

「非選好型尺度で測定されたデータから選好型尺度で測定されたデータへのマッピングにおける概念的妥当性検証」

研究分担者

下妻晃二郎（立命館大学 生命科学部 生命医科学科 教授、同大学 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（CHEERS）ユニット長）

研究協力者

兼安貴子（立命館大学 生命科学部 生命医科学科 助教）

1. はじめに

わが国においては 2019 年から医療技術評価（health technology assessment: HTA）の政策応用の仕組みが開始された。そこでは評価対象の医薬品・医療材料の費用効果分析が行われるが、分析における効果の指標としては、生存期間と QOL の統合指標である質調整生存年(quality-adjusted life year (QALY))が専ら用いられる。またそこで使用される、質、すなわち quality weight は、健康状態効用値(health-state utility values)（以下「効用値」）（日本の費用対効果評価の分析ガイドライン¹⁾上の用語では「QOL 値」）は、理論的には、質の高い臨床試験や臨床研究内などにおいて選好型尺度（preference-based measure: PBM）を用いて測定された値を使用することが望ましいとされる。実際、上記のガイドラインの「8.2」では PBM を、中でも特に EQ-5D-5L の使用を推奨している¹⁾。さらに「8.2.2」では、「8.2」に該当するデータが存在しない場合は、その他の適切な健康関連 QOL(HRQOL)尺度で測定したデータから効用値へマッピングしたデータを使用してもよいが、その場合は「適切な手法を用いて効用値に変換していること」を説明しなければならない、としている。

さて上記で述べられている「適切な手法」とは何か、であるが、マッピングについては代表的なガイダンスが2つ出版されている。MAPS 声明²⁻⁷⁾や ISPOR 研究班による実施基準⁸⁾などが公表されている。前者は 23 項目のチェックリストであり、後者は 23 の推奨が示されている。後者の内訳は、モデリング前に関することが 6 項目、モデリングとデータ分析に関することが 5 項目、報告書に関することが 12 項目である。

いずれのガイダンスも、マッピング前後の数学的な当てはまりに関する内容が主で、概念的妥当性(conceptual validity)、すなわち、マッピング前に用いられた尺度の概念構成と

マッピング後の尺度の概念構成が適合しているか、そもそもそれらを用いてマッピングをしてもいいのか、に関する記述は多くない。例えば MAPS 声明では 23 項目のうち 1 項目で、ISPOR 実施基準では、モデリング前の 6 項目中の 2-3 項目と報告書に関する 12 項目中の 2 項目程度で触れられていたりするにすぎない。

また、これらのガイダンスには必ずしも触れられていないが、基本的にはマッピングで扱うデータや変換式(algorithm)は、日本の HTA で用いるのが目的であれば、日本人や日本在住の人を対象に測定されたデータを用いることも大事である。

以上の背景から、本報告では、マッピングの概念的妥当性についての、近年の議論に焦点をあてて紹介する。

2. マッピングに関する論文の系統的レビュー

本題の前に、近年のマッピングの手法や議論について、萩原の講演資料⁹⁾から、Mukuria C, et al の最新の系統的レビュー¹⁰⁾の概要を紹介する。2007-18 年の 180 研究がまとめられており、マッピング後の PBM としては、EQ-5D が 63%、SF-6D が 19%、HUI が 6%の割合で使用され、マッピングモデル作成時のサンプルサイズは、30~130,000 (100 以下は 10%) であり、マッピングモデルの種類と割合は、直接マッピングが 82%、レスポンスマッピングが 18%であったとされる。線型回帰は 75%以上で利用され、2 つ以上の回帰モデルが 61%の論文で使用されていた。共変量としては、年齢 51%、性別 55%、臨床指標が 20%で使われており、さらに、予測誤差指標の報告は 82%の論文で認められていた。

3. マッピング前後で使用した尺度間の概念の重なり (conceptual overlap) の有無と程度の確認

Mukuria C, et al¹⁰⁾によると、マッピング前後で使用した尺度間の概念の重なりや程度について方法論は未開発であるとされる。そこで、Round J and Hawton A は、これらの検証方法についての考え方や提案を発表している¹¹⁾ので次に紹介する。

3-1. マッピングの概念的妥当性の考え方

マッピング前と後と同じものを測定しているという確信がない場合、前に測定したスコアが後で測定しているスコアをそもそも高く予測しているとは言えない。すなわち、適切な統計学的手法を用いて信頼性の高いマッピング性能を見積もっていることは、マッピングの過程が妥当であることにとって必要ではあるが、十分とはいえない。

マッピング前に用いられる non-PBM の多くは、疾患特異的な状態に興味があったり、介入に関する感度が高かったりすることを重視して使用されていることが多いが (臨床現場で役立つ HRQOL を測定したいのであるので当然)、それが一般的な健康度や健康に関する選好、すなわち PBM で測定されるスコアの良い指標とは限らない。言い換えれば、

マッピング前の尺度と後の尺度はそもそも異なる目的を持った尺度であることが多く、概念的な妥当性に欠けることがしばしばある。

上記のように、マッピング前後の2つの尺度が全く同じドメイン（要素、次元）を同じようにカバーしていることはないため、マッピング前のデータと後のデータの間で概念的な首尾一貫性(coherence)の程度を検討することは重要である。より具体的には、概念構造に殆ど重複が認められない場合はマッピング性能も妥当なものにはならない。

そもそも、EQ-5DなどのPBMで測定する値には、想像できる最高の健康が1、死が0、と上限と下限が定められている一方、non-PBMでは、各質問項目やドメインの最高点や最低点はそれぞれ最高や最低の健康度やQOLとは限らない。つまり、臨床的に興味がある部分を拡大して測定した状態になっている。また、ドメイン間でも、例えばAQLQ（喘息のPRO尺度）では、主な関心事である呼吸に関するドメインは他のドメインよりも選好がより大きくなる。現時点では、その過剰な大きさの程度を測定することは容易ではない。

マッピング前に使用されるnon-PBMの中でも特に疾患や状態特異的尺度はマッピングされる後によく使用されるgeneric PBMよりも、狭い範囲の概念を測定している場合が多く、単に数学的な当てはまりだけを見ていると、失われた情報に気づかないこともあり得る。つまり、マッピング前後で測定している概念の範囲が異なっているのに、単に数学的な当てはまりだけを重視すると、失われた健康の価値と維持されている健康の価値に関する系統的なバイアスが生じることになる。これは無視できない課題である。

3-2. マッピングにおける概念的妥当性改善法に関する preliminary な提案

Round J and Hawton A は、上記の課題を踏まえて、概念的妥当性を改善するための preliminary な提案を行っている¹¹⁾。

主な提案は、マッピングモデルとして、直接マッピングではなく、レスポンスマッピングを用いることである。後者は前者と異なり、2段階のプロセスを踏む。すなわち、マッピング前の尺度で得られたスコアを、マッピング後にEQ-5Dのスコアに変換したい場合、まず、EQ-5Dの次元のスコアに回帰し、次にそれを、一般人のサンプルから得られた変換式を用いて最終的なEQ-5Dのインデックスの値に割り当てるのである。

後者は、前者と比べて明らかな統計学的有意性は提供しないようにみえるが、個々のスコアとスコアの範囲を予測するにはよい働きをする。すなわち、レスポンスマッピングは、前後の両尺度間の関係性に関して、より大きな概念的明確性を与えることになる。つまり、何が何にマッピングされているか、そして、マッピングの過程で何が失われ何が維持されているのかがより明らかになる。

このことは、HTAにおける資源配分の意味決定の妥当性(decision validity)の点からも大事である。概念的妥当性を考慮していないマッピングデータを用いた場合、得られた資源配分の意味決定が妥当かどうかとも評価できない。

個々の尺度の妥当性は伝統的に一連の確立された4つの基準、すなわち、face, content, criterion, construct validity によって検証されているが、これらはマッピング性能における妥当性とは異なり、ある尺度から別のある尺度への変換の過程の妥当性を評価するのに必要である。

概念的な意思決定妥当性の検討にも、上記の4つの妥当性が必要とされる。

現時点では、マッピングの概念的妥当性に特化したガイダンスは存在しない。

4. おわりに

HTAにおいて、マッピング過程の概念的妥当性の検証が適切に行われていない場合、資源配分に関する意思決定の妥当性も確保されないことになる。従って、概念的妥当性が十分検証されないマッピングによって得られた効用値は安易に使用するべきではない。

文献

1. 中央社会保険医療協議会における費用対効果評価の分析ガイドライン第3版 国立保健医療科学院 保健医療経済評価研究センター (C2H). 2022年1月19日 中医協総会了承 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Ffc2h.niph.go.jp%2Ftools%2Fguideline%2Fguideline_ja.pdf&clen=1024931&chunk=true
2. Petrou S, et al: Preferred reporting items for studies mapping onto preference-based outcome measures: The MAPS statement. *J Med Econ* 2015;18(11):851-857.
3. Petrou S, et al: Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Int J Technol Assess Health Care*. 2015;31(4):230-235.
4. Petrou S, et al: Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Pharmacoeconomics* 2015;33(1):985-991.
5. Petrou S, et al: Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Appl Health Econ Health Policy*. 2015;13(5):437-443.
6. Petrou S, et al: The MAPS reporting statement for studies mapping onto generic preference-based outcome measures: explanation and elaboration. *Pharmacoeconomics* 2015;33(10):993-1011.
7. 白岩健ら：医療経済評価におけるマッピングと MAPS 声明について. *保健医療科学* 2018;67(4):422-426.
8. 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;20(1):18-27.

9. 萩原康博：QOL 値へのマッピング手法の現状と課題. 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(CHEERS) 特別シンポジウム 2022年3月5日(web開催)
10. Mukuria C, et al: 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 2019;17(3):295-313.
11. Round J and Hawton A: Statistical alchemy: Conceptual validity and mapping to Generate Health State Utility Values. Pharmacocon Open 2017;1(4):233-239.