

Comment on “Checklist to operationalize measurement characteristics of patient-reported outcome measures”

Terwee CB, de Vet HCW, Prinsen CAC, Mokkink LB.

VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam, the Netherlands

As the developers of the COSMIN checklist, we were obviously very interested in the recent article by Francis et al. published in the journal *Systematic Reviews*, on the development of a new checklist to operationalize measurement characteristics of patient-reported outcome measures.¹ The authors aimed to develop a set of simplified criteria, in the form of a checklist, that, with minimal instruction, researchers and clinicians with limited methodological expertise with regard to PRO measures, can use in assessing the measurement properties and usefulness of PRO measures for particular circumstances. The main reason for developing this checklist was that a more rigorous checklist, such as the COSMIN checklist², is complex to use for researchers with limited PRO measure methodological expertise.

We agree with the authors that the COSMIN checklist is complex and requires at least some expertise in psychometrics or clinimetrics as basis for PRO measure development and evaluation. This may be unavoidable because the methodological principles of PRO measure development and evaluation are complex. Although we understand the temptation to try to simplify the criteria, we think a simplified method will increase the risk of biased ratings of the quality of PRO measures. The checklist presented by Francis et al. looks simple and clear, but there are several important methodological shortcomings, which may introduce bias in the ratings of the PRO measures:

First, when assessing the quality of PRO measures, it is important to take the methodological quality of the studies on the measurement properties into account. When the quality of the study is not adequate, the results (i.e. the reliability coefficient, or the correlation of a PRO measure with another instrument) may be biased. The COSMIN checklist is a tool for assessing the methodological quality of studies on measurement properties.²⁻⁴ It is recommended to first rate the quality of the studies on measurement properties before rating the results of the studies. This corresponds to Cochrane methodology for systematic reviews, where the methodological quality of the trials is distinguished from the effect sizes found in the trials. In the checklist presented by Francis et al. some questions refer to methodological aspects of the study, or even only if certain things have been done, while other questions refer to the quality of the PRO measure itself. This should be separated.

Note that the COSMIN checklist is a tool for assessing the *quality of studies* on measurement properties. It is not a tool for assessing the *quality of the PRO measure* itself. For assessing the quality of a PRO measure, the COSMIN group developed a systematic approach, in which the PRO measure is rated based on the number of studies on its measurement properties, the quality of the available studies, the total sample size of the studies, the results of the studies, and the consistency of the results across studies. This methodology was based on general Cochrane methodology for systematic reviews and is described in a protocol for systematic reviews of measurement instruments (see www.cosmin.nl).

Second, due to its simplicity, the criteria presented by Francis et al. are not detailed enough to provide a transparent and systematic rating of the quality of a PRO measure. For example, the criteria for content validity only consider *whether* certain things have been done (involvement of patients and content experts) or reported (methodology for deriving items) but not *how* they were done (e.g. whether adequate qualitative methods were used). Furthermore, no criteria are provided for what constitutes good content validity. Also, no criteria are provided for what constitutes good dimensionality (i.e. what are good results of a factor analyses or IRT analyses) or good responsiveness of a PRO measure. Leaving it up to researchers or clinicians with limited methodological expertise to decide upon what constitutes good measurement properties introduces a serious risk of bias.

The authors evaluated the reliability of the new checklist by comparing the ratings of six PRO measures by six clinicians without expertise in measurement theory with the ratings of two more experienced practitioners. It is encouraging to see a quite acceptable Kappa value after very limited training. However, we do not know how the scores of the PRO measures (not provided in the paper) would compare to a more rigorous approach, thus the validity of the checklist is unclear. We greatly appreciate the attention given by the authors to the quality of PRO measures. The authors did a great job in identifying and explaining the most important attributes of PRO measures. We especially appreciate the emphasis on the conceptual model and patient input in the development of PRO measures, which is of great importance in PRO measure development. We also appreciate the attention for the importance of systematic reviews of measurement instruments. Such systematic reviews are an important tool for instrument selection. The number of systematic reviews of measurement instruments is increasing (<http://database.cosmin.nl/>) to more than 100 reviews per year being published in the most recent years. However, the quality varies widely and there is clear room for improvement.^{5,6}

However, considering the limitations of this new checklist and the risk of biased ratings, we do not recommend using this checklist in systematic reviews of PRO measures, even though its simplicity may be considered attractive. The methodological principles of PRO measure development and evaluation are complex and therefore clinimetrics/psychometrics is considered a specific methodological discipline. It may not be fair to expect that researchers and clinicians working in other disciplines are able to make evidence-based recommendations about the quality of PRO measures without some form of clinimetric training or help from clinimetric/psychometric experts. We think that trying to simplify the criteria is not the best way to go. Instead, we plea for more collaboration with clinimetric/psychometric experts, in order to perform high quality systematic reviews of PRO measures using rigorous methods.

References

1. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Syst Rev* 2016;5(1):129.
2. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19(4):539-49.
3. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
4. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21(4):651-7.
5. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res* 2009;18(3):313-33.
6. Terwee CB, Prinsen CA, Ricci Garotti MG, Suman A, de Vet HC, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res* 2016;25(4):767-79.