準のいくつかの側面を強調しておく。たとえば、対象となる効用値尺度の分布のプロットは、効用値の直接的な分析にあたって、潜在的に適切なモデリン

グ法を検討する出発点となる。分析者は、理論的妥当性を持ち、その主要な仮定が適用でき、マッピングの論文において妥当性を裏付ける経験的根拠が存在するモデルを使用すべきである。これらの基準を満たさないモデルを使用する場合は追加的な正当化が必要になり、結果は精査されるべきである。追加的な正当化は、特定の適用でマッピングがバイアスを伴わないことを示す根拠、またはそのバイアスの性質がCEAでのマッピングの使用において問題とならないことを示す根拠の形をとりうる。たとえば、分析者が集計しようとするコホート決定モデルにおいて、少数の健康状態のみが定義され、これらの健康状態が極端な不健康/健康ではない場合、マッピングからのバイアスは、推計される費用対効果にわずかな影響しか及ぼさないかもしれない。それでも、事前になんらかの潜在的バイアスが及ぼす影響を評価することは困難である。

ほとんどの状況では、前述の基準を満たすマッピング法の使用が極めて重要である。これは、費用対効果において、バイアスのかかった推計値の範囲と影響はとて大きく、また、実現可能な効用値の範囲外に予測される可能性があるためである。たとえば、疾患重症度の両極端での健康状態におけるモデルに基づくCEA、患者個人のシミュレーションモデル、または単一の臨床試験とともに実施されたCEAのような個人レベルのデータに基づく分析はすべて、不適切なマッピング法が適用された場合、実質的にバイアスのかかった費用対効果の推計となるリスクがある。

いくつかの種類モデルは、反復推計を要する。分析者は、古典的³⁰⁾またはベイズの枠組み³¹⁾によって実施するかどうかにかかわらず、推計したアルゴリズムの適切な収束を保証する必要がある。

一般的に、候補のデータセットは同一の個人を時間の経過とともに複数観察した内容から構成される。通常、すべての観察内容を活用すべきである。マルチレベルモデルは、これらの観察値の相関を反映するために使用されうる。少なくとも、クラスタ化された標準誤差が計算されるべきである。共変量とPBMの関係に断絶があると信じるに足る理由がある場合、分離モデルを推計し、パラメータの安定性を検証するべきである。

共変量の選定

ほとんどの状況で、マッピングが行われるデータセットは幅広い潜在的な説明変数を含む。分析者にとっての最初の決定は、臨床的な有効性データと選好にもとづくデータの間の主要な繋がりとなる、非PBMの選択に関係する。非PBMは、臨床研究で使用される主要なアウトカム尺度、あるいは副次的アウトカムの中で唯一のQOL測定法であるため、多くの状況で一目瞭然である。しかし、これらの尺度は、しばしば個別質問の形式で、次元スコアまたは単一の要約スコアとして報告される。一般的には、アウトカム尺度から分解された情報を説明変数として使用する回帰モデルにより大きな説明力がある。しかし、これは説明変数の数を増やすだけでなく、幅広い利用が可能な形で臨床上の根拠との関連をも提供しない可能性がある(たとえば、Longworthら³²⁾による、欧州がん研究・治療機構による36の個々の質問への回答をモデル化した研究)。これは、リウマチ関節炎を例に説明できる。一般的に費用対効果研究は、選好に基づく測定法に対応付けた健康状態質問票を使用する³³⁾。この質問票は、2~3の個々の質問からなる8つの下位区分から得られる、0~3の範囲をとる機能障害の要約スコアである。分析者は、0~3の要約スコアではなく、個別項目または次元スコアを説明変数として使用することで性能に優れたモデルを見つけることができるかもしれないが、これらが共変量の選定のための唯一の基準となるべきではない(たとえば、Bansbackら³⁴⁾を参照)。マッピング関数が個別の質問または各構成要素のスコアから健康効用を推計するために使用され、臨床試験とともに経済評価が実施されるケースにおいて、こうしたアプローチは有用である。しかし、複数の臨床試験からのデータを統合する決定モデルは、一般に要約スコアのみを報告する公表結果に依存している。

他の状況では、分析者は複数の疾患特異的なアウトカムを選択できる。たとえば、強直性脊椎炎では、一般的に、臨床研究では疾患活動性および機能障害に関するBath Ankylosing Spondylitis Disease Activity Index (BASDAI) および Bath Ankylosing Spondylitis Functional Index (BASFI) アウトカム尺度の両方を報告している。選好に基づく測定法との概念的重複

表2 モデリングとデータ分析に関する推奨事項の概要

1. 費用効果分析に正式な回帰によるマッピングモデルが必要なのか、それとも患者のサブサンプルの平均値をとるのが適しているのかを検討する。
2. 回帰が必要な場合、下記に基づいてモデルを選定する。
 - a. モデルの仮定が対象とする効用値の測定方法に対応しうる、もっとも単純な統計モデルの種類を検討する。効用値のデータ分布のプロットを使用して、その選択を公表する。
 - b. さまざまな手法の性能に関する既存の経験的根拠。どんな特定の疾患領域からの根拠においても、これが制限される理由はない。
 - c. マッピングが使用される費用効果分析の種類、およびバイアスのかかった推計が結果に影響を及ぼす程度。
3. 応答マッピングの場合、記述システムでのカテゴリカルデータの順序性を尊重するモデルを選択するべきである。そして、期待値は解析的に計算されるべきである。
4. 好ましいマッピングモデルの選定は、すべての回帰分析に共通する実施基準に適合する反復プロセスである。
5. 共変量は事前に理論的に正当化されるべきである。共変量の除外は、費用対効果モデルで使用されない場合でも、モデル定式化の誤りのリスクとなる。

は、複数の測定法の組み入れと、それによるモデルの適合によって改善される可能性がある。

共変量には社会人口学的変数、疾患特性および治療法もありうる。モデルの定式化の誤りを回避するために共変量を組み込むことが、実施基準である（誤差項およびバイアスのある係数の推計値に配分される欠落変数の影響が起こる）。経済評価は、これらそれぞれの説明変数の直接的な使用のためにデザインされていないかもしれないが、それでもこの点は変わらない。分析者は、マッピングを使用し、状況に適切な説明変数の値を単純に設定することができる。これは、説明変数が欠落するよりも好ましい。もちろん、儉約性のあるマッピングモデルを開発するために高い相関を持った共変量の組み入れを回避する判断が、ここでは必要である。

共変量は事前に理論的に正当化され、問題の共変量が特定のCEAで直接的に使用される/されないにかかわらず、分析者が結果を使用できる方法で報告されるべきである。たとえば、CEAのマッピング関数を使用する際、多くの場合共変量として年齢の組み入れが必要であり、統計的に有意でなくても、好ましいモデルに留め置かれるべきである。これによって、臨床アウトカム尺度の一部として取り込まれる内容から独立し、どんな年齢の影響も適切に反映されることになる。マッピングが臨床試験と並んでCEAに使用される場合、マッピングデータセットおよび臨床試験の両方に共通する共変量を用いて、互いの一般化可能性を改善できる。本セクションからの推奨事項を表2にまとめた。

マッピング研究の報告

マッピング研究は、医療技術評価、価格設定または償還の意思決定を行う機関に提出される根拠の重要な要素を形成する。その知見は個人レベルのデータへアクセスできない人々がマッピングの品質と関連性を完全に評価できる方法で報告されなければならない。この透明性の要件に加え、CEAの結果を使用するために十分な情報が報告されれば、他の分析者にとっても役立つだろう。

データセット

2つ以上のデータセットをマッピングに使用することが許容される場合、少なくとも選定根拠の定性的評価が提供されるべきである。研究の鍵となる特徴とデザインが記述されなければならない。推計データセットで使用されるサンプルの特性も、完全に説明する必要がある。マッピングモデルで 사용되는すべての変数と同様に、潜在的に重要な臨床的、症状的、人口統計学的または他のPROMsの変数は、中心傾向と分布の尺度の面から記述されるべきである。説明変数として使用される疾患に特異的な尺度によって記述されるように、疾患重症度の両端の患者の観察記録のすべての分布に、特に注意を払う必要がある。これは、CEAの対象となる患者とサンプルが重複する範囲を表示し、観察データ外に必要な外挿の範囲を示している。

研究全体、そしてマッピング研究での使用のために選ばれたサブサンプルの両方において、患者のサ

ンプリング方法に関する完全な情報が提供される必要がある。

数多くの研究が、同一個人に対する時間経過に伴う複数の観察を含む。この状況で、こういった複数の縦断的な観察の経時的な測定値におけるパターンと、これらの観察期間にわたって変化する患者の特徴の報告が重要である。たとえば、相当のフォローアップ期間がある場合、年齢はベースラインを実質的に変える重要な変数となる。利用可能な観察の数は、選択した共変量の組み合わせによって異なり、これによって最終分析と研究サンプル全体の記述に実質的な相違が生じる可能性がある。これは、適合の尺度や罰則つき尤度統計量を使用するモデル間の比較可能性も示す。

統計モデルの種類と正当化

先に概要を示したように、分布の特徴から生じる効用値データの分析に固有の統計上の課題が多く存在する。分析者は、事前に、問題となる対象のPBMで別の手法を試した既存の文献、推計データセットの分布の特徴の精査、およびあらゆる将来的な費用対効果研究での利用を提案されたマッピング関数を参照し、選択肢となる手法を選定し、正当化しなければならない。

モデルの代数的記述は、透明、簡潔、明確で、どんな有能な分析者によっても、その結果は正確に使用されることを確実にする。他で記述されていない標準的でないモデルは、常にそのような記述を含んでいなければならない。マッピングの回帰モデルによる数組の共変量の予測値の例が報告されるべきである。いくつかの公表された研究の中には、ユーザーが定義した入力値からの予測を計算する追加的なソフトウェアが提供されている^{26, 34)}。

共変量使用の正当化と定式化方法

マッピングに使われるデータセットは、一般的に、分析者に広範囲の潜在的説明変数を提供する。これらは、しばしば複数の構成要素または要約インデックススコア(2つ以上の可能性がある)として採点される可能性があり、疾患特異的なアウトカム尺度、臨床的尺度、症状に特異的な情報、そして、特に人口統計学的変数を扱う。理論的正当化は、統計解析

において精査されるセット内の全変数を含めて行う必要がある。疾患特異的なアウトカムにおいて扱われるQOLの次元を評価し、効用に基づいた尺度によって扱われるものに対比することは有益である。

潜在的に大規模な説明変数の集合から、少数を特定の形式で含む傾向のある好ましいモデルへの移行に使用される方法については、詳細に検討する必要がある。これらの回帰モデルの決定には、数多くの方法がある¹⁰⁾。

モデル選定の結果と性能

モデルの種類を選定する際の理論的な正当化には、対象の効用値尺度に特に焦点を当て、先行文献や実施されるマッピングに固有の特徴が提供されるべきである。回帰モデルは、明示的に認められ、そして、妥当性を検証または評価されるべき仮定を置く。マッピングの用途がわかっている場合にはそれも議論されるべきである。関連する側面には、結果が使用される疾患の範囲、不確実性の検討方法、共変量を条件とした期待効用値のみが分析に必要とされるのかどうか(主にコホート決定モデルにおけるケースとして)、またはシミュレートしたデータ(試験に基づく分析または患者レベルのシミュレーションモデルとして)が必要なのか、が含まれる。

結果は、透明性のある形で報告される必要がある: 結果の読み手は、実行可能なモデルの中から好ましいモデルを選定された過程を認識しなければならない、そして、そのプロセスの妥当性を判断するために十分な情報が提供されなければならない。これは、読み手が好ましいモデルの性能を十分に評価できる必要があることを意味する(そして、あまり好ましくないモデルの性能に関して、少なくとも複数の側面に関する内容を必要とする)。モデルを組み立てるプロセスの各段階での判断が必要となる; 報告では、これらの判断と根拠を強調する必要がある。読み手が将来の費用対効果研究にマッピングモデルの結果を使用できるように、十分な情報提供がなされるべきである。

性能の一側面として特に重要なのが、モデルの適合性—モデル化された値がデータ内で観察されるものとどれだけ一致するか—である。好ましいモデルへの移行は、全体の適合性に基づいた、いくつ

かのルールに機械的に従うべきではない。特別な共変量を含むべきかどうかといった、状況に沿った個別の判断が必要である。しかし、最終的な好ましいモデルに関しての適合性に関する詳細も必要である。特に単独で示される場合、 R^2 等の適合性に関する要約値は非常に限定的な価値しかなく、後にCEAで使用されるマッピングの妥当性にわずかな情報しか提供しない。QOLの(当然な)主観的な性質を考えれば、患者間のばらつきの程度はQOLのデータにおいて本質的に高い。この結果、これは、比較的低い R^2 の統計値となる。赤池情報量基準やベイズ情報量基準のような罰則つき尤度統計量は、モデルの種類における定式化の比較のための、より適切な手段を提供する。平均絶対誤差および二乗平均平方根誤差等の適合性に関する他の要約尺度は、マッピングに関する文献において一般的に適用されている。これらの尺度は予測の分野に由来している。そのため、従属効用値の限定的な範囲と患者アウトカムデータにおける固有のばらつきの程度ゆえに、マッピングの分野に適用される場合、これらの尺度は非常に感度が低く見えることを認識しておく必要がある。モデルの適合性のあらゆる尺度において、臨床的な説明変数に関して定義されたサンプル全体と具体的なデータ範囲の両方が報告されるべきである。臨床的な変数を条件とした平均予測効用値と平均観察効用値は、あらゆる系統的なバイアスの存在および位置(たとえば、Wailooら³⁵⁾を参照)ならびにいつそのバイアスが発生するかを特定するのに役立つ。

モデルの適合性は、データと比較して予測値の点推計の参照によってのみ評価されるべきではない。この後で記述するように、患者のばらつきが組み入れられたならば、予測値周辺の不確実性とモデルのアウトプットを検討する必要がある。

結果の報告

すべての係数値は、正確な推計を可能にする十分な小数点以下の桁数が報告される必要がある。サンプル平均値周辺での再スケーリングと、中心化した共変量がこれを促進する。推計された係数と関連した相関における不確実性は、CEAのパラメータの不確実性を反映させるために必須である。そのため、

共分散行列を定常的に報告することで³⁶⁾、確率的感度分析の実施が可能となる。パラメータの不確実性に加え、個人レベルでデータを帰属させるためのマッピング関数の使用(たとえば、臨床試験とともに分析を実施する場合は、個人レベルのばらつきの反映を必要とする。現実世界のデータでは、同一の観察特性を持つ個人から同一の健康効用が報告されないことははっきりしている。もし、マッピング回帰モデルが単にこれらの個人に対し、同一の条件つき期待値を帰属させるために使用されたならば、個人レベルの説明のつかないばらつきが無視され、臨床研究とマッピングの結果の両方で誤った報告がなされることになる。この想定されるばらつきの程度と形態に関する情報は、マッピングの回帰の誤差項の分布に含まれ、これを反映するシミュレーション方法の基礎として使用できる。そのため、誤差項の詳細に関する定常的な報告も重要である。オンラインで入手可能な資料によって、公表されるマッピング研究にはこれらの重要項目を含めるべきである。

モデル選定に関する前述の指針は、効用値尺度の許容範囲外の推計が可能なモデルを選定すべきではないことを示唆している。しかし、このようなモデルが選定されている場合、不確実性またはばらつきの分析用にマッピング関数からサンプリングをする際に、これらのサンプルが許容範囲外にある頻度を報告する必要がある。これらの許容不能な値がその後どのように使用され、あるいはCEAで修正されるかについても報告されなければならない。具体的なCEAを想定しないままマッピングが行われる場合、モデルからシミュレーションしたデータセットの結果を報告することが有用である。これは、将来のCEAに知らせるのに役立つ、オリジナルデータの分布をモデルからシミュレーションしたデータと比較する手段を形成するのに役立つ(そして、モデル選定プロセスの一部として利用可能である)。

経験的なバリデーション

他の統計モデルと同様、マッピングモデルにはバリデーションが必要である。ここで報告した指針の多くは、この要件に基づいている。使用されるデータセットと決定問題の記述、モデル構築のプロセス、

表3 マッピング研究の報告に関する推奨事項の概要

1. マッピング推計の候補であるデータセット間の関連する差異を記述する。
2. 選定したデータセットの十分な詳細を提供する。どのように研究が行われ、患者がサンプリングされたかについて記述する。患者の疾患重症度の分布を含む、ベースラインおよびフォローアップ時の特性を提供する。応答の縦断的な欠測について記述すべきである。
3. 効用値データの分布をプロットする。
4. 対象の効用値分布の特性およびマッピング関数の用途案を参照し、選定されたモデルの種類を正当化する。
5. 対象の効用値の推計方法によってカバーされる健康の次元と、説明的な臨床的尺度によってカバーされるものを比較する。
6. 最終的なモデルを決定するためのアプローチを記述する。実施された検定および行った判断を含める。
7. 適合性の要約尺度は、全体のサンプルにとって限定的な価値しか持たない。臨床上のアウトカム尺度によって測定される疾患重症度を条件とした適合性に関する情報を提供する。臨床的な変数を条件とした平均観察効用値に対する平均予測効用値のプロットを含めるべきである。
8. 係数値、誤差項、分布、分散および共分散が必要である。
9. いくつかの共変量のセットによる予測値の例を提供する。使用者が定義した入力値に対する予測値を計算するプログラムの提供を検討する。
10. マッピングの回帰におけるパラメータの不確実性は、PSAの標準的手法を使用して反映されるべきである。費用対効果分析でのモデルの適切性の評価は、PSAのための効用値の分布を検討し、特にこれらがPBMの許容可能な効用値の範囲外にあるかどうかについて特に焦点を当てるべきである。
11. マッピング関数からデータを帰属する場合、シミュレーション方法と誤差項の分布に関する情報を使用して、個人レベルのばらつきを組み込む必要がある。これらのシミュレートしたデータは、PBMの許容可能な範囲と比較した値の評価を含む、生の観察データとの比較が可能である。
12. 別個のデータセットにおけるマッピングの結果の再推計、あるいは他のバリデーションの形式は、定常的に必要とされるわけではない。

PBM, 選好にもとづく尺度; PSA, 確率的感度分析

そして最終的な好ましいモデルの性能といったこれらの各要素が、バリデーションに関する情報を提供する。関連する患者集団内でモデルの予測が正確であることとわれわれはどの程度、自信を持てるか、そして、その予測は他の同様の患者集団とどの程度関連性を持つのか？ 英国における既存のマッピング指針は、経験的なバリデーション¹⁵⁾を推奨しており、この点において2つの異なる研究からなる(外的バリデーション)、または1つのデータセットを分割した(内的バリデーション)、2つのデータセットによるモデルの推計として記される。数多くの手法が内的バリデーションに使用可能である(たとえば、ブートストラップによる手法)。多くの場合、これらの経験的バリデーション技術は分割を検討するには不十分なサンプルサイズのただ1つの候補データセットしかないことから、選択肢にはならない。

これらのバリデーション方法のいずれかが実施可能であったとしても、マッピングの設定およびこれらの分析が提供する情報の追加的な値において、使用可能な手法のうちどれが最適かという不確実性が

残る。サンプルの分割は推計に対して縮小したサンプルサイズというさらなるペナルティを課す。これらの理由から、すべてのマッピング研究に経験的なバリデーションを推奨することは時期尚早と考える。これは、CEAに他の回帰に由来する入力値を引き受けるアプローチとも一致している。

効用値データの分析への代替の方法論的アプローチのバリデーションは、現実世界のhead-to-headの反復試験と異なる疾患領域からシミュレートしたデータセットを通じて得られるだろう。しかし、上記で述べた限界から、定常的な多標本のバリデーション法は、標準的な応用的マッピング研究には必要とされない。本節からの推奨事項は、表3にまとめた。

マッピングモデルから得た結果の使用

費用対効果研究のためのマッピングモデルの選定

分析者は費用対効果モデルにデータを追加するために、おそらく公表されている文献から、既存のマッピングを選定する必要があることが多いだろう。あ

る状況では、関心のある集団に合致する既存のマッピングが存在しないかもしれない。これは、患者の厳密な特性が人口統計学的変数、疾患のステージまたは疾患重症度に関して一致しないことが理由かもしれない。別の状況では、マッピングが異なる疾患の患者に基づいているなどのより根本的な相違がある。たとえば、欧州がん研究・治療機構 (QLQ30) は、どんな種類のがん患者にも PROM を使用する。このマッピングは乳癌患者のサンプルに基づいて推計されている³⁷⁾。CEA におけるマッピング研究の適切性を判断する際は、問題の患者間または疾患間の違いの評価に基づくべきである。これらの違いは、マッピングの共変量と対象の PBM の関係を一般化不可能にするだろうか？

予測値

経済評価のためのマッピングの主な用途は、一連の説明変数に対する平均の健康状態の効用値、つまり共変量を条件とした期待値を予測することである。ここで提示された指針に従うことで、モデルの定式化および推計された係数に十分な理解を与えられ、必要な期待値を得る方法が明確になる。将来的な参照のために、所与の共変量の期待効用値と標準誤差を報告することは、マッピング研究にも役立つかもしれない。公表された研究の中には、さらに進んで追加ファイルとして表計算プログラムを提供しているものもある^{26, 34)}。

ばらつき

誤差項の分布を含む、統計モデルの仕様のすべてとその推計された結果は、分析者が個人レベルのばらつきを反映させるために必要な情報を提供する。もっともシンプルなもの、単一の平均 0 (ゼロ) の正規分布に従う誤差項と報告された分散を含む。そのため、必要とされるあらゆる健康状態 / 患者特性のばらつきを反映するために、関連する条件付き分布から単純にサンプリングされている。

不確実性

確率感度分析は、医療経済モデルにおけるパラメータの不確実性を反映する標準的に認められた方法である。モンテカルロシミュレーションは、モデ

ルの定式化、係数と分散共分散の推計値が報告されるならば、マッピング研究に伴う、回帰モデル入力値に関連する同時分布からのサンプリングに使用される。

結 論

臨床研究への適切な PBM の組み入れは常に推奨されているが(この課題に関するガイダンスとして、Wolowacz らによる ISPOR 実施基準ガイド¹²⁾を参照)、これは経済評価のニーズにとって常に適しているか、あるいは十分というわけではない。そのため、マッピングは分析者が臨床的なエビデンスと経済評価に必要なエビデンスの間を橋渡しするために必要とされている。マッピング研究が適切に取り組み、透明性を持って報告され、その結果が適切に使用されるなら、意思決定者はこの方法で得られた推計値の妥当性に自信を持つことができる。

謝辞 Sorrel Wolowacz と Richard Willke の個人的な貢献に深く感謝する。ISPOR ミラノとアムステルダムの欧州学会でのわれわれのフォーラムにコメントしたすべての査読者に感謝する。われわれは、報告の草稿と投稿されたコメントを査読した次の方々に特に感謝する：Tony Ades, René Allard, Gouri Shankar Bhattacharyya, Lance Brannman, Michael Carter, David Cella, Akriti Chahar, Stephanie Yanjing Chen, Shiva Devarakonda, Salah Ghabri, Linda Gore Martin, Inigo Gorostiza, Thomas Grochtdreis, Michael Hagan, Nadine Hillock, Cynthia Holmes, Shrividya Iyer, Steve Kay, Jeanette Kusel, Ramanath K.V., Dawn Lee, Joanna Leśniowska, Sophia E. Marsh, Alan Martin, Nicholas Mitsakakis, Sharanya Murty, Nneka Onwudiwe, Guilhem Pietri, G.M. Rabiul Islam, Ghabri Salah, Mihail Samnaliev, Carsten Schousboe, Sarah Shingler, Fatema Turkistani, Uday Venkat. 彼らのフィードバックは本報告の質と ISPOR 班報告書の専門家の同意形成の両方を進めた。最後に、Theresa Tesoro の本班報告書の発展への支援に大いに感謝する。

参考文献

- 1) Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977;296:716-21.
- 2) The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
- 3) Brazier JE, Roberts JR. The estimation of a preference-based index from the SF-12. *Med Care* 2004;42:851-9.

- 4) Horsman J, Furlong W, Feeny D, Torrance D. The health utilities index (HUI): concepts, measurement properties, and applications. *Health Qual Life Outcomes* 2003;1:54.
- 5) National Institute for Health and Care Excellence. NICE Guide to the Methods of Technology Appraisal. London: National Institute for Health and Care Excellence, 2013.
- 6) Lopez-Bastida J, Oliva J, Antoñanzas F, et al. Spanish recommendations on economic evaluation of health technologies. *Eur J Health Econ* 2010;11:513–20.
- 7) Choices in methods for economic evaluation. 2013. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf. [Accessed December 9, 2016].
- 8) PHARMAC. Prescription for pharmacoeconomic analysis version 2.2. 2015. Available from: <https://www.pharmac.govt.nz/assets/pfpa-2-2.pdf>. [Accessed September 14, 2016].
- 9) International Society for Pharmacoeconomics and Outcomes Research. Pharmacoeconomic guidelines around the world. Available from: <http://www.ispor.org/PEguidelines/index.asp>. [Accessed February 1, 2016].
- 10) Mortimer D, Segal L, Sturm J. Can we derive an “exchange rate” between descriptive and preference-based outcome measures for stroke? Results from the transfer to utility (TTU) technique. *Health Qual Life Outcomes* 2009;7:33.
- 11) Kearns B, Ara R, Wailoo A, et al. Good practice guidelines for the use of statistical regression models in economic evaluations. *Pharmacoeconomics* 2013;31:643–52.
- 12) Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force report. *Value Health* 2016;19:704–19.
- 13) Pennington B, Davis S. Mapping from the Health Assessment Questionnaire to the EQ-5D: the impact of different algorithms on cost-effectiveness results. *Value Health* 2014;17:762–71.
- 14) Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. *Med Decis Making* 2014;34:919–30.
- 15) Devlin N, Parkin D, Browne J. Using the EQ-5D as a Performance Measurement Tool in the NHS (Report No. 09/03). London: Department of Economics, City University London, 2009.
- 16) Longworth L, Rowen D. NICE DSU Technical Support Document 10: the use of mapping methods to estimate health state utility values 2011. Available from: <http://www.nicedsu.org.uk/TSD%2010%20mapping%20FINAL.pdf>. [Accessed December 9, 2016].
- 17) Petrou S, Rivero-Arias O, Dakin H, et al. Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Health Qual Life Outcomes* 2015;13:106.
- 18) Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ* 2009;11:215–25.
- 19) Basu A, Manca A. Regression estimators for generic health-related quality of life and quality-adjusted life years. *Med Decis Making* 2012;32:56–69.
- 20) Khan KA, Madan J, Petrou S, Lamb SE. Mapping between the Roland Morris Questionnaire and generic preference-based measures. *Value Health* 2014;17:686–95.
- 21) Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D health state utility values. *Value Health* 2012;15:550–61.
- 22) Acaster S, Pinder B, Mukuria C, et al. Mapping the EQ-5D index from the cystic fibrosis questionnaire-revised using multiple modeling approaches. *Health Qual Life Outcomes* 2015;13:33.
- 23) Pullenayegum EM, Tarride J, Xie F, et al. Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? *Value Health* 2010;13:487–94.
- 24) Huang I, Frangakis C, Atinson MJ, et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. *Health Serv Res* 2008;43:327–39.
- 25) Hunger M, Baumert J, Holle R. Analysis of SF-6D index data: is beta regression appropriate? *Value Health* 2011;14:759–67.
- 26) Tsuchiya A, Brazier J, McColl E, Parkin D. Deriving preference-based single indices from non-preference-based condition specific instruments: converting AQLQ into EQ5D indices. Discussion Paper, Health Economics and Decision Science, University of Sheffield 2002.
- 27) Wailoo A, Hernández M, Philips C, et al. Modeling health state utility values in ankylosing spondylitis: comparisons of direct and indirect methods. *Value Health* 2015;18:425–31.
- 28) Gray A, Rivero-Arias O, Clarke P. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med Decis Making* 2006;26:18–29.
- 29) Conigliani C, Manca A, Tancredi A. Prediction of patient-reported outcome measures via multivariate ordered probit models. *J R Stat Soc Ser A Stat Soc* 2015;178:567–91.
- 30) Cramer JS. *Econometric Applications of Maximum Likelihood Methods*. New York, NY: Cambridge University Press, 1986.
- 31) Gelman A, Shirley K. Inference from simulations and monitoring convergence. In: Steve Brooks, Andrew Gelman, Galin Jones, Meng Xiao-Li, eds. *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman and Hall/CRC, Taylor and Francis Group, 2011. p. 163–74.
- 32) Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decisionmaking: a systematic review, statistical modeling and survey. *Health Technol Assess* 2014;18:1–224.
- 33) Tosh J, Brennan A, Wailoo A, Bansback N. The Sheffield rheumatoid arthritis health economic model. *Rheumatology* 2011;50(Suppl. 4): iv26–31.
- 34) Bansback N, Marra C, Tsuchiya A, et al. Using the Health Assessment Questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. *Arthritis Rheum* 2007;57:963–71.
- 35) Wailoo A, Hernandez Alava M, Escobar Martinez A. Modeling the relationship between the WOMAC Osteoarthritis Index and EQ-5D. *Health Qual Life Outcomes* 2014;12:37.
- 36) Ara R, Wailoo AJ. NICE DSU Technical Support Document 12:

the use of health state utility values in decision models. 2011:
Available from:
<http://www.nicedsu.org.uk/TSD12%20Utilities%20in%20modelling%20FINAL.pdf>. [Accessed December 9, 2016].

37) Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur J Health Econ* 2010;11:427-34.

本稿は Wailoo AJ, et al. Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health*. 2017 Jan;20(1):18-27. doi: 10.1016/j.jval.2016.11.006. の日本語訳である。翻訳と掲載にあたり、Copyright Clearance Center (License No: 5439110996837) の許諾を得た。

本成果物は、厚生労働科学研究費補助金「関連学会の取組と連携した PRO ガイドラインの作成 (研究代表者: 立命館大学 下妻 晃二郎, 課題番号: 20AC1003)」および「医薬品・医療機器等の費用対効果評価における公的分析と公的意思決定方法に関する研究 (研究代表者: 国立保健医療科学院 保健医療経済評価研究センター 福田敬, 課題番号: 21AA2005)」の支援により作成した。