Overview of psychometrics to support the development and use of patient-reported outcome measures in migraine clinical trials

Carrie R. Houts[1], James S. McGinley[1], Tracy K. Nishida[1], Richard B. Lipton[2], & R.J. Wirth[1]

[1] Vector Psychometric Group, Chapel Hill, North Carolina, USA, [2] Albert Einstein College of Medicine, Bronx, New York, USA

Summary

- A variety of psychometric/statistical evidence is needed to support the development and use of patient-reported outcome measures for use in migraine clinical trials

Background

- Patient-centeredness has gained importance in migraine clinical trials
- Patient-reported outcome (PRO) measures have become increasingly critical to demonstrate that acute and preventive migraine treatments meaningfully impact outcomes that are important to patients
- Demonstrating that a PRO measure is fit-for-purpose (valid) in a given context of use (i.e., adult migraine clinical trials) is necessary for claims regarding treatment
  - e.g., Drug X improves physical functioning; Device Y reduces impact on everyday activities to be approved by regulatory bodies.
- Using current FDA guidance and personal experience in supporting the use of PROs, we provide an overview of the statistical information that is typically necessary to rigorously support a PRO as fit-for-purpose in a given context of use.

Methods

- We provide a high-level overview analyses for developing migraine related PRO measures once items have been generated from a literature review and qualitative work.
- These steps include data handling, item-level descriptives, dimensionality analyses, and validity evidence, and determining meaningful score different (MSD) / meaningful within-person change (WPMC) thresholds
- A running hypothetical, 8-item physical function (PF) PRO measure consistent with on-going work of the Migraine Clinical Outcomes Assessment System (MiCOAS) project is used, in which higher scores indicate better PF

Results

Item-Level Descriptives

- Frequency tables for the observed responses of each candidate PF item
- Item-level summaries are examined for floor effects, ceiling effects, and missing data to identify items that are performing sub-optimally
  - If many observed responses occur in the least severe [floor effect] or most severe [ceiling effect] response category, this may indicate that item is not informative/well-calibrated to the sample
  - Collapsing over response options may be necessary to avoid sparseness

Dimensionality Assessment

- In our hypothetical 8-item PF PRO measure example, a single underlying concept of interest/latent variable (that is, PF) is assumed to exist (Figure 1)

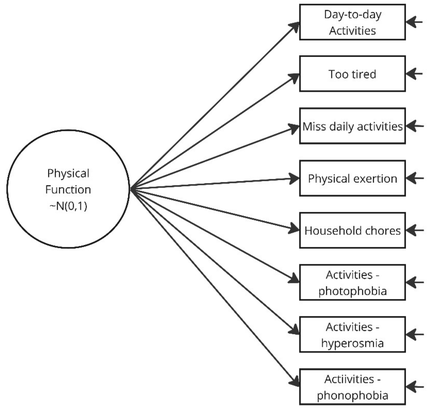
Item Factor Analysis (IFA)

- Confirmatory IFA is used to assesss the fit of the a priori model
- If the *a priori* model does not achieve good fit, exploratory IFA models may be used or items which do not load strongly on the single factor may be trimmed

Item Response Theory (IRT)

- IRT is used to examine individual items and the associated scores.
  - Item trace lines curves and test reliability functions are typically reported to visualize results

Figure 1. Example Path diagram



Reliability

- CTT analyses (i.e., coefficient alpha, alpha with item i removed, and item-total correlations) evaluate internal consistency reliability
- Test-retest reliability is evaluated using uncorrected Pearson correlations and intraclass correlation coefficients (ICCs)

Validity Evidence

- Convergent/discriminant evidence, how PF scores correlate with theoretically related constructs (stronger correlations) and more distal construct (near-zero or weak correlations)
- Known-groups evidence shows that clinically distinct groups (e.g., chronic vs episodic migraine [CM vs EM] patients) have differential PF scores (e.g., CM would be expected to have lower [worse] PF PRO measure scores than EM)
- Patient changes in PF should be reflected in PF scores (Sensitivity to Change) and PF change scores should correlate with related variables

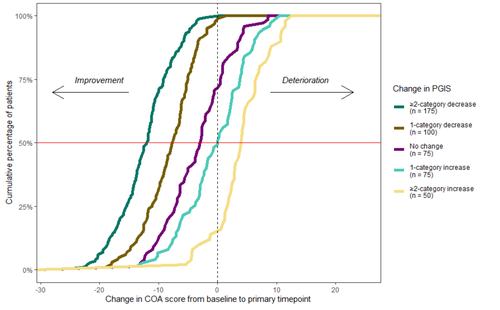
MSD / MWPC

- A single value (or range of values) to define “meaningful” change on the PF PRO measure is determined by triangulating across anchor- and distribution-based methods
- Candidate anchor variable should correlate at least 0.3 with PF scores
- Plot empirical cumulative distribution functions (eCDFs; Figure 2) and empirical probability distribution functions (ePDFs) across the levels of the anchors to support the evaluation of candidate MWPC thresholds

Conclusion

- In addition to qualitative information, extensive psychometric evidence is necessary to demonstrate that scores from a PRO measure are reliable and valid in a given context.

Figure 2. Example eCDF plot



Sponsorship

This presentation was supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award (UH3FD006795) totaling $3,986,552 with 100% funded by FDA/HHS. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the U.S. Government.

Acknowledgements

The authors would like to thank the FDA including Robyn Bent, MS; CHAMP for assisting in recruitment; Elizabeth Nicki Bush, MHS, Roger K. Cady, MD, David W. Dodick, MD, Peter J. Goadsby, MD, PhD., Katie M. Golden, BA, Jason Sico, MD, and Walter F. Stewart, PhD, MPH for serving as advisors to the MiCOAS project.