

本稿はSpringer社の許諾のもとに、MAPS (MAPPING onto Preference-based measures reporting Standards : 選好に基づく尺度へのマッピングの報告基準) 声明について、その内容に解説を加えた文献 (Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. The MAPS reporting statement for studies mapping onto generic preference-based outcome measures: Explanation and elaboration. Pharmacoeconomics. 2015;33(10):993-1011.) を翻訳したものである。

## <解説>

### 選好に基づくアウトカム尺度へのマッピング研究のための MAPS 声明 —解説と詳細—

訳 白岩健<sup>1)</sup>, 船越大<sup>2)</sup>, 村澤秀樹<sup>3)</sup>, 下妻晃二郎<sup>2,3)</sup>, 斎藤信也<sup>4)</sup>, 福田敬<sup>1)</sup>

<sup>1)</sup> 国立保健医療科学院保健医療経済評価研究センター

<sup>2)</sup> 立命館大学大学院生命科学研究科

<sup>3)</sup> 立命館大学生命科学部

<sup>4)</sup> 岡山大学大学院保健学研究科

#### 抄録

**背景:** 医療経済評価において、「マッピング」のプロセスは、他の健康指標や尺度のデータから健康効用値 (health utilities) を推定する際に使用されることが多くなっている。しかし現在、マッピング研究を報告するためのガイドラインはない。

**目的:** 本研究の全体的な目的は、マッピング研究を報告する際に考慮すべき必須項目のチェックリストを作成することである。このMAPS (MAPPING onto Preference-based measures reporting Standards : 選好に基づく尺度へのマッピングの報告基準) 声明は、完全に透明性のある報告を促進するためのチェックリストである。本稿では、MAPS声明に含まれる項目の解説と詳細を提供する。

**方法:** これまでに報告チェックリストや報告ガイドラインが公表されていなかったため、新たな報告項目のリストとそれに関する解説を作成した。この広範なリストから必須となる報告項目を明らかにするため、アカデミア、コンサルタント、医療技術評価機関、そして生物医学雑誌の編集者コミュニティの代表者とともに、2ラウンドのデルファイ変法による調査を行った。

**結果:** 当初の29個の候補項目リストから、23個の必須報告項目を作成した。これらの項目に番号をつけ、6つのセクションに分類した。具体的には、(i) タイトルと抄録、(ii) 緒言、(iii) 方法、(iv) 結果、(v) 考察、(vi) その他である。各項目には、それぞれの推奨を要約し、過去の公開論文の中から優れた報告を具体例として示して、この推奨に詳細な説明を加えた。

**結論:** MAPS声明は、マッピングの研究において、明確で透明で完全な報告を促進することが期待されている。本声明の対象は、マッピングアルゴリズムを開発する研究者、マッピング研究の原稿査読プロセスに関わる査読者や編集者、研究の資金提供者である。MAPSの作業グループは、この報告チェックリストの更新の必要性について、5年後に検討する予定である。

## 1 緒言

選好に基づくアウトカム尺度に「マッピング (mapping)」する手法は、医療経済評価に用いる健康効用値 (health utility) を算出する手段として用いられることが多くなっている[1]。マッピングでは、他の健康指標・尺度のデータから選好に基づくアウトカム尺度、すなわち健康効用値を予測するためのアルゴリズム開発や使用などを行うものである。マッピングの元となる尺度は、非選好型の健康アウトカム指標や尺度、例外的にはその地域の医療技術評価機関が使用を好まない、選好に基づくアウトカム尺度であるかもしれない。一度開発されたアルゴリズムは、対象となる選好に基づく尺度の測定値がない状況で健康効用値を予測するために、マッピング元になる尺度を含む臨床試験や観察研究、経済モデルに適用できる。予測効用値は、個人レベルデータを扱う標準的手法を用いて解析されるか (例えば、試験に基づく経済評価においてなど)、決定分析モデルにおける各健康状態のために要約統計量を求めることになる。

ここ数年、健康効用値を予測するためのマッピングを用いた研究発表が急増し、この分野の研究発表をまとめたデータベースが現れ始めている[2]。技術評価に関わる著者ら[3]と組織[4]の一部は、マッピング研究実施のための技術的なガイドラインを示している。しかし、マッピング研究報告のためのガイドラインはいまだ存在しない。健康関連の研究に則して[5]研究の相対的価値を評価できるよう、マッピング研究は十分かつ透明性を持って報告される必要がある[6]。さらに、マッピング研究からの誤った知見に基づき、新技術に対する規制や償還を決定すれば大きな機会費用が生じることになるだろう。このことが、本稿で解説し詳述する、MAPS (MApping onto Preference-based measures reporting Standards: 選好に基づく尺度へのマッピングの報告基準) 声明の作成に関連している。

このMAPS声明の目的は、マッピング研究を報告する際に考慮すべき必須項目のチェックリスト形式で、推奨を提供することである。このチェックリストは、研究者による完全に透明性のある報告を促進することが期待される。したがって、マッピング研究の実施ではなく、その報告の質を向上させることに重点が置かれているが、本声明が研究の方法論的厳密性を間接的に向上させる可能性もある[7]。本MAPS声明は、マッピングアルゴリズムを開発する研究者、研究への資金提供者、そしてマッピング研究の原稿査読プロセスに関わる査読者や編集者を主な対象としている[5,6]。この声明の作成にあたり、「マッピング」という用語は、他の健康指標や尺度のデータを用いて選好に基づくアウトカム尺度の結果を予測するすべてのアプローチを網羅して用いられており、一部の研究者が使用している「クロスウォーク (cross-walking)」や「効用値への変換 (transfer to utility)」などの関連する用語を含む[1,8]。同様に、「アルゴリズム」とい

う用語は、統計的関連性およびより複雑な演算を含む最も広い意味で用いている。

## 2 MAPS 声明の作成

本MAPS声明の作成にあたっては、近年出された医学研究の報告ガイドライン[5]とその他の報告ガイドライン[9-14]をモデル化して用いた。マッピング研究の報告に関する声明について学術誌から作成要請があり、6名の医療経済学者 (SP, ORA, HD, LL, MO, AG) と1名のデルファイ専門家 (RF) からなる作業グループが構成された。作業グループメンバーの1人 (HD) は、臨床あるいは健康関連QOL (quality-of-life) 尺度をEQ-5Dにマッピングする研究のシステムティックレビューを以前に行っていた[2]。入手していた他の関連研究と報告書に加えて、このシステムティックレビューの検索語を用いて、マッピング研究の報告ガイドラインに関する、幅広い文献検索を行った。この作業から、これまでに報告ガイダンスが発表されていないことが確認された。そのため、作業グループのメンバーはたたき台となる29個の報告項目からなる最初のリストとそれにもなう解説を作成した。作業グループメンバーによるさらなるレビューの後に、ここから25個の報告項目と解説が抽出された。

作業グループメンバーは、この分野の研究者とステークホルダーの中から62名のデルファイパネル候補者を選定した。この候補者の中には、マッピング研究で多くの経験を有するアカデミアやコンサルタント業界に属する個人、マッピング研究がもたらすエビデンスを常に評価している医療技術評価機関の代表者、そして生物医学雑誌の編集者などがいた。また、このデルファイパネルにはMAPS作業グループの医療経済学者も含まれていた。この62名のうち48名 (77.4%) が、マッピング研究の報告に求められる最低限の基準と、それにもなう報告チェックリストを作成するためのデルファイ調査への参加に同意した。

デルファイパネル参加者には、作業グループのメンバーからウェブ調査への個別リンクが送付された。回答をしていない者には、14日後と21日後の2回、督促メールを送信した。各パネリストは研究期間を通じて互いに匿名で参加し、個人を特定できるのは作業グループメンバーの中の1人だけであった。パネリストは、25個の候補項目の重要性について、作業グループが定めた9段階の評価尺度 (1:「重要でない」から9:「非常に重要である」) で点数をつけ、その評点に対する自信の度合い (「自信がない」、「いくぶん自信がある」、「非常に自信がある」) を回答した。そして、候補項目とその説明へのコメントをつけ、パネリストが次回検討すべき追加項目をあげ、その他の一般的コメントを行うよう依頼された。報告の候補項目は6つのセクション、(i) タイトルと抄録、(ii) 緒言、(iii) 方法、(iv) 結果、(v) 考察、(vi)

選好に基づくアウトカム尺度へのマッピング研究のための MAPS 声明 —解説と詳細—

その他にならべられた。また、パネリストは勤務地域と性別、主と副の業務環境に関する情報について回答した。

初回の回答の分析には、「ランド研究所 (RAND) /カリフォルニア大学ロサンゼルス校 (UCLA)」による適切性評価法の修正版を使用した[15]。これは、各被評価項目 (i) に対して、中央値、パーセンタイル範囲 (IPR)

(30%と70%) と、対称性を調整したパーセンタイル範囲 (IPRAS) の算出を行うものである。IPRASには非対称的な評価の補正因子が含まれ、 $IPR_i > IPRAS_i$ の場合にパネルに不一致があると判定される[15]。本研究では RAND/UCLAアプローチを修正し、パネリストには「適切性」よりも「重要性」を尋ねるなどした。重要性の評

表1 マッピング研究の報告に含めるべき項目のチェックリスト

セクション/トピック	項目番号	推奨	報告ページ 番号/行番号
<b>【タイトルおよび抄録】</b>			
タイトル	1	アウトカム尺度間でのマッピング研究であることを明示する。研究で使用した元の尺度と対象となる選好に基づく尺度を記述する。	
抄録	2	可能な場合は、構造化抄録を推奨する。目的、データソースとその主な特徴、使用したアウトカム尺度、推定・バリデーション戦略などを含む方法、モデル性能の指標などの結果、結論、主な研究成果の意味を含める。	
<b>【緒言】</b>			
研究実施の根拠	3	マッピング研究を実施する根拠を、幅広いエビデンスに基づいて記述する。	
研究の目的	4	使用した元の尺度と対象とする尺度、研究における疾患や集団の状況に言及し、リサーチクエスチョンを明示する。	
<b>【方法】</b>			
推定用サンプル	5	推定用サンプルをどのように同定したか、それを選択した根拠、対象者募集とデータ収集の方法、研究の場所や状況を記述する。	
外的妥当性用サンプル	6	外的妥当性用サンプルを利用する場合、それを選択した根拠、対象者募集とデータ収集の方法、研究の場所や状況を記述する。	
元の尺度および対象とする尺度	7	元の尺度と対象とする尺度、マッピング研究に用いた方法を記述する。	
探索的データ解析	8	元の尺度と対象とする尺度の概念的重なりを評価するために用いた方法を記述する。	
欠測データ	9	データが欠測した程度や、分析に用いたサンプルの欠測をどのように扱ったかを明記する。	
モデルリング方法	10	マッピングアルゴリズムの開発に用いた統計モデルを記述し、説明する。	
予測スコアあるいは予測効用値の推定	11	予測スコアあるいは予測効用値を、各モデルからどのように推定するかを記述する。	
バリデーション法	12	マッピングアルゴリズムのバリデーションに用いた方法を記述し、説明する。	
モデル性能の統計量	13	よりよいモデルを決定するためのモデル性能を評価する統計量を明記、説明し、これらの統計量をどのように推定し、用いたかを記述する。	
<b>【結果】</b>			
最終サンプル数	14	分析で用いた推定用サンプル数とバリデーション用サンプル数を明記する (対象人数と観察値数をともに含む)。	
記述的情報	15	サンプルにおける対象者の特性を記述する (あるいは、同様の情報を示す過去の出版物に言及する)。元の尺度と対象とする尺度の要約スコアを示し、両尺度の重なり合う部分の評価に用いた分析結果をまとめる。	
モデル選択	16	推奨するモデルを明記し、そのモデルを選んだ理由を説明する。	
モデル係数	17	選択したモデルにおけるすべてのモデル係数と標準誤差を示す。選択したモデルの結果を用いて、どのように効用値を算出できるのか、明確なガイドラインを示す。	
不確実性	18	予測効用値の平均まわりの標準誤差と個人レベルのばらつきを推定できる情報を報告する。	
モデル性能と表面的妥当性	19	選択したモデルの予測精度や適合度の統計量など、モデル性能の結果を表あるいは文章により示す。選択したモデルの表面的妥当性の評価を提示する。	
<b>【考察】</b>			
先行研究との比較	20	先行研究で、同じ尺度間のマッピングアルゴリズムを開発したものがあつた場合、その詳細を示し、アルゴリズムの違いを、モデル性能や予測、係数の点から記述する。	
研究の限界	21	マッピングアルゴリズムの限界となりうるものについて概要を示す。	
適用範囲	22	マッピングアルゴリズムが利用可能な臨床および研究の状況について概要を示す。	
<b>【その他】</b>			
追加情報	23	研究の資金源、非金銭的支援、研究の設計・実施・報告における資金提供者の位置づけを記述する。著者らと資金提供者の間に何らかの利益相反があれば報告する。	

価は古典的なRAND/UCLAの定義に従い, 評価の中央値により, 1から3(重要でない), 4から6(どちらでもない), 7から9(重要である)に単純に分類した[15].

作業グループの対面会議で, 初回デルファイ調査の結果をレビューした. 参加に同意した48名のうち46名(95.8%)が初回の調査を完了した. 25項目のうち, 24項目は「重要」と評価され, 1項目(「資金調達法」)は「重要」「重要でない」のどちらでもないと評価された. RAND/UCLA法に基づけば, どの項目の評価においても不一致の根拠はなかった. これらの結果は, MAPS作業グループの回答を除外しても変わらなかった. 初回の定性的なフィードバックに基づき, 「モデリングアプローチ」と「反復測定」の項目を統合し, 「モデル診断」と「モデル妥当性(plausibility)」も同様に統合した. 加えて, いくつかの推奨とその説明は, パネリストの定性的フィードバックから, 語句を修正した.

初回ラウンドに参加したパネリストには, 2ラウンド目のデルファイ調査への参加を依頼した. 初回ラウンド後に加えられた修正の概要が示された. これには, 各推奨と説明の修正が履歴つきで見られる文書も含まれていた. 2回目に参加するパネリストには, 1回目の結果(および統合した項目を分解した結果)をまとめたグループの結果(平均点とその標準偏差, 中央値とそのIPR, ヒストグラム, RAND/UCLAの重要性ラベルと同意レベル)が示された. 初回ラウンドの各項目に対する自身の評点(および統合した項目の各評点)も見ることができた. 2ラウンド目に参加したパネリストは, 各項目について重要性の評価を修正することができ, それをしない場合は1ラウンド目の評価のままであることが伝えられた. 統合した項目については, 新たに評価するよう求められた. また, 2ラウンド目に参加したパネリストは, 各項目について追加でコメントし, またグループにとって有用な追加情報を提供する機会が与えられた. 2ラウンド目のデルファイ調査参加者のうち, 14日後と21日後に督促メールが送付された者はいなかった. 2ラウンド目のデータ分析は1ラウンド目と同じ手法が用いられた.

2ラウンド目のデルファイ調査の結果についても, 作業グループでの対面会議によりレビューされた. 1ラウンド目に参加した46名のうち, 39名(84.8%)が2回目の調査を完了した. 2回目に含まれた23項目全てが「重要」と評価され, RAND/UCLA法に基づけば, いずれの項目の評価についても不一致の根拠はなかった. 2回目に参加したパネリストからの定性的フィードバックにより, 少数の推奨とその説明の語句に小さな修正が加えられた. これは2ラウンド目の回答者にフィードバックされ, 最終版となる推奨とその説明の読みやすさについてコメントする最後の機会があった.

以上の方法により, 合意した23個の報告項目からなるリストが作成された(表1). MAPS作業グループのメンバーにより作成された本稿は, 23個の報告項目の解説と詳細を提供する.

### 3 本論文の利用法

解説と詳細に関する本論文は, 他の報告ガイドラインのために作られた同種の論文をモデルにしている[9-14]. この報告における23個の各項目は, 過去の文献から引用した優れた報告方法の具体例とともに示した. いくつかの例では, 2次引用を削除し, あるいはテキストの一部を省略して“…”と表記した. 可能な場合は, 各項目に対してその理由と関連するエビデンスに基づき, 推奨とともに説明を加えた. マッピング研究を報告する際に不可欠な要件となるリストに焦点を置くが, 追加情報が報告の質を改善する可能性がある箇所も強調した. これら23項目は数字が付され, 6つのセクションに分類された. 具体的には, (i) タイトルと抄録, (ii) 緒言, (iii) 方法, (iv) 結果, (v) 考察, (vi) その他である. しかし, 報告では必ずしもここに示した順序で各項目にふれる必要はない. それぞれの推奨項目が, 報告の本文あるいはAppendixのいずれかに含まれていることが重要である.

### 4 MAPS チェックリスト

#### 4.1 タイトルと抄録

##### 4.1.1 項目1: タイトル

推奨: アウトカム尺度間でのマッピング研究であることを明示する. 研究で使用した元の尺度と対象となる選好に基づく尺度を記述する.

例: 「クッシング症候群におけるヨーロッパレジストリ(ERCUSYN) データを用いたCushing QOLスコアからEQ-5D効用値へのマッピング」[16]

解説: アウトカム尺度間でのマッピング研究に関する報告であることをタイトルに明示する必要がある. 報告がMEDLINEやCentre for Reviews and Dissemination(CRD)のようなデータベースに, 適切な索引が収載されるよう, 「マッピング」, 「クロスウォーク」, 「効用値への変換」など本分野に特有の用語を用いることが推奨される. 本分野の研究で最も一般的な用語は「マッピング」である[2]. この用語は, 外部の測定データを利用して健康効用値を予測するアルゴリズムの開発者が, 今後も引き続き使用する可能性が高い. 文字数が許せば, 元の尺度名と選好に基づく対象の尺度名をタイトルに含めるべきであり, 目的とする集団あるいは疾患名を記載することが有用である. タイトルに不明確な用語を用いると, 報告に誤った索引がつけられ, データベース検索から漏れてしまうリスクが生じる.

##### 4.1.2 項目2: 抄録

推奨: 可能な場合は, 構造化抄録を推奨する. 目的, データソースとその主な特徴, 使用したアウトカム尺度, 推定・バリデーション戦略などを含む方法, モデル性能の指標などの結果, 結論, 主な研究成果の意味を含める.

例：「目的：Roland Morris Disability Questionnaire (RMQ) は、腰痛 (LBP) に広く使われる健康状態の尺度である。しかし、それは選好に基づくものではない…。腰痛治療におけるランダム化比較試験のデータを用いて、本研究ではRMQスコアと、EQ-5DあるいはSF-6Dにより得られる健康効用値のマッピングアルゴリズム開発を検討した。

方法：本研究は、Back Skills Training Trial (BeST) のデータに基づいている。患者701名から、RMQ、EQ-5D、SF-12(これからSF-6D効用値が算出される) など広範なアウトカム尺度により、ランダム化後、ベースラインと3か月、6か月、12か月後時点で収集されたデータである。EQ-5DとSF-6Dの健康効用値と次元回答 (dimension response) を予測するために、ベースライン時点の試験データを利用し、直接 (direct) および回答マッピング (response mapping) 法の両者を用いて、モデルを推定した。モデルの予測精度を検討するために、多段階のモデル選択プロセスを使用した。次に、経時的観察データを利用する別の手法とマッピングモデルを検討した。推定されたマッピングアルゴリズムは、UK Back Pain Exercise and Manipulation (BEAM) 試験の外部データを用いてバリデーションした。

結果：健康効用値を正確に予測する多数のモデルを開発した。最も性能の良かったRMQからEQ-5Dへのモデルは、経時データに基づくベイジアン擬似尤度推定を用いたベータ回帰であった (平均二乗誤差 (MSE) : 0.0380) …。RMQからSF-6Dへのマッピングにおいて選択したモデルは経時データに基づく有限混合モデルであった (MSE:0.0114) …。

結論：回帰的手法を用いて、RMQスコアとその回答から、EQ-5DおよびSF-6Dの健康効用値を適切に予測することは可能である。本研究における回帰式は、EQ-5DやSF-6Dのデータが入手できない際に、健康効用値を推定する実証的根拠を提供する…。」

[17]

解説：抄録は、マッピング研究の目的と方法、結果やその意味を理解できるようにする必要がある。抄録は、報告全体にアクセスするかを決めるフィルターの役割を果たすことが多い。編集者や査読者が、目的、過程、研究結果の重要性をすばやく読み取る手段にもなる。加えて、環境によっては、その研究の潜在的読者が報告全体を入手できない場合もある。そのため、抄録では出版者が定める語数内で報告の主要部分をまとめ、マッピング研究に関する最適な情報を示す必要がある。報告の本文に書かれていない情報を記載してはならない。

抄録の詳細な構成は、出版者の規定により定められていることが多い。抄録が、研究の目的、方法、結果、結論等の小見出しがついた形で構成されていると、構造化されていない抄録よりも、情報としてより過不足なく、理

解もしやすい[18]。マッピング研究の構造化抄録は必要に応じて、以下の情報を含むべきである。すなわち、研究の目的、データソースおよびその重要な特性 (場所や状況、集団あるいは臨床的な特徴、サンプル数)、使用したアウトカム尺度および推定・バリデーションの戦略を含む方法、モデルの性能の指標を含む結果、結論、主要な研究結果の意味である。現在利用できるマッピング研究のためのMeSH項目はない。そのため、「マッピング」のような複数の総合的な記述語 (descriptor) を報告のキーワードに含めるべきである。

## 4.2 緒言

### 4.2.1 項目3: 研究実施の根拠

推奨：マッピング研究を実施する根拠を、幅広いエビデンスに基づいて記述する。

例：「高齢者へのサービスを評価する際に、とくにそれが疾患や状態に特異的な場合は、効用に基づく尺度ではなく、非効用型のアウトカム尺度により収集することがある。このような集団にはその方が適していると考えられているからである…。経済評価において非効用型アウトカム尺度を用いると、より広範な介入間における結果の比較可能性が確保されない。これは、効用に基づくアウトカム尺度を用いた場合には、解決するものである。そのため、回帰分析に基づき非効用型のアウトカム尺度から効用に基づく尺度を予測するアルゴリズムやマッピング関数は、最終目的がCUAを実施することである場合に有用である…。本論文では、非効用型だが、通常用いられている機能的自立の臨床尺度パーセルインデックス (BI) を、回帰によって選好に基づく尺度EQ-5Dに対応づけた分析結果を報告するものである。これまで、van Exelらの研究のみがBIをEQ-5Dにマッピングしていた…。本論文は、このvan Exelらの研究に基づくが、より大規模なデータを用いており、それは中間的ケアサービスの評価として、イギリスで実施・公表されたものとしては、最大のものである。」[19]

解説：緒言では、マッピング研究の根拠と、それがより広範なエビデンスの基盤に何を加えるかを伝えなければならない。背景を設定する構成としては以下を含むものが有用である。すなわち、(i) 目的とするアウトカム尺度間の新しいマッピングアルゴリズムが必要であることの記述。該当する場合は、その地域における技術評価のための方法論ガイドラインの記述、(ii) 特定の疾患や集団のタイプにおいて、目的とするアウトカム尺度間のマッピングアルゴリズムを開発、あるいは重要な関係を発見した先行研究の概要、(iii) この新たなマッピングアルゴリズムが、規制や償還の決定を行う機関にどのような情報を提供するかに関する見通し、である。新たな

研究で評価される元の尺度と対象とする尺度間のマッピングアルゴリズムが、すでに開発されている場合は、新たな研究の必要性を説明する必要がある。

#### 4.2.2 項目4: 研究の目的

推奨: 使用した元の尺度と対象とする尺度, 研究における疾患や集団の状況に言及し, リサーチクエスチョンを明示する。

例: 「本研究の目的は, PDQ-8からEQ-5Dの健康効用値を計算するアルゴリズムを開発することである。このことにより, PDの研究においてPDQ-8やPDQ-39によってのみ健康アウトカムが評価されている場合にも, 費用効用分析が実施可能になる。」 [20]

解説: 緒言では, マッピング研究に取り組む目的や仮説を明示すべきである。リサーチクエスチョンを明確にするには, 元の尺度と対象とする尺度, 研究における疾患や集団の状況などマッピング研究の基礎となる情報について詳細な記述が必要になる。また, 用いたのが(インデックススコアあるいは効用値への)直接マッピングか, (次元回答への)間接(あるいは回答)マッピングか, その両方かを記載できるだろう。したがって, この項目の報告は, 以下のチェックリストの推奨7(元の尺度および対象とする尺度)と推奨10(モデリング方法)の報告とあわせて考えるべきである。研究で複数の目的に取り組む際は, 主要な目的が明確に示され, 重要な副次的目的についても述べる必要がある。

### 4.3 方法

#### 4.3.1 項目5: 推定用サンプル

推奨: 推定用サンプルをどのように同定したか, それを選択した根拠, 対象者募集とデータ収集の方法, 研究の場所や状況を記述する。

例: 「このデータは, 11歳から15歳までの子供のHRQoL, 身体活動, 食事と肥満の状態の関係を調査した研究から取得した。イギリスにおいて4つの中学校の横断調査を実施した…。これらの学校は, 試験の成績, 無料学校給食児童の割合, 特別教育の必要がある児童の割合をマッチングして選ばれた。2,858名の児童を対象に冬に1回と夏に1回の2時点で行う匿名の調査に参加するよう, 依頼した。冬期調査では869名が回答し, 夏期調査では1,000名が回答したため, 合計1,869件の回答を得た。本研究の詳細は他の文献に記述されている…。夏期調査のサンプル数の方が多いため, 本報告のモデリングにはこの1,000名の回答を用いて, このサンプルを地域で2つに分け, 推定用(イギリス北西部の2校の生徒)とバリデーショ用(イギリス南西部の2校の生徒)サンプルとした。」 [21]

解説: マッピングアルゴリズムの推定に使用するデータは, 既存のデータセットあるいは当該マッピング研究のために収集されたサンプルからのものであるかもしれない。研究では, その推定用サンプルがどのように, そして, なぜ選択されたのか詳細を十分に示し, 理解できるようにする必要がある。記載すべき詳細には, サンプル選択の根拠, 組入と除外基準, サンプル数の根拠, 対象者募集とデータ収集の方法, サンプルの場所や状況などが含まれるべきである(提示すべき記述統計の推奨については, 推奨15を参照)。マッピングアルゴリズムの推定に既存データセットを利用する場合は, 詳細のわかる適切な情報源を示す必要がある。

#### 4.3.2 項目6: 外的妥当性用サンプル

推奨: 外的妥当性用サンプルを利用する場合, それを選択した根拠, 対象者募集とデータ収集の方法, 研究の場所や状況を記述する。

例: 「最もあてはまりのよかったマッピングアルゴリズムの外的妥当性は, Elective Orthopaedics Centre (EOC) のデータセットを利用して検討された。これは, 2004年以降ロンドン南西部の4つのNHSトラストをカバーするNHS治療センターで, 股関節あるいは膝関節の置換術を受けた患者からなる大規模な観察コホートであり, 最終モデルを選択するまでは利用不可能だったものである。患者は, 術前と術後6ヶ月, 12ヶ月および/あるいは24ヶ月時点でEQ-5DとOKSに回答した。患者募集は引き続き行っているが, PROMsとの重複を避けるため, 本分析では2009年3月31日以前に膝関節の初回あるいは再置換術を受けた患者のみを含めている。外的妥当性用データセットには, OKSおよび/あるいはEQ-5Dのデータが不完全な患者を除く4,505名, 10,002件の観察データが含まれている。」 [22]

解説: 外的妥当性用サンプルを利用する場合, 評価を行った状況に対する外的妥当性用サンプルの一般化可能性を判断できるよう, そのサンプルを選択した根拠を述べる必要がある。対象者募集とデータ収集の方法, 場所や状況に関する情報を示さなければならない(提示すべき記述統計の推奨については, 項目15を参照)。マッピングアルゴリズムのバリデーショに既存データセットを用いる際には, 詳細のわかる適切な文献を提示する必要がある。推定用と外的妥当性用サンプルの類似性および差異は, これらの意味するところとあわせて記述する必要がある。バリデーショ法の報告に関する推奨は項目12を参照せよ。

#### 4.3.3 項目7: 元の尺度および対象とする尺度

推奨: 元の尺度と対象とする尺度, マッピング研究に用いた方法を記述する。

例：「患者は、Health Assessment Questionnaire Disability Index (HAQ), 0~100のスコアで痛みを評価するvisual analogue scale (VAS), EQ-5Dなどに回答した。HAQは、過去1週間に患者が経験した8つの項目、衣服の着脱と身支度、起立、食事、歩行、衛生、手の届く範囲、握力、家事や雑用における困難の程度について患者からの報告を基にしている。スコアは0から3までの0.125刻みになっており（一般的には完全に連続と考えられている）、点数が高いほど機能的障害の程度が高いことを示す。これをRAの臨床試験に含めることは事実上必須とされており、多くの経済モデルの構築を促進するものとして広く使用されている。この分析には、イギリスのEQ-5Dタリフ値（あるいは「インデックススコア」）が用いられ、先行研究の結果との比較に有用である。」 [23]

解説：報告では、マッピング研究に用いた元の尺度と対象とする尺度を明確に記述する必要がある。健康関連QOLや健康状態の測定法に関する記述を含めるべきであり、これには、何らかの臨床尺度と同時に使用したバージョン（例えば、EQ-5Dの言語バージョン）などがある。該当する場合、それぞれの尺度の次元（あるいはドメイン）およびレベル（あるいはファセット（facet））の数とその文（codification）を示す必要がある。尺度が要約尺度（summary scale）に集約される詳細な方法と、高いスコアがアウトカムの良し悪しのどちらを示すのかについて記述しなければならない。データに選好値のタリフが用いられている場合は、具体的な価値づけ（valuation）法と出典を明記する必要がある。報告は、推定用サンプル、該当する場合は外的妥当性用サンプルに関して、元の尺度と対象とする尺度の測定実施方法、これらが集められた時点を明記しなければならない。既存のデータセットを利用した場合は、その詳細を示す適切な情報源を提供する必要がある。

#### 4.3.4 項目8：探索的データ解析

元の尺度と対象とする尺度の概念的重なりを評価するために用いた方法を記述する。

例：「マッピングアプローチの厳密性は、「元の」尺度と「対象の」尺度の記述システム（description system）間に相当の重なりが存在することに依存している…。それぞれの尺度は同じHRQOLを評価するため、MSIS-29とEQ-5Dの重なり、そしてMSIS-29とSF-6Dの重なりは、かなり大きいことが期待される。共通範囲（joint coverage）の領域を線図であらわしたものを図1に示す。」 [24]

「本研究では、この2つの尺度間の（非）類似性を見出すためのデータを、スピアマン相関を用いて探索することから分析を始めた。この相関行列は、2つ

の質問票の全項目に対する項目間相関からなる。次に、この2つの尺度が持つ次元構造を発見・比較するために、探索的・検証的主成分分析（PCA）を行った。探索的PCAにおいては、固有値が1を超える構成概念を選択した。検証的PCAでは、構成概念の数を標的尺度の構成概念数に制約した。より解釈可能な因子を得るために、バリマックス回転—因子軸の直行回転—を用いて探索的・検証的両方のPCAの因子を回転した。」 [25]

解説：健康アウトカム指標あるいは尺度と選好に基づく尺度間のマッピングアルゴリズムの推定は、元の尺度と対象とする尺度の次元（あるいはドメイン）の概念的重複に依存する。重なりの評価を行った場合は、それを報告する必要がある。これには、内容の重なりを定性的評価、あるいは定量的・統計的評価がある。統計的手法を用いた場合は、どの方法を選択し、それがなぜ適切なのかを報告する。統計的手法については詳細を十分に記述をする必要がある [例えば、主成分分析（PCA）を用いた場合は、そのタイプ（例えば、探索的、検証的など）、抽出した構成概念の選択基準（例えば、固有値が1を超えたもの、5つの構成概念など）、該当する場合はその回転法（例えば、バリマックス回転など）]。

#### 4.3.5 項目9：欠測データ

推奨：データが欠測した程度や、分析に用いたサンプルの欠測をどのように扱ったかを明記する。

例：「PDQ-39における8つのドメインスコアの欠測（3.0%）は、多変量正規分布を仮定して、EMアルゴリズムにより補完した。」 [26]

「全レコードのうち、2,471件（16.0%）は、EQ-5DとSF-12で完全に妥当な回答ではなかつたため除外され、解析対象として12,967件のサンプルが残った。」 [27]

解説：欠測データは、推定用およびバリデーション用データセットの特徴を示しているかもしれない。関連したデータセットにおける欠測データの量と、欠測データの取り扱い方法について明確に記述する必要がある。完全データに基づく（complete case）解析は、マッピングアルゴリズムの開発において新たな困難が生じることを避けるために、マッピング研究で広く用いられている。しかし、完全データに基づく解析を行うと、利用可能なサンプルサイズが大きく減少する可能性がある。よって、マッピングアルゴリズムの推定に用いる元のデータ/対象のデータ、あるいはその他の変数を補完することもあるだろう。どの情報が補完されたかを明確にする必要がある。補完法を用いた場合、その決定を説明し、補完した変数と使用した手法およびソフトウェアを明示することが重要である。

#### 4.3.6 項目10: モデリング方法

推奨: マッピングアルゴリズムの開発に用いた統計モデルを記述し, 説明する.

例: 「本論文ではまず, 4つの関数の形式を用いてOKSの各設問に対する回答をEQ-5D効用値に回帰することにより, 直接的な効用値のマッピングモデルを推定した…。観察値の9.6% (17,184/179,482) はEQ-5Dにおける完全な健康 (効用値=1) であり, それを考慮するためにtwo-partモデルを用いた…。術前の質問で45.9% (48,318/105,235) が1つ以上のEQ-5Dドメインで重度の問題を示し, その結果, EQ-5DタリフのN3項によって非常に低い効用値となったため, three-partモデルも開発, 評価をした…。また, EQ-5Dの5つの各ドメインに対して患者が選択した回答レベルを予測するため, 回答マッピングも使用した。これらは前述したように, EQ-5Dの各ドメインに対して, 多項ロジット (mlogit) あるいは順序ロジスティック回帰 (ologit) モデルを別々にあてはめることにより推定を行った…。すべてのモデルの説明変数には, 患者がOKSの各質問について特定の回答レベルをもつかを示す48個のダミー変数を含んでいる。なお, 応答レベル4 (問題なし) は参照項である。しかし, すべてのモデルは, 別の2組の説明変数を用いた評価もなされた。OKSの12の質問スコア (0から4の評価) と全OKSスコア (質問スコアの非加重合計に基づき0から48で評価される) である…。Stataバージョン11 (Stata-Corp, College Station, TX) を用いてすべてのモデルを推定した。すべてのモデルにおいて, 観察値の患者内クラスタリングを考慮し, 標準誤差を調整するために, Stataのクラスターオプションを使用した。」[22]

解説: 元の尺度と対象とする尺度の関係を調べるための統計モデルの選択は, 明確に記述し説明する必要がある。マッピング研究に用いる統計モデルは, (インデックスあるいは効用値への)「直接」法, そして (各次元の回答への)「間接」あるいは「回答マッピング」に分類できる。それぞれの方法および推定量には利点と欠点があるが, 異なる環境においてどちらを使うべきかという明確なガイドラインはない[1]。したがって, モデリング方法に関する十分な情報を提供し, これら推定戦略全体の頑健性を評価できるようにする必要がある。用いた推定法 (例えば, 最小二乗法 (OLS), ベータ回帰, two-partモデルなど) を明確にし, 説明しなければならぬ。元の尺度の構成要素や, 人口統計学的・臨床的特性, 1次/2次の多項式や交互作用項など, 各モデルで使用した説明変数を記述する必要がある。最終モデルにおいて, 有意でない変数あるいは直感に反すると思われる符号の変数を除外するために用いた変数選択法を記述すべきである。ステップワイズ回帰 (例えば, 変数増加法, 変数減

少法, 変数増減法や統計的手法によらない (manual) 手法など) を適用した場合は, それについても記述する必要がある。数名あるいは全員の経時測定 (ベースラインとその後の観察値) データセットにおいて, 経時測定に対して何らかの調整をした場合, (もしあれば) どのような調整を行ったのか記述すべきである。縦断研究のある一時点 (例えばベースラインなど) のみのデータを用いてモデルを推定した場合, それについて明記すべきである。利用した統計分析ソフトウェアについても報告する必要がある。利用者が一般的に使用できない推定のコマンドを用いる場合, アクセス可能な統計パッケージにおけるプログラミングコード作成を検討する必要がある。

#### 4.3.7 項目11: 予測スコアあるいは予測効用値の推定

推奨: 予測スコアあるいは予測効用値を, 各モデルからどのように推定するかを記述する。

例: 「直接マッピングモデルによる予測値は, 「推定後の予測」のコマンドを用いてGLMとfractional logitモデルから得られる予測値の逆変換により, 推定された。OLSモデルでは, 1を超える予測された効用値を1に設定した。two-partモデルでは, 各患者の期待効用値を次のように推定した:

$$\text{Utility} = \Pr(\text{Utility} = 1) + (1 - \Pr(\text{Utility} = 1)) U \quad (1)$$

ここでUは, 完全でない健康状態の予測効用値,  $\Pr(\text{Utility} = 1)$  は完全に健康な状態であることの予測確率である。」[22]

解説: マッピング研究では, 異なるモデルをどのように適用したのか理解できるよう, 詳細を十分に示す必要がある。予測値を計算するために使われた推定後のコマンドや, オプション (とくに複数が利用可能な場合) を明記し, 逆変換を行った場合はその詳細 (例えば, 効用値の減少分 (disutility) を効用値に, あるいは対数効用値を自然尺度に変換するなど) を示すことが望ましい。OLSのような観察範囲外の予測値を出力するモデルにおいては, その尺度の最大値を超えて予測された値を, その尺度の最大値に設定したかを明記する (例えば, 効用値が1を超える場合は1に設定するなど)。一般に期待値法[28]は, two-partモデルと回答マッピングモデルの予測効用値を推定するのに最も適した方法である。これは無限回のモンテカルロドロウ (draw) の使用と同じであり[28], (最大あるいは最尤確率法と異なり) 不偏予測値となる。

#### 4.3.8 項目12: バリデーション法

推奨: マッピングアルゴリズムのバリデーションに用いた方法を記述し, 説明する。

例: 「本研究では, それぞれの統計モデルを異なる



データセットにどの程度一般化できるかを評価するため、サンプル内交差検証法とサンプル外検証の手法を用いた…。これらの方法は、モデルの予測精度に関して、適合度の統計量R2を使用するよりも優れている。サンプル内での交差検証法は、主要なデータセットをk組のサブサンプル (k=5) に無作為に分割するk-分割法を用いて行った。サブサンプルのひとつは予測モデルを検証するためのバリデーション用データとして取り置き、残り4サンプルをトレーニングデータとして用いた。その後、5個のサブサンプルをバリデーション用データとしていずれも1度だけ使い、交差検証法プロセスを5回繰り返した…。サンプル外バリデーションは、EQ-5D効用値とNEIVFQ 25の次元スコアの両方を含んだ異なるバリデーション用データセットを用いて実施した…。」 [29]

解説：新しいマッピングアルゴリズムは、その作成に使用したものと異なるデータセットでバリデーションするのが理想的である[30]。バリデーションを行わなかった場合、それについて明示し、説明する必要がある。マッピングアルゴリズムのバリデーションに使用した方法とデータセットはすべて記述しなければならない。内部あるいはサンプル内バリデーション（アルゴリズムの推定に使用したものと同一データセットにおけるサブセットでモデル性能を評価する）は、最良のモデルを決定する際の過剰適合（overfitting）を避けるために有用な場合がある[31,32]。方法としては、ホールドアウトアプローチ（例えば50%あるいは25%など、人数や観察数の一部をバリデーションのみに用いる）、反復k-分割交差検証法あるいはleave-one-out交差検証法（あるデータセットから繰り返し抽出することにより、複数の重複した推定用サンプルを作成、このデータから統計モデルを推定し、残った観察値でバリデーションする）などがある。外部あるいはサンプル外バリデーションは、最終モデルの評価と他のデータセットで得られる可能性の高い予測精度を評価するために用いられる[22]。可能であれば、外部バリデーション用データは推定および内部バリデーション用サンプルとは異なる方法で収集され、最終モデル決定後に、マッピングアルゴリズムを開発した研究者が評価する[31]。使用される用語は様々で、異なる種類のバリデーションの方法とデータセットを示すために、しばしば同じ用語が用いられる。そのため、実施したバリデーションのタイプと、それがモデル選択にどのようにつながったのかを十分に説明する必要がある。

#### 4.3.9 項目13: モデル性能の統計量

推奨：よりよいモデルを決定するためのモデル性能を評価する統計量を明記、説明し、これらの統計量をどのように推定し、用いたかを記述する。

例：「EQ-5DとSF-6Dの推定されたインデックススコアの平均値、平均絶対誤差（MAE）、平均二乗誤差（MSE）と平均平方二乗誤差（RMSE）を示す。MAEが観察スコアと推定スコアにおける絶対誤差の平均値であるのに対し、MSEは観察スコアと推定スコアの二乗誤差の期待値である…。MAEとMSEはいずれも個人レベルの平均精度を測るものであるが、MSEは大きな誤差の重みが大い、RMSEが低い方がモデル性能は優れている。最もよい性能のモデルは、RMSEが低いという基準により選ばれた。その後、選択されたモデルの性能は、インデックススコアの観察値と予測値のMAEに基づき、モデルの適合度にはR2を用いた。MAE、MSE、RMSEはいずれもモデル性能を評価する基準だが、本研究ではRMSEが最小のモデルを提示する。これは、RMSEがデータと同じ単位で測定され、「典型的」な誤差の大きさをあらわし、そして偶発的な大きな誤差に対する感度が高いためである。」 [33]

解説：モデル性能を評価する、あるいは複数のモデルから選択するために様々な統計量を用いることができる。推定モデルのすべてを同じ重みで報告する場合を除き、推奨するモデルを選択した際に用いた統計量とそれを評価したデータセットを明確に報告すべきである。可能であれば、分析を始める前にこのモデル性能の主要な統計量をあらかじめ詳述しておく必要がある。モデル評価は、モデル適合度の統計量（例えば、R2、情報量基準など）よりも、予測尺度（例えば、平均絶対誤差（MAE）や平均二乗誤差（MSE）など）に基づくべきであると一般に考えられている[1]。モデル性能における理想的な統計量は、点推定とともに推定値の分布を考慮するものであろう。予測精度の全側面をとらえる単一の統計量は開発されていないため、複数の尺度や基準（表面的妥当性および/あるいは単純化の原理（parsimony）を含む場合もある）によりモデル選択を行っている。統計量の算出方法について、結果を理解し計算を再現できるよう、詳述する必要がある。特に、R2は調整前か調整後か、推定モデルのモデル適合度か予測値と観察値を相互に関連づける別のモデルにおける適合度かを明記しなければならない。正確には、「平均絶対誤差」（あるいは「平均絶対偏差」）は、各観察値と予測値の差を計算し、その絶対値（すなわち、マイナス記号を無視する）をとって、全観察値にわたり平均して算出する。同様に、「平均二乗誤差」は、各観察値に対する予測値の差を推定し、これらの差の平方をとって全観察を通じた平均を算出する統計量である。これら両方の統計量とも、常に正の値をとる。

## 4.4 結果

### 4.4.1 項目14: 最終サンプル数

推奨：分析で用いた推定用サンプル数とバリデーション

用サンプル数を明記する (対象人数と観察値数をともに含む).

例: 「905名の患者から, 時点の異なる複数の観察値を収集した (患者一人あたりの観察数は平均 5.365, 最小 3, 最大 8). 実際の観察数は3,425から 3,945の範囲であり, 対比較可能な観察数は3,230から 3,640であった (表 2).」 [34]

表 2 ENACT-1とENACT-2をまとめたデータセットにおける対になった観察値のスピアマン相関係数

	EQ-5D	SF-6D
CADI		
N (ペア数)	3375	3640
相関係数	-0.62	-0.66
IBDQ		
N (ペア数)	3320	3230
相関係数	0.76	0.85

すべての相関係数は $P < 0.0001$ であった.  
 CDAI: Crohn's Disease Activity Index, ENACT: Efficiency of Natalizumab as Active Crohn's Therapy, IBDQ: Inflammatory Bowel Disease questionnaire  
 Elsevier社[34]の許諾を得て掲載した.

解説: 分析で用いた推定用サンプルとバリデーション用サンプルの大きさ, そしてそれがモデル間で異なる場合は, 各モデルにおけるサンプル数も報告する必要がある. 報告に有用な情報は, 1人あたり観察数, 各尺度で利用可能な人数と観察数, そして異なる尺度の組み合わせについて対となる観察値の数などである. 観察数は総サンプル数より少ない場合があり, 欠測データの取り扱いによりモデル間で異なることもある.

#### 4.4.2 項目15: 記述的情報

推奨: サンプルにおける対象者の特性を記述する (あるいは, 同様の情報を示す過去の出版物に言及する). 元の尺度と対象とする尺度の要約スコアを示し, 両尺度の重なり合う部分の評価に用いた分析結果をまとめる.

例: 「[[推定用] データセットは, 滲出型AMDである以外は健康との診断を受けた男女55歳以上の対象者154名を含む… [バリデーション用] データセットは, カナダ, フランス, ドイツ, スペイン, イギリスの50歳以上の男女401名を含む… 合計151名の対象者のデータを用いて, NEI-VFQ25とEQ-5D効用値のマッピングを行った (表 2). 対象者の多くは女性だった (2.4:1). 視力が良い方の眼のVAに基づく, 対象者のほとんどは軽度の視覚障害であった ( $n=77$ ). 推定されたEQ-5D効用値の平均値は0.7711 ( $SD=0.21$ ) だった. 対象者の約24%は完全に健康な状態であると報告された ( $n=36$ ). 32種類のEQ-5D健康状態が報告された. NEI-VFQ 25

スコアでは, 全体的視力, 近くのものがよく見えないことによる困難, 視力による精神症状, 運転, 視力による役割制限のドメインにおいて, 対象者は60未満のスコアを報告した. 対象者のスコアはNEI-VFQ 25における運転の次元で最も低かった (平均値=43.18,  $SD=35.82$ ) (表 3). 推定されたEQ-5D効用値は歪度が負であり [-1.3445, 標準誤差 ( $SE$ ) =0.017], 天井効果 (ceiling effect) が生じていることが示唆された. NEI-VFQ25における11の視力次元と全体的健康観の分布, EQ-5D効用値の分布を図 2に示した.」 [29]

解説: 推定用サンプルとバリデーション用サンプルにおける各個人の重要な特性を理解し, 評価できるよう, 十分に詳細を報告する必要がある. 年齢と性別のような標準的な人口統計学的要因に加え, 関連する特性として疾患特性, 国籍, 人種などがある.

推定用サンプルとバリデーション用サンプルにおける元の尺度と対象とする尺度に関する情報として, 平均値, 標準偏差, 範囲を記載する必要がある. 紙面が許すならば, これらの情報を視覚的に, 例えば元の尺度と対象とする尺度の値の分布を図示するなど是有用であると思われる. このような分布情報は, 元の尺度と対象とする尺度の測定可能範囲すべてにわたって結果が頑健であるか, 他の集団にどれだけ一般化可能であるか, モデルで用いた分布の仮定が適切かなどについての評価に有用である. このような状況における, 正式な統計的検定についても報告する必要がある.

元の尺度と対象とする尺度間の重なりの評価に用いた分析結果 (項目 8の推奨を参照) は, 分析方法によって記述的情報や統計的情報を示すことが求められることもある.

#### 4.4.3 項目16: モデル選択

推奨: 推奨するモデルを明記し, そのモデルを選んだ理由を説明する.

例: 「予測精度の主要な統計量であるMSEに基づく, 最も優れた予測モデルは, mlogitを用いる回答マッピングアルゴリズムであり ( $MSE:0.0356$ ; 表 2), 次いでthree-partモデルであった ( $MSE:0.0358$ ). ただし, three-partモデルにおいてMAEはmlogitより低い (0.1338 vs 0.1341). ologit応答マッピング ( $MSE:0.0359$ ), two-partモデル ( $MSE:0.0360$ ) そしてOLS ( $MSE:0.0363$ ) もよい性能を示した. しかし, fractional logitとGLMモデルは相対的に予測が悪く ( $MSE:0.0367-0.0397$ ), 平均で0.00063-0.0025だけ系統的に過小評価していた. mlogitモデルも, 0.5未満の効用値については他のどのモデルよりも過大評価は小さい (平均残差: 0.160 vs 0.162-0.170) が, 0.5以上の患者の効用値はologitとガンマ分布のGLM以

外のどのモデルよりも大きく過小評価（平均残差：-0.078, vs -0.075 to -0.076）した。」[22]

解説：モデル性能を評価する、あるいは複数のモデルから選択するために様々な統計量を用いることができる。よって、どのモデル性能を示す統計量により、推奨されるモデルを選定したのか、そしてそれを評価したデータセットについて、明確に報告する必要がある（項目13の推奨を参照）。すべてのモデルで予測精度の統計量を報告すべきであり、モデルの適合度をあらわす統計量を提示してもよい。また、例えば、疾患の重症度が異なる患者カテゴリーや上述の例にあるように、健康状態の高低で分けたグループなどのサブセットに対してこのようなモデル性能の統計量を報告することも有用である。

#### 4.4.4 項目17: モデル係数

推奨：選択したモデルの結果を用いて、どのように効用値を算出できるのか、明確なガイドラインを示す。

例：Khanら[21]の表6：最もあてはまりのよい2つのモデルの結果 N = 896。

「例えば、ドメインモデルを使用して、HUI3効用値は  $1 - (0.06315 \times \text{DEV} + 0.02621 \times \text{PHYS} + 0.02349 \times \text{RHD} - 0.00346 \times \text{HL} - 0.00059 \times \text{AGE} - 0.01363 \times \text{FE-MALE})$  と計算される。年齢は月数、HLはdBであり、

DEV・PHYS・RHDは、OM8-30ドメイン/ファセットスコアであり、TARGETおよびEurotitisのデータセットから標準的スコアリング法を用いて計算した。FEMALEはダミー変数で、患者が女性なら1、男性なら0である。」[35]

解説：報告の結果部分では十分な情報を提供し、選択されたモデルの結果に基づき自らのデータセットで各個人の予測効用値を算出できるようにすべきである。回帰手法を用いたすべての研究において、モデル係数とその標準誤差や選択されたモデルの95%信頼区間を提示する必要がある（項目18の推奨も参照）。その他の方法を用いる研究（例えば、ベイジアンネットワークあるいはクロス集計など）では、可能であれば予測するために必要なすべてのデータやコードを、必要に応じてオンラインAppendixに提示する必要がある。報告された係数から予測効用値と個人レベルのばらつきを算出できるよう、すべての変数のコード化方法などについて、十分な詳細を提供しなければならない（項目18の推奨を参照）。また、選択したモデルに基づき、健康状態の具体例においてどのように効用値を算出できるかを、本文に含めることも推奨される。複雑なモデル（例えば、回答マッピングモデルなど）では、予測効用値を計算できるよう、オンラインAppendixで独立したプログラム（syntax）（例 [28]）を提供することも適切かもしれない。

表6 最もあてはまりのよい2つのモデルの結果

	OLS (6)		OLS (5)	
	係数	標準誤差	係数	標準誤差
年齢	-0.006136	(0.004741)		
性別	-0.009385	(0.012292)		
PedsQL PF	0.009067***	(0.002571)	0.009127***	(0.002568)
PedsQL EF	0.006807**	(0.002533)	0.006611**	(0.002530)
PedsQL SF	0.005630*	(0.002831)	0.005705*	(0.002829)
PedsQL SchF	0.005802*	(0.002371)	0.006011*	(0.002367)
PedsQL PF Squared	0.00002	(0.000025)	0.00002	(0.000025)
PedsQL EF Squared	-0.000049**	(0.000018)	-0.000048**	(0.000018)
PedsQL SF Squared	0.000011	(0.000016)	0.000011	(0.000016)
PedsQL SchF Squared	-0.000017	(0.000015)	-0.000017	(0.000015)
PedsQL PF * EF	-0.000005	(0.000027)	-0.000004	(0.000027)
PedsQL PF * SF	-0.000053	(0.000029)	-0.000055	(0.000029)
PedsQL PF * SchF	-0.000066*	(0.000030)	-0.000066*	(0.000030)
PedsQL EF * SF	-0.000011	(0.000023)	-0.000009	(0.000023)
PedsQL EF * SchF	0.000061**	(0.000021)	0.000059**	(0.000021)
PedsQL SF * SchF	-0.000026	(0.000022)	-0.000027	(0.000022)
Constant	-0.335861**	(0.118035)	-0.428496***	(0.094210)
Observations	896		896	
Adjusted R2	0.287		0.2868	
MSE	0.0315		0.0316	
MAE	0.1063		0.1067	

カッコ内は標準誤差, \*p<0.05, \*\*p<0.01, \*\*\* p<0.001  
PF: 身体的機能, EF: 感情の機能, SF: 社会的機能, SchF: 学校  
Springer Science and Business Media社の許諾を得た。[21]

4.4.5 項目18: 不確実性

推奨：予測効用値の平均まわりの標準誤差と個人レベルのばらつきを推定できる情報を報告する。

例：「PCSとMCSの平均スコアの共分散は0.0133だった。導出サンプルからの残差分散は0.02295だった」[36]。「確率感度分析を用いて、最もよい性能を示すFACT-Gモデルのマッピング係数における不確実性を検討した。ばらつきと変数間の相関を考慮するため、回帰係数は正規分布にしたがうと仮定し、またモデルの共分散行列を用いた。平均値が収束するために10万回のシミュレーションを行う必要があった。各シミュレーションの平均値ごとにEQ-5Dスコアを計算し、パーセンタイル値を用いて平均の推定値周辺のばらつきをまとめた。」[37]

解説：予測効用値の平均まわりの標準誤差や個人レベルの予測値とそれに関連するばらつきが正確に推定できるための情報を示す必要がある。これまで発表されたマッピング研究のほとんどはこの情報を十分詳細に報告していないため、不確実性の過小評価につながっている可能性がある[30]。このことは、その後の臨床研究における治療効果の信頼区間推定あるいは経済評価における増分費用効果比の正確さに大きく影響しうる。マッピング研究においてばらつきの推定方法には、パラメトリック法[36]、ノンパラメトリック法[38]、確率感度分析がある[39]。必要な情報は手法によって様々だが、回帰に基づくすべてのマッピング研究では、分散共分散行列と残差に加えて、選択したモデルのMAE、MSEあるいはRMSEのいずれかを報告する必要がある。また、最も

単純なモデル以外では、分散共分散行列はオンラインAppendixとして提供することが望ましい。

4.4.6 項目19: モデル性能と表面的妥当性

推奨：選択したモデルの予測精度や適合度の統計量など、モデル性能の結果を表あるいは文章により示す。選択したモデルの表面的妥当性の評価を提示する。

例：Longworthらの表17[37]

「11項目のうち、(有意ではないが) 予想と異なる負の統計量となったのは、役割感情機能 (RE) の次元における最初の項目だけだった。このため、縮約モデルでは同項目を除外した。心の健康 (MH) 次元の2項目についても、一貫性 (consistency) のある回帰式を得るために、2つの回答選択肢を1つにまとめる必要があった。」[40]

解説：選択されたモデルの予測精度については、その推定用データセットとすべてのバリデーション用データセットにおけるMAEおよび/あるいはMSEの推定値などを、表あるいは本文中で提示する必要がある。選択したモデルの適合度統計量、例えばR2値、あるいは赤池情報量基準 (AIC) やベイズ情報量基準 (BIC) のような情報量基準を提示してもよい。例えば、よい健康状態の人々と悪い状態の人々、対象尺度の効用値により層別された個人など、異なるサブグループにおいてモデル性能をあらわす統計量を示すことは、別の健康状態における選択したモデルの性能を示しているかもしれない[30]。観察効用値と選択したモデルから得られた予測効用値の相関を示す散布図も有用である。選択したモデルの表面

要約統計量とモデル性能	n	観察値	OLS model	OLS model	OLS model	OLS model	OLS model	OLS model
			2	3	4	6	7	8
			全次元	有意な次元のみ	有意な次元 + 二乗項	有意な項のみ	有意な項 + まとめ	有意な項 + 年齢
平均 (SD)	771	0.5793 (0.3423)	0.5793 (0.2797)	0.5793 (0.2792)	0.5793 (0.2830)	0.5793 (0.2863)	0.5793 (0.2844)	0.5793 (0.2866)
中央値		0.691	0.6281	0.6244	0.6451	0.6498	0.6557	0.6502
範囲		-0.5940 to 1	-0.1846 to 1.02	-0.1915 to 1.031	-0.3712 to 0.9419	-0.4078 to 0.9713	-0.3670 to 0.9430	-0.4046 to 0.9714
R <sup>2</sup>			0.668	0.665	0.684	0.700	0.691	0.701
調整済み R <sup>2</sup>			0.662	0.662	0.681	0.689	0.683	0.690
AIC			-286	-294	-340	-338	-330	-339
BIC			-216	-257	-307	-207	-237	-205
Ramsey RESET			F <sub>3,753</sub> = 12.57, p < 0.001	F <sub>3,761</sub> = 13.09, p < 0.001	F <sub>3,761</sub> = 1.56, p = 0.198	F <sub>3,737</sub> = 1.00, p = 0.3945	F <sub>3,736</sub> = 0.58, p = 0.6310	F <sub>3,736</sub> = 0.88, p = 0.449
MAE			0.149	0.151	0.143	0.139	0.142	0.139
Shrinkage			0.836	0.996	0.997	1.060	1.072	1.042

continued

NIHR Journal Library より転載許可を得た。 [37]

的妥当性について示すために、モデル係数のすべてが期待した符号であるかを明記する必要がある。これは、元の尺度における悪い健康状態が目的の尺度において低い効用値に関連していることを示す。モデル係数の相対的大きさについて、仮説を形成するのに十分な情報を事前に得ている場合、これらの仮説が支持されるかを明記しなければならない。選択したモデルで、係数の符号や大きさが予想と異なる結果となった変数を含める（あるいは除く）根拠を示し、それが意味するところを議論して、考えうる説明を加える必要がある。

## 4.5 考察

### 4.5.1 項目20: 先行研究との比較

推奨：先行研究で、同じ尺度間のマッピングアルゴリズムを開発したものがあつた場合、その詳細を示し、アルゴリズムの違いを、モデル性能や予測、係数の点から記述する。

例：「患者データに対してどのマッピング法がより信頼できるかを決めるために、本研究のモデルを先行研究と比較した…。文献上の既存モデルは、本研究で用いたデータセットにより再推定をするのではなく、公表されている結果とアルゴリズムにより推定された。この方法を用いたのは、臨床試験に含まれる健康状態尺度がSF-36（あるいはSF-12）のみであった時に、経済評価においてマッピングによりEQ-5D値を推定するためである。このため、実際の適用においては公表されている結果とアルゴリズムを用いるしかなく、このモデルを再推定することはできない…。図2に観察EQ-5D効用値、モデル（3）と既存法における予測EQ-5D効用値を示した…。すべての方法においてマッピングの関係は類似しており、より重度のEQ-5D健康状態のすべてに対して過大な予測をした。表3では、EQ-5D効用値範囲別に、実際の効用値と比較した既存法での予測効用値の平均誤差、平均絶対誤差、平均平方誤差を示した…。図2で示したように、すべてのモデルでより重度の健康状態の誤差が大きい。平均誤差、平均絶対誤差、平均平方誤差の点からも、本研究のモデルは既存モデルよりも優れている。」[41]

解説：元の尺度と対象とする尺度間のマッピングアルゴリズムが、他の研究により別に開発されている場合、類似点と相違点に関する情報は、新たなアルゴリズムの相対的利点を判断するために有用である。モデル性能における先行研究との一貫性の程度を考察する必要がある。対象とする尺度における予測値の範囲の違い、悪い健康状態における効用値の過大評価と良い健康状態における効用値の過小評価の程度は、モデル性能の比較により検討する必要がある。過去に同一の、あるいは類似した疾患領域や集団で研究が行われている場合、これらの研究

間にみられる、同じドメインの係数値の系統的な差異も報告すべきである。

可能であれば、過去に公表されたマッピングアルゴリズムを推定用データセットに適用し、既存モデルによる予測と新たなアルゴリズムによる予測とを比較検討することも必要である。マッピングアルゴリズムの差異についての考察は、可能性のある原因について考慮しなければならない。

### 4.5.2 項目21: 研究の限界

推奨：マッピングアルゴリズムの限界となりうるものについて概要を示す。

例：「本研究で示したMIC研究における分析と結果には、以下の5つの限界がある。1つ目は、このデータが、調査会社のパネルに登録した回答者から得たものであるため、少なくともオンライン質問票に回答する意思の強さが、標準的集団と異なっている可能性がある…。2つ目は、回答の負担を最小限にするため、現在、心疾患と診断されているか、心疾患は回答者の最も重篤な症状であるのかに限って質問している。心疾患のタイプやその罹患期間に関する情報は得ていない。そのため、追加情報があればクロスワーク関数において、より高い精度の予測が得られた可能性がある。3つ目は、6つのMAU尺度それぞれの効用値を計算するために、各国で同一のアルゴリズムを使用した。原則として、尺度ごとの効用値を計算する際には、各国におけるアルゴリズムをそれぞれ用いることが望ましい…。4つ目は、マッピングアルゴリズムを開発するために、3つの回帰推定値を用いているが、他の手法として、打ち切り最小絶対偏差モデルやtwo-partモデルなども考慮できたかもしれない…。最後の限界として、このマッピングアルゴリズムでは、外部のデータセットを用いたバリデーションを行っていないことがある。」[42]

解説：マッピングアルゴリズム研究の限界は、このアルゴリズムの利用者が、研究や意思決定への適用可能性を判断するのに役立つよう、そして適切な注釈により、それらを実際に用いる際に役立つよう議論する必要がある。これらの限界には、予測精度の低さや、推定用サンプルにおける潜在的なバイアス、特定の臨床あるいは一般集団のサブグループにおけるマッピングアルゴリズムのあてはまりの悪さ、などがある。同様に、他の方法やモデルが使用できたか、最終アルゴリズムをどの程度バリデーションできたか、特定の患者や集団に対して一般化できない可能性についても考慮する必要がある。

### 4.5.3 項目22: 適用範囲

推奨：マッピングアルゴリズムが利用可能な臨床および

研究の状況について概要を示す。

例：「これらの結果が一般化可能かという疑問が生じる。使用したデータは食道癌の患者から収集したものである。このデータを用いる利点は、EQ-5Dの5つの次元におけるそれぞれの水準に十分な数の患者がいたことである。しかし、この患者集団は、「平均的な」癌患者集団を代表していない可能性が高い。癌のタイプやステージと同様に、年齢や性別などがモデルの予測性能に影響する可能性がある。異なるタイプの癌患者集団、すなわち乳癌においてこのモデルは優れた予測値を示したが、患者の平均年齢は2つのデータセットで類似していた。EORTC QLQ-C30データをEQ-5D値に変換するにあたって、このモデルの使用が推奨されるためには、別の患者集団における予測性能を明らかにする今後の研究が必要である。」[43]

解説：マッピングアルゴリズムが、臨床や一般集団において、その使用が推奨できる患者や集団などの状況を示す必要がある。同時に、提示したアルゴリズムを使用すべきではない状況についても記述しなければならない。

#### 4.6 その他

##### 4.6.1 項目23: 追加情報

推奨：研究の資金源、非金銭的支援、研究の設計・実施・報告における資金提供者の位置づけを記述する。著者らと資金提供者の間に何らかの利益相反があれば報告する。

例：「この研究は、Acorda Therapeuticsから研究費助成を受けている。著者らは、研究の計画と実施、データの収集と管理・分析・解釈、本稿の作成とレビューにおいて、完全に中心的な役割をはたした。本稿の公表前に、Acorda Therapeuticsがその内容を確認した。」[44]  
「著者らは、同社との間に利益相反は存在しない。」[45]

解説：マッピング研究の計画と結果に、資金提供者がバイアスを与える可能性について示唆するエビデンスはないものの、研究の資金提供者に関して透明性を確保すべきである。同様に、研究に対する現物給付やその他何らかの援助があった場合は、それを報告すべきであり、例えば、著者に含まれていない個人やグループによる統計解析の支援や、より広範な研究支援、執筆への支援がある。マッピング研究への資金提供者や非金銭的支援をめぐる透明性については、その他最近の健康関連報告ガイドライン[12,14]の定めるところに従う必要がある。マッピング研究の計画、実施、報告における資金提供者の役割について概要を説明する必要がある。さらに、研究に関する自身や資金提供者の役割について、利益相反があ

る、あるいは利益相反があると思われる場合は、それを報告しなければならない。

#### 5 考察

ここ数年、健康効用値の予測にマッピングを用いる研究発表が急増している。最近のレビュー論文によると、2013年までに90件の研究が公表され、臨床あるいは健康関連QOL尺度とEQ-5Dのマッピングアルゴリズムは121件におよんでいる[2]。例えばSF-6D[46]やHealth Utilities Index (HUI) [47]などは多くのマッピングアルゴリズム(例[1,42,48-52])の対象となっているが、健康効用値を測定できるEQ-5D以外の選択に基づく尺度へのマッピングアルゴリズムはこのレビュー論文には含まれていない。多くの医療経済評価において、一次データの収集は困難であり、そのような状況を考えると、健康効用値を推定するためのマッピングアルゴリズムの流行は今後も続くだろう。しかし、マッピングは新たな不確実性をもたらすことから、適切な効用値の尺度を用いて一次データを収集することが望ましい。

このMAPS声明は、必須項目のチェックリストという形で推奨を提供するために開発されたものであり、マッピング研究を報告する際に考慮する必要がある。これは方法論に関するガイドラインとなるよう意図されたものでなく、研究方法の質を評価するツールでもない。最適に近い報告により、読者、究極的には政策決定者によって、誤った結論が導かれることを避けるのが目的である。この声明は、最近の医学研究における他の報告ガイドラインにそって、MAPS声明に含まれる23項目の理解をより深めるための説明と詳細に関する記述で構成されている。これが、マッピング研究を報告する研究者の教育的枠組みとして働くことを願っている。本声明の解説・詳細に関する構成は、最近の他の報告解説書に準じている[9-14]。

このMAPS声明とその解説・詳細の記述は、最近発行された医学研究報告ガイドライン[5]が骨格となっている。デルファイパネルは、関連専門家と雑誌編集者からなる学際的な多国籍チームで構成された。パネルにはマッピング研究の経験を持つメンバーもおり、EQ-5Dマッピング研究に関する最近のレビュー [2]には、84名の第一著者が含まれるが、そのうち31名(36.9%)がパネルリストに選ばれた。より大きなパネルであっても推奨される最終的な内容は変わらなかっただろうと考える。本研究で使用したデルファイ法には、医療の報告ガイドラインの作成者[15]によりごく最近用いられた分析方法のみを含めた。このMAPSチェックリストに厳密にしたがうことにより、マッピング報告の語数が増えるかについては判断できないところである。ただ、雑誌によるオンラインAppendixのさらなる利用により、報告本文の語数制限が厳しい状況でも、包括的な報告が可能になるだろうと思われる。

医学研究における他の報告ガイドラインでは、報告

チェックリストの公表以降、報告の質が向上した[53-55]との結果が示されている。しかし、MAPSではこれまでのところ、その使用によりマッピング研究の報告の質が改善したとの実証的な結果は出ていない。MAPS作業グループが計画している今後の調査には、MAPS声明導入の利点（そして、実際には有害な影響の可能性）に関する事前と事後の評価などがある。この分野の概念的、方法論的、実用的進展に対応するため、将来においてMAPS声明を更新する必要が生じるだろう。更新の際に反映される可能性のある方法論的進展には、より複雑なモデルへの移行、不確実性をより適切に扱う方法、MAEやMSEなど予測精度の統計量を適切に用いるガイダンスなどがあるかもしれない。MAPS作業グループは、この報告チェックリストの更新の必要性について、5年後に検討する予定である。

最後に、本稿はMAPS声明の項目の説明と詳細を提供するものである。医療経済誌やQOL関連雑誌がMAPSを受け入れ、査読における利用を促進し、編集上の要件と「投稿規定 (instruction to authors)」が更新されていくよう努力する。

**謝辞** デルファイパネルへのコメントや参加を公表することを許可した以下の人々に謝意を表す：Roisin Adams, Roberta Ara, Nick Bansback, Garry Barton, Andy Briggs, Chris Carswell, Gang Chen, Doug Coyle, Ralph Crott, Richard Edlin, Alastair Fischer, Juan Manuel Ramos-Gon<sup>o</sup>i, Ola Ghatnekar, Patrick Gillard, Ning Yan Gu, Annie Hawton, Haomiao Jia, Kamran Khan, Billingsley Kaambwa, Nick Kontodimopoulos, Quang Le, Jason Madan, Katherine Mac-Gilchrist, David Meads, Duncan Mortimer, Bhash Naidoo, Angus Ng, Nalin Payakachat, Simon Pickard, Rafael Pinedo-Villanueva, Eleanor Pullenayegum, Jeff Richardson, Janet Robertson, Jeff Round, Donna Rowen, Sean Rundell, Paulos Teckle, Matthijs Versteegh, Feng Xie, and Tracey Young.

## 倫理基準への準拠

本研究実施にあたり、外部からの研究費は受けなかった。作業グループのすべてのメンバーは自由意思に基づいて行動した。

RFは、デルファイ研究のためにサービスを提供する企業 (Clinvivo Limited) の株主であり、取締役である。しかし、本研究でこれらのサービスは利用しなかった。したがって、上記の関係にかかわらず、RFは利益相反がないことを宣言している。残る著者 (SP, ORA, HD, LL, MO, AG) にも開示すべき利益相反はない。この研究の計画、実施、結果の解釈、報告書作成に著者らは全員関係した。SPはこの研究の保証人を務めている。

## 参考文献

- [1] 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.* 2010;11(2):215-225. doi:10.1007/s10198-009-0168-z.
- [2] Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes.* 2013;11:151. doi:10.1186/1477-7525-11-151.
- [3] Chuang LH, Whitehead SJ. Mapping for economic evaluation. *Br Med Bull.* 2012;101:1-15. doi:10.1093/bmb/ldr049.
- [4] Longworth L, Rowen D. NICE DSU technical support document 10: the use of mapping methods to estimate health state utility values. Sheffield: Decision Support Unit, SCHARR, University of Sheffield; 2011.
- [5] Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med.* 2010;7(2):e1000217. doi:10.1371/journal.pmed.1000217.
- [6] Moher D, Simera I, Schulz KF, Hoey J, Altman DG. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Med.* 2008;6:13. doi:10.1186/1741-7015-6-13.
- [7] Narahari SR, Ryan TJ, Aggithaya MG, Bose KS, Prasanna KS. Evidence-based approaches for the Ayurvedic traditional herbal formulations: toward an Ayurvedic CON-SORT model. *J Altern Complement Med.* 2008;14(6):769-776. doi:10.1089/acm.2007.0818.
- [8] 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. doi:10.1186/1477-7525-7-33.
- [9] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med.* 2003;138(1):W1-12.
- [10] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care.* 2007;19(6):349-357. doi:10.1093/intqhc/mzm042.
- [11] Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4(10):e297. doi:10.1371/journal.pmed.0040297.
- [12] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche

- PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-34. doi:10.1016/j.jclinepi.2009.06.006.
- [13] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010;340:c869. doi:10.1136/bmj.c869.
- [14] Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*. 2013;16(2):231-250. doi:10.1016/j.jval.2013.02.002.
- [15] Fitch K, Bernstein S, Aguilar M, Burnand B, LaCalle J, Lazaro P, van het Loo M, McDonnell J, Vader J, Kahan J. *The RAND/UCLA Appropriateness Method User's Manual*. Santa Monica; 2001.
- [16] Badia X, Roset M, Valassi E, Franz H, Forsythe A, Webb SM. Mapping CushingQOL scores to EQ-5D utility values using data from the European Registry on Cushing's syndrome (ERCUSYN). *Qual Life Res*. 2013;22(10):2941-2950. doi:10.1007/s11136-013-0396-7.
- [17] Khan KA, Madan J, Petrou S, Lamb SE. Mapping between the Roland Morris Questionnaire and generic preference-based measures. *Value Health*. 2014;17(6):686-695. doi:10.1016/j.jval.2014.07.001.
- [18] Wong HL, Truong D, Mahamed A, Davidian C, Rana Z, Einarson TR. Quality of structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*: a 10-year follow-up study. *Curr Med Res Opin*. 2005;21(4):467-473.
- [19] Kaambwa B, Billingham L, Bryan S. Mapping utility scores from the Barthel index. *Eur J Health Econ*. 2013;14(2):231-241. doi:10.1007/s10198-011-0364-5.
- [20] Cheung YB, Tan LC, Lau PN, Au WL, Luo N. Mapping the eight-item Parkinson's Disease Questionnaire (PDQ-8) to the EQ-5D utility index. *Qual Life Res*. 2008;17(9):1173-1181. doi:10.1007/s11136-008-9392-8.
- [21] Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL generic core scales. *Pharmacoeconomics*. 2014;32(7):693-706. doi:10.1007/s40273-014-0153-y.
- [22] Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Qual Life Res*. 2013;22(3):68-94. doi:10.1007/s11136-012-0189-4.
- [23] Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. *Rheumatology (Oxford)*. 2013;52(5):944-950. doi:10.1093/rheumatology/kes400.
- [24] Hawton A, Green C, Telford CJ, Wright DE, Zajicek JP. The use of multiple sclerosis condition-specific measures to inform health policy decision-making: mapping from the MSWS-12 to the EQ-5D. *Mult Scler*. 2012;18(6):853-861. doi:10.1177/1352458511429319.
- [25] Oppe M, Devlin N, Black N. Comparison of the underlying constructs of the EQ-5D and Oxford Hip Score: implications for mapping. *Value Health*. 2011;14(6):884-891. doi:10.1016/j.jval.2011.03.003.
- [26] Young MK, Ng SK, Mellick G, Scuffham PA. Mapping of the PDQ-39 to EQ-5D scores in patients with Parkinson's disease. *Qual Life Res*. 2013;22(5):1065-1072. doi:10.1007/s11136-012-0231-6.
- [27] Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med Decis Mak*. 2006;26(1):18-29. doi:10.1177/0272989X05284108.
- [28] Ramos-Gonçá lves J, Rivero-Arias O, Dakin H. Response mapping to translate health outcomes into the generic health-related quality of life instrument EQ-5D: introducing the mrs2eq and oks2eq commands. *Stata J*. 2013;13:474-491.
- [29] Payakachat N, Summers KH, Pleil AM, Murawski MM, Thomas J 3rd, Jennings K, et al. Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Qual Life Res*. 2009;18(7):801-813. doi:10.1007/s11136-009-9499-6.
- [30] Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010.
- [31] Eriksson L JE, Kettaneh-Wold N, Trygg J, Wikström C, Wold S. Appendix I: model derivation, interpretation and validation. Multi- and megavariate data analysis part 1: basic principles and applications. 2nd ed. Umea: MKS Umetrics AB; 2006.
- [32] Blum A KA, Langford J, editor. Beating the hold-out: bounds for K-fold and progressive cross-validation. In: 12th annual conference on computational learning theory. New York: ACM; 1999.
- [33] Browne C, Brazier J, Carlton J, Alavi Y, Jofre-Bonet M. Estimating quality-adjusted life years from patient-reported visual functioning. *Eye (Lond)*. 2012;26(10):1295-1301. doi:10.1038/eye.2012.137.
- [34] Buxton MJ, Lacey LA, Feagan BG, Niecko T, Miller DW, Townsend RJ. Mapping from disease-specific measures



- to utility: an analysis of the relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. *Value Health*. 2007;10(3):214-220. doi:10.1111/j.1524-4733.2007.00171.x.
- [35] Dakin H, Petrou S, Haggard M, Bengt S, Williamson I. Mapping analyses to estimate health utilities based on responses to the OM8-30 Otitis Media Questionnaire. *Qual Life Res*. 2010;19(1):65-80. doi:10.1007/s11136-009-9558-z.
- [36] Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. *Med Decis Mak*. 2004;24(2):160-169. doi:10.1177/0272989X04264015.
- [37] Longworth L, Yang Y, Young T, Mulhern B, Hernandez Alava M, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision making: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014;18(9):1-224. doi:10.3310/hta18090.
- [38] Chan KK, Willan AR, Gupta M, Pullenayegum E. Underestimation of uncertainties in health utilities derived from mapping algorithms involving health-related quality-of-life measures: statistical explanations and potential remedies. *Med Decis Mak*. 2014;34(7):863-872. doi:10.1177/0272989X13517750.
- [39] Ara R, Wailoo A. Using health state utility values in models exploring the cost-effectiveness of health technologies. *Value Health*. 2012;15(6):971-974. doi:10.1016/j.jval.2012.05.003.
- [40] Lundberg L, Johannesson M, Isacson DG, Borgquist L. The relationship between health-state utilities and the SF-12 in a general population. *Med Decis Mak*. 1999;19(2):128-140.
- [41] Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes*. 2009;7:27. doi:10.1186/1477-7525-7-27.
- [42] Chen G, McKie J, Khan MA, Richardson JR. Deriving health utilities from the MacNew Heart Disease Quality of Life Questionnaire. *Eur J Cardiovasc Nurs*. 2014. doi:10.1177/1474515114536096.
- [43] McKenzie L, van der Pol M. Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: the potential to estimate QALYs without generic preference data. *Value Health*. 2009;12(1):167-171. doi:10.1111/j.1524-4733.2008.00405.x.
- [44] Limone BL, Sidovar MF, Coleman CI. Estimation of the effect of dalfampridine-ER on health utility by mapping the MSWS-12 to the EQ-5D in multiple sclerosis patients. *Health Qual Life Outcomes*. 2013;11:105. doi:10.1186/1477-7525-11-105.
- [45] Goldsmith KA, Dyer MT, Buxton MJ, Sharples LD. Mapping of the EQ-5D index from clinical outcome measures and demographic variables in patients with coronary heart disease. *Health Qual Life Outcomes*. 2010;8:54. doi:10.1186/1477-7525-8-54.
- [46] Brazier J, Roberts J, Deverill M. The estimation of a preference based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271-292.
- [47] Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. *Health Utilities Index*. *Pharmacoeconomics*. 1995;7(6):490-502.
- [48] Yang Y, Wong MY, Lam CL, Wong CK. Improving the mapping of condition-specific health-related quality of life onto SF-6D score. *Qual Life Res*. 2014;23(8):2343-2353. doi:10.1007/s11136-014-0668-x.
- [49] Lee L, Kaneva P, Latimer E, Feldman LS. Mapping the gastrointestinal Quality of Life Index to short-form 6D utility scores. *J Surg Res*. 2014;186(1):135-141. doi:10.1016/j.jss.2013.08.006.
- [50] Roset M, Badia X, Forsythe A, Webb SM. Mapping CushingQoL scores onto SF-6D utility values in patients with Cushing's syndrome. *Patient*. 2013;6(2):103-111. doi:10.1007/s40271-013-0010-7.
- [51] Payakachat N, Tilford JM, Kuhlthau KA, van Exel NJ, Kovacs E, Bellando J, et al. Predicting health utilities for children with autism spectrum disorders. *Autism Res*. 2014;7(6):649-663. doi:10.1002/aur.1409.
- [52] Goldfeld KS, Hamel MB, Mitchell SL. Mapping health status measures to a utility measure in a study of nursing home residents with advanced dementia. *Med Care*. 2012;50(5):446-451. doi:10.1097/MLR.0b013e3182407e0d.
- [53] Moher D, Jones A, Lepage L, Group C. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001;285(15):1992-1995.
- [54] Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, Doig CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care*. 2005;9(5):R575-582. doi:10.1186/cc3803.
- [55] Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 2006;185(5):263-267.