VALUE IN HEALTH 17 (2014) 491–492

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

EDITORIAL

Review of the Task Force Report on PRO Data Collection in Clinical Trials Using Mixed Modes

Congratulations to the ISPOR PRO Mixed Modes Good Research Practices Task Force [1] for tackling a challenging question that many patient-reported outcome (PRO) researchers have pondered either publicly or privately as they review PRO results tables based on data collected through a combination of data collection methods: "Would these results and the study conclusions be different if the same mode had been employed?" The report's general recommendation to "Just say no!" to mixing modes within a single study (or set of studies that may be pooled) is good advice and, if followed, will save many headaches and resources. If ignored and data are combined without consideration of the guidance offered in this report, post hoc evaluations may be necessary to backfill evidence to support the fact that the PRO data and conclusions are not compromised, especially to support medical product labeling.

The task force report is well written and provides an enormous amount of information about the current technologies available for PRO data collection. There is a deep dive into the issues of mixed modes from a technology, formatting, and qualitative perspective. The recommendations for a "faithful migration" are detailed in these areas and reflect the vast expertise of the task force members. Guidance is provided for the types of qualitative and quantitative data important for the evaluation of measurement equivalence before mixing modes in clinical trials. Multiple viable designs to gather the necessary data are described for both qualitative and quantitative studies. For evaluations based on qualitative data, guidelines are offered about what results deserve increased concern before mixing modes and recommendations for next steps to resolve these concerns. For evaluations based on quantitative data, the report is light, mentioning only one (the intraclass correlation coefficient [ICC]) of many analyses that could be conducted to assess measurement equivalence. The omission of greater detail for the quantitative evaluations is disappointing, especially given the amount of information provided about statistical evaluations within the previous ePRO task force report by Coons et al. [2], which focused on the evidence needed to support measurement equivalence when migrating from paper to an electronic platform.

Review of the Coons et al. [2] report provides additional and essential details. A randomized crossover design with patients completing both modes within an appropriate time period is the best design to provide data for the ICC evaluation. In addition, an ICC should reach at least 0.70 to make group comparisons and 0.85 to 0.95 to support patient-level decisions [3,4]. Furthermore, the previous task force pointed out the importance of considering the measure's test-retest reliability in its original mode when evaluating the ICC in order not to hold the alternate modes' evaluation to a higher standard than the original mode's test-retest stability. Finally, if the planned quantitative study design does not support the computation of an ICC (e.g., in the context of a randomized parallel-group design), then the comparison of the internal consistency [5] (e.g., Cronbach's alpha) for each mode should be considered, along with methods such as item response theory or differential item functioning.

Given the high-stakes nature of the present task force's objective to "address the use and mixing of data collection modes within and between trials where the PRO endpoints are intended to be used to support medical product labeling," increased focus on the effect of combining scores is necessary. Throughout the report, recommendations are provided with measurement equivalence as the goal. As defined by the task force, however, measurement equivalence is simply "measuring the same thing"—a lower hurdle that may more adequately be considered "measurement comparability." To support "measurement equivalence" for potential PRO labeling, the recommendations should include evaluations that patients identified as "responders" on one mode are identified as "responders" on another mode. One potential approach is to compare cumulative distribution functions by mode using confidence intervals. A more robust approach involves evaluating the measurement precision across the entire score distributions using techniques such as Rasch measurement theory or item response theory.

What if the results from these evaluations do not support score equivalence? The recommendations in the report include an example of the addition of mode in the planned statistical model. But what if the planned evaluation is a responder analysis? In this case, it may be necessary to consider equating scores across modes and then identifying responders on the basis of equated scores [6]. Although this approach would increase the complexity of the scoring, the benefits may be worth the effort, especially in the growing area of rare diseases for which pooling data is often a necessity.

Overall, the task force report provides practical guidance to avoid most pitfalls when designing studies involving multiple modes of data collection. In addition to this sound advice, however, additional work is needed to provide guidance when there is evidence that the PRO scores are comparable but not exactly equal. There are analytical techniques that offer potential solutions beyond those outlined in the current task force report (e.g., differential item functioning and equating). As a next step, task force members and PRO researchers in general should pursue a reasonable and scientifically grounded process for both

Source of financial Support: The authors have no other financial relationships to disclose.

the development of a thoughtful mixed mode approach and evaluations of the resulting PRO scores (and conclusions), so that together we can put forth the best evidence to support high-stakes PRO end points.

Lori McLeod, PhD
RTI Health Solutions, Research Triangle Park, NC, USA.

1098-3015/$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2014.06.003

REFERENCES

[1] Eremenco S, Coons SJ, Patty J, et al. PRO data collection in clinical trials using mixed modes: report of the ISPOR PRO Mixed Modes Good Research Practices Task Force. Value Health 2014;17:501–16.
[2] Coons SJ, Gwaltney CJ, Hays RD, et al. 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 Health 2009;12:419–29.
[3] Nunnally JC, Bernstein IH. Psychometric Theory. (3rd ed). New York: McGraw-Hill, 1994.
[4] Weiner EA, Stewart BJ. Assessing Individuals. Boston: Little Brown, 1984.
[5] Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:294–334.
[6] Lord FM. Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Erlbaum, 1980.