

International guidance on the selection of patient-reported outcome measures in clinical trials

Crossnohere, Norah L. ; Brundage, Michael; Calvert, Melanie; King, Madeleine; Reeve, Bryce B.; Thorner, Elissa ; Wu, Albert W ; Snyder, Claire

DOI:

[10.1007/s11136-020-02625-z](https://doi.org/10.1007/s11136-020-02625-z)

License:

None: All rights reserved

Document Version

Peer reviewed version

Citation for published version (Harvard):

Crossnohere, NL, Brundage, M, Calvert, M, King, M, Reeve, BB, Thorner, E, Wu, AW & Snyder, C 2020, 'International guidance on the selection of patient-reported outcome measures in clinical trials: a review', *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*.
<https://doi.org/10.1007/s11136-020-02625-z>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is a post-peer-review, pre-copyedit version of an article published in Quality of Life Research. The final authenticated version is available online at: <https://doi.org/10.1007/s11136-020-02625-z>

This document is subject to Springer Nature re-use terms:

<https://www.springer.com/gp/open-access/publication-policies/aam-terms-of-use>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

International guidance on the selection of patient-reported outcome measures in clinical trials: A review

Norah L. Crossnohere^{1†}

Michael Brundage²

Melanie J. Calvert³

Madeleine King⁴

Bryce B. Reeve⁵

Elissa Thorner⁶

Albert W. Wu^{1,7}

Claire Snyder^{1,6,7}

1. Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
2. Queen's Cancer Research Institute, Queen's University, Kingston, Canada
3. Centre for Patient Reported Outcomes Research (CPROR), Institute of Applied Health Research, and NIHR, Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK
4. The University of Sydney, Sydney, Australia
5. Duke University School of Medicine, Durham, NC
6. Johns Hopkins Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD
7. Johns Hopkins School of Medicine, Baltimore, MD

[†]Corresponding author:

ncrossn1@jhu.edu

Word Count: 4245

Tables/Figures: 5

References: 46

Keywords: Outcome assessment, quality of life, guidance as topic, health care, clinical trials, oncology

Introduction

Patient-reported outcome (PRO) measures (PROMs) are a critical aspect of clinical trials given their ability to directly capture the experiences and perspectives of patients in regard to symptoms, health-related quality of life, and satisfaction with treatments [1-6]. Despite growing appreciation of the value of PRO data to improve clinical science and health outcomes [7, 8], important information regarding PRO methods (e.g. rationale/hypothesis, data collection approaches, training, management and analysis) may be missing from clinical trial protocols and reports [9-11]. Further, the process of translating PRO knowledge from clinical trials into clinical practice has not been adequately established [5].

Optimizing the value of PROMs from clinical trials requires that PROs are measured appropriately, reported clearly, and applied appropriately in practice. To meet this need for PRO optimization, the PROTEUS Consortium (Patient-Reported Outcomes Tools: Engaging User & Stakeholders) was formed with international representation from key patient, clinician, research, and regulatory groups [12]. The primary objective of the PROTEUS Consortium is to ensure that patients, clinicians, and other decision-makers can use PRO data from clinical trials to make the best decisions they can about treatment options, particularly in oncology.

The PROTEUS Consortium has identified six “core” documents that provide guidance on PRO use across the clinical research lifecycle, from trial conceptualization to result dissemination including: selecting PRO measures [13]; writing PRO protocols [14]; analyzing PRO data [15] reporting PRO findings [16, 17]; and interpreting PRO papers [18]. These documents generally address unique phases of the clinical research continuum, and where there is overlap the documents offer consistent recommendations. However, the PROTEUS Consortium acknowledges that other published guidance documents also provide recommendations on the use of PROs in clinical trials, most notably with regards to selecting an appropriate PROM. In particular, various professional societies, consortia, and regulatory agencies have published recommendations on the selection of PROMs for use in clinical trials [13, 19-21].

With the proliferation of guidance documents aiming to aid the selection of PROMs for research studies, it is critical to evaluate whether the advice provided is consistent or conflicting. Consistent guidance provides reassurance that agreement exists in the field. Conflicting guidance could create confusion and uncertainty, limiting the usefulness of any individual recommendations. This review identifies guidance documents that provide input on the selection of a PROM for a clinical trial and explores the consistency of these recommendations across guidance documents.

Methods

Using an identification approach consistent with those used by others in the comparison of international guidance [22, 23], we conducted a targeted review of prominent guidance documents that offer recommendations on the selection of PROMs for clinical trials. Structured (i.e. checklists, rating scales, etc.) and unstructured (i.e. narrative-style) guidance documents offering recommendations on the selection of a PROM for a clinical trial were eligible for inclusion. Guidance were excluded if they related to how to develop/validate a measure, or if they focused on only a single PROM selection criterion. Guidance offering recommendations on the selection of PROMs for specific clinical conditions were not included, with the exception of guidance regarding oncology, as the primary focus of the PROTEUS Consortium is oncology clinical trials. Included documents are collectively referred to as 'guidance documents' or 'guidance.' Formal quality appraisal was not conducted as the goal of the review was to collate and compare recommendations in the field rather than to evaluate the quality of the science [24]. The time period of the review was 2000-2019. In instances where guidance had been updated over the time period of the review only the most recent version was included.

Identification of PROM Selection guidance documents

Guidance documents for review were identified through expert referral from the PROTEUS Steering Group (i.e. named authors) and the Consortium membership of 26 stakeholders from the US, Canada, UK, Europe, and Australia working with PROMs in fields including psychometrics, health services research, clinical trial design, clinical trial reporting, regulatory affairs, professional societies, national research funders, patient advocacy, and medical journal editors.

Expert referral of guidance was supplemented with a targeted literature/grey literature search. Key search terms for both searches included 'patient-reported outcome measures', 'psychometrics', and 'outcome assessment', as well as their relevant derivatives. The literature search was conducted in PubMed, and the grey literature search was conducted by searching the websites of regulatory agencies, international professional societies, and prominent public-private partnerships working in health services and outcomes research. The reference lists of the included guidance as well as searched citations of all included guidance were screened for additional relevant resources. The literature search was conducted by one author (NLC) with guidance from other authors.

Data abstraction and coding

Guidance documents were assessed qualitatively using content analysis. Content analysis identifies patterns and analyzes occurrences of messages embedded within texts [25]. A conventional content analysis approach was used to derive coding categories directly from the text data. This approach is appropriate when there is little pre-existing theory upon which to base coding schema [26]. Codes for the present analysis were based on domains identified as meaningful PROM properties in the International Society for Quality of Life Research (ISOQOL) published guidance on minimum standards for PROMs in patient-centered outcomes research (hereon referred to as “ISOQOL Minimum Standards”) [13]. The PROTEUS Consortium selected the ISOQOL Minimum Standards as the “core” guidance for PROM selection.

A total of nine content categories were ultimately established through the content analysis: conceptual/measurement model, reliability, content validity, construct validity, responsiveness, interpretability of scores, translation, burden, and an umbrella category of ‘additional domains.’ The ISOQOL Minimum Standards are parsimonious by design, so in addition to extracting data on the domains reported in the Minimum Standards, we also noted considerations for PRO selection featured in other guidance documents that could not be coded with the initial codes from the ISOQOL Minimum Standards and later determined if these represented a new category. Table 1 defines the domains that served as coding content categories.

A structured abstraction process was used to collect general information about each guidance, including the stated purpose of the guidance, organizational affiliation of the guidance, country of organizational affiliation, and format of guidance (rating scale, checklist, or narrative). Data regarding any of the identified domains was abstracted from the original document and compiled into a single table so as to compare all recommendations for each content category across the guidance documents. Data abstraction and content analysis was conducted by one author (NLC).

Synthesis

General information about the guidance documents was aggregated and summarized descriptively. Domains from the content analysis were synthesized qualitatively wherein all recommendations were compared and contrasted across each other document.

Expert validation

A stakeholder involved with each of the included guidance documents (“guidance experts”) was contacted via email and invited to comment on the findings of the data abstraction and the characterization of the guidance document in an earlier version of the review manuscript. Guidance experts were also provided with the opportunity to indicate any additional guidance documents that might be relevant to the current review.

Results

A total of eight guidance were evaluated for inclusion in the review (Figure 1). The guidance identified via literature search – a series of Good Research Practice papers developed by ISPOR, the professional society for health economics and outcomes research, was excluded from the current review as it related to developing a measure rather than selecting a measure [27, 28], and only described one measurement property (content validity) [29]. Seven guidance were ultimately included in the review.

Table 2 reports the final guidance included in the review, their stated purpose, and general information regarding each guidance. Three out of the seven sets of guidance were developed by international professional consortia (ISOQOL Minimum Standards[13]; COnsensus-based Standards for the selection of health Measurement INstruments [COSMIN] Initiative documents [30-33]; Medical Outcomes Trust [MOT] Review Criteria [21]); three were published by regulatory agencies in the European Union and United States (European Medicines Agency [EMA] Appendix 2, [19] US Food and Drug Administration [FDA] PRO Guidance, [20] FDA PFDD [Patient-focused drug development] Discussion Document 3 [34]). Of note, the 2018 FDA PFDD discussion document 3 is not a formal guidance; as of June 2020 it is a discussion document only, and will continue to undergo development and change before becoming first a draft guidance and eventually a final FDA guidance for industry. Once finalized, the resulting document will replace the 2009 PRO guidance and serve as the FDA's new guidance for clinical outcome assessments (COAs; including PROMs). EMA Appendix 2 was not developed to provide 'specific recommendation regarding valid instrument selection' but does 'outline broad principles of scientific best practice and...guidance on the value of PRO data in the development of medicinal products for the treatment of cancer.' The seventh guidance document was published by an organization compiling health-related quality of life measures available in Spanish (Spanish Cooperative Investigation Network for Health and Health Service Outcomes Research, Evaluating the Measurement of Patient-Reported Outcomes [Red-IRYSS EMPRO]).[35] The Red-IRYSS EMPRO guidance was developed to convert the MOT Review Criteria into individual items.[36] Therefore the two are nearly identical in their recommendations. Publication dates for the documents ranged from 2002 (MOT Guidance) to 2018 (FDA PFDD and COSMIN Initiative documents).

Two of the seven guidance are formatted in a checklist style, wherein desired characteristics of PROMs are listed and users of the guidance are urged to assess the use of a given PROM based on the criteria provided (ISOQOL Minimum Standards, MOT Review Criteria). Two of the guidance use a rating approach wherein users rate adherence of the measure of interest against specific criteria and thresholds of acceptability (COSMIN Initiative, Red-IRYSS EMPRO).

The remaining guidance, those produced by the FDA and EMA, are narrative in nature and provide general recommendations on how to select PROMs for clinical trials for regulatory submission.

Table 3 includes a brief description of each guidance's recommendation for each measurement property. Three guidance explicitly endorse content validity as the most important measurement property (FDA PRO, FDA PFDD, COSMIN Initiative). Neither ISOQOL, EMA Appendix 2 nor Red-IRYSS EMPRO indicate any measurement property as priority. While ISOQOL Minimum Standards does not indicate the priority of any specific property, conceptual/measurement model was the most highly endorsed instrument attribute for inclusion in the minimum standards during the Delphi process used to establish the standards.

Below we compare the guidance documents recommendations for PROM selection on each of eight core properties of PROMs.

Conceptual and measurement model

All guidance documents discuss the value of a conceptual and measurement model as a criterion for PROM selection. The conceptual model provides a description and framework for the targeted construct(s) to be included in a PRO measure. The measurement model maps the individual items in the PRO measure to the construct. The EMA guidance discusses the role of a conceptual model as a tool to demonstrate how PROs might add value to clinical research. Other documents, such as the ISOQOL Minimum Standards, FDA PRO, FDA PFDD, MOT Review Criteria, and Red-IRYSS EMPRO each describe conceptual/measurement models of the PROM itself, such as how items used in the measure relate to one another. The COSMIN Initiative includes a conceptual model in the rating of PROM development [33].

Reliability

All guidance documents reference the importance of reliability in PROM selection. ISOQOL Minimum Standards, MOT Review Criteria, Red-IRYSS EMPRO, and COSMIN Initiative [33, 37, 38] indicated 0.70 as a threshold for internal consistency and reproducibility/interrater reliability. ISOQOL Minimum Standards note that these values may be lower "if justifiable." Neither FDA PRO, FDA PFDD, nor EMA Appendix 2 set specific thresholds for assessing reliability. While FDA PRO emphasizes the use of test-retest as an important evaluation of reliability, ISOQOL Minimum Standards places less emphasis on test-retest specifically, instead indicating that a variety of methods can be used to assess reliability - including test-retest, internal consistency, and item response theory - but that each approach should be justified.

Content validity

All guidance documents indicate the importance of content validity. COSMIN/COMET, FDA PRO, and FDA PFDD explicitly state that content validity is the most important measurement property to consider when selecting a PROM. To underscore this importance, the COSMIN initiative published a stand-alone paper explaining methods for assessing content validity [32], highlighting the importance of a PROM's "track record" as a tool to gage a measure's content validity. The EMA guidance took a similar approach in Appendix 2. The MOT Review Criteria highlight the importance of reviewing the validity of PROMs, including content, construct, and criterion validity, based on four attributes: clarity, comprehensiveness, relevance, and redundancy of items and scales. These are merged into one item assessing content validity in the Red-IRYSS EMPRO, which was created based on the MOT Review Criteria. The ISOQOL Minimum Standards, FDA PRO guidance, and FDA PFDD document each highlight that qualitative data can serve as supporting information in establishing content validity. The ISOQOL Minimum Standards and FDA PFDD guidance include recall period as part of the content validity assessment.

Construct validity

All guidance documents with the exception of EMA Appendix 2 discuss construct validity as a criterion in PROM selection. The ISOQOL Minimum Standards, FDA PRO, FDA PFDD, and MOT Review Criteria all describe construct validity as being determined by the ability of a measure to detect differences among known patient groups consistent with a priori hypotheses. FDA PRO and FDA PFDD also indicate that not only should the measure overall detect changes between groups but so too should individual items or domains of the measure. In the COSMIN Initiative construct validity is included under the taxonomy of structural validity, hypotheses testing, cross-cultural validity/measurement invariance [33, 37]. It is also described that construct validity is a conceptual consideration for the selection of a PROM [38]. The Red-IRYSS EMPRO uses 3 items to assess construct validity based on the MOT Review Criteria, including: that the methods to assess construct (& criterion) validity are described and appropriate, that the sample used to test construct (& criterion) validity is described and appropriate, and that the hypotheses are described and results are consistent with them.

Responsiveness

All guidance documents describe responsiveness as the ability of a PROM to detect true changes in patients' health status, although they do so with somewhat different words and emphasis. FDA PRO, FDA PFDD, and EMA Appendix 2 focus more attention than other guidance documents on the importance of working with patients to define a 'meaningful change.' MOT Review Criteria, ISOQOL Minimum Standards, and Red-IRYSS EMPRO focus greater attention on describing responsiveness as an aspect of validity, highlighting the importance of hypothesis

testing as a means of demonstrating responsiveness. While ISOQOL Minimum Standards and MOT Review Criteria allude to responsiveness in terms of change at a population level, FDA PRO indicates that these changes can be either within-person or at the population level. The COSMIN Initiative guidance provide a specific threshold for having established responsiveness, which is defined as requiring at least 75% of changes in score to be in accordance with the hypothesis.[37, 38]. COSMIN also supports assessing responsiveness as a potential risk of bias in systematic reviews[33].

Interpretability of scores

All guidance documents indicate that interpretability is an important property. ISOQOL Minimum Standards describes interpretability in terms of having a meaningful understanding of scores, such as what high scores versus low scores represent, and what degree of change is small versus large. The COSMIN Initiative includes interpretability as a feasibility concern in the selection of outcome measures, but does not include it as a component in any of its rating systems, given that interpretability is not a measurement property [32, 33, 37, 38]. EMA Appendix 2 indicates that the meaning of any differences in PRO scores should be easily understood. FDA PRO and FDA PFDD describe interpretability of scores as the ability to describe scores and their changes within individual patients. FDA PRO, along with MOT Review Criteria and Red-IRYSS EMPRO, also describe the need for anchor-based analyses to facilitate interpretation, where anchors are measures external to the PRO scale, that are themselves well understood by clinicians and other users, and are correlated with PROM scores conceptually and empirically.

Translation

All guidance documents discuss the importance of translating PRO measures and specifically ensuring cultural equivalence of measures across linguistic/cultural groups. The ISOQOL Minimum Standards recommend that methods for translating a measure be well-documented, and that all translated measures be evaluated through at least qualitative methods such as cognitive testing. ISOQOL as well as the COSMIN Initiative [38], FDA PRO, FDA PFDD, MOT Review Criteria, and subsequently Red-IRYSS EMPRO describe the importance of demonstrating that PRO measurement properties are statistically comparable across language groups, which can be done using psychometric methods such as differential item functioning. The COSMIN Initiative purposefully omitted translation from their risk of bias assessment checklist, however, indicating that it was not a measurement property [33]. EMA Appendix 2 does not provide specific recommended tests or methods for assessing cultural equivalence but does highlight its importance.

Burden

All guidance documents include some reflection on instrument burden. The COSMIN Initiative guidance includes burden as a feasibility concern but does not include it as a component of its rating systems [32, 33, 37, 38]. There is a greater emphasis on respondent burden than investigator / administrative burden in FDA PRO and EMA Appendix 2 as compared to ISOQOL Minimum Standards. Respondent burden specifically is also one of the 8 key attributes of the MOT Review Criteria and is included in Red-IRYSS EMPRO. EMA Appendix 2 provides among the most specific recommendations regarding respondent burden, suggesting that baseline PRO completion take respondents no more than 20 minutes, with subsequent assessments limited to 10-15 minutes.

Additional considerations for PRO selection

In addition to the eight core measure properties posed by the ISOQOL Minimum Standards, several other considerations for the selection of a PROM were mentioned in other guidance documents. First, alternative modes of administration (e.g. pen-and-paper, electronic, self-report, interviewer, proxy) is discussed as a consideration in the EMA Appendix 2, FDA PRO, FDA PFDD, Red-IRYSS EMPRO, and MOT Review Criteria. Second, FDA PRO, FDA PFDD, and EMA Appendix 2 all discuss the appropriateness and validation of PROs for special populations (e.g., pediatrics, rare diseases) as a selection criterion. Third, FDA PFDD and EMA Appendix 2 also contextualize PROs within the broader field of clinical outcome assessments (COAs). This conceptualization may help researchers to identify the appropriate type of outcome assessment, be it a PRO or functional outcome, clinician-reported outcome, etc., to fit their research question. Fourth, the COSMIN Initiative considers the quality of evidence used to endorse measurement properties for PROMs in both the content validity and selecting outcome measures guidance documents[37, 38]. Fifth, both FDA PFDD and FDA PRO include an assessment of the context of use of PROMs, e.g. the specific population and sample. Sixth, the Red-IRYSS EMPRO guidance also includes an item allowing the instrument rater to provide a subjective, global assessment of the instrument's appropriateness.

Expert validation

Feedback from guidance developers fell into three overarching suggestions. The first suggestion was to provide more details regarding the search strategy/inclusion. To address this concern we added specificity to the description of the search strategy in the Methods section and also included Figure 1 to walk readers through review process. The second suggestion was to better describe the nuances between all of the tools. We address this concern by describing all guidance documents' recommendations for all measurement properties, whereas previously we had only included a description of how tools differed from the ISOQOL Minimum Standards. The third suggestion was to, in some cases, modify how to we presented and characterized

guidance documents and their inclusion of the various measurement properties. In these cases we reviewed tool developers comments, re-reviewed the original guidance to resolve any discrepancies between our original interpretation and the expert's feedback, and made changes as appropriate.

Discussion

We reviewed existing documents that provide recommendations and/or guidance regarding the selection of PROMs for use in clinical trials and found general consistency across the seven guidance documents. A snap-shot of this consistency is found in table 4, which denotes whether the core domains are included in the guidance documents. This convergence of recommendations points to the growing unity of the field. There is greater regulation for the use of PROMs in clinical trials – both top-down from regulatory agencies internationally who are specifying standards for PROMs in labeling claims [39] and bottom-up from patient groups and PRO career-scientists and psychometricians who have fought for their inclusion in decision-making and improved their methodological rigor.[40]

Nuances in the guidance emerge when considering their publishing organizations and format. While scientific consortia produced all the checklist-style and rating-style guidance, regulatory agencies produced narrative-style guidance. These differences are also aligned with differences in theoretical orientation, with some (predominantly those produced by scientific consortia) relying heavily on psychometrics as the determinants of an appropriate PROM whereas others (predominantly those produced by regulatory agencies) considering psychometric criteria but also highlighting the important of PROMs that are relevant to the lived experiences of patients. In general, there appears to be growing interest from regulators and scientists to engage patients and their advocates in research developing and selecting PROMs.

Although the guidance documents included in the current review were generally consistent, they were not identical in their recommendations. The documents vary in the level of detail they provide. For instance, in looking at ISOQOL Minimum Standards, a PROM with a Pearson correlation of 0.70 appears sufficient. However, EMPRO offers greater specification, indicating that researchers should also account for time interval between the administrations and the stability of patients. They also vary in the context for which they provide recommendations, and when selecting a guidance, it would be important to consider the context in which the guidance will be applied. For example, FDA PROs/FDA PFDD guidance would likely be most appropriate to inform the selection of a PROM that will be included in a regulatory application, whereas COSMIN might be more relevant if conducting a systematic review of PROMs.

The identification of potential guidance documents for inclusion in the current targeted review benefited from international expert referral including from groups conducting clinical trials, regulatory agencies, medical editors who oversee the publication of clinical trials, PRO scientists, and PRO guidance developers themselves – nearly 40 individuals in total. Although the current review features key guidance documents used to inform the selection of PROMs for clinical trials it may not have captured all guidance documents. Future research may consider using systematic review methods to ensure that all guidance documents are included. Such a review might even need to include environmental scanning/scoping review methods to develop a search strategy that identifies guidance published in the grey literature. We opted to include the FDA PFDD discussion document on methods to identify what matters to patients, which is not formal guidance, but a document used to inform the development of a forthcoming methodological guidance for PROMs and other COAs. The points made in this document have not been enacted, but do provide some insight into the FDA’s “current thinking” on the topic.

As the PRO landscape continues to evolve, so too have the guidance documents used to assess them. Many of the older guidance documents have evolved into newer guidance, e.g. the FDA’s PROs guidance from 2009 helped inform the FDA PFDD discussion document, and the EMPRO checklist is based on the MOT review criteria. This evolution is particularly noteworthy within the COSMIN initiative which first published a checklist on evaluating the methodological quality of studies evaluating PRO measurement properties in 2010 [41], updated it in 2018 [33], and has developed specific documents regarding content validity [32] selecting outcomes for core outcomes sets [38], and guidelines for systematic reviews of PROs [31]

When evaluating the quality of research overall, it is important to differentiate the quality of a study using PROMs from the quality of the PROM itself. PROMs themselves may be psychometrically validated and robust, but this does not mean they are appropriate for use in all research settings. Sub-optimal PRO data collection or analysis of PRO data impair the usefulness of PRO data irrespective of the quality of instrument. In the current review we also observed a trend for newer guidance to consider contextual factors in the appropriateness of a PROM. For instance, the FDA’s Draft PFDD discussion document advises using instruments that are fit-for-purpose and appropriate for the context of use. Specific criteria on how to demonstrate that an instrument is fit-for-purpose have not been established. It is also important to note that the quality of studies used to develop and validate PROMs is also an important consideration and the COSMIN Initiative has provided tools to evaluate the psychometric study quality [33, 42]. Appropriate PROM selection is only one aspect of conducting a high-quality study. A detailed protocol, appropriate analytic approach, and clear reporting are also critical for a high-quality PRO study and are addressed by other core guidance documents from the PROTEUS Consortium [14-18].

It is notable that only one guidance document eligible for inclusion in this review was cancer-specific rather than disease-agnostic in recommendations. The EMA guidance on PROs is oncology specific, which is notable given that this is the only formal guidance regarding the use of PROs within the EMA aside from a 2005 brief reflection paper on health-related quality of life measures [43]. The field of oncology was an early-adopter of PROMs in clinical trials and has contributed to the increasing rigor and acceptability of symptom, satisfaction, and quality of life endpoints [7, 44]. Despite being oncology-specific the EMA guidance still largely overlaps with the general guidance in the current review. PROs are increasingly being included and embedded in other EMA guidance documents that are disease specific, e.g. in Crohn's disease [45]. There is also a push for EMA to coordinate its approach to evaluating PROs [46]. Moving forward it is unclear if EMA will continue to pursue condition-specific guidance on the use of PROs or if it will embrace disease-agnostic guidance, as is done by US FDA.

As a group tasked with optimizing the use of PROs in clinical trials, the PROTEUS Consortium is most interested in encouraging end-users to be thoughtful and systematic in their approach to using PROMs, regardless of the guidance used. PROTEUS has highlighted the ISOQOL Minimum Standards within its own materials given the existing popularity and relative parsimony of the Minimum Standards. However, users aiming for regulatory submissions or other applications should review the relevant guidance from the appropriate governing bodies. Regardless, the high level of consistency of recommendations observed across guidance indicates that application of any of the guidance, in addition to the Minimum Standards, will likely improve the quality and interpretation of PRO endpoints in clinical trials, thereby providing valuable insights from the patient perspective with potential to inform regulatory approval and clinical practice.

Declarations:

Ethics approval: Not applicable

Availability of data and material: All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests: MC reports grants from Macmillan Cancer Support, Innovate UK, Health Data Research UK, UCB Pharma the NIHR, NIHR Birmingham Biomedical Research Centre and NIHR Surgical Reconstruction and Microbiology Research Centre (SRMRC) at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, and personal fees from Astellas, Takeda, Glaukos and Merck outside the submitted work.

Funding: Funded by a Patient-Centered Outcomes Research Institute (PCORI) Eugene Washington Engagement Award (12565-JHU). Dr. Snyder is a member of the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins (NCI P30 CA006973).

Acknowledgements: We would like to thank the many guidance authors who responded to our requests for feedback.

References

1. Acquadro, C., et al., *Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001*. Value Health, 2003. **6**(5): p. 522-31.
2. Au, H.J., et al., *Added value of health-related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG*. Expert Rev Pharmacoecon Outcomes Res, 2010. **10**(2): p. 119-28.
3. Till, J.E., et al., *Research on health-related quality of life: dissemination into practical applications*. Qual Life Res, 1994. **3**(4): p. 279-83.
4. Lipscomb, J., C.C. Gotay, and C. Snyder, *Outcomes assessment in cancer: measures, methods and applications*. 2004: Cambridge University Press.
5. Brundage, M., et al., *A knowledge translation challenge: clinical use of quality of life data from cancer clinical trials*. Qual Life Res, 2011. **20**(7): p. 979-85.
6. Bezjak, A., et al., *Oncologists' use of quality of life information: results of a survey of Eastern Cooperative Oncology Group physicians*. Qual Life Res, 2001. **10**(1): p. 1-13.
7. Basch, E., et al., *Patient-reported outcomes in cancer drug development and us regulatory review: Perspectives from industry, the food and drug administration, and the patient*. JAMA Oncology, 2015. **1**(3): p. 375-379.
8. Basch, E., et al., *Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment*. Jama, 2017. **318**(2): p. 197-198.
9. Kyte, D., et al., *Systematic evaluation of Patient-Reported Outcome protocol content and reporting in cancer trials*. JNCI: Journal of the National Cancer Institute, 2019.
10. Ahmed, K., et al., *Systematic evaluation of patient-reported outcome (PRO) protocol content and reporting in UK cancer clinical trials: the EPiC study protocol*. BMJ open, 2016. **6**(9): p. e012863.
11. Mercieca-Bebber, R., et al., *The patient-reported outcome content of international ovarian cancer randomised controlled trial protocols*. Quality of Life Research, 2016. **25**(10): p. 2457-2465.
12. *The PROTEUS Consortium*. 2020; Available from: www.TheProteusConsortium.org.
13. Reeve, B.B., et al., *ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research*. Qual Life Res, 2013. **22**(8): p. 1889-905.
14. Calvert, M., et al., *Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension*. Jama, 2018. **319**(5): p. 483-494.
15. Coens, C., et al., *International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium*. The Lancet Oncology, 2020. **21**(2): p. e83-e96.
16. Calvert, M., et al., *Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension*. Jama, 2013. **309**(8): p. 814-822.

17. Snyder, C., et al., *Making a picture worth a thousand numbers: recommendations for graphically displaying patient-reported outcomes data*. *Quality of Life Research*, 2019. **28**(2): p. 345-356.
18. Wu, A.W., et al., *Clinician's checklist for reading and using an article about patient-reported outcomes*. *Mayo Clinic Proceedings*, 2014. **89**(5): p. 653-661.
19. EMA, *Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. The use of patient-reported outcome (PRO) measures in oncology studies*. 2016.
20. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009, US Food and Drug Administration.
21. Lohr, K.N., *Assessing health status and quality-of-life instruments: attributes and review criteria*. *Quality of Life Research*, 2002. **11**(3): p. 193-205.
22. Eichenfield, L.F., et al., *Current guidelines for the evaluation and management of atopic dermatitis: A comparison of the Joint Task Force Practice Parameter and American Academy of Dermatology guidelines*. *J Allergy Clin Immunol*, 2017. **139**(4s): p. S49-s57.
23. Makady, A., et al., *Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies*. *Value Health*, 2017. **20**(4): p. 520-532.
24. Grant, M.J. and A. Booth, *A typology of reviews: an analysis of 14 review types and associated methodologies*. *Health Information & Libraries Journal*, 2009. **26**(2): p. 91-108.
25. Hsieh, H.F. and S.E. Shannon, *Three approaches to qualitative content analysis*. *Qual Health Res*, 2005. **15**(9): p. 1277-88.
26. Neuendorf, K.A., *The content analysis guidebook*. 2016: Sage.
27. Patrick, D.L., et al., *Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding*. *Value Health*, 2011. **14**(8): p. 978-88.
28. Patrick, D.L., et al., *Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument*. *Value Health*, 2011. **14**(8): p. 967-77.
29. Rothman, M., et al., *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 Health*, 2009. **12**(8): p. 1075-83.
30. *COSMIN Initiative*. 2020 [11 June 2020]; Available from: <https://www.cosmin.nl/finding-right-tool/select-best-measurement-instrument/>.
31. Prinsen, C.A., et al., *COSMIN guideline for systematic reviews of patient-reported outcome measures*. *Quality of Life Research*, 2018. **27**(5): p. 1147-1157.
32. Terwee, C.B., et al., *COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study*. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 2018. **27**(5): p. 1159-1170.
33. Mokkink, L.B., et al., *COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures*. *Qual Life Res*, 2018. **27**(5): p. 1171-1179.

34. *Patient-Focused Drug Development: Collecting Comprehensive and Representative Input, in Draft Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders.* . 2018, US Department of Health and Human Services.
35. Valderas, J.M., et al., *Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures.* Value in Health, 2008. **11**(4): p. 700-708.
36. *Evaluating Measures of Patient-Reported Outcomes (EMPRO).* 16 Dec 2019]; Available from:
<https://medicine.exeter.ac.uk/research/healthresearch/healthservicesandpolicy/projects/proms/theemprotool/>.
37. Prinsen, C.A.C., et al., *COSMIN guideline for systematic reviews of patient-reported outcome measures.* Qual Life Res, 2018. **27**(5): p. 1147-1157.
38. Prinsen, C.A.C., et al., *How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline.* Trials, 2016. **17**(1): p. 449.
39. Kluetz, P.G., D.J. O'Connor, and K. Soltys, *Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada.* Lancet Oncol, 2018. **19**(5): p. e267-e274.
40. Mercieca-Bebber, R., et al., *The importance of patient-reported outcomes in clinical trials and strategies for future optimization.* Patient Relat Outcome Meas, 2018. **9**: p. 353-367.
41. Mokkink, L.B., et al., *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, 2010. **19**(4): p. 539-549.
42. *COSMIN Study Design checklist for Patient-reported outcome measurement instruments.* 2019; Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf#.
43. *Reflection paper on the regulatory guidance for the use of health-related quality of life measures in the evaluation of medicinal products.* 2005, European Medicines Agency.
44. Aaronson, N.K., et al., *Quality of life research in oncology. Past achievements and future priorities.* Cancer, 1991. **67**(S3): p. 839-843.
45. *Guideline on the development of new medicinal products for the treatment of Crohn's Disease.* 2018, European Medicines Agency.
46. *EMA Regulatory Science to 2025.* 2018, European Medicines Agency.

Table 1. Measure properties and their definitions in relation to PRO measures*

Domain	Definition
Conceptual and measurement model	The conceptual model provides a description and framework for the targeted construct(s) to be included in a PRO measure. The measurement model maps the individual items in the PRO measure to the construct.
Reliability	The degree to which measure is free from measurement error.
Content validity	The extent to which the PRO measure includes the most relevant and important aspects of a concept in the context of a given measurement application.
Construct validity	The degree to which scores on the PRO measure relate to other measures (e.g., patient-reported or clinical indicators) in a manner that is consistent with theoretically derived a priori hypotheses concerning the concepts that are being measured.
Responsiveness	The extent to which a PRO measure detects changes in the construct of interest over time.
Interpretability of scores	The degree to which one can assign easily understood meaning to a PRO measure's scores.
Translation of the PRO measure	The evidence of equivalence of measurement properties for a culturally or linguistically translated version of a PRO instrument.
Burden	The time, effort, and other demands placed on respondents completing the instrument, or investigators/administrators administering the instrument.

*Table adapted from Reeve et al. [13]

Table 2. General information about PROM selection guidance documents

Organization	Guidance	Year	Country	Format	Purpose	Characterization	Priority properties
ISOQOL	Minimum Standards	2013	Int	Checklist	Identify minimum standards for the design and selection of PROs for PCOR and CER.	Presence/absence of evidence for each property	None indicated – conceptual /measurement model most highly endorsed in Delphi process
COSMIN	COSMIN/COMET Guidance for selecting outcome measures	2016		Rating	Develop a guideline on how to select outcome measurement instruments for outcomes included in a core outcome set.	9 properties assessed by one criterion each. Properties rated as acceptable (+), not indicated (?), not acceptable (-)	Content validity
	Mokkink et al. Risk of bias checklist	2018		Rating	Develop guidance on how to assess risk of bias in systematic reviews of PROMs.	10 properties assessed by multiple criteria each. Rated as very good, adequate, doubtful, inadequate.	
	Prinsen et al. Guideline for systematic reviews of PROMs	2018	Int	Rating	Develop a guideline for conducting systematic reviews of PROMs including search strategy, quality evaluation, and reporting.	8 properties assessed by one criterion each. Properties rated as acceptable (+), not indicated (?), not acceptable (-). Overall evidence quality is graded.	
	Terwee et al. Methods for evaluating content validity	2018		Rating	Develop consensus-based guidelines for evaluating the content validity of PROMs.	1 property (content validity) assessed by 10 items. Properties rated as acceptable (+), not indicated (?), not acceptable (-)	
FDA	PRO Guidance for Industry	2009	USA	Regulatory guidance	Document how FDA evaluates existing, modified, or newly created PROs for regulatory decision making.	Narrative; lists properties that FDA reviews	Content validity
FDA	PFDD Discussion Document 3*	2018	USA	Document to inform regulatory guidance	Foster discussion on communicating FDA’s current thinking regarding the selection, development, or modification of clinical outcome assessments.	Narrative	Content validity
EMA	Appendix 2: PROs in oncology trials	2016	EU	Regulatory guidance	Outline broad principles of scientific best practice and provide guidance on the value of PRO data in the development of medicinal products for cancer.	Narrative	None indicated
MOT	Review Criteria	2002	Int	Checklist	Offer a conceptualization of eight attributes of health status/QoL instruments and the criteria by which attributes of those instruments should be reviewed	Presence/absence of evidence for each attribute	None indicated
Red-IRYSS	EMPRO‡	2008	ESP	Rating	Develop an evaluation guidance for the standardized assessment of PROs to assist in instrument selection.	8 properties assessed using 39 items total. Items rated 1 (lowest) to 4 (highest). Summative scoring.	None indicated – Domains are equally weighted

Organizations: ISOQOL = International Society for Quality of Life; Red-IRYSS = Spanish Cooperative Investigation Network for Health and Health Service Outcomes Research; COSMIN = Consensus-based Standards for the selection of health Measurement Instruments; COMET = Core Outcome Measures in Effectiveness Trials Initiative; US FDA = United States Food and Drug Administration; EMA = European Medicines Agency; MOT = Medical Outcomes Trust

Guidance: EMPRO = Evaluating the Measurement of Patient-Reported Outcomes; PRO = Patient reported outcomes; PFDD= Patient-focused drug development; *Countries:* Int = international; ESP = Spain; USA = United States of America; EU = European Union

* Of note, the 2018 FDA PFDD 3 document is not a formal guidance but rather a discussion document used to help develop a draft guidance for industry and subsequently a final guidance. Recommendations provided by this document have not been finalized or enacted in practice.

‡ Note that EMPRO is directly based on criteria presented in MOT Review Criteria

Table 4. PROM selection criteria included in guidance documents at-a-glance

Domain	ISOQOL Minimum Standards	COSMIN Initiative	EMA Appendix 2	FDA PRO	FDA PFDD*	MOT Review Criteria	Red-IRYSS EMPRO‡
Conceptual & measurement model	✓	✓	✓	✓	✓	✓	✓
Reliability	✓	✓	✓	✓	✓	✓	✓
Content validity	✓	✓	✓	✓	✓	✓	✓
Construct validity	✓	✓	X	✓	✓	✓	✓
Responsiveness	✓	✓	✓	✓	✓	✓	✓
Interpretability of scores	✓	+/-	✓	✓	✓	✓	✓
Translation	✓	✓	✓	✓	✓	✓	✓
Burden	✓	+/-	✓	✓	✓	✓	✓
Additional domains		<ul style="list-style-type: none"> • Structural validity • Quality assessment 	<ul style="list-style-type: none"> • Alternative mode of admin. • COAs • Special patient populations 	<ul style="list-style-type: none"> • Alternative mode of admin. • Special patient populations • Context of use 	<ul style="list-style-type: none"> • COAs • Context of use, fit-for-purpose • Special patient populations • Alternate modes of admin 	<ul style="list-style-type: none"> • Alternative mode of admin 	<ul style="list-style-type: none"> • Alternative mode of admin. • Global assessment of instrument by rater

✓ Denotes domain is discussed in guidance

+/- Denotes domain is discussed but not included as core component in rating system (only applicable to rating-style documents, i.e. COSMIN Initiative and EMPRO)

X Denotes domain is not discussed in guidance document

COA = Clinical outcome assessments

* Of note, the 2018 FDA PFDD 3 document is not yet a formal guidance but rather a discussion document that is used to help develop a draft guidance for industry and subsequently a final guidance. Recommendations provided by this document have not been finalized or enacted in practice.

‡ Note that EMPRO is directly based on criteria presented in MOT Review Criteria

