

ISOQOLは患者中心アウトカム研究と 効果比較研究で用いる患者報告アウトカム尺度の ための最小基準を推奨する

Bryce B. Reeve, Kathleen W. Wyrwich, Albert W. Wu, Galina Velikova, Caroline B. Terwee, Claire F. Snyder, Carolyn Schwartz, Dennis A. Revicki, Carol M. Moinpour, Lori D. McLeod, Jessica C. Lyons, William R. Lenderking, Pamela S. Hinds, Ron D. Hays, Joanne Greenhalgh, Richard Gershon, David Feeny, Peter M. Fayers, David Cella, Michael Brundage, Sara Ahmed, Neil K. Aaronson, Zeeshan Butt

受理日: 2012年12月17日/オンライン掲載日: 2013年1月4日
© Springer Science+Business Media Dordrecht 2013

訳 堺琴美¹⁾ 星野絵里¹⁾ 兼安貴子^{1,2)} 下妻晃二郎^{1,2)} 鈴鴨よしみ³⁾

¹⁾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) ²⁾立命館大学 生命科学部 ³⁾東北大学大学院医学系研究科障害科学専攻肢体不自由学分野

本研究は国際QOL研究学会 (International Society for Quality of Life Research: ISOQOL) を代表して実施した。
電子補足資料: 本論文のオンライン版 (doi:10.1007/s11136-012-0344-y) には承認されたユーザーが利用できる補足資料が含まれる。

B. B. Reeve: Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 1101-D McGavran-Greenberg Building, 135 Dauer Drive, CB 7411, Chapel Hill, NC 27599-7411, USA
e-mail: bbreeve@email.UNC.edu

B. B. Reeve, J. C. Lyons: Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

K. W. Wyrwich, D. A. Revicki • W. R. Lenderking: United BioSource Corporation, Bethesda, MD, USA

A. W. Wu, C. F. Snyder: Johns Hopkins School of Medicine, Baltimore, MD, USA

G. Velikova, J. Greenhalgh: University of Leeds, Leeds, UK

C. B. Terwee: VU University Medical Center, Amsterdam, The Netherlands

C. Schwartz: DeltaQuest Foundation, Inc., Concord, MA, USA

C. M. Moinpour: Fred Hutchinson Cancer Research Center, Seattle, WA, USA

L. D. McLeod: Research Triangle Institute Health Solutions, Durham, NC, USA

P. S. Hinds: Children's National Medical Center, Washington, DC, USA

P. S. Hinds: The George Washington University School of Medicine, Washington, DC, USA

R. D. Hays: David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

R. Gershon, D. Cella • Z. Butt: Northwestern University Feinberg School of Medicine, Chicago, IL, USA

D. Feeny: University of Alberta, Alberta, Canada

P. M. Fayers: University of Aberdeen, Aberdeen, UK

P. M. Fayers: Norwegian University of Science and Technology (NTNU), Trondheim, Norway

M. Brundage: Queen's University, Kingston, ON, Canada

S. Ahmed: McGill University, Montreal, QC, Canada

N. K. Aaronson: The Netherlands Cancer Institute, Amsterdam, The Netherlands

N. K. Aaronson: University of Amsterdam, Amsterdam, The Netherlands

抄 録

目的：患者中心アウトカム研究 (patient-centered outcomes research: PCOR) や効果比較研究 (comparative effectiveness research: CER) のもっとも重要な側面は、介入を評価するために患者の観点や経験と臨床データとを統合することである。したがって、PCOR と CER では、研究、医療提供、政策に用いる情報を提供するために患者報告アウトカム (patient-reported outcome: PRO) データを適切に収集する必要がある。本イニシアティブの目標は PCOR や CER で用いる PRO 尺度の設計と選択のための最小基準を特定することであった。

方法：PRO 尺度の選択のための既存のガイドラインを特定するために文献レビューを行った。また、国際 QOL 研究学会 (International Society for Quality of Life Research: ISOQOL) のメンバーを対象としたオンライン調査も行い、PRO 基準についての意見を求めた。回答者の 50% 超が「最小基準として必要である」として支持した基準を「推奨」に指定

した。

結果：文献レビューにより 387 件の論文が特定された。調査回答率は ISOQOL メンバー 506 名中 120 名であった。回答者の PRO 研究経験は平均 15 年であり、89% の人がフィードバックを提供するのに自分は「適格である」または「極めて適格である」と回答した。PRO 尺度の測定基準についての最終的な推奨は以下のとおりであった：概念および測定モデルの文書化、信頼性や妥当性 (内容的妥当性、構成概念妥当性、反応性) のエビデンス、スコアの解釈可能性、質の高い翻訳、実施可能な患者と研究者の負担。

結論：これらの最小基準の開発は、PCOR と CER に用いる情報を提供するための PRO 尺度の適切な使用を促進し、それによって医療提供の有効性と効率が改善することを意図したものである。次の段階は、意思決定に関連する PRO 尺度の選択のためのベストプラクティスを特定するために、これらの最小基準を拡張することである。

キーワード 患者報告アウトカム, 効果比較, 患者中心アウトカム研究, 計量心理学, 質問票

緒 言

患者中心アウトカム研究 (patient-centered outcomes research: PCOR) と効果比較研究 (comparative effectiveness research: CER) のもっとも重要な側面は、医療介入の安全性や有効性を評価するために、患者の観点から評価した健康状態と、臨床的および生物学的データとを統合することである。これらを統合することにより、健康関連 QOL (HRQOL) と、それが疾患や治療によってどのような影響を受けるかが、生存期間や抗腫瘍効果などの従来の臨床エンドポイントを補完する、ということが認識される。HRQOL のエンドポイントについては、患者が経験していることに関する最良の情報源は患者自身の報告であることが広く受け入れられている。PCOR と CER の課題は、どのようにすれば、医療提供、研究、政策立案における意思決定に役立つ

形で患者報告データをもっともよく収集できるかにある。

観察研究や介入研究では、患者報告アウトカム (PRO) 尺度の使用が増加している。米国食品医薬品局 (FDA) は、PRO 尺度を「臨床医や他の誰の解釈も介さず、患者から直接得られる患者の健康状態に関するすべての報告¹⁾」と定義している。患者は、自覚症状 (例: 疼痛, 疲労, 悪心), 機能状態 (例: 性的機能, 腸・排尿機能), ウェルビーイング (例: 身体的, 精神的, 社会的), 生活の質 (QOL), ケアや治療に対する満足度などの、介入または疾病負担を評価するうえで重要な多数の領域について正確な報告を行うことができる¹⁻⁴⁾。患者こそがそのような領域を評価するためのゴールドスタンダードの情報源であることはほぼ間違いない。患者中心アウトカムに関する妥当な研究上の結論を導くためには、PRO 測定は十分に頑健な測定特性を示す評価尺度

を用いた標準化された方法で行う必要がある⁴⁻⁹⁾。

本研究の目標は、PCOR や CER で使用する PRO 尺度を選択するための最小基準を特定することであった。最小基準の定義は、ある PRO 尺度がこれらの基準を満たさない場合、PCOR での研究で使用するには不適切であると判断されるような基準とした。この基準一式の開発の主眼は、PCOR 研究で用いる PRO 尺度を判断するための重要特性を明確に定めることであった。これらの基準の特定には2つの相補的なアプローチを用いた。1 番目の方法は公表および未公表のガイダンス文書を含む文献の広範なレビューであった。2 番目は、正式な調査を通じて、PRO 測定および PCOR の専門家の国際的グループである国際 QOL 研究学会 (International Society for Quality of Life Research: ISOQOL) のメンバーからの意見をj得ることであった¹⁰⁾。本研究の主目的ではないものの、これらのアプローチにより、必要な最小基準とはみなされないが、PRO 尺度の「ベストプラクティス」とみなされる基準を特定することもできた。

最小基準の特定は、医療提供を促進し、最終的に患者の健康とウェルビーイングを改善するという目標を PCOR や CER が達成することを可能にするための第 1 段階である。科学的に確実で、意思決定に関連する PRO 尺度へのアクセスは、研究者が、患者の観点からみた介入のさまざまな益に関する経験的エビデンスを収集することを可能とする^{6,9,11,12)}。次に、この情報を患者、医療提供者および政策決定者に普及することにより、患者に対して意味のあるエンドポイントを用いた、患者の生活への介入の影響に関するより豊富な観点を提供することができる¹³⁾。

方 法

本論文は米国患者中心アウトカム研究所 (Patient-Centered Outcomes Research Institute: PCORI) が資金提供した研究に基づいている¹⁴⁾。本論文は PCORI が別途公表した PCORI の方法論委員会 (Methodology Committee) による基準を代表するものではないが、これらの基準の一部はこの作業によって情報提供された¹⁵⁾。推奨基準の原案作成と最終選択を導くために、本論文の著者で構成される

ISOQOL 科学的アドバイザリー・タスクフォース (scientific advisory task force: SATF) が設立された。SATF が推奨案の作成をするための文献レビューを行った後、正式な調査において ISOQOL メンバーが推奨案のレビューを行った。文献レビューと、ISOQOL メンバーからの回答とフィードバックは、本論文に記載した最終的な推奨に考慮された。

文献レビュー

PRO 尺度に関連する既存のガイダンス文書を特定するために、公表および未公表の文献の系統的レビューを行った。このレビューでは、PCOR と CER における PRO 尺度の選択、関連評価尺度の特性 (例: 信頼性, 妥当性, 回答の負担, 解釈可能性), ならびにこれらの特性を評価するための定性的および定量的方法の使用を特定した。コンセンサス声明, ガイドライン, エビデンスに基づく論文に焦点を置き, 特に, 広く一般化可能な原則を記述した論文や文書に焦点をあてた。ただし, ある集団や評価尺度に特異的であったいくつかの論文は計量心理学的方法が厳格であったため対象とした。

文献レビューでは、公表されている MEDLINE 検索方法を採用し、PRO 尺度の測定特性を特定した¹⁶⁾。公表されている検索方法を基本にして、MEDLINE のシソーラス、米国国立医学図書館が定める生命科学用語集 (Medical Subject Headings: MeSH)、米国心理学会 (American Psychological Association: APA) のオンライン版心理学用語シソーラスを用いて適用した。MEDLINE, PsycINFO, Combined Index to Nursing and Allied Health Literature (CINAHL) など複数の関連電子データベースの検索も並行して行った (データベースにおける検索用語は付録 1, 電子的補足資料を参照)。公表日または対象者の年齢による制約はあらかじめ設けなかった。また、ISOQOL メンバーのメーリングリストへの依頼を通じて関連論文を入手した。

1 名の著者 (ZB) が特定された論文とガイドラインのタイトルと抄録のレビューを行った。関連論文の全文を入手し、レビューした。選択された論文で引用されている参考文献のレビューを行い、追加の関連論文を特定した。ZB が本研究に必要な情報を抽出し、作業の一貫性を保証するために、他の 2 名

の著者 (DC と RG) が独立していくつかの論文のレビューを行った。

文献レビューにより収集した PRO 測定基準に基づき、ISOQOL SATF が推奨案を作成し、ISOQOL メンバーが以下に記載した調査によりレビューを行った。SATF による一連の電子メールと電話会議の繰り返し実施を通して、系統的な文献レビューによって特定した候補となる基準を話し合い、議論した。候補となる基準間の重複は最小限に抑え、類似の項目は統合した。メンバー間で見解の相違が生じた場合は、調査において異なる選択肢を保持し、メンバー全員が候補となる基準を評価し、コメントできるようにした。その結果として得られた調査は 23 の最小基準の候補からなり、ISOQOL メンバーがそれらの評価を行った。

ISOQOL メンバー調査

ISOQOL は、効果的な介入を特定し、医療の質を高め、人々の健康を促進するために、HRQOL やその他の患者中心アウトカムの科学研究の推進に従事している¹⁰⁾。ISOQOL は 1993 年に設立された国際的な協同ネットワークであり、研究者、臨床家、患者支援者、政府系科学者、業界代表者、政策決定者から構成される。ISOQOL メンバーの多くが PRO 方法論の専門家であり、研究、医療提供、集団調査における患者報告データの特定と適用を改善するために、最新の定性的・定量的な方法の使用に焦点を置いている。研究で広く用いられている PRO 尺度や、PRO 尺度の開発と評価のためのガイドラインの多くは、ISOQOL メンバーにより開発されたものである。調査時点で、506 名の ISOQOL メンバーがメーリングリストに登録されていた。

ウェブベースの調査では、文献上でコンセンサスが得られていないように見える領域について特に注意したうえで、最小基準に対する見解を ISOQOL メンバーに求めた。たとえば、多項目の PRO 尺度の信頼性 (再テスト信頼性、内的整合性) を評価するために、種々のアプローチの相対的重要性の順位を付けるように ISOQOL メンバーに依頼した。さらに、PRO 尺度の以下の 6 つの重要な特性について、推奨の合意を求めた：(1) 概念および測定モデル、(2) 信頼性、(3) 妥当性、(4) スコアの解釈可能性、

(5) 翻訳、(6) 患者と研究者の負担。

調査では、回答者が最小基準についての明確な定義をもつことが重要であると考えられた。調査における 2 枚目の画面に以下のガイダンスを表示した：「あなたは、患者中心アウトカム研究 (PCOR) で用いる PRO 尺度の選択と設計のための最小基準を確立する調査の質問に回答していることにご留意ください。すなわち、最小基準に合致しない PRO 尺度は調査研究に適切とみなすべきではないと考えています。」この記載は、成熟モデルの開発の一環として PRO 尺度の検証および強化を継続しない可能性を示すためのものではなかった。調査では PCOR に直接言及したが、これらの推奨は CER にも適合すると SATF は考えている。簡略化するため、結果の記述では単に「PCOR」を用いる。

SATF が文献レビュー結果を統合することにより作成した各推奨について、参加者は以下の回答選択肢のいずれか 1 つを選択することができた：最小基準として必須、望ましいが最小基準としては必須ではない、まったく必須ではない (PRO 尺度には必要ない)、よくわからない、または意見なし。結果の分析では、一般規則として、回答者の 50% 以上が最小基準として必須であることに合意した推奨を採択した。回答者の 50% 未満が合意した場合は、ISOQOL SATF がその推奨をレビューし、推奨が不明であった可能性、あるいは PRO 尺度の「最小基準」ではなく、「ベストプラクティス」(または「理想的な基準」) であるとみなした方がよいかどうかを判断した。また、回答者には、各推奨の後に設けたフリーテキスト欄を用いてコメントすることを奨励した。調査結果からこのテキストを抽出し、ISOQOL SATF の決定と最終的な推奨に役立てた。

調査内容と調査方法の説明をノースカロライナ大学チャペルヒル校 (UNC) の倫理審査委員会 (IRB) による審査のために提出し、UNC の Office of Human Research and Ethics による IRB 審査は免除されることが決定した。オンライン調査は UNC のサイトライセンスの下で Qualtrics Software System を用いて設計し実施した¹⁷⁾。

2012 年 2 月 20 日、ISOQOL メンバーのメーリングリスト (n = 506) 宛に調査のリンクを送付した。調査に関する指示では、PCORI との契約の締切り

に間に合うように、9日以内に調査に回答するようにメンバーに依頼した。しかし、より多くのISOQOLメンバーの回答を募るために、2012年3月20日まで回答期間を延長した(29日間)。調査の目的、プロジェクトの目標、資金源に関する情報を含めた。回答はすべて匿名で行われ、個人を特定できる情報は収集しなかった。回答期間中に2回のリマインダーを送信した。

以下の理由により、ISOQOLメンバー全員からの回答は期待しなかった：(1)調査では、自身がPRO測定分野で必須の専門性を有すると考えるISOQOLメンバーを特に対象とした、(2)専門家の意見を求めるには短期間であった。そのような専門性を有するメンバーに適格性を制限しなかったが、調査の一部として、回答者には専門性のレベルを自己報告するように求めた。

結 果

文献レビューにより特定されたガイダンス

よく知られているガイダンス文書が多数特定された。これには以下のガイダンスが含まれた：FDAガイダンス^{1, 18-20}、Medical Outcomes Trustによる優れたHRQOL尺度の特性に関する2002年のガイドライン²、COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments)からの広範な国際的な専門家による推奨^{3, 4, 21-25}、欧州がん研究・治療機構(European Organization for Research and Treatment of Cancer: EORTC)による質問票開発のためのガイドライン²⁶、慢性疾患治療の機能評価(Functional Assessment of Chronic Illness Therapy: FACIT)アプローチ²⁷、国際医薬経済・アウトカム研究学会(International Society for Pharmacoeconomics and Outcomes Research: ISPOR)タスクフォースの推奨文書²⁸⁻³¹、米国心理学会(American Psychological Association: APA)による教育・心理テスト基準³²、その他³³⁻³⁸。また、われわれは国立衛生研究所(National Institutes of Health)の患者報告アウトカム測定情報システム®(Patient-Reported Outcomes Measurement Information System®: PROMIS®)ネットワークが最近作成した基準も入手した。この文書からPRO尺度のための

最小基準に関する有用な情報が得られると考えられた。さらに、ISOQOLは最近、この全研究領域の網羅的な論文レビューと関連のある2つのガイダンス文書(効果比較研究におけるPRO尺度の使用に関する文書、医療提供の現場でのPRO尺度の統合に関する文書)を作成した^{5, 39}。

ISOQOLメンバーは、本タスクと関連のある合計301の参考文献をさらに特定した。MEDLINEデータベースの正式な検索により821の参考文献が特定され、個々にレビューを行った結果、最終的に60の関連論文が追加された。PsycINFO検索結果から関連の可能性のある172の文献が特定され、レビューにより、最終的に22の関連論文が追加された。CINAHLにより特定された126の抄録のレビュー後に、独自の参考文献がさらに4件特定された。

文献レビューから特定された28の主要なガイダンス文書を表1に示す。これらの文書は、ISOQOLメンバーへの調査で評価するISOQOL SATFの最小限のガイドライン案の情報源として役立つ。ISOQOL SATFによるさらなるレビューおよび議論のために選択された文書はPCORにおけるPRO尺度の選択のためのガイドラインと基準の模範的な記載となるものであった。文献レビューの一環として、さらに多くの関連参考文献を特定したが、われわれの焦点は多くの部分が関連している既存のガイダンス文書にあった。同一のガイドラインについて述べた複数の公表論文は別文書として引用しなかった。

ISOQOL調査に回答した参加者の特性

本調査に回答したISOQOLメンバー120名(23.7%)の特性の要約を表2に示す。約64%がPhD(または同様の博士号)、18%がMDであった。68%が学術研究者、21%が臨床医、8%が企業代表者、23%が企業コンサルタント、6%が連邦政府職員であった。地理的分布はさまざまであり、回答者の48%が北米(そのうち86%が米国)、33%が欧州であった。

参加者は定性的および定量的方法のスキルを有し、PRO測定基準の推奨に関するガイダンスを十分提供できると報告した。回答者の約81%が定量的方法について中程度からかなりの訓練を受けてお

表1 特定された患者報告アウトカム尺度に関するガイドライン

著者, 年	ガイドライン	研究デザイン	説明
Acquadro et al. 48)	国際共同臨床試験で用いる健康関連QOL質問票の翻訳方法の文献レビュー	正式な文献レビュー	翻訳方法論に関するより実証的な研究の提唱, いくつかの既存のガイドラインのレビュー, 翻訳のための多段階プロセスの推奨
Cella 27)	慢性疾患治療の機能評価(Functional Assessment of Chronic Illness Therapy: FACIT)マニュアル	方法に関する説明	FACIT尺度の開発と翻訳方法論を要約し, 既存の尺度に対する基礎的な計量心理学的な方法の情報を提示
Coons et al. 28)	電子版および紙版の患者報告アウトカム尺度の測定同等性を裏付けるために必要なエビデンスに関する推奨	専門家の見解, 文献レビュー	紙版から電子版にPRO尺度の提供方法を移行するためのサポートに必要なエビデンスに関する決定のための全般的フレームワークの提供
COSMIN group, 2010 24)	COSMIN研究: CONsensus-based Standards for the selection of health Measurement INstruments (健康測定評価尺度の選択のためのコンセンサスペースの基準)	系統的文献レビューおよびデルファイ法を用いた繰返しプロセスにより確立されたガイドライン	内的整合性, 信頼性, 測定誤差, 内容的妥当性, 構成概念妥当性, 基準関連妥当性, 反応性, 解釈可能性の評価のためのデザイン要件および望ましい統計手法についてコンセンサス
Crosby et al. 49)	健康関連QOLにおける臨床的に意味のある変化の定義	文献レビュー	健康関連QOLにおける臨床的に意味のある変化を定義するための最新アプローチのレビュー, およびそれらの使用に関するガイドラインの提供
Dewolf et al. 36)	翻訳の手順	専門家の見解	EORTC QoL質問票(QLQ)の翻訳のための方法論に関するガイダンスの提供
Erickson et al. 19)	概念の分類体系と評価尺度の階層構造: 医薬品の製品表示に適用する患者報告アウトカム(PRO)評価尺度の概念的フレームワークを確立・評価するためのツール	専門家の見解	医薬品開発におけるPRO尺度の効能を示すうえで(そのような目的でPRO尺度の試験が行われていない場合であっても)有用となる可能性のあるPRO概念の分類体系と評価尺度の階層構造の提案
Frost et al. 50)	患者報告アウトカム尺度の信頼性と妥当性のための十分なエビデンスは何か?	文献レビュー	FDAガイダンスを特に参照しつつ, 具体的な統計学的閾値の指針として文献を用いて, PRO尺度に必要な計量心理学的特性に関する具体的なガイダンスを提供する論文
Hays et al. 51)	健康関連QOL研究における臨床的に意味のある変化の概念: それにはどのような意味があるのか?	専門家の見解	臨床的に意味のある最小の差を定義するための単一の閾値に対する議論
Johnson et al. 26)	質問票モジュール開発のためのガイドライン	専門家の見解	問題の作成, 項目リストの構成, プレテストやフィールドテストに関連するEORTCの方法論に従ったPRO尺度モジュール開発の詳細な説明の提供
Kemmler et al. 52)	個々の患者におけるQOLの変化の評価のために臨床的意義と統計学的有意性を統合するための新しいアプローチ	化学療法試験からの縦断的データ	この試験のデータを用いて, 個々の参加者における変化の評価(vs. 群間比較), 統計学的有意性と臨床的意義に基づく評価の重要性の強調
Kottner et al. 53)	信頼性と合意に関する研究報告のためのガイドライン(Guidelines for reporting reliability and agreement studies: GRRAS)の提案	文献レビュー, 専門家コンセンサス	医薬品分野での評価者間一貫性および評価者間信頼性の報告のための一連のガイドラインの提案
Magasi et al. 33)	患者報告アウトカム尺度の内容的妥当性: PROMISミーティングからの観点	専門家の発表と議論	内容的妥当性に焦点を置いて, PROMISミーティングでの所見を記載した論文。その結果として, コンセンサスに基づくガイドラインの必要性等のいくつかの推奨の概略の説明(具体的な提案はなし)
Norquist et al. 42)	患者報告アウトカム尺度のための想起期間の選択: 考慮点の基準	文献レビュー	PRO尺度のための想起期間の選択は疾患の性質, 症状の安定性, 症状の経時的推移に依存する。
Revicki et al. 12)	米国での製品表示および公告宣伝表示をサポートする健康関連QOL研究に関する推奨	レビュー	医薬品表示および公告宣伝表示に関する効能・効果を主張するためのエビデンス基盤の重要性の概略説明
Revicki et al. 7)	製品表示と公告宣伝表示のための患者報告アウトカムの理論および計量心理学的特性の記載: PROエビデンス提出資料	報告	PRO尺度のエビデンス資料の目的と内容, および規制当局の審査における潜在的な役割に関する説明
Revicki et al. 34)	患者報告アウトカムにおける反応性と最小重要差を決定するために推奨される方法	文献レビュー, 専門家の見解	患者ベースおよび臨床的アンカーと, 複数のアプローチや方法を通しての収斂に基づくべき最小重要差(MID)の推定に関する具体的な推奨
Rothman et al. 30)	既存の患者報告アウトカム(PRO)評価尺度の使用とその修正	専門家の見解	既存の評価尺度の内容的妥当性の評価と文書化に関する重要な問題についての考察, 内容的妥当性に対する潜在的脅威と改善方法についての考察

表1 特定された患者報告アウトカム尺度に関するガイドライン (続き)

著者, 年	ガイドライン	研究デザイン	説明
Schmidt et al. ⁵⁴⁾	異文化間QOL評価尺度の開発における現在の問題点	文献レビュー	PRO尺度の異文化間適応の概要の提供, 広範な開発ガイドラインの提示, 国際研究へのさらなる推進の呼びかけ
Schunemann et al. ⁸⁾	臨床試験における患者報告アウトカム尺度の結果の解釈: 臨床医の観点	事例に基づく報告	PROMスコアの閾値やスコアの差に意味付けを行う方法を説明するためのいくつかの事例の提供
Scientific Advisory Committee of Medical Outcomes Trust ²⁾	健康状態とQOL評価尺度の評価: 特性と評価レビュー基準	専門家の見解	PRO尺度の8つの重要な特性 (概念および測定モデル, 信頼性, 妥当性, 反応性, 解釈可能性, 回答者と測定者の負担, 代替形式, 文化的および言語適応)の説明
Sprangers et al. ⁵⁵⁾	QOLの意味のある経時的变化の評価: 臨床医のためのユーザーガイド	文献レビュー, 専門家の見解	治療上の意思決定プロセスにおけるPROデータの使用方法に関して臨床医の指針となる一連のガイドライン/疑問の提案
Snyder et al. ⁵⁾	臨床実践における患者報告アウトカム評価の実施: 選択肢と考慮点のレビュー	文献レビュー	ISOQOLグループによる, 臨床実践におけるPRO使用の指針となる一連の選択肢と考慮点の作成, 代替アプローチの長所と短所
Turner et al. ⁵⁶⁾	患者報告アウトカム: 評価尺度の開発および選択に関する問題点	文献レビュー	PRO尺度の開発と選択に際して考慮すべき概念や論点の広範な要約
米国食品医薬品局 ¹⁾	業界向け指針, 患者報告アウトカムの測定法: 医薬品/医療機器における適応申請のための方法	専門家の見解	「本指針は, 医薬品/医療機器の適応の認可申請に使われる, 既存の, 修正した, または新規に作られた患者報告アウトカム尺度を, 食品医薬品局 (FDA) がどのように審査・評価するかが記載されている」。この指針では, 概念的枠組, 内容妥当性, 信頼性, 妥当性, 変化検出能力, PRO尺度の修正, 特定集団を対象としたPRO尺度の使用について扱っている。
Wild et al. ²⁹⁾	患者報告アウトカム尺度の翻訳と文化的適応プロセスの優れた実践のための原則	文献レビュー, 専門家の見解/コンセンサス	ISPORタスクフォースによる, PROMSの翻訳と文化的適応のための種々の方法の長所と短所の批評
Wild et al. ³¹⁾	国際共同試験-必要な翻訳, 異なる国々において同一言語を用いる方法, データ統合をサポートするための方法に関する推奨	専門家の見解, 文献レビュー	PRO尺度に必要な翻訳に関する意思決定ツールの提供, 複数の国で同一言語が話される場合に用いるアプローチ, 異なる言語バージョンでのデータ統合をサポートするためのエビデンス収集方法
Wyrwich et al. ³⁸⁾	患者報告アウトカム尺度の経時的变化の解釈方法	文献レビュー	意味のあるPROスコアの変化の閾値の記述, 伝達に用いられる方法, 用語の進化をレビューした論文

表2 参加者の特性 (参加者の報告による)

サンプルの特性	% (n = 120)	サンプルの特性	% (n = 120)
学位 ^a		PRO尺度のデザインと評価における定量的トレーニング	
MD	18%	膨大なトレーニング	37%
PhD/その他の博士号 (例: ScD)	64%	中程度のトレーニング	44%
RN/NP	5%	わずかなトレーニング	16%
理学療法士/作業療法士	7%	トレーニングなし	3%
MA, MSc, MPHまたはその他の修士号	43%	PRO尺度のデザインと評価における定性的トレーニング	
役割 ^a		膨大なトレーニング	18%
学術研究者	68%	中程度のトレーニング	35%
臨床家	21%	わずかなトレーニング	40%
企業代表者	8%	トレーニングなし	7%
企業コンサルタント/CRO従業員	23%	能力	
連邦政府職員	6%	極めて能力がある	50%
患者支援者	2%	能力がある	39%
その他	8%	やや能力がある	8%
地理的居住地		あまり能力がない	3%
北米	48%	健康関連QOL (HRQOL) または患者報告アウトカム (PRO) 分野での平均年数	
米国	(86%)	HRQOLまたはPRO分野での平均年数	15年間 (範囲: 1-40年間)
欧州	33%		
南米	5%		
アジア	10%		
アフリカ	1%		
オーストラリア	3%		

^aこの特性では複数回答を可とした。

表3 PRO尺度の特性の定義

概念および測定モデル	概念モデルはPRO尺度に含めるターゲットとなる構成概念の説明とフレームワークを提供する。測定モデルはPRO尺度の個々の項目を構成概念にマッピングする。
信頼性	PRO尺度が測定誤差を含まない程度 ^{2, 4, 40, 41)} ・内的整合性信頼性 多項目のPRO尺度における項目間の相互関連性の程度 ^{2, 4)} ・再テスト信頼性 評価尺度の再現性の評価基準。すなわち、安定した集団において経時的に一貫したスコアを与える能力 ²⁾
妥当性	PRO評価尺度が測定を意図するPRO概念を測定する程度 ^{2, 4, 41)} ・内容的妥当性 特定の測定実施の状況において、PRO尺度がもっとも関連性が高く、重要な概念の側面を含む程度 ⁵⁰⁾ ・構成概念妥当性 PRO尺度のスコアが、現在測定している概念に関して理論的に導出した先行仮説と一貫して、他の尺度(例:患者報告または臨床的尺度)と関連する程度 ⁴⁰⁾ ・基準関連妥当性 PRO尺度のスコアが「ゴールドスタンダード」の適切な反映である程度 ⁴⁾ ・反応性 PRO尺度が経時的に測定される構成概念の変化を検出できる程度 ^{2, 37)}
スコアの解釈可能性	PRO尺度のスコアに対して容易に理解できる意味を持っている程度 ^{2, 4)}
最小重要差(MID)	説明を受けた患者または代理人が重要(有益または有害)であると認識し、患者または臨床家が管理方法の変更を検討する可能性のある、関心のあるアウトカムにおけるスコアの最小の差 ^{44, 57, 58)}
負担	評価される者(回答者の負担)または評価尺度を実施する者(研究者または実施者の負担)に課せられる時間、労力およびその他の負担 ²⁾

り、53%が定性的方法について中程度からかなりの訓練を受けていると報告した。全体で、89%がガイダンスを提供する「能力がある」または「極めて能力がある」と報告した。感度分析として、ガイダンスを提供する能力が「やや能力がある」または「あまり能力がない」と報告した11%の回答者を除外した場合の推奨についても検証したが、最終的な推奨に変更はなかった。この分野でのPRO測定と研究の経験の平均期間は15年であった。

PCORで使用するPRO尺度の選択のための最小基準

PRO尺度の特性の定義を表3、最小基準の推奨案に関するISOQOLの調査結果の概要を表4に示す。これらの結果とISOQOLメンバーからのフィードバックに基づく最終的な推奨を表5に示す。文献レビューと調査の結果を以下に述べる。

概念および測定モデル

ISOQOLメンバーは表4(#1)に記載した最小基準を全面的に支持し、回答者の90%が、PRO尺度はPROの構成概念を定義し、意図する集団における当該尺度の意図する適用方法について記文書を有すべきであるとの言明を支持した。また、回答者の61%が、その文書では、測定する概念を測定モデルにおいて操作可能にする方法について述べるべきであることに同意した。

PRO尺度の信頼性

ISOQOL回答者の大部分が、最小基準として、多項目のPRO尺度の内的整合性の信頼性、単一項目のPRO尺度の再テスト信頼性を評価すべきであることに同意した(表4, #2参照)。しかし、彼らは、最小基準として、多項目のPRO尺度は再テスト信頼性のエビデンスを必要とすることには同意しなかった。彼らは、主にPCORで検討される一部の集団は安定しておらず、そのHRQOLは変動し得ることから、再テスト信頼性に関する臨床的な懸念を示した。この現象は再テスト信頼性の推定値を減少させ、経時変化を正確に検出する際のPRO尺度の信頼性を低下させる可能性がある。さらに、2回の調査時点が互いに近接している場合、記憶効果(memory effect)は再テスト信頼性に対して必ず影響を及ぼす。

回答者はグループレベルの比較の最小の信頼性レベルとして0.70を支持した。これは、この分野で広く受け入れられている値である^{2, 40, 41)}。この信頼性レベルでの測定の標準誤差は約0.55の標準偏差である。しかし、絶対的なカットオフ値を設けること(例:信頼性係数の推定値が0.69であるPRO尺度は信頼性が低いとみなす)は規範的すぎるとの懸念もあった。一部の回答者(36%)は「信頼性の最小レベルについて言及すべきではないが、使用予定のPRO測定を適用する状況において信頼性を適切に正当化すべきである」との言明を支持した。

表 4 推奨案についての ISOQOL 調査の結果

最小基準のための推奨案	調査結果 (n = 120)
1 概念および測定モデル	
PRO尺度はそれに含まれる概念と使用を意図する集団を定義し記述した文書を有するべきである。	最小基準として必須—90% 望ましいが、最小基準としては必須ではない—9% 必須ではない—0% 不明—1% 意見なし—0%
それに加え、測定モデルにおける概念の構成方法(尺度の次元性に関するエビデンス、項目が測定される各概念とどのように関連するか、PRO尺度に含まれる概念間の関係、など)に関する文書を有するべきである。	最小基準として必須—61% 望ましいが、必須ではない—35% 必須ではない—3% 不明—1% 意見なし—0%
2 信頼性	
群間比較のためのPRO尺度の信頼性は理想的には0.70以上であるべきである。	はい、0.70以上であるべきである—54% いいえ、_(数値を記入)_以上であるべきである—8%(回答の範囲:0.50-0.80) 提案された適用の文脈において信頼性の最低基準は適切に正当化すべきではない—36% 意見なし—2%
多項目の一次元尺度の信頼性は内的整合性の評価を含むべきである。	最小基準として必須—79% 望ましいが、必須ではない—14% 必須ではない—2% 不明—3% 意見なし—2%
多項目の一次元尺度の信頼性は再テスト信頼性の評価を含むべきである。	最小基準として必須—43% 望ましいが、必須ではない—51% 必須ではない—3% 不明—3% 意見なし—0%
単一項目の尺度の信頼性は再テスト信頼性により評価すべきである。	最小基準として必須—60% 望ましいが、必須ではない—34% 必須ではない—2% 不明—3% 意見なし—1%
3 妥当性	
3a 内容的妥当性	
PRO尺度はその内容的妥当性を裏付けるエビデンス(患者や専門家が、PRO尺度の内容が概念、集団、測定の適用目的に関連があり、包括的であると判断するエビデンス等)を有するべきである。	最小基準として必須—78% 望ましいが、必須ではない—19% 必須ではない—2% 不明—0% 意見なし—1%
測定適用に関連するPROの特性(項目により測定される概念)を求め、確認するために用いる定性的・定量的方法に関する文書	最小基準として必須—53% 望ましいが、必須ではない—44% 必須ではない—2% 不明—1% 意見なし—0%
評価に含まれる参加者の特性に関する文書(例:人種/民族、文化、年齢、社会経済的状态、識字能力)	最小基準として必須—52% 望ましいが、必須ではない—47% 必須ではない—0% 不明—0% 意見なし—1%
PRO尺度の開発プロセスにおける項目の派生、修正、優先順位付けの情報源に関する文書	最小基準として必須—46% 望ましいが、必須ではない—46% 必須ではない—7% 不明—0% 意見なし—1%
測定適用のための想起期間の根拠	最小基準として必須—41% 望ましいが、必須ではない—52% 必須ではない—5% 不明—1% 意見なし—1%

表4 推奨案についての ISOQOL 調査の結果 (続き)

最小基準のための推奨案	調査結果 (n = 120)
3b 構成概念妥当性	
PRO尺度はその構成概念妥当性を裏付けるエビデンス(測定されたPROと類似または非類似の尺度間で予想される関連に関する事前に設定した仮説を裏付ける実証的所見の文書等)を有するべきである。	最小基準として必須—55% 望ましいが、必須ではない—44% 必須ではない—1% 不明—0% 意見なし—0%
PRO尺度はその構成概念妥当性を裏付けるエビデンス(「既知」グループ間で予想されるスコアの差に関する事前に設定した仮説を裏付ける実証的所見の文書等)を有するべきである。	最小基準として必須—41% 望ましいが、必須ではない—57% 必須ではない—2% 不明—0% 意見なし—0%
3c 反応性	
縦断的調査研究で用いるPRO尺度は反応性のエビデンス(研究対象集団におけるスコアの変化に関して事前に設定した仮説と一致するスコアの変化の実証的エビデンス等)を有するべきである。	最小基準として必須—57% 望ましいが、必須ではない—42% 必須ではない—1% 不明—0% 意見なし—0%
PRO尺度が信頼性(内的整合性)、内容的妥当性、構成概念妥当性に関して十分なエビデンスを提供する横断的データを有するが、経時的な反応性(経時的に測定された構成概念の変化をPRO尺度が検出する能力)に関するデータはまだない場合、その他のPRO尺度が利用できないとしたら、縦断的研究において妥当な経時的データ提供するためにPRO尺度の使用は受け入れるか?	はい—65% いいえ、それを受け入れる前に反応性のエビデンスが必要—32% 意見なし—0% コメント(ブランクの回答欄に記入)—22%
4 スコアの解釈可能性	
PRO尺度は、測定された概念に対してスコアの低値や高値が表すもの等のスコアの解釈を裏付ける文書を有するべきである。	最小基準として必須—64% 望ましいが、必須ではない—35% 必須ではない—1% 不明—0% 意見なし—0%
PRO尺度は、参照集団における代表的な平均値と標準偏差等のスコアの解釈を裏付ける文書を有するべきである。	最小限基準として必須—39% 望ましいが、必須ではない—57% 必須ではない—4% 不明—0% 意見なし—0%
PRO尺度は、患者や臨床家の観点から意味があると考えられる群間や経時的なスコアの最小重要差に関するガイダンス等のスコアの解釈を裏付ける文書を有するべきである。	最小基準として必須—23% 望ましいが、必須ではない—72% 必須ではない—5% 不明—0% 意見なし—0%
5 PRO尺度の翻訳	
1つ以上の言語に翻訳されるPRO尺度は、異なる言語における比較またはデータの併合が可能となるように、翻訳されたバージョンの測定特性が同等であることのエビデンスを有するべきである。	最小基準として必須—47% 望ましいが、必須ではない—49% 必須ではない—4% 不明—0% 意見なし—0%
翻訳に関わる個人の背景や経験に関する文書	最小基準として必須—43% 望ましいが、必須ではない—49% 必須ではない—8% 不明—0% 意見なし—0%
各言語でのPRO尺度の翻訳・評価に用いる方法に関する文書	最小限の標準として必須—81% 望ましいが、必須ではない—16% 必須ではない—3% 不明—0% 意見なし—0%
異なる言語バージョン間での一致の程度に関する文書	最小基準として必須—38% 望ましいが、必須ではない—53% 必須ではない—7% 不明—2% 意見なし—0%

表4 推奨案についての ISOQOL 調査の結果 (続き)

最小基準のための推奨案	調査結果 (n = 120)
6 患者と研究者の負担	
一般集団からの成人の回答者が参加する研究のためのPRO尺度の読解力レベルは少なくとも...であるべきである。	4年生程度の教育レベル—7% 6年生程度の教育レベル—23% 8年生程度の教育レベル—6% その他の教育レベル—8% PRO尺度の識字能力についての最小限の要件は設けるべきではないが、その提案する適用場面の状況において適切に正当化すべきである—43% 不明—9% 意見なし—4%

PRO 尺度の妥当性

最小基準と考えられる妥当性のうちもっとも多く用いられる種類のものは、内容的妥当性、構成概念妥当性、反応性であった。反応性はしばしば妥当性の一側面とみなされる^{4,37)}。しかし、縦断的研究での PRO 測定における反応性の重要性から、反応性は分けて議論されることも多い⁴⁾。基準関連妥当性については、それに対して PRO 尺度を比較するための「ゴールドスタンダード」が通常ないことから、考慮されなかった。ISOQOL メンバーの調査では、横断的または縦断的研究での PRO 尺度における基準関連妥当性が重要であると回答した者はそれぞれ 7%と 10%のみであった。APA 標準マニュアル³²⁾では、妥当性は妥当性のすべての側面を含む単一の概念であることを提案していることに注意はしたほうがいい。しかし、アウトカム研究の分野では、おそらく異なる種類の妥当性に言及するには異なる方法論が必要であることから、今なお上記の用語が区別して用いられている。

内容的妥当性は PRO 尺度のために評価すべきもっとも重要な種類の妥当性であると判定され、ISOQOL メンバーは、PRO 尺度は横断的または縦断的研究で使用する前に内容的妥当性のエビデンスを有するべきであると、それぞれ 58%と 61%が回答した (表 4 ではデータ未記載)¹⁾。ISOQOL メンバーにより内容的妥当性に対する最小基準の推奨は支持されたが (表 4, #3a 参照), 想起期間 (recall period) に関しては不一致がみられた。想起期間とは、患者が測定された PRO を用いて自身の経験を記述する際の参照期間、(例: 現在, 過去 24 時間, 過去 7 日間, 過去 4 週間) である。回答者の多く (52%) が、想起期間の根拠を示すことが望ましいが、PRO 尺

度のための最小基準としては必須ではないと考えた。最終的な推奨では、研究参加者が妥当な回答を行えるように、参照期間は慎重に検討することを推奨した。ただし、想起期間は測定する PRO の領域、研究の背景、検討する集団に応じて変わることから、単一の想起期間は推奨しなかった⁴²⁾。

内容的妥当性のもう 1 つの側面は項目の起源に関係する。最小基準とみなされたが ISOQOL メンバーにより支持されなかった記述の 1 つが、「PRO 尺度の開発プロセスにおける項目の派生、修正、優先順位付けの情報源に関する文書」であった。回答者の大部分 (46%が「最小基準として必須」、46%が「望ましいが、必須ではない」と回答) がこの基準は重要であると回答したことから、われわれは、この文書化は PRO 尺度のための「ベストプラクティス」とみなすが、最小基準とはしないことを推奨した。

構成概念妥当性も妥当性の重要な構成要素であると判断された。回答者の大部分 (55%) が、類似および非類似の尺度間での予想される関連性に関する先行仮説を裏付ける実証的な所見に関する文書が PRO 尺度の最小基準であると判断した (表 4, #3b 参照)。われわれの当初の推奨のもう 1 つの部分では、「既知」グループ間でのスコアの予想される差についての事前に設定した仮説を裏付ける実証的所見を記載する、「既知グループ」妥当性について文書化されたエビデンスを考慮した。われわれは、既知グループ妥当性は、グループ間の差について過去の実証的エビデンスがある場合に、PRO 尺度があるグループを他のグループと弁別する能力を示すことから、構成概念妥当性評価の重要な部分であると判断した。しかし、ISOQOL メンバーの大部分 (57%) は、これを「望ましいが、必須ではない標準」

と判定した。したがって、われわれは、これを最小基準ではなく「ベストプラクティス」のための基準とみなした。

反応性（縦断的妥当性とも呼ばれる）は構成概念妥当性の一側面である^{23,37,43}。ISOQOL 回答者の大部分が、最小基準として、縦断的研究において PRO 尺度を用いる前に、事前に設定した仮説と一致するスコアの変化についての実証的エビデンスを得ることを支持した（表 4, #3c 参照）。しかし、回答者の 65%が、当該評価尺度の反応性を裏付ける先行研究がなくても、PRO 尺度の信頼性、内容的妥当性および構成概念妥当性についての横断的研究での科学的証拠があれば、縦断的研究においても PRO 尺度を用いるであろうと報告した。

スコアの解釈可能性

PCOR での使用において PRO 尺度が広く受け入れられるためには、患者、臨床家、研究者、政策決定者等のさまざまなステークホルダーが容易に解釈できるスコアを提供する必要がある³⁸。文献レビューの結果、標準的とみなされるスコアの解釈可能性を高めるいくつかの方法が明らかになった。エンドユーザーはスコアの高値または低値が何を表すかを知ることができなくてはならない。さらに、あるグループから他のグループ（またはある時点から他の時点）へのスコアの意味のある差または変化が何から構成されるかを知ることが、測定されるアウトカムの理解を促すであろう。PRO 尺度のスコアの解釈可能性を高めるもう 1 つの方法は、ある研究でのスコアを母集団（例：米国の一般集団、または特定の疾病を有する集団）の既知スコアと比較することである。そのようなベンチマークが利用できると、いくつかの参照または基準となるグループと比べることで、研究対象グループのスコアについての理解を高める。

回答者の大部分が、最小基準として、PRO 尺度はスコアの解釈を裏付ける文書（スコアの低値または高値が何を表すかの説明、等）を有するべきであることを支持した（表 4, #4 参照）。しかし、一般基準や参照となるスコア、または最小重要差（MID）の推定値のようなより有用なメトリクスは必須とは考えられなかったが、極めて望ましいと考えられ

た^{34,44,45}。

PRO 尺度の翻訳

PCOR と CER は、多国間または多文化の状況でしばしば実施され、異なる言語への PRO 尺度の翻訳を要する。これらの集団間での HRQOL の結果を比較または統合するためには、測定する HRQOL の概念や測定に用いる質問票の文言が翻訳により同じように解釈されることが極めて重要である^{29,46}。

本調査でレビューを行った最初の推奨案（表 4, #5 参照）のうち、ISOQOL メンバーは、最小基準として、「PRO 尺度を各言語に翻訳・評価するための方法に関する文書」の記載を支持した。フォローアップとしての質問への回答（表 4 には要約せず）では、翻訳された PRO 尺度を用いる前に翻訳の品質をレビューするために定性的方法（例：認知インタビュー）を用いる点について、回答者の 41%が必要と回答したが、40%が期待されるが、必須ではないと回答した。PRO 尺度を使用する前に翻訳の品質をレビューするための定量的方法（例：特異項目機能テスト）が必須であると回答したのは 24%のみであり、回答者の 42%が、翻訳された PRO 尺度を用いる前に定量的評価を含めることは望ましい（ただし、必須ではない）と回答した。これらの所見に基づき、ISOQOL SATF は、翻訳された PRO 尺度のための最小限の標準として定性的エビデンスを含めることを推奨した（表 5）。

患者と研究者の負担

委員会は、PCOR 研究のための PRO 尺度を選択する際には患者と研究者の負担を考慮すべきであることに同意した。PRO 尺度は、患者は健康状態が悪いことが多いため、過度の負担になってはならず、生活を損なうような長すぎる質問票や頻回のデータ収集が行われるべきではない。調査回答者の 92%が「回答者の負担」は PCOR における PRO 尺度を選択する際の「重要」または「極めて重要」な考慮点であることを支持した。

同様に、回答者の 90%が、PCOR における PRO 尺度を選択する際には識字能力が「重要」または「極めて重要」な考慮点であることを支持した。PRO 尺度から収集したデータは、被験者が質問内容を理

表 5 患者中心アウトカム研究または効果比較研究で用いられる患者報告アウトカム (PRO) 尺度のための最小基準についての最終的な推奨

1	概念および測定モデル PRO尺度はそれに含まれる概念と使用を意図する集団を定義・記述した文書を有するべきである。それに加え、測定モデルにおける概念の構成方法(尺度の次元性に関するエビデンス、項目が測定される各概念とどのように関連するか、PRO尺度に含まれる概念間の関係、など)に関する文書を有するべきである。
2	信頼性 群間比較のためのPRO尺度の信頼性は0.70以上であることが望ましいが、適切な根拠がある場合は、それを下回ってもよい。信頼性は内的整合性の信頼性、再テスト信頼性、または項目反応理論等の種々の方法により推定することができる。各方法の根拠を示すこと。
3	妥当性
3a	内容的妥当性 PRO尺度はその内容的妥当性を裏付けるエビデンス(患者や専門家が、PRO尺度の内容が概念、集団、測定の適用目的に関連があり、包括的であると判断するエビデンス等)を有するべきである。これは以下に関する文書を含むべきである:(1)測定適用に関連するPROの特性(項目により測定される概念)を求め、確認するために用いる定性的・定量的方法、(2)目標集団に関する類似性または差異に特に力を置いた、評価に含まれる参加者の特性(例:人種/民族、文化、年齢、社会経済的状态、識字能力)、(3)測定適用のための想起期間の根拠
3b	構成概念妥当性 PRO尺度はその構成概念妥当性を裏付けるエビデンス(測定されたPROと類似または非類似の尺度間で予想される関連に関する事前に設定した仮説を裏付ける実証的所見の文書等)を有するべきである。
3c	反応性 縦断的調査研究で用いるPRO尺度は反応性のエビデンス(研究対象集団におけるスコアの変化に関して事前に設定した仮説と一致するスコアの変化の実証的エビデンス等)を有するべきである。
4	スコアの解釈可能性 PRO尺度は、測定された概念に対してスコアの低値や高値が表すもの等のスコアの解釈を裏付ける文書を有するべきである。
5	PRO尺度の翻訳 1つ以上の言語に翻訳されるPRO尺度は各言語でのPRO尺度の翻訳・評価方法に関する文書を有するべきである。研究は少なくとも翻訳を評価するため定性的方法(例:認知的テスト)からのエビデンスを含むべきである。
6	患者と研究者の負担 PRO尺度は患者と研究者の過度の負担になってはいけない。PRO尺度の長さは、評価に含まれる他のPRO尺度、PROデータ収集の頻度、試験対象集団の特性の文脈で考慮すべきである。PRO尺度に含まれる項目に必要な識字能力は通常6年生(年齢12歳)以下の教育レベルとすべきであるが、提案する適用の状況において適切に正当化すべきである。

解し、自身の経験や観点を正確に反映した回答を行うことができる場合にのみ妥当である。PRO尺度の開発者が質問と回答選択肢を明確かつ容易に理解できるようにすることは極めて重要である。PRO尺度の定性的テスト(例:認知インタビュー)には、質問を評価するために、識字能力の低い個人を含むべきである⁴⁷⁾。回答者の23%がPRO尺度は6年生程度の教育レベル(年齢11~12歳)で書かれるべきであると回答し、43%が特定の研究用途に応じて識字能力のレベルを適切に正当化すべきであると回答した。

考 察

既存のガイドラインの文献レビューとPRO測定・研究の専門家の調査に基づき、われわれはISOQOLを代表して、患者中心アウトカム研究や効果比較研究で用いるPRO尺度のための最小基準を提示した。これらの推奨事項は、概念および測定モデルの特性の文書化、スコアの信頼性、妥当性、解釈可能性のエビデンス、質の高い翻訳、実施可能な患者と研究者の負担であった(表5に要約)。PRO尺度が本報告で述べた基準をどの程度遵守するかはPRO測定

の品質を反映する。

PRO尺度がこれらの測定特性を満たす、または上回ることのエビデンスを適切に文書化することにより、PCORやCERで用いるPRO尺度がより広く受け入れられるようになる。この文書化には、PRO尺度の測定特性についての焦点を絞った方法論的に厳密な研究、またはPCORやCER研究においてPRO尺度から収集したHRQOLデータの解析等が含まれる。このような文書は査読文献や公的にアクセス可能なウェブサイトで見ることができるようにすべきである。研究対象集団と類似の集団からエビデンスが得られた範囲で、研究者は、患者の経験や観点を収集するためのPRO尺度により大きな信頼を置くことができる。

PCORとCERにこれらの最小基準を適用する際には、多くの点を考慮する必要がある。PCORやCERに参加する集団は、第II相または第III相臨床試験に通常組み入れられる集団よりも不均一である可能性が高い。この集団の不均一性は、PRO尺度の測定特性評価に含めるサンプルに反映されるべきである。たとえば、定性的・定量的研究はいずれも、研究対象集団での有病率を反映した人種/民族、性別、年齢群に基づく割当抽出法(quota sampling)

を必要とすることがある。

研究者は、PRO 尺度の測定特性を裏付けるエビデンスの強さを慎重に考慮しなくてはならない。すべての集団や適用場面において妥当であるか否かを判定する評価尺度の閾値はない。さらに、単一の研究のみでは、すべての研究の文脈における測定特性をすべて確認することはできない。他のすべての科学的原則と同様に、測定科学は、異なる状況で再現される、一連のエビデンスの反復的な蓄積に依存している（成熟モデル）。したがって、PRO 尺度の適切性を評価するために用いられる情報は、エビデンスの重み（研究の数と質、結果の一貫性）である。古い PRO 尺度にはときに、新しい尺度よりも多くのエビデンスを有するという利点があるので、この点は今回の基準に反映された。

本研究の限界は、SATF の個々のメンバーのバイアスが調査内容に影響を及ぼす可能性があることである。用いたプロセスの透明性、およびメンバー内での広範囲の専門性と観点は、多くのバイアスが導入される可能性を低減した。さらに、調査回答率は低く、これもバイアスの可能性であることが示唆される。ただし、調査中に収集した人口統計学的データからは、回答者が専門家としてさまざまな観点を有する経験豊富な ISOQOL メンバーであり、大部分の回答者が自分自身を調査項目に対する評価と回答を行う能力があると回答した点をわれわれは指摘しておく。

ISOQOL は、PRO 尺度が特定の PCOR 研究に対して適切または不適切であると判断できるように、これらの最小標準を作成した。それにより PRO 尺度が受け入れ可能であると判断される最小基準を設けることが目的であった。これらの基準はより厳格な基準を有する「理想的な基準」または「ベストプラクティス」を反映するものではない^{2,3,40)}。たとえば、PRO 尺度についての確立された最小重要差はスコアの解釈可能性を高め、意思決定の際に情報を提供する。他の例としては、異なる評価手段（例：紙媒体、コンピュータ、携帯機器、電話）における PRO 測定の同等性を確立することは、より広範な患者が PCOR に参加することを促す可能性がある。ISOQOL による PCOR と CER における PRO 尺度のための「ベストプラクティス」の推奨事項は、

HRQOL 測定の技術を向上させるための ISOQOL の戦略的イニシアティブにおける次のステップである。

PCORI の方法論委員会 (Methodology Committee) は、患者中心アウトカム研究に関連する基準とガイドラインのレビューの一環として、本研究内容のレビューを行った。今回提示した ISOQOL の推奨事項は、PRO 測定特性に関して、PCORI の方法論委員会 (Methodology Committee) による基準よりも具体的な情報に焦点を当てている¹⁵⁾。

ISOQOL が現在推奨しているこれらの最小基準を満たす、または上回る PRO 尺度を特定し選択することで、PCOR と CER で生成されるエビデンスにより、健康関連アウトカムに対する患者の観点が信頼性や妥当性のある方法で示される可能性が高まる。次に、堅実な測定特性をもつ評価尺度に基づくこの PRO のエビデンスを用いて、さまざまな健康介入に伴う益と害に関する臨床および健康政策の意思決定に情報を提供し、集団の健康状態をモニタリングすることができる。

謝辞 本研究は患者中心アウトカム研究所 (PCORI-SOL-RMWG-001: PIs: Zeeshan Butt, PhD, Northwestern University; Bryce Reeve, PhD, University of North Carolina at Chapel Hill) の資金提供を受けた。本論文中で表明された見解は著者の見解であり、PCORI の見解を必ずしも反映するものではない。

参考文献

- 1) US Food and Drug Administration. (2009). Patient-reported outcome measures: Use in medical product development to support labeling claims. Guidance for industry. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071975.pdf>. Accessed November 26, 2011.
- 2) Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, 11(3), 193-205.
- 3) Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2006). Protocol of the COSMIN study: COnsensus-based standards for the selection of health measurement INstruments. *BMC Medical Research Methodology*, 6, 2. doi:10.1186/1471-2288-6-2.
- 4) Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737-745. doi:10.1016/j.jclinepi.2010.02.006.

- 5) Snyder, C. F., Aaronson, N. K., Choucair, A. K., Elliott, T. E., Greenhalgh, J., Halyard, M. Y., et al. (2011). Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations. *Quality of Life Research*, doi:10.1007/s11136-011-0054-x.
- 6) Basch, E. M., Reeve, B. B., Mitchell, S. A., Clauser, S. B., Minasian, L., Sit, L., et al. (2011). Electronic toxicity monitoring and patient-reported outcomes. *Cancer Journal*, 17(4), 231-234. doi:10.1097/PPO.0b013e31822c28b3.
- 7) Revicki, D. A., Gnanasakthy, A., & Weinfurt, K. (2007). Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: The PRO Evidence Dossier. *Quality of Life Research*, 16(4), 717-723. doi: 10.1007/s11136-006-9153-5.
- 8) Schunemann, H. J., Akl, E. A., & Guyatt, G. H. (2006). Interpreting the results of patient reported outcome measures in clinical trials: The clinician's perspective. *Health and Quality of Life Outcomes*, 4, 62. doi:10.1186/1477-7525-4-62.
- 9) Deyo, R. A., & Patrick, D. L. (1989). Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Medical Care*, 27(3 Suppl), S254-S268.
- 10) International Society for Quality of Life Research. <http://www.isoqol.org/>. Accessed July 30, 2012.
- 11) Guyatt, G., & Schunemann, H. (2007). How can quality of life researchers make their work more useful to health workers and their patients? *Quality of Life Research*, 16(7), 1097-1105. doi: 10.1007/s11136-007-9223-3.
- 12) Revicki, D. A., Osoba, D., Fairclough, D., Barofsky, I., Berzon, R., Leidy, N. K., et al. (2000). Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Quality of Life Research*, 9(8), 887-900.
- 13) Lipscomb, J., Donaldson, M. S., Arora, N. K., Brown, M. L., Clauser, S. B., Potosky, A. L., et al. (2004). Cancer outcomes research. *Journal of the National Cancer Institute Monographs* (33), 178-197. doi:10.1093/jncimonographs/lgh039.
- 14) US Patient-Centered Outcomes Research Institute <http://www.pcori.org>. Accessed 26 November, 2011.
- 15) Methodology Committee of the Patient-Centered Outcomes Research, I. (2012). Methodological standards and patient-centeredness in comparative effectiveness research: The PCORI perspective. *Journal of the American Medical Association*, 307(15), 1636-1640. doi:10.1001/jama.2012.466.
- 16) Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, 18(8), 1115-1123. doi: 10.1007/s11136-009-9528-5.
- 17) Qualtrics Labs Inc. Why choose qualtrics survey software? <https://www.qualtrics.com/why-survey-software>. Accessed November 26, 2011.
- 18) US Food and Drug Administration. (2010). Qualification process for drug development tools. Draft Guidance for Industry. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>. Accessed November 26, 2011.
- 19) Erickson, P., Willke, R., & Burke, L. (2009). A concept taxonomy and an instrument hierarchy: Tools for establishing and evaluating the conceptual framework of a patient-reported outcome (PRO) instrument as applied to product labeling claims. *Value in Health*, 12(8), 1158-1167. doi:10.1111/j.1524-4733.2009.00609.x.
- 20) Patrick, D. L., Burke, L. B., Powers, J. H., Scott, J. A., Rock, E. P., Dawisha, S., et al. (2007). Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in Health*, 10(Suppl 2), S125-S137. doi:10.1111/j.1524-4733.2007.00275.x.
- 21) Angst, F. (2011). The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Medical Research Methodology*, 11, 152; author reply 152. doi:10.1186/1471-2288-11-152.
- 22) Mokkink, L. B., Terwee, C. B., Gibbons, E., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). Inter-rater agreement and reliability of the COSMIN (COnsensus-based standards for the selection of health status measurement instruments) checklist. *BMC Medical Research Methodology*, 10, 82. doi:10.1186/1471-2288-10-82.
- 23) Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, 10, 22. doi:10.1186/1471-2288-10-22.
- 24) Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, 19(4), 539-549. doi:10.1007/s11136-010-9606-8.
- 25) Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, 21(4), 651-657. doi:10.1007/s11136-011-9960-1.
- 26) Johnson, C., Aaronson, N., Blazey, J. M., Bottomley, A., Fayers, P., Koller, M., et al. (2011). EORTC Quality of life group: Guidelines for developing questionnaire modules. http://groups.eortc.be/qol/sites/default/files/archives/guidelines_for_developing_questionnaire_final.pdf. Accessed November 26, 2011
- 27) Cella, D. (1997). *Manual of the functional assessment of chronic illness therapy (FACIT) measurement system*. Evanston, IL: Northwestern University.
- 28) Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., et al. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value in Health*, 12(4), 419-429. doi:10.1111/j.1524-4733.2008.00470.x.
- 29) Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, 8(2), 94-104. doi:10.1111/j.1524-4733.2005.04054.x.
- 30) Rothman, M., Burke, L., Erickson, P., Leidy, N. K., Patrick, D. L.,

- & Petrie, C. D. (2009). Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value in Health*, 12(8), 1075-1083. doi: 10.1111/j.1524-4733.2009.00603.x.
- 31) Wild, D., Eremenco, S., Mear, I., Martin, M., Houchin, C., Gawlicki, M., et al. (2009). Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR patient-reported outcomes translation and linguistic validation good research practices task force report. *Value in Health*, 12(4), 430-440. doi:10.1111/j.1524-4733. 2008. 00471.x.
 - 32) Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1998). *Standards for educational and psychological testing* Washington DC: American Psychological Association.
 - 33) Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R. D., Brod, M., et al. (2011). Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. *Quality of Life Research*,. doi:10.1007/s11136-011-9990-8.
 - 34) Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102-109. doi:10.1016/j.jclinepi. 2007.03.012.
 - 35) Valderas, J. M., Ferrer, M., Mendivil, J., Garin, O., Rajmil, L., Herdman, M., et al. (2008). Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value in Health*, 11(4), 700-708. doi:10.1111/j.1524-4733. 2007.00309.x.
 - 36) Dewolf, L., Koller, M., Velikova, G., Johnson, C., Scott, N., & Bottomley, A. (2009). EORTC quality of life group: Translation procedure. http://groups.eortc.be/qol/sites/default/files/archives/translation_manual_2009.pdf. Accessed November 26, 2011.
 - 37) Hays, R. D., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research*, 1(1), 73-75.
 - 38) Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., Acaster, S., & the Industry Advisory Committee of International Society for Quality of Life, R. (2012). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research*, 2012 Apr 17. [Epub ahead of print.]. doi:10.1007/s11136-012-0175-x.
 - 39) Ahmed, S., Berzon, R. A., Revicki, D., Lenderking, W., Moinpour, C. M., Basch, E., et al. (2012). The use of patient-reported outcomes (PRO) within comparative effectiveness research: Implications for clinical practice and healthcare policy. *Medical Care*, 50(12), 1060-1070.
 - 40) Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42. doi:10.1016/j.jclinepi.2006.03.012.
 - 41) Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
 - 42) Norquist, J. M., Girman, C., Fehnel, S., Demuro-Mercon, C., & Santanello, N. (2011). Choice of recall period for patient-reported outcome (PRO) measures: Criteria for consideration. *Quality of Life Research*,. doi:10.1007/s11136-011-0003-8.
 - 43) Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., & Aaronson, N. K. (2006). Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*, 4, 70. doi:10.1186/1477-7525-4-70.
 - 44) Brozek, J. L., Guyatt, G. H., & Schunemann, H. J. (2006). How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health and Quality of Life Outcomes*, 4, 69. doi:10.1186/1477-7525-4-69.
 - 45) Norman, G. R., Sridhar, F. G., Guyatt, G. H., & Walter, S. D. (2001). Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Medical Care*, 39(10), 1039-1047.
 - 46) Koller, M., Kantzer, V., Mear, I., Zarzar, K., Martin, M., Greimel, E., et al. (2012). The process of reconciliation: Evaluation of guidelines for translating quality-of-life questionnaires. *Expert Review of Pharmacoeconomics & Outcomes Research*, 12(2), 189-197. doi:10.1586/erp.11.102.
 - 47) Jordan, J. E., Osborne, R. H., & Buchbinder, R. (2011). Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *Journal of Clinical Epidemiology*, 64(4), 366-379. doi:10.1016/j.jclinepi. 2010.04.005.
 - 48) Acquadro, C., Conway, K., Hareendran, A., & Aaronson, N. (2008). Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value in Health*, 11(3), 509-521. doi:10.1111/j.1524-4733. 2007.00292.x.
 - 49) Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, 56(5), 395-407.
 - 50) Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., Hays, R. D., & Mayo, F. D. A. P.-R. O. C. M. G. (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health*, 10(Suppl 2), S94-S105. doi: 10.1111/j.1524-4733.2007.00272.x.
 - 51) Hays, R. D., & Woolley, J. M. (2000). The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics*, 18(5), 419-423.
 - 52) Kemmler, G., Zabernigg, A., Gatteringer, K., Rumpold, G., Giesinger, J., Sperner-Unterweger, B., et al. (2010). A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient. *Journal of Clinical Epidemiology*, 63(2), 171-179. doi:10.1016/j.jclinepi.2009.03.016.
 - 53) Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hrobjartsson, A., et al. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1), 96-106. doi:10.1016/j.jclinepi. 2010.03.002.
 - 54) Schmidt, S., & Bullinger, M. (2003). Current issues in cross-cultural quality of life instrument development. *Archives of*

- Physical Medicine and Rehabilitation, 84(4 Suppl 2), S29-S34. doi:10.1053/apmr.2003.50244.
- 55) Sprangers, M. A., Moinpour, C. M., Moynihan, T. J., Patrick, D. L., Revicki, D. A., & Clinical Significance Consensus Meeting Group. (2002). Assessing meaningful change in quality of life over time: A users' guide for clinicians. *Mayo Clinic Proceedings*, 77(6), 561-571. doi:10.4065/77.6.561.
- 56) Turner, R. R., Quittner, A. L., Parasuraman, B. M., Kallich, J. D., Cleeland, C. S., & Mayo, F. D. A. P-R. O. C. M. G. (2007). Patient-reported outcomes: Instrument development and selection issues. *Value in Health*, 10(Suppl 2), S86-S93. doi:10.1111/j.1524-4733.2007.00271.x.
- 57) Schunemann, H. J., & Guyatt, G. H. (2005). Commentary—goodbye M(C)ID! Hello MID, where do you come from? *Health Services Research*, 40(2), 593-597. doi:10.1111/j.1475-6773.2005.00374.x.
- 58) Schunemann, H. J., Puhan, M., Goldstein, R., Jaeschke, R., & Guyatt, G. H. (2005). Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). *Copd*, 2(1), 81-89.

本稿は Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013 Oct;22(8):1889-905. doi: 10.1007/s11136-012-0344-y. の日本語訳である。翻訳と掲載にあたり、Copyright Clearance Center (License No: 5379070610506) の許諾を得て、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)」の支援により作成した。