AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION

Ad Hoc Committee on Patient-Reported Outcomes

October 2015

Report

Feasibility and Selection of Patient-Reported Outcomes into NOMS

Kathryn Yorkston, PhD, CCC-SLP (Chair)
Carolyn Baylor, PhD, CCC-SLP
Gerasimos Fergadiotis, PhD, CCC-SLP
William Hula, PhD, CCC-SLP
Joseph Murray, PhD, CCC-SLP
Nina Simmons-Mackie, PhD, CCC-SLP
David Tulsky, PhD
Howard Goldstein, PhD, CCC-SLP (Board of Directors Liaison)
Rob Mullen, Ex Officio
Janet Brown, CCC-SLP (ASHA Staff)
Rebecca Venediktov, CCC-SLP (ASHA Staff)
Margaret Rogers, PhD, CCC-SLP (Chief Staff Officer for Science and Research)

Report of the ASHA Ad Hoc Committee on Patient-Reported Outcomes

CHARGE STATEMENT

In 2014, ASHA formed an Ad-Hoc Committee to make recommendations for evaluating, selecting, and implementing Patient Reported Outcome Measures.

- (a) the proposed method to be used in selecting patient-reported items from extant item banks;
- (b) the proposed method of testing the selected items in the National Outcomes Measurement System (NOMS) 2.0 environment;
- (c) the proposed implementation platform, timeline, and budget;
- (d) recommendations concerning the details of administration, such as the mechanism(s) for responding (e.g., phone, mail, web), patient privacy issues, patient versus proxy reporting, and the timing of administration; and
- (e) recommendations about the inclusion and integration of information on functional outcomes from multiple sources within the NOMS 2.0 environment.

Abbreviations

CAT – Computer Adaptive Testing
EMPRO – Evaluating Measurement of Patient-Reported Outcomes
ICF - International Classification of Functioning, Disability and Health
IRT – Item Response Theory
NOMS – National Outcomes Measurement System
PRO – Patient-Reported Outcomes
PROMIS – Patient-Reported Outcomes Measurement Information System

HOW THE COMMITTEE ACCOMPLISHED ITS WORK

The committee conducted its work with a two-day face-to-face meeting at ASHA headquarters on 3/10-11/2014 and via a series of 7 conference calls from Jan thru Dec, 2014. The March launch meeting provided background and reviewed the committee's charge. The background included:

- Policy and clinical practice landscape (Janet Brown from SLP Practices and Lisa Satterfield from Government Relations)
- Overview and brief introduction to NOMS (Mullen)
- Brief overview of IRT including common terms so that we are using a common language (Tulsky, Hula and Baylor)[1-3]
- Update on PROMIS and NIH Toolbox (Tulsky)[4, 5]
- Discussion of constructs that should be measured (Simmons-Mackie)[6]
- History of available self-report measures in the field of speech-language pathology (Hula and Baylor)[7-11]

NEED FOR AND FEASIBILITY OF INCLUSION OF PROS INTO NOMS

After discussion of the background information, it was the consensus of the committee that there is a need to provide clinicians and clinical researchers with easy access to a range of psychometrically-sound patient-reported outcomes (PRO) measures related to treatment of communication and swallowing disorders. Further it is feasible to integrate these measures into the ASHA National Outcomes Measurement System (NOMS). The need for such measures is spurred by a number of factors. First, the health care reimbursement landscape will increasingly be one in which providers are called upon to submit measures related to patient experience. Ideally, these measures will be limited in number and applicable both across multiple disciplines and multiple disabling conditions. These measures will be used to determine payment. Second, rigorous PROs are becoming increasingly available, in part, because of the NIH initiative called Patient Reported Outcomes Measurement Information Systems (PROMIS). This initiative stresses standardized measures developed using Item Response Theory (IRT) which allow for precise measurement in conjunction with reduced response burden. Finally, widely accepted models of disablement provide definitions of constructs that can best be measured via patient self-report [12].

PROPOSED METHOD OF SELECTING INSTRUMENTS

Goals of Selection

The selection of instruments should be guided by the following general principles:

- Broad coverage: The goal in selecting instruments is to provide coverage of a broad range of outcomes including language, speech, voice, and swallowing conditions in adults in post-acute, chronic health care settings.
- Psychometric Rigor: Instruments selected should be considered to be the "best currently available."
- Development Pathways: The selection process should reflect the dynamic nature of the field of instrument development. The selection process should help to identify well-developed scales, scales that are in the process of development but not current in final form, and constructs that represent gaps in our ability to measure outcomes. The selection of measures should be updated regularly.

Method of Evaluation

Our task was to identify a framework within which current PRO measures could be evaluated. To achieve this goal, criteria were needed to identify the strengths and weaknesses of the PRO measures. The criteria had to cover the theoretical model on which the measure was based and its psychometric properties. To identify minimum standards for the evaluation of PROs the committee considered several sources including the minimum standards proposed by Reeve et al. (2013) based on input from the International Society of Quality of Life Research; the methods used in the PROMIS item analyses and calibration of item banks [14]; the Consensus-based Standards for the selection of Health Measurement Instruments [15]; the minimum criteria included in the Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) [16]; the US Food and Drug Administration guidelines on PRO measures and their use in medical product development[17]; and, the S*tandards for Educational and Psychological Testing [18]*.

The committee adopted a modified version of the EMPRO tool to evaluate PROs that fall within the scope of practice of speech language pathologists. The EMPRO tool covers eight key PRO attributes: conceptual and measurement model, reliability, validity, responsiveness, interpretability, burden, alternative modes of administration, and cross-cultural and linguistic adaptations. These areas are well aligned with the current guidelines reported in the literature with respect to the minimum standards necessary to evaluate PROs including the recommendations of the Scientific Advisory Committee of the Medical Outcomes Trust [19]. Further, EMPRO includes a four point rating system of each question that could be used to rank order the tools under evaluation [16]; and, it has guidelines for scoring each item. In addition, EMPRO has proved useful in recent investigations of PROs for heart failure [20], prostate cancer [21], and shoulder injury [22].

To make the task more manageable, particularly given the potentially large number of tools to be evaluated, the committee opted to eliminate a number of questions from the EMPRO. Further, the sub scales were organized such that the sum of the first two

subscales (Conceptual Model & Content Validity) along with a description of each tool would allow us to identify tools that are promising and suitable for our purposes despite the relevant lack of quantitative psychometric information at this point. See Appendix A for a listing of the items to be rated.

Initial Screening: Identification of Candidate Measures

ASHA staff conducted an initial broad search of published articles for PROs related to communication/swallowing. This search yielded 86 instruments. The committee reviewed and discussed this list with a goal of creating a more specific set of inclusion/exclusion criteria that can serve as the initial screening mechanism. The following criteria were generated.

- Inclusion Criteria:
    - The instrument must be based on self-report of patients.
    - It must be appropriate for adults with medical (e.g., neurologic or cancer-related) communication or swallowing disorders (either developmental or acquired) in the post-acute phase of rehabilitation.
    - The construct being measured must be clearly defined and expected to change with treatment.
    - A history of how items were developed must be present.
    - The instruments must be of high quality developed either with IRT or classical test theory according to the criteria specified by the modified EMPRO procedure described below [16].
- Exclusion Criteria
    - Scales where a specific communication- or swallowing-related score is not available, e.g. single item or subset of communication/swallowing items from a larger instrument that does not generate a separate communication/swallowing scale. Generally, single items should be avoided because they lack the necessary psychometric properties (e.g. reliability, responsiveness to change, and so on).
    - Articles containing checklists or questionnaires (surveys) rather than summary scores.
    - Instruments that contain only caregiver or other proxy reports
    - Instruments that do not report psychometric development and properties, e.g. scales designed for studying the effect of treatment such as asking about voice compared with prior to treatment with no discussion of the psychometric properties of the scale.

Gathering Psychometric-Data about Candidate Instruments

In order to gather information about the various instruments as efficiently and accurately as possible, the authors of those instruments meeting the eligibility criteria (candidate instruments) will be invited to provide ASHA with information reflecting the instrument's development and psychometric properties. If the author is not available to provide information, ASHA will abstract information from available published literature.

Final Selection of Instruments

ASHA staff in consultation with members of the PRO committee will review the information provided by the authors and place the measure in one of the following categories: (1) Strongly recommended, (2) Recommend with provisions, or (3) Not recommended at this time. The category of recommended with provisions should be used for instruments that are under development and require more testing to confirm their psychometric properties. This category should also be used for instruments with a narrow focus, e.g. applicable for a single patient group. The process of instrument selection should be updated regularly, e.g. by a search for new instruments in the published literature or by nomination of appropriate instruments.

TOWARD INTEGRATION IN NOMS

As soon as an instrument has been evaluated and placed in the "strongly recommended" category, integrating in the NOMS platform can be piloted with the following features:

- An author supplied User's Manual should be available that includes:
    - A statement regarding the construct being measured the intended population(s), and the interpretations or purposes intended for the resulting scores.
    - A narrative description of instrument development and psychometric properties
    - Administration procedures
    - Explanation of alternative forms
    - Guideline for interpretation
- Alternative forms, e.g. tailored short forms, computer adaptive testing (CAT) and so on
- A set of options for administration, e.g. paper and pencil, online, via interview.
- A link to CAT engine platform for administration and scoring if available
- A link to other PROMIS and Neuro QoL measures that might be of interest to SLPs, e.g. self-efficacy, social functioning, fatigue and so on.

IMPLEMENTATION

With the long-term goal of integration in NOMS, several phases of implementation should be conducted. First, authors of candidate instruments will be asked to provide detailed information about the development and psychometric properties of their instrument. The authors should be given the opportunity to provide feedback in order to clarify/modify items being rated. Second, once the initial integration into NOMS has been completed (perhaps before it is publically available), it should be field tested in sites representing a range of health care settings and patient populations. The goal is to see if system works (including support material). Feedback from both the authors and initial users should be used to modify the process both in terms of reducing potential barriers to implementation and enhancing usefulness of the platform. This feedback will provide valuable information to direct future development. For example, the need for more in-depth training or more extensive supplementary material may arise for the feedback.

DEVELOPING A RESEARCH BASE

Once the PRO instruments have been successfully integrated in the NOMS platform, then it is important to disseminate information about the instruments and to encourage research projects that use information in the database.

Dissemination

Information about the availability and usefulness of PRO instruments on the NOMS platform can be disseminated in a variety of ways. Options for dissemination include but are not limited to the following:

- Seminar on PROs at the ASHA Convention
- Brief article for ASHA Leader
- Systematic review of PROs in communication perhaps for AJSLP – This review may help to identify gaps in measurement where important constructs are not measured
- Systematic review of PROs in swallowing perhaps for AJSLP
- ASHA sponsors workshop for students, junior faculty related to measurement development
- Small grants to clinicians or researchers for creative dissemination campaigns.
- Teaching guides or "modules" for use in graduate level clinical courses

Research Development

The NOMS platform that integrates performance-based measures with PROs will provide an excellent opportunity for a variety of clinical research projects related to specific patient populations. Possible research questions include but are not limited to the following:

- What's the relationship between performance-based measures and PROs for specific patient populations?
- How do PROs compared with gold standard measures?
- What variables are associated with the PROs for specific patient populations?
- What supports are needed for patients with severe language problems? How do supported results compare to non-supported administration; what is the impact on psychometric properties of the tool?
- What is the relationship between PROs and proxy ratings for patient with severe impairment? For those with mild to moderate impairment?
- How do PROs compare if they are administered using various modes, e.g. online, paper and pencil, via telephone interview?
    - Investigate which factors (e.g., linguistic, cognitive, psychological, motoric, cultural) at the human-computer interaction create barriers to outcomes assessment for people with disorders.
    - Study if and how inaccessible measurement compromises the validity, reliability and interpretability of outcome measures.

- Identify solutions to evaluate and enhance accessibility of outcome measurement systems.

One mechanism for encouraging this research would be for ASHA to develop a request for small grant proposals that utilize the NOMS data. These grants would be appropriate for student or junior faculty projects and might be developed in conjunction with an ASHA sponsored workshop.

SUMMARY OF RECOMMENDATIONS

This committee recommends that ASHA

- Systematically identify and evaluate existing PRO instruments related to communication and swallowing disorders
- Select psychometrically-sound instruments that reflect broad coverage of communication and swallowing outcomes in adults with medical conditions
- Integrate a set of strongly recommended measures into the NOMS platform, with links to the PROMIS or other platforms that provide assessment protocols.
- Encourage the development of a research base related to PROs including both dissemination and support of small clinical research projects.

TIMELINES

2015

- Initial Screening of Instruments
- Pilot test Author Assessment Survey
- Invitation to authors of candidate instruments
- Evaluation of instruments as surveys are completed

2016

- Build the modified NOMS platform with links to PROMIS Assessment Center
- Integrate “Strongly Recommended” Instructions into NOMs
- Pilot test the NOMS/PRO platform
- Initial disseminations efforts
- Distribute call for proposals

2017

- Continue selection of PROs as instruments are published and surveys are completed
- Continued Dissemination
- Development of a small research grant program/workshop
- Fund 3-5 small research grants

Appendix A: Content of the Table of Evidence for Describing PRO measures (Adapted from EMPRO). [16]

DESCRIPTIVE INFORMATION

1) Intended Use/Interpretation
    a. Intent
    b. Population
    c. Construct
    d. Manner
    e. Content
2) Content Appropriateness
    a. Logical Structure
    b. Empirical Analysis (e.g. frequency, importance)
    c. Professional Opinion
    d. Patient Input
    e. Potential Advantage of Specific Subgroups
3) Description of Item Development Process
4) Burden
    a. Number of Items
    b. Time of administration
    c. Mode
5) Language/Cultural Adaptation

CONCEPTUAL AND MEASUREMENT MODEL

1) The concept to be measured is clearly stated.
2) The conceptual and empirical basis for obtaining the items for the instrument and for combining them into one or more dimensions is clearly stated and appropriate.
3) The dimensionality and distinctiveness of the scales is specifically described and well-supported.
4) The involvement of the target population for obtaining the final content of the instrument is clearly described, the methods are appropriate and the results are satisfactory.
5) Evidence of scale variability in the population is specifically described and appropriate to its intended use.

VALIDITY

1) Sufficient evidence is presented regarding content-related validity of the instrument for its intended use.

2) The hypotheses regarding construct validity are specifically described and the results are consistent with them.
3) A clear rationale and support for the choice of criteria measures or gold standards for criterion validity is provided.
4) The testing of the validity of the instrument for each population of interest is clearly described.

RELIABILITY

1) Internal Consistency: The precision of a scale, based on the homogeneity of the scale's items at one point in time.
    a. Cronbach's coefficient alpha and/or KR-20 values are acceptable.
    b. Reliability estimates employing the IRT approach are clearly reported and acceptable.
    c. Internal consistency data for each population of interest are clearly described.
2) Reproducibility: The stability of an instrument over time (test–retest) and inter-rater agreement at one point in time.
    a. Well-argued rationale is provided to support the design of the test-retest comparison and the interval between first and subsequent administrations.
    b. Coefficients for test-retest reliability and/or inter-rater reliability are specifically described and adequate for all scores.
    c. Item parameter estimates, using IRT applications, are adequately described and appropriate.

RESPONSIVENESS

1) The estimated magnitude of change is clearly described and the results are acceptable.
2) The magnitude of change in a group that is expected to change has been compared with that of a group that is expected to remain stable in longitudinal studies.

INTERPRETABILITY

1) The rationale for the selection and evaluation of the external criteria are specifically described and well-supported.
2) The strategies to facilitate interpretation are clearly described and appropriate.

BURDEN

1) Respondent Burden
    a. The skills and time needed to complete the instrument are clearly described and acceptable.

b. Indications as to when or under what circumstances the instrument is not suitable for respondents.

2) Administrative Burden
    a. The resources required for administration of the instrument are specified.
    b. The time required of a trained interviewer to administer the instrument is clearly indicated and acceptable.
    c. The amount of training and level of education or professional expertise and experience needed to administer the instrument are clearly defined and acceptable.

ALTERNATIVE MODES OF ADMINSTRATION

1) The psychometric characteristics and use of each alternative mode of administration are specifically described and adequate.
2) Information is provided concerning comparability of alternative modes of administration and the original and the results are acceptable.

REFERENCES

1. Baylor, C., et al., *An Introduction to Item Response Theory and Rasch Models for Speech-Language Pathologists.* American Journal of Speech-Language Pathology, 2011. **20**(3): p. 243-259.

2. Cella, D., R. Gershon, and J. Lai, *The future of outcome measurement: Item banking, tailored short-forms, and computerized adaptive testing.* Quality of Life Research, 2007. **16**: p. 133-141.

3. Revicki, D.A. and D.F. Cella, *Health status assessment for the twenty-first century: Item response theory, item banking and computer adaptive testing.* Quality of Life Research, 1997. **6**: p. 595-600.

4. Amtmann, D., et al., *The PROMIS Initiative: Involvement of Rehabilitation Stakeholders in Development and Examples of Applications in Rehabilitation Research.* Archives of Physical Medicine and Rehabilitation, 2011. **92**(10, Supplement): p. S12-S19.

5. Cella, D., et al., *The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008.* Journal of Clincial Epidemiology, 2010. **63**(1179-1194).

6. Kagan, A., et al., *Counting what counts: A framework for capturing real-lide outcomes of aphasia intervention.* Aphasiology, 2008. **22**(3): p. 258-280.

7. Baylor, C.R., et al., *An introduction to Item Response Theory for speech-language pathologists.* American Journal of Speech-Language Pathology, 2011. **20**(3): p. 243-259.

8. Hula, W.D., G. Fergadiotis, and N. Martin, *Model Choice and Sample Size in Item Response Theory Analysis of Aphasia Tests.* Am J Speech Lang Pathol, 2012. **21**(2): p. S38-50.

9. Doyle, P., et al., *Measuring communicative functioning in community-dwelling stroke survivors: Conceptual foundation and item development.* Aphasiology, 2008. **22**(7): p. 718-728.

10. Baylor, C., et al., *The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form.* Journal of Speech Language and Hearing Research, 2013. **56**: p. 1190-1208.

11. Donovan, N.J., C.A. Velozo, and J.C. Rosenbek, *The communicative effectiveness survey: Investigating its item-level psychometrics.* Journal of Medical Speech-Language Pathology, 2007. **15**(4): p. 433-447.

12. World Health Organization, *International Classification of Functioning, Disability and Health (ICF).* 2001, Geneva, Switzerland: Author.

13. Reeve, B., et al., *ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research.* Quality of Life Research, 2013. **22**(8): p. 1889-1905.

14. Reeve, B.B., et al., *Psychometric evaluation and calibration of health-related quality of life item banks.* Medical Care, 2007. **45**(5 Suppl 1): p. S22-S31.

15. Mokkink, L.B., 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): p. 1471-2288.

16. 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.

17. Food and Drug Administration, *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009, FDA: Silver Springs, MD.

18. American Pscyhological Association, American Education Research Association, and National Council on Measurement in Education, *Standard for educational & psychological tests*. 1999, Washington DC: American Pscyhological Association,

19. 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.

20. Garin, O., et al., *Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures.* Heart failure reviews, 2014. **19**(3): p. 359-367.

21. Schmidt, S., et al., *Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments.* Quality of Life Research, 2014: p. 1-13.

22. Schmidt, S., et al., *Evaluation of shoulder-specific patient-reported outcome measures: A systematic and standardized comparison of available evidence.* Journal of Shoulder and Elbow Surgery, 2014. **23**(3): p. 434-444.