患者報告アウトカム尺度のシステマティックレビューのためのCOSMINガイドライン

■

COSMIN guideline for systematic reviews of patient-reported outcome measures

■

C. A. C. Prinsen[1, 4]，L. B. Mokkink[1]，L. M. Bouter[1]，J. Alonso[2]，D. L. Patrick[3]，H. C. W. de Vet[1]，C. B. Terwee[1]

受理日：2018 年 1 月 23 日 / オンライン掲載日：2018 年 2 月 12 日
© C. A. C. Prinsen, et al. 2018.　本稿はオープンアクセス論文である。

訳　宮崎貴久子（京都大学大学院医学研究科社会健康医学系専攻 健康情報学分野）
兼安　貴子（立命館大学 生命科学部，総合科学技術研究機構）
齋藤　信也（岡山大学大学院保健学研究科）
下妻晃二郎（立命館大学 生命科学部，総合科学技術研究機構）

抄　録

目的：患者報告アウトカム尺度（patient reported outcome measure: PROM）のシステマティックレビューは，介入試験や診断検査の精度研究のレビューとは異なり複雑である。実際，1 つ以上の PROM のレビューの実施は複数のレビュー（各 PROM の各測定特性につき 1 回のレビュー）からなる。測定特性のレビューのために特化して設計されたガイダンスがないことから，われわれの目的は PROM のシステマティックレビュー実施のためのガイドラインを開発することであった。

方法：文献レビューおよび専門家の見解に基づき，既存のガイドラインに従い COnsensus-based Standards for the selection of health Measurement INstruments（COSMIN）運営委員会は PROM のシステマティックレビューのためのガイドラインを開発した。

結果：PROM のシステマティックレビューを実施するための連続的な 10 ステップによる手順を提案する。ステップ 1 ～ 4 は文献検索の準備および実施，ならびに関連研究の選択に関するものである。ステップ 5 ～ 8 は適格な研究の質，測定特性，解釈可能性および実行可能性の評価に関するものである。ステップ 9 および 10 は推奨の定式化およびシステマティックレビューの報告に関するものである。

結論：PROM のシステマティックレビューのための COSMIN ガイドラインは，測定特性に関する研究の方法論的質を PROM 自体の質（すなわち，その測定特性）と結び付ける方法論を含む。これにより，レビュアーが透明性のある結論を導き出し，PROM の品質に関するエビデンスに基づく推奨を行うことが可能となり，研究および臨床診療で用いるための PROM のエビデンスに基づく選択をサポートする。

キーワード：COSMIN，システマティックレビュー，測定特性，PROM，アウトカム測定尺度，アウトカム尺度，方法論

C. A. C. Prinsen
c.prinsen@vumc.nl

[1]Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, De Boelelaan 1089a, 1081 HV Amsterdam, The Netherlands; [2]Health Services Research Unit, IMIM-Hospital del Mar Medical Research Institute; CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; [3]Department of Health Services, University of Washington, Seattle, WA, USA; [4]Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University, Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

略　語

AUC：Area under the curve（曲線下面積）
CFI：Comparative fit index（比較適合度指標）
COMET：Core Outcome Measures in Effectiveness Trials（有効性試験におけるコアアウトカム尺度）
COSMIN：COnsensus-based Standards for the selection of health Measurement INstruments（健康測定尺度の選択に関する合意に基づく指針）
CTT：Classical test theory（古典的テスト理論）
DIF：Differential item functioning（特異項目機能）
GRADE：Grades of Recommendation, Assessment, Development and Evaluation（エビデンスの質と推奨の強さのグレーディング）
ICC：Intraclass correlation coefficient（級内相関係数）
IRT：Item response theory（項目反応理論）
LoA：Limits of agreement（誤差の許容範囲）
MIC：Minimal important change（最小重要変化）
PRISMA：Preferred Reporting Items for Systematic Reviews and Meta-Analyses（システマティックレビューおよびメタアナリシスのための優先的報告項目）
PROM：Patient-reported outcome measure（患者報告アウトカム尺度）
SEM：Standard error of measurement（測定の標準誤差）
SDC：Smallest detectable change(検出可能な最小限の変化)
TLI：Tucker–Lewis index（Tucker–Lewis 指標）

緒　　言

患者報告アウトカム（patient reported outcome: PRO）は，臨床家や他の誰からも修正や解釈されない患者の健康状態のあらゆる側面について，患者から直接得られる測定結果である[1]。PRO は，患者報告アウトカム尺度（PRO measure: PROM）としても知られ，自己記入式質問票を用いて評価されることがもっとも多い。しかし，用いる PROM の品質はかなり異なることが知られており，信頼性および妥当性がもっとも高い PROM が選択されているかどうかは通常明らかではない[2-5]。

PROM のシステマティックレビューは，特定の研究対象集団において対象となる構成概念の測定にもっとも適した PROM を選択するための重要なツールである。高い質のシステマティックレビューは PROM の測定特性の包括的な概観を提供し，特定の目的（研究または臨床診療，判別，評価，予測への適用）にもっとも適した PROM を選択する際のエビデンスに基づく推奨をサポートする。異なった PROM は，それぞれの目的に応じて異なることがあり，実行可能性の側面も左右する。また，PROM のシステマティックレビューは検討中の PROM の測定特性に関する知識のギャップを特定し，測定特性に関する新たな研究の設計に用いることができる。

PROM のシステマティックレビューの数は増加しており，1990 年代初頭はかろうじて年に 1 件であったのが，現在では毎年 100 件を超える[6]。健康関連アウトカム測定尺度のシステマティックレビューの質に関する最近のレビューでは，かなり改善の余地があることが示された[7]。

COnsensus-based Standards for the selection of health Measurement INstruments（COSMIN）イニシアティブは，研究および臨床診療のための高い質の PROM の選択を容易にすることを目的とする。COSMIN が開発したツールの一つに PROM のシステマティックレビューのためのプロトコルがあり，2011 年以降，COSMIN のウェブサイト（http://www.cosmin.nl）で入手可能となった[8]。PROM のシステマティックレビューのための広範な公表ガイドラインがないことから，COSMIN 運営委員会（本論文の著者ら）はこのプロトコルを拡張し，PROM のシステマティックレビューのための包括的な方法論的ガイドラインとすることを目指した。本ガイドラインは，評価目的で使用される PROM のうち，少なくとも測定特性に関する情報が得られているものについて，システマティックレビューの方法論を一連の 10 のステップで記述し，特定の目的のための PROM の選択を支援する。システマティックレビューの実施を支援する詳細情報は，付属の「PROM のシステマティックレビューのための COSMIN 方法論—ユーザーマニュアル」および「PROM の内容妥当性評価のための COSMIN 方法論—ユーザーマニュアル」に記載されている（いずれも COSMIN のウェブサイトで入手可能）[8-10]。これらのユーザーマニュアルは本ガイドラインの補足文書であり，PROM のシステマティックレビューを行

うレビュアーの支援を目的とする。「PROM のシステマティックレビューのための COSMIN 方法論―ユーザーマニュアル」は PROM のシステマティックレビューのおのおののステップについて異なるシナリオの複数の例でサポートされた詳細な情報を提供する。

方　　法

経験的なエビデンスがない中で，PROM のシステマティックレビューのための本 COSMIN ガイドラインは，われわれ（COSMIN 運営委員会）が数年にわたり実施した，PROM のシステマティックレビューの実施[11,12]，システマティックレビューを行う他のレビュアーの研究支援[13,14]および COSMIN 方法論の開発[15,16]の経験に基づいている。さらに，連続した 2 本のレビュー[7,17]において，PROM のシステマティックレビューの品質を検討した。また，COSMIN 方法論を用いたレビューについては，レビュー著者による COSMIN 方法論に関連したコメントを特に検索した。さらにまた，COSMIN 運営委員会では，対面会議（CP，WM，HdV および CT）や電子メールによる討議を繰り返した。PROM の内容妥当性に関する最近のデルファイ研究の結果[18]およびコアアウトカムセット（core outcome set: COS）に含めるアウトカムのための測定尺度の選択に関する以前のデルファイ研究の結果[19]からも経験を積んだ。さらに加えて，以下の既存のレビューガイドライン：介入のシステマティックレビュー[20]および診断検査精度のレビュー[21]のためのコクランハンドブック，PRISMA 声明[22]，米国医学研究所の効果比較研究のシステマティックレビューの基準（the Institute of Medicine standards for systematic reviews of comparative effectiveness research）[23]，エビデンスの質と推奨の強さのグレーディング（Grading of Recommendations Assessment, Development and Evaluation：GRADE）の原則[24]とも用語を統一して開発した。

結　　果

PROM のシステマティックレビューを実施するための一連の 10 のステップによる手順を推奨する（**図 1**）。これらのステップは 3 つのパート（A，B および C）に分けられる。

パート A．文献検索の実施

パート A はステップ 1 〜 4 からなる。通常，これらのステップはシステマティックレビューを行う際の標準的な手順であり，既存のレビューガイドラインとも一致する[20,21]。

ステップ 1．レビュー目的の定式化

PROM のシステマティックレビューの目的は PROM の質に焦点を当てる。これには以下の 4 つの重要要素を含む必要がある：(1) 構成概念，(2) 対象集団，(3) 尺度の種類，(4) 対象となる測定特性。例:「われわれの目的は多発性硬化症（MS），パーキンソン病（PD）または脳卒中患者のためのすべての自己報告式疲労質問票の測定特性の質を批判的に評価，比較および要約することである」[25]。

ステップ 2．適格基準の定式化

適格基準はレビュー目的の以下の 4 つの重要要素と一致すべきである：(1) PROM は対象となる構成概念の測定を目的とする，(2) 調査サンプル（例: または任意の過半数，例：50%以上）は対象となる集団を代表する，(3) 研究は PROM に関するものである，(4) 研究の目的は 1 つ以上の測定特性の評価，PROM の開発（内容妥当性の評価）または対象となる PROM の解釈可能性の評価（例：研究対象集団におけるスコアの分布の評価，欠測項目の割合，床効果および天井効果，関連する［部分］集団でのスコアおよび変化スコアの利用可能性，最小重要変化［minimal important change: MIC］または最小重要差[26]）。PROM をアウトカム測定尺度として用いただけの研究は除外することを推奨する。これらの研究は PROM の測定特性についての非直接的なエビデンスを提供する。たとえば，アウトカムを評価するために PROM を用いた研究（例：ランダム化比較試験）や他の尺度の妥当性検証研究において PROM を用いた研究に関するものである。さらに，レビューには論文本文のみを含めることを推奨する。抄録はしばしば研究デザインに関する情報が

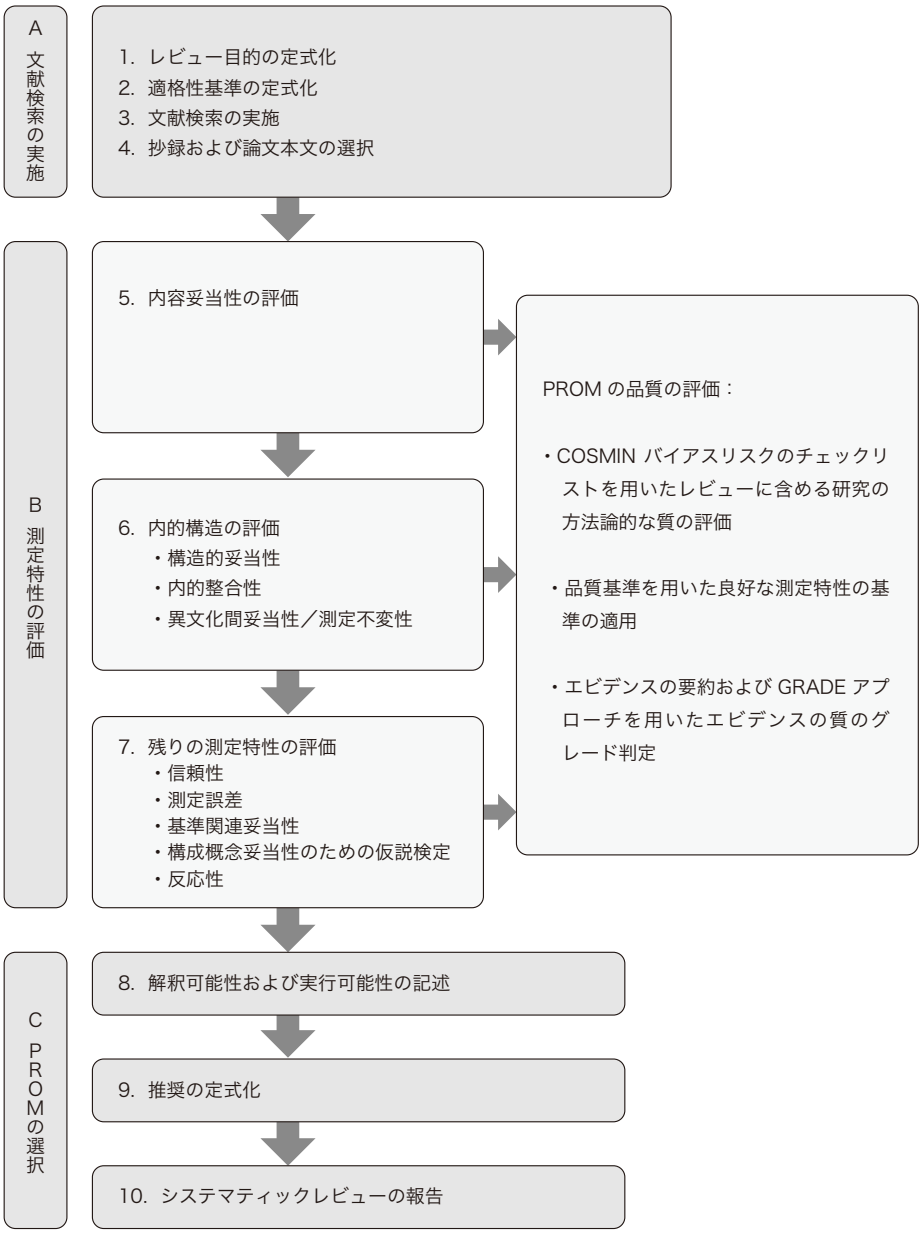

図 1　PROM のシステマティックレビューを実施するための 10 のステップ

極めて限られており，ステップ 5 ～ 7 での研究の質評価および測定特性の結果の妨げとなる。

ステップ 3.　文献検索の実施

コクラン方法論[20, 21]と一致し，およびコンセンサス[19]に基づき，MEDLINE と EMBASE を検索すべき最低限のデータベースとする。それに加え，対象となる構成概念や集団に応じて，他の（内容特異的な）データベース（例：Web of Science, Scopus，CINAHL，PsycINFO）を検索することが推奨される。

適切な検索方法は，以下のレビュー目的の 4 つの重要要素：(1) 構成概念，(2) 対象集団，(3) 尺度の種類，(4) 測定特性のための検索用語（索引語，フリーテキスト用語）の包括的な収集からなる。医療司書や対象となる構成概念および研究対象集団の専門家に相談することが推奨される。

Oxford 大学の患者報告アウトカム評価グループ

（Patient-Reported Outcomes Measurement Group）により，PubMed 用の包括的な PROM フィルターが開発されている。これは，測定尺度の種類のための検索フィルターとして使用することができ，COSMIN のウェブサイトで入手可能である[8]。測定特性に関する検索用語については，測定特性に関する研究を検索するための感度の高い検証済みの検索フィルターを用いることを推奨する[27]。PubMed および EMBASE 用の検索フィルターが利用可能であり，COSMIN のウェブサイトから入手することができる[8]。PubMed 検索方法の例については，COSMIN ユーザーマニュアル[9] を参照のこと。

コクラン方法論と一致し，開始日から現在までのデータベースを検索することが推奨される[20,21]。言語制限の使用はステップ 2 で定めた選択基準による。一般には，レビューのために論文を翻訳するリソースがない場合であっても，検索方法において言語制限を使用しないことが推奨される。このようにして，レビュー著者は少なくともそれらの存在を報告することができる。

ステップ 4.　抄録および論文本文の選択

通常，2 名のレビュアーが独立して抄録および論文本文の選択を行うことが推奨される[20,21]。抄録に基づき，少なくとも 1 名のレビュアーが，関連性がある研究であると判断した場合，または疑わしい場合は，論文本文を入手し，スクリーニングを行う必要がある。違いについて議論し，2 名のレビュアーの間で合意に達しない場合は，3 人目のレビュアーに相談することが推奨される。レビューに含めた論文の参考文献をすべて確認し，関連する可能性のある追加の研究を検索することも推奨される。新たな論文が多数認められた場合，最初の検索方法は十分に包括的ではなかった可能性があり，改善し，再度検索を行う必要があろう。

パート B．測定特性の評価

パート B はステップ 5 ～ 7 からなり，レビューに含める PROM の測定特性の評価に関するもので，3 つのサブステップを含む（**図 1**）。最初に，COSMIN バイアスリスクのチェックリスト[28] を用いて，測定特性に関するおのおのの研究の方法論的な質を評価する。各研究を「極めて良好（very good）」「適切（adequate）」「疑わしい（doubtful）」または「不適切な質（inadequate quality）」のいずれかに評価する。2 番目に，良好な測定特性のための更新された基準[29]（コンセンサスが得られ[19]，最近の新たな洞察に基づいてわずかに変更された基準）（**表 1**）を用いて，測定特性に関するおのおのの研究の結果を評価する。それぞれの結果を「十分（sufficient）（+）」「不十分（insufficient）（−）」または「不確定（indeterminate）（?）」のいずれかに評価する。3 番目に，エビデンスを要約し，GRADE アプローチを用いてエビデンスの質のグレードを判定する。測定特性に関する利用可能なすべての研究の結果を定量的に統合または定性的に要約し，良好な測定特性のための基準と比較し，PROM の（全体的な）測定特性が「十分（sufficient）（+）」「不十分（insufficient）（−）」「非一貫（inconsistent）（±）」または「不確定（indeterminate）（?）」かどうかを判定する。前のサブステップでの焦点は単一研究であったが，ここでの焦点は PROM である。研究あたりの判定がすべて「十分（sufficient）」（またはすべて「不十分（insufficient）」）の場合，結果を統計学的に統合することが可能であり，総合判定は良好な測定特性の基準に基づき「十分（sufficient）（+）」（または「不十分（insufficient）（−）」）となる。結果に一貫性がない場合，非一貫の説明（例：異なる研究対象集団または方法）を探索すべきである。説明が見いだせた場合，一貫性のある結果が得られた関連のあるサブグループ（例：成人対小児，急性疾患患者対慢性疾患患者, PROM の異なる［言語］バージョンなど）について総合評価を行う。説明が見いだせない場合，総合評価は「非一貫（inconsistent）（±）」となる。十分な情報が利用できない場合の総合評価は「不確定（indeterminate）（?）」とする。COSMIN ユーザーマニュアルには，良好な測定特性のための基準を用いて，統合または要約された測定特性の結果を判定する方法に関する詳細情報がある[9]。

各測定特性の総合評価（十分［sufficient］［+］，不十分［insufficient］［−］, 非一貫［inconsistent］［±］）はエビデンスの質のグレードの判定を伴う。これは，統合した結果または総合評価がどの程度の確信を

表 1　良好な測定特性のための更新された基準

測定特性	判定	基準
構造的妥当性	＋	**CTT** CFA：CFI, TLIまたは同等の測定> 0.95, RMSEA < 0.06もしくはSRMR < 0.08[a] **IRT/Rasch** 一次元性[b]の逸脱がない：CFI, TLIまたは同等の測定> 0.95, RMSEA < 0.06もしくはSRMR < 0.08 かつ 局所独立の逸脱がない：支配因子について調整後の項目間の残差相関 < 0.20またはQ3's < 0.37 かつ 単調性の逸脱がない：見た目が適切なグラフ, または項目のスケーラビリティ > 0.30 かつ 適切なモデルの当てはまり IRT：$\chi^2 > 0.001$ Rasch：インフィットおよびアウトフィット平均平方 ≥ 0.5 および ≤ 1.5, またはZ-標準化値 > –2 および < 2
	?	CTT：「+」に関する情報すべてが報告されたわけではない IRT/Rasch：モデルの当てはまりに関する報告がない
	–	「+」の基準を満たさない
内的整合性	＋	十分な構造的妥当性[d]についての「低」以上のエビデンス[c]があり, かつ各一次元スケールまたはサブスケールのクロンバックのα係数 ≥ 0.70[e]
	?	『十分な構造的妥当性dについての「低」以上のエビデンス[c]』の基準を満たさない
	–	十分な構造的妥当性dについての「低」以上のエビデンス[c]があり, かつ各一次元スケールまたはサブスケールのクロンバックのα係数 < 0.70[e]
信頼性	＋	ICCまたは重み付けκ係数 ≥ 0.70
	?	ICCまたは重み付けκ係数の報告がない
	–	ICCまたは重み付けκ係数 < 0.70
測定誤差	＋	SDCまたはLoA < MIC[d]
	?	MICが定義されていない
	–	SDCまたはLoA > MIC[d]
構成概念妥当性のための仮説検定	＋	結果が仮説と一致する[f]
	?	（レビューチームにより）仮説が定義されていない
	–	結果が仮説と一致しない[f]
異文化間妥当性/測定不変性	＋	多群因子分析において群因子（年齢, 性別, 言語など）間で重要な差が認められない, または群因子について重要なDIFがない（McFadden R^2 < 0.02）
	?	多群因子分析またはDIF分析が行われていない
	–	群因子間で重要な差がある, またはDIFが認められた
基準関連妥当性	＋	ゴールドスタンダードとの相関 ≥ 0.70またはAUC ≥ 0.70
	?	「+」に関する情報すべてが報告されたわけではない
	–	ゴールドスタンダードとの相関 < 0.70またはAUC < 0.70
反応性	＋	結果が仮説と一致する[f], またはAUC ≥ 0.70
	?	（レビューチームにより）仮説が定義されていない
	–	結果が仮説と一致しない[f], またはAUC < 0.70

基準はたとえば Terwee et al. [29] および Prinsen et al. [19] に基づく。
AUC, 曲線下面積；CFA, 確認的因子分析；CFI, 比較適合度指標；CTT, 古典的テスト理論；DIF, 特異項目機能；ICC, 級内相関係数；IRT, 項目反応理論；LoA, 誤差の許容範囲；MIC, 最小重要変化；RMSEA, 近似の二乗平均平方根誤差；SEM, 測定の標準誤差；SDC, 検出可能な最小限の変化；SRMR, 標準化残差平均平方；TLI, Tucker–Lewis指標
「+」= 十分（sufficient），「–」= 不十分（insufficient），「?」= 不確定（indeterminate）
[a]要約スコアの質を判定するためには, 因子構造は研究間で等しくなくてはならない。
[b]一次元性は下位尺度あたりの因子分析を指すが, 構造的妥当性は（多次元）患者報告アウトカム尺度の因子分析を指す。
[c]GRADEアプローチに従ったエビデンスのグレード判定により定めたとおり。
[d]このエビデンスは異なる研究から得られる場合もある。
[e]PROMの開発段階では関連性があるが, 既存のPROMの評価に際しては関連性がないため,「クロンバックのα係数 < 0.95」の基準を削除した。
[f]すべての研究の結果を総合した後, 結果の75%が仮説と一致するかどうかを判断する。

表 2　エビデンスの質のグレード判定のための修正 GRADE アプローチ

エビデンスの品質	以下の場合，低くなる
高	バイアスリスク
中	−1 深刻
低	−2 非常に深刻
非常に低	−3 極めて深刻
	非一貫性
	−1 深刻
	−2 非常に深刻
	不精確性
	−1 合計 $n = 50 \sim 100$
	−2 合計 $n < 50$
	非直接性
	−1 深刻
	−2 非常に深刻

出発点は，エビデンスが高い質であることと仮定する。次に，バイアスリスク（質の低い研究），結果における（説明のつかない）非一貫性，または非直接性の結果が認められる場合は，各押印（バイアスリスク，非一貫性，不精確性，非直接性）につき，エビデンスの品質のグレードを中，低，または非常に低まで1または2段階下げる[44]。グレードを下げる方法に関する情報は，COSMINユーザーマニュアル[9]に詳述されている。
n = サンプルサイズ

表 3　質レベルの定義

質レベル	定義
高	真の測定特性が測定特性の推定値に近いことに大きな確信がある。
中	測定特性の推定値に対し中程度の確信がある：真の測定特性は測定特性の推定値に近いと考えられるが，大きく異なる可能性もある。
低	測定特性の推定値に対する確信性には限界がある：真の測定特性は測定特性の推定値とは大きく異なるかもしれない。
非常に低	測定特性の推定値に対し，ほとんど確信がない：真の測定特性は測定特性の推定値とは大きく異なる。

これらの定義はGRADEアプローチ[24]から改変した。グレードを下げる方法に関する詳細情報については，COSMINユーザーマニュアル[9]を参照のこと。

もって信頼できるかを表す。特定の測定特性についての総合評価が「不確定（indeterminate）（?）」の場合，その PROM の質を評価できないため，エビデンスの質についてのグレード判定をしないことに留意する。介入研究のシステマティックレビューのための GRADE アプローチでは，5 要因：バイアスリスク，非直接性，非一貫性，不精確性，および出版バイアスの有無に応じてエビデンスを 4 段階（高，中，低，非常に低）に分類する[24]。ここで，PROM のシステマティックレビューにおけるエビデンスの質のグレードの判定のための修正 GRADE アプローチを導入する。結果の信頼性について懸念がある場合，GRADE アプローチを用いて，エビデンスの質のグレードを下げる。COSMIN 方法論では，GRADE の 5 要因のうち以下の 4 つを採用している：バイアスリスク（すなわち，研究の方法論的な質），非一貫性（すなわち，研究間での結果の説明のつかない非一貫性），不精確性（すなわち，利用可能な研究の総サンプルサイズ），非直接性（すなわち，レビュー対象となる集団とは異なる集団からのエビデンス）（**表 2**）。各測定特性および各 PROM について，個別にエビデンスの質を評価する。はじめは常に，統合または全体の結果が高い質であると仮定する。次に，バイアスリスク，（説明のつかない）非一貫性，不精確性または非直接的な結果が認められる場合は，各要因につき，エビデンスの質のグレードを中，低，または非常に低（定義は**表 3** を参照）まで 1 または 2 段階下げる。グレードを下げる方法に関する具体的な詳細は COSMIN ユーザーマニュアルに説明がある[9]。2 名のレビュアーが独立に質を評価し，必要な場合は 3 人目のレビュアーの協力を得て，レビュアー間での合意を得ることが推奨される。

レビューでは，PROM の各バージョン（すなわち，患者のサブグループごとに異なるバージョン，異なる言語バージョンなど）を個別に検討すべきであることに留意する。

ステップ 5.　内容妥当性の評価

内容妥当性とは，PROM の内容が測定する構成概念を適切に反映する程度を意味する[30]。対象となる構成概念および研究対象集団について，PROM の項目が適切で包括的であり，理解可能であるかどうかが内容妥当性によって明らかにされるため，内容妥当性はもっとも重要な測定特性であると考えられる。内容妥当性の評価はレビュアーによる主観的判断を要する。この判断では，PROM の開発研究，PROM に関する追加の内容妥当性研究の質および結果（利用可能な場合），ならびにレビュアーによる PROM 内容の主観的判定を考慮すべきである。PROM の内容妥当性の評価方法に関するガイダンスはほか[10]を参照のこと。

PROMの内容妥当性が不十分であることを示す質の高いエビデンスがある場合，そのPROMはシステマティックレビューのステップ6～8での検討は行わず，ステップ9でこのPROMのための推奨を直接導き出すことができる。

ステップ6．内的構造の評価

内的構造はPROM内の異なる項目がどのように関連しているかを意味し，複数の項目をスケールまたは下位尺度に統合する方法を決定するために明らかにすることが重要である。このステップは構造的妥当性（一次元性を含む），内的整合性，異文化間妥当性，および測定不変性その他の評価に関するものである。ここでは，既存のPROMのさらなる改良または新たなPROMの開発ではなく，既存のPROMの検証について述べる。これら3つの測定特性は，ステップ7での残りの測定特性とは対照的に，個々の項目の質および項目間の関係に焦点を当てる。PROMの内容妥当性を評価した後に，これらの測定特性を直接評価することが推奨される。

スケールまたは下位尺度の構造的妥当性（一次元性）のエビデンスは内的整合性の解析（すなわち，クロンバックのα係数）の解釈の前提条件であるため，最初に構造的妥当性（ステップ6.1）を評価した後，内的整合性（ステップ6.2）および異文化間妥当性/測定不変性（ステップ6.3）を評価することを推奨する。

ステップ6は，スケールまたは下位尺度の全項目が単一の基礎となる構成概念の表現であり，互いに相関することを仮定する反映的モデルに基づくPROMにのみ関連する。反映的モデルの例に不安の測定がある。不安は悩み，パニック，落ち着きのなさなどの具体的な特性において現われる。これらの特性について患者に質問することにより，不安の程度（この項目は構成概念の反映である）を評価することができる[31]。スケールまたは下位尺度の項目が互いに相関すると想定されない場合（形成的モデル），これらの解析に意味はなく，ステップ6は省略することができる。PROMが反映的モデルまたは形成的モデルに基づくかどうかの報告がない場合，レビュアーはPROMの内容が反映的モデルまたは形成的モデルに基づく可能性が高いかどうかを判断する必要がある[32]。

ステップ6.1．構造的妥当性の評価

構造的妥当性は，PROMのスコアが，測定される構成概念の次元性を適切に反映する程度を示し[30]，通常，因子分析または項目反応理論（item response theory: IRT）/Rasch解析により評価される。システマティックレビューでは，構造的妥当性を評価するために因子分析を行った研究と各下位尺度の一次元性を下位尺度ごとに個別に評価するために因子分析を行った研究を区別することが有用である。構造的妥当性を評価するためには，PROMの全項目について因子分析を行い，PROMの下位尺度の（仮定的な）数，および下位尺度内の項目のクラスター化を評価する（構造的妥当性研究）。下位尺度ごとに一次元性を評価するためには，各下位尺度の項目について個別に複数の因子分析を行い，各下位尺度自体が単一の構成概念を測定するかどうかを評価する（一次元性研究）。これらの分析は内的整合性の解析の解釈（ステップ6.2）およびIRT/Rasch解析には十分であるが，構成概念妥当性の一部としての構造的妥当性のためのエビデンスは提供しない。

構造的妥当性の評価はパートBで述べた以下の3つのサブステップからなる：(1) レビューに含める研究の方法論的質の評価，(2) 良好な測定特性のための基準の適用，(3) エビデンス，およびエビデンスの質のグレード判定の要約。

PROMの構造的妥当性が不十分という質の高いエビデンスがある場合，以降のステップにおいて，このPROMのさらなる評価を再考すべきである。

ステップ6.2．内的整合性の評価

内的整合性は項目間の相互関連性の程度を示し，しばしばクロンバックのα係数により評価される[30,33]。構造的妥当性の評価と同様に，内的整合性の評価も上述の3つのサブステップからなる。

ステップ6.3．異文化間妥当性/測定不変性の評価

異文化間妥当性/測定不変性は，翻訳または文化的に適したPROMに対する項目の性能が元のバージョンのPROMの項目の性能を適切に反映する程度を示す[30]。PROMを異なる「文化的」集団（民族，言語，性別，または年齢群が異なる集団）で用いる場合，異文化間妥当性/測定不変性を評価すべきであるが，ここでは異なる患者集団も考慮すべきであ

表 4　構成概念妥当性および反応性を評価するための一般的仮説

一般的仮説
1　類似の構成概念を測定する尺度（の変化）との相関が 0.50 以上
2　関連はあるが，類似しない構成概念を測定する尺度（の変化）との相関が低い（0.30〜0.50）
3　関連のない構成概念を測定する尺度（の変化）との相関が 0.30 未満
4　類似の構成概念を測定する尺度（の変化）との相関と関連はあるが，類似しない構成概念を測定する尺度（の変化）との相関の差が 0.10 以上 関連はあるが，類似しない構成概念を測定する尺度（の変化）との相関と関連のない構成概念を測定する尺度（の変化）との相関の差が 0.10 以上
5　関連のある（部分）集団（例：対象となる構成概念のレベルが高いまたは低いと予想される患者）間での意味のある変化
6　反応性については，AUC ≥ 0.70 であるべきである。

AUC：「代表的な基準」として用いられる外的尺度の変化との曲線下面積

る[9]。異文化間妥当性 / 測定不変性の評価では，たとえばロジスティック回帰分析を用いて特異項目機能（differential item functioning: DIF）が生じるかどうか，または多集団確証的因子分析（multigroup confirmatory factor analysis: MGCFA）を用いて因子構造および因子負荷量がグループ間で等しいかどうかを評価する。測定不変性および non-DIF は，潜在特性が同一レベル（グループ間の違いは許容する）である異なるグループの回答者が特定の項目に対して類似の反応を示すかどうかを意味する[34]。異文化間妥当性 / 測定不変性の評価も上記の 3 つのサブステップからなる。

ステップ 7．残りの測定特性の評価

次に，上記の 3 つのサブステップを再び用いて，残りの測定特性（信頼性，測定誤差，基準関連妥当性，構成概念妥当性のための仮説検定，反応性）を評価する。内容妥当性および内的構造とは異なり，これらの測定特性の評価は項目レベルではなくスケールまたは下位尺度全体の質に関する情報を提供する。

PROM を含む測定特性の評価では，考慮すべきいくつかの重要な問題がある。良好な測定誤差の基準を適用するためには，検出可能な最小限の変化（smallest detectable change: SDC）または誤差の許容範囲（limits of agreement: LoA），および MIC に関する情報が必要である。この情報は他の研究から得られることがある。MIC はアンカーに基づく経時的アプローチを用いて決定しておくべきである[35-38]。MIC は，複数の研究から複数のアンカーを用いて算出するのが最善である[39,40]。SDC または LoA が MIC よりも小さいかどうかを判断するための情報が十分利用できない場合は，エビデンスの質のグレード判定は行わずに，SDC または LoA に関する情報のみを報告することを推奨する（MIC に関する情報のみで PROM の解釈可能性に関する情報を提供することに留意する）。

構成概念妥当性および反応性のための仮説検定については，レビュアーは，それに対して結果を評価する仮説自体を定式化することが推奨される[9,28]。これらの仮説をレビュー目的に沿って定式化し，予想される関係（例：レビュー中の PROM と PROM との比較に用いる比較尺度の関係，相関の予想される方向および大きさ）を含める。一般的仮説の例を**表 4** に示す。このようにして，研究で見いだされたすべての結果を同一セットの仮説と比較することができる。結果の 75％以上が仮説と一致する場合，要約結果を「十分（sufficient）」と判定する。このようにして，PROM の構成概念妥当性に関するより頑強な結論を導き出すことができる。

パート C．PROM の選択

パート C はステップ 8 〜 10 からなり，PROM の解釈可能性および実行可能性の評価，推奨の定式化，およびシステマティックレビューの報告に関してである。

ステップ 8．解釈可能性および実行可能性の記述

解釈可能性は，PROM の定量的スコアまたはスコアの変化に対して定性的意味（臨床的または広く理解された含意）を与える程度として定義される[30]。たとえば，一部の測定特性を解釈するためにはスコアの分布に関する情報が必要であり，それはスコアのクラスター化を明らかにし，これが床効果・天井

効果を生じるかどうかを示す[31]。実行可能性は，時間または資金などの制約が与えられた場合に意図する状況でのPROMの適用の容易さとして定義される[41]。実行可能性は，記入時間，尺度にかかる費用，尺度の長さ，実施の種類および容易さなどの側面を表す[19]。実行可能性はPROM（自己記入式）に記入する患者，および面接を行い，患者にPROMを手渡す研究者や臨床家に適用される。解釈可能性および実行可能性はPROMの質には言及しないことから，測定特性ではない。しかし，両者は，十分考慮されたPROMの選択のための重要な側面とみなされる。2つのPROMがあり，質の面で区別することが極めて困難な場合，もっとも適した尺度の選択に際しては，実行可能性の側面を考慮することが推奨される。レビュアーは時間枠および予算の範囲内で何が実施可能かどうかを判断すべきである[19]。

ステップ9．推奨の定式化

評価的な適用での使用にもっとも適したPROMに関する推奨を，対象となる構成概念および研究対象集団に関して定式化する。根拠に基づいた十分に透明性のある推奨を行うために[31]，含めたPROMを以下の3カテゴリーに分類することを推奨する：（A）構成概念および対象となる集団にとってもっとも適したPROMとして推奨される可能性のあるPROM（十分な内容妥当性［すべてのレベル］のエビデンスおよび十分な内的整合性の低以上のエビデンスを有するPROM），（B）推奨される可能性があるが，さらなる妥当性検証研究を要するPROM（AまたはCに分類されないPROM），（C）推奨すべきでないPROM（不十分な測定特性についての高い質のエビデンスのあるPROM）。PROMを特定のカテゴリーに分類した根拠を述べ，該当する場合は，今後の妥当性検証の作業についての方向性を与えるべきである。もっとも適したPROMを1つ勧告することが推奨される[19]。この推奨は測定特性の評価に基づくだけでなく，解釈可能性および実行可能性の側面にも依存する可能性がある。

ステップ10．システマティックレビューの報告

PRISMA声明[22]に従い，以下の情報を報告することを推奨する：（1）PRIMSAフロー図に示された，文献検索，ならびに研究およびPROM選択の結果（レビューに含めた論文およびPROMの最終的な数を含む），（2）レビューに含めたPROMの特性（尺度の名称，測定する構成概念，開発されたPROMの研究対象集団，意図する使用文脈，PROMの言語バージョン，スケールまたはサブスケールの数，項目数，回答選択肢，想起期間，解釈可能性の側面，実行可能性の側面など），（3）研究対象集団の次のような特性：地理的位置，言語，疾患領域，対象集団，サンプルサイズ，年齢，性別，設定，国など，（4）各研究の測定特性およびPROMごとの方法論的な質，（5）測定特性ごとの結果の要約（summary of findings: SoF）テーブル。これには，測定特性，総合判定（十分［+］不十分［–］，不整合性［±］，不確定［?］），およびエビデンスの質のグレード（高，中，低，非常に低）について集められたまたは要約された結果を含む。最終的に，これらのSoFテーブル（測定特性ごとに1つ）を用いて，特定の目的または使用状況のためにもっとも適したPROMの選択に関する推奨を提供する。アウトカム測定の標準化（例：COSの開発）に取り組み，メタ解析を容易にするために，もっとも適したPROMを1つ勧告することが推奨される[19]。この推奨は解釈可能性および実行可能性の側面にも依存する。報告および公表に使用できるテーブルについては，COSMINユーザーマニュアル[9]を参照のこと。これらのテーブルは，レビュー全体を通してデータ抽出プロセスに用いることができることに留意する。さらに，たとえばウェブサイトや検討中の論文の（オンライン）補足資料のように利用できる検索方法を用いることが推奨される。

考　　察

経験的な実証がない中で，COSMIN運営委員会は，このガイドラインで述べたPROMのシステマティックレビュー実施のための方法論を開発した。PROMのシステマティックレビューを実施するための一連の10のステップによる手順が推奨される。PROMのすべての測定特性をさらに評価すべきか，あるいはPROMをさらなる評価から除外できるかどうかを決定する際には，事前に定めた順序で測定

特性を評価することが有用である。本ガイドラインはPROMのシステマティックレビューのために開発されたものであるが，ステップ5～7が当てはまるPROM以外のレビューのためのガイダンスとしても用いることができる。

本研究にはいくつかの限界があることをわれわれは認識している。本ガイドラインの開発はデルファイ法やノミナル・グループ技法（専門家パネル）のような構造化プロセスに基づくものではなく，コンセンサス会議の方法に従った[42]。われわれは内容妥当性および構造的妥当性についてのシステマティックレビューにおける方法論のみを適用し[43]，その他のレビューについては未着手である。次に，PROMのシステマティックレビューの方法は完全には開発されておらず，一部の側面はさらなる探索を要する。最初に，われわれは複数のデータベースを検索することを推奨する。しかし，PROMレビューのためのPubMedやEMBASE以外のデータベースの付加価値は限られており，系統的に評価されていない。2番目に，測定特性に関する研究を検索するための，MEDLINEやEMBASE以外のデータベース用の検索フィルターを開発すべきである。3番目に，測定特性を統計学的に統合併合する方法は少なく，さらなる開発を要する。4番目に，エビデンスの質の表に含めたサンプルサイズの要件は経験則である（サンプルサイズの要件に関するさらなる情報は，COSMINユーザーマニュアルにある）。5番目に，エビデンスの質をグレード判定する方法はいまだ完成していない。測定特性に関する研究のレジストリがないため，PROMのシステマティックレビューでは，GRADEアプローチに従って出版バイアスを評価することは難しい。また，エビデンスの質のグレードを下げる基準は現時点で存在するものの，グレードを上げる基準（例：測定特性が極めて良好なため）は（いまだ）定まっていない。最後に，今後の研究では，信頼性または妥当性に関するわれわれの方法を評価するであろう。

結　　論

この方法論のガイドラインの目的は，レビュー著者が透明性のある標準化された方法でPROMのシステマティックレビューを実施するのを支援することである。本ガイドラインは，これらのレビューの質およびエビデンスに基づくPROMの選択に寄与するであろう。

謝辞：本研究は，European Union Seventh Framework Programme［FP7/2007-2013］の一部としてCore Outcome Measures in Effectiveness Trials［COMET，http://www.comet-initiative.org］イニシアティブと共同で実施された。

著者の貢献度：LB Mokkink，LM Bouter，J Alonso，DL Patrick，HCW de VetおよびCB TerweeはCOSMIN分類およびオリジナルのCOSMINチェックリストを開発した。著者全員がCOSMINバイアスリスクのチェックリストの開発に関与している。

資金提供：CPEuropean Union's Seventh Framework Programme［FP7/2007-2013］からGrant AgreementNo. 305081として資金提供を受けた。

倫理基準の遵守

倫理審査の承認：本稿は，著者のいずれかがヒト被験者を用いて実施した研究を含まない。

オープンアクセス：本稿は，クリエイティブ・コモンズ表示4.0国際ライセンス（http://creativecommons.org/licenses/by/4.0/）の下に提供されている。原著者および出典を適切に明示し，クリエイティブ・コモンズ・ライセンスへのリンクを提供し，本稿に対する改変があれば，これを表示する限り，いかなる媒体でも自由に使用し，配布し，複製することができる。

参 考 文 献

1）J1. U.S. Food and Drug Administration (FDA). (2009). Guidance forIndustry. U.S. Department of Health and Human Services. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. https://www.fda.gov/downloads/drugs/guida nces/ucm193282.pdf. Accessed 7 Jan 2018.

2）Griffiths, C., Armstrong-James, L., White, P., Rumsey, N., Pleat, J., & Harcourt, D. (2015). A systematic review of patient reported outcome measures (PROMs) used in child and adolescent burn research. Burns, 41(2), 212–224.

3）Hermans, H., van der Pas, F. H., & Evenhuis, H. M. (2011). Instruments assessing anxiety in adults with intellectual disabilities: A systematic review. Research in Developmental Disabilities, 32(3), 861–870.

4）Keage, M., Delatycki, M., Corben, L., & Vogel, A. (2015). A systematic review of self-reported swallowing assessments in progressive neurological disorders. Dysphagia, 30(1), 27–46.

5）Ritmala-Castren, M., Lakanmaa, R. L., Virtanen, I., & Leino-Kilpi, H. (2014). Evaluating adult patients' sleep: An integrative literature review in critical care. Scandinavian Journal of Caring

Sciences, 28(3), 435–448.
6）COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) database of systematic reviews of outcome measurement instruments. Amsterdam. http://www.cosmin.nl/database-of-systematic-reviews.html. Accessed 5 Feb 2018.
7）Terwee, C. B., Prinsen, C. A., Ricci Garotti, M. G., Suman, A., de Vet, H. C., & Mokkink, L. B. (2016). The quality of systematic reviews of health-related outcome measurement instruments. Quality of Life Research, 25(4), 767–779.
8）COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) website. http://www.cosmin.nl. Accessed 5 Feb 2018.
9）Mokkink, L. B., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) – user manual. http://www.cosmin.nl/.
10）Terwee, C. B., Prinsen, C. A., Chiarotto, A., de Vet, H. C., Bouter, L. M., Alonso, J., Westerman, M. J., Patrick, D. L., & Mokkink, L. B. (2018). COSMIN methodology for assessing the content validity of PROMs – user manual. http://www.cosmin.nl/.
11）Collins, N. J., Prinsen, C. A., Christensen, R., Bartels, E. M., Terwee, C. B., & Roos, E. M. (2016). Knee Injury and Osteoarthritis Outcome Score (KOOS): Systematic review and metaanalysis of measurement properties. Osteoarthritis Cartilage, 24(8), 1317–1329.
12）Gerbens, L. A., Prinsen, C. A., Chalmers, J. R., Drucker, A. M., von Kobyletzki, L. B., Limpens, J., Nankervis, H., Svensson, Å., Terwee, C. B., Zhang, J., Apfelbacher, C. J., Spuls, P. I. & Harmonising Outcome Measures for Eczema (HOME) initiative. (2017). Evaluation of the measurement properties of symptom measurement instruments for atopic eczema: a systematic review. Allergy, 72(1), 146–163.
13）Chinapaw, M. J., Mokkink, L. B., van Poppel, M. N., van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for youth: A systematic review of measurement properties. Sports Medicine, 40(7), 539–563.
14）Speksnijder, C. M., Koppenaal, T., Knottnerus, J. A., Spigt, M., Staal, J. B., & Terwee, C. B. (2016). Measurement properties of the quebec back pain disability scale in patients with nonspecific low back pain. Systematic review. Physical Theraphy, 96(11), 1816–1831. Quality of Life Research (2018) 27:1147–1157 1157
15）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.
16）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.
17）Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., et al. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Quality of Life Research, 18(3), 313–333.
18）Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., et al. (2017). COSMIN standards and criteria for evaluating the content validity of patient-reported outcome measures: A Delphi study. (Submitted to Quality of Life Research).
19）Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., et al. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - A practical guideline. Trials, 17(1), 449.
20）Higgins, J. P. T., & Green, S. (Eds.). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration (2011). http://handbook.cochr ane.org/. Accessed 5 Feb 2018.
21）Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. (2013). http://metho ds.cochr ane.org/sdt/handbook-dta-reviews. Accessed 5 Feb 2018.
22）Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement. (2016). http://www.prisma-statement.org/. Accessed 5 Feb 2018.
23）Eden, J., Levit, L., Berg, A., & Morton, S. (Eds.). (2011). Institute of Medicine; Board on Health Care Services; Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. Finding what works in health care: Standards for systematic reviews. Retrieved February 27, 2017, from https://www.nap.edu/catalog/13059/finding-what-works-in-health-care-standards-for-systematic-reviews.
24）GRADE Handbook. (2013). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. http://gdt.guidelinedevelopment.org/app/handbook/handbook.html. Accessed 5 Feb 2018.
25）Elbers, R. G., Rietberg, M. B., van Wegen, E. E., Verhoef, J., Kramer, S. F., Terwee, C. B., & Kwakkel, G. (2012). Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: A systematic review of measurement properties. Quality of Life Research, 21(6), 925–944.
26）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.
27）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.
28）Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2017). COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. Quality of Life Research. https://doi.org/10.1007/s11136-017-1765-4.
29）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.

30）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.
31）de Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). Measurement in medicine. Cambridge: Cambridge University Press.
32）Fayers, P. M., Hand, D. J., Bjordal, K., & Groenvold, M. (1997). Causal indicators in quality of life research. Quality of Life Research, 6(5), 393–406.
33）Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78, 98–104.
34）Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.
35）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.
36）de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. Health and Quality of Life Outcomes, 4, 54.
37）de Vet, H. C., Ostelo, R. W., Terwee, C. B., van der Roer, N., Knol, D. L., Beckerman, H., et al. (2007). Minimally important change determined by a visual method integrating an anchorbased and a distribution-based approach. Quality of Life Research, 16(1), 131–142.
38）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.
39）Yost, K. J., Eton, D. T., Garcia, S. F., & Cella, D. (2011). Minimally important differences were estimated for six patientreported outcomes measurement information system-cancer scales in advanced-stage cancer patients. Journal of Clinical Epidemiology, 64(5), 507–516.
40）van Kampen, D. A., Willems, W. J., van Beers, L. W., Castelein, R. M., Scholtes, V. A., & Terwee, C. B. (2013). Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). Journal of Orthopaedic Surgery and Research, 8, 40.
41）Outcome Measures in Rheumatology (OMERACT) Handbook. (2017). https://www.dropbox.com/s/kkph9e3jdwctewi/OMERACT % 20Handbook % 20Dec % 2020 % 202017.pdf?dl=0. Accessed 7 Jan 2018.
42）Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. British Medical Journal, 311(7001), 376–380.
43）Chiarotto, A., Ostelo, R. W., Boers, M., & Terwee, C. B. (2017). A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in low back pain. Journal of Clinical Epidemiology, S0895–S4356(17), 30543–30547. https://doi.org/10.1016/j.jclinepi.2017.11.005.
44）Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., et al. (2011). GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology, 64(4), 383–394.

本稿は Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research (2018) 27:1147–1157. https://doi.org/10.1007/s11136-018-1798-3 の日本語訳である。

本稿は，Creative Commons CC BY ライセンスのもとで配布されるオープンアクセス論文であり，無制限の利用を許可するものである。このライセンスは原著作物を適切に引用することを条件に，いかなる媒体でも無制限の使用，配布，および複製を許可する。

本成果物は，厚生労働科学研究費補助金「関連学会の取組と連携した PRO ガイドラインの作成（研究代表者：立命館大学 下妻晃二郎，課題番号：20AC1003)」の支援により作成した。