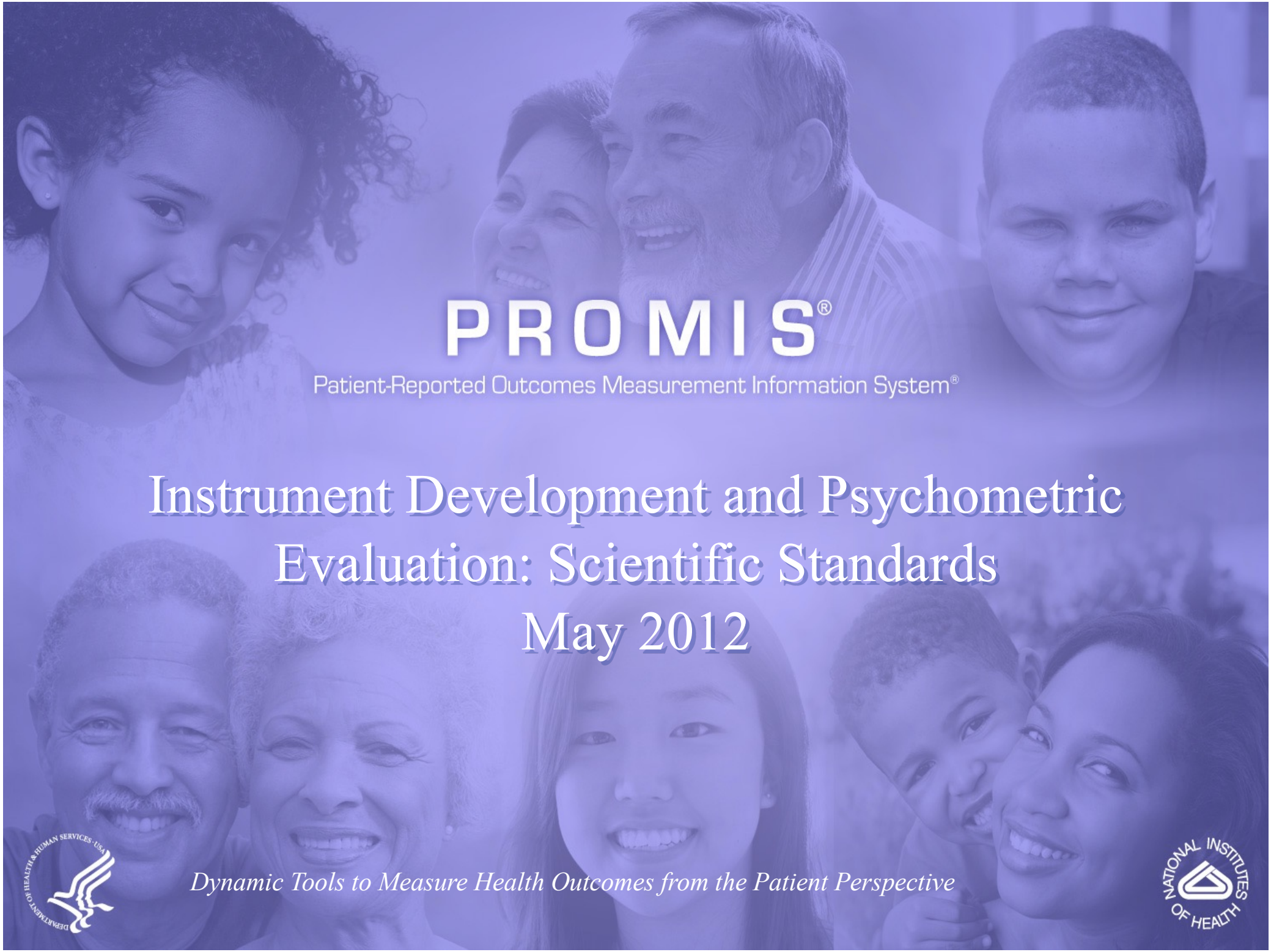

PROMIS®

Patient-Reported Outcomes Measurement Information System®

Instrument Development and Psychometric Evaluation: Scientific Standards

May 2012

Dynamic Tools to Measure Health Outcomes from the Patient Perspective

PROMIS® Standards document

- Patient-Reported Outcome Measurement Information System (PROMIS®) provides clinicians and researchers access to efficient, valid and responsive self-reported measures in health, including symptoms, function and well-being.
- *Instrument Development and Psychometric Evaluation: Scientific Standards* describes a set of standards that serve as
  - the scientific foundation and guidance for the development of PROMIS
  - evaluation of PROMIS item banks and instruments.

PROMIS® Standards document

- Practices are based on measurement science, experience of PROMIS® investigators, published literature on methodology of PROMIS ® and other PRO measurement development
- Standards are operationalized by series of guidelines that provide detailed guidance for item bank development and psychometric evaluation

PROMIS® Scientific Standards Address:

1) Defining target concept and conceptual model
2) Generating and design of individual items
3) Constructing item pool
4) Determining item bank properties
5) Field testing and instrument formats
6) Validity
7) Reliability
8) Interpretability
9) Language translation and cultural adaptation

1.Defining the Target Concept and Conceptual Model

- Conceptual model incorporating target concept(s) should be defined and based on extant literature with input from patients, content and measurement experts, clinicians, end users and stakeholders

- Placement of the instrument within PROMIS® framework should be defined

2. Composing Individual Items

- **Individual items should be refined through cognitive interviewing to ensure:**
  - **The meaning is understood as intended**
  - **The item is clear and contains one concept**
- **Items also should be reviewed for:**
  - **Translatability**
  - **Literacy**
  - **Readability**

2. Composing Individual Items, cont'd.

- **Level of life course and cultural harmonization should be addressed**
- **Existing PROMIS® item formats should be considered and utilized as appropriate**
- **The language used should be clear to patients and consistent with results of formative research**

3. Constructing Item Pool

- **Item pool should cover a full breath of target construct**
- **Consideration should be paid to overlap of facets already covered by extant PROMIS® domains**
- **Consideration should be paid to intellectual property issues**

4. Determining Item Bank Properties

- **Psychometric characteristics of items within an item bank should be determined based on a representative sample of respondents drawn from a relevant and interpretable population**
- **Item banks should have good measurement characteristics:**
  - **–Well-characterized and modeled dimensionality**
  - **–High degree of information and low standard error**
  - **–Model fit, item and scale properties**
- **Differential item functioning (DIF) should be assessed for key groups and its impact on measurement properties identified**

5. Testing and Instrument Formats

- **Instrument formats should be appropriately defined based on intended use and item bank properties including:**
  - **–CATS**
  - **–Fixed length short-forms**
  - **–Profiles**
  - **–Screeners**
- **Adequate scale properties and performance should be demonstrated and include assessment of respondent burden**
- **Instruments that use different modes and methods of administration should demonstrate:**
  - **–comparability of scale properties and performance**
  - **–assessment of respondent burden for each mode**

6. Validity

- **Validity for construct, content and criterion should be addressed relative to a priori hypothesized relationships with related measures**
  - **Description of methods and sample used to evaluate validity, including hypotheses tested and rationale for choice of "gold standard" and measures should be provided**
- **Final instrument should be re-reviewed by experts and end-users to assess consistency with or identify differences between original definitions and final product**

6. Validity, cont'd.

- If an instrument is purported to be responsive, relevant anchor-based methods in representative populations should be provided
- Longitudinal data collected will compare a group expected to change with a group expected to remain stable
- Rationale should be provided for the external anchors used to document change
- Rationale should be provided for the time intervals used for assessment

7. Reliability

- **Reliability of an instrument should be described, including methods used to collect data and estimate reliability**
- **Internal consistency reliability estimates may consist of:**
  - **Information and standard errors at different locations of the scale (item response theory)**
  - **Reliability estimates and standard errors for all score elements (classical test theory)**

7. Reliability, cont'd.

- **Reproducibility of the measure should be described, providing**
  - **–Rationale to support the design of the study**
  - **–The interval between initial and subsequent administration to support the assumption that the population is stable**

8. Interpretability

- The degree to which one can assign easily understood meaning to the instrument's quantitative scores should be described
- Rationale should be provided for the external anchors used to facilitate interpretability of scores
- Information should be provided on the ways in which data from an instrument should be reported or displayed

8. Interpretability, cont'd.

- **Availability of comparative data from the general population and/or group-specific scores should be described**
- **Guidance should be provided on the meaningfulness of scores and changes in scores for use by researchers and clinicians**

9. Translation and Cultural Adaptation

- **Translation of items and instruments should include both forward and backward translations of all items, response choices, and instructions**
- **Translation of items, response choices, and instructions should be obtained through**
  - **Iterative process of forward and back translation**
  - **Bilingual expert review**
  - **Pre-testing with cognitive debriefing**

9. Translation and Cultural Adaptation, cont'd

- **Harmonization across all languages and a universal approach to translation should guide the process**

Appendix

- Each standard refers to guideline documents in the appendix for further description of processes for performing recommended practices
- Appendix also includes the PROMIS® Instrument Maturity Model that describes the stages of instrument development from conceptualization through evidence of psychometric properties in multiple diverse populations.

PROMIS Instrument Maturity Model Stages
Stage Descriptions

1		2		3		4		5
1A	**1B**	**2A**	**2B**	**3A**	**3B**	**4A**	**4B**	**5**
Item Pool	Preliminary Item Bank	Calibrated Item Bank	Item Bank, Profile or Global Health Measure - Prelim Validated	Instruments Validated	Instruments - Prelim responsive-ness	Maturing Instruments	Item Bank Expansion	Fully Mature Instruments: Score Interpretations
Conceptualized	Ready for Calibration	Dimension-ality Assessed & Calibrated	Validity (Construct & Concurrent - limited)	Cross-Sectional, Population Specific	Prelim responsive-ness	Extensive validity & responsiveness in general and pertinent population samples	Item bank modifications - population specific or expansion/ refinement	How scores can be used to understand and respond to health care needs and differences in health is determined and documented

Internal Psychometric Criteria

Criteria	Applies To
QUALITATIVE: Conceptual documentation and evidence supporting content validity	All stages
Item response theory (IRT): Dimensionality Specified; Item calibration; information and DIF analyses	All but stage 1A
Classical test theory (CTT): Evidence supporting dimensionality, reliability and validity (e.g. concurrent validity with legacy)	All but stage 1A
POPULATION: Sample variability reflects variability in construct	Stages 2, 3, and 4
FORMAT: CAT and short form measures; Computer, paper forms	Stages 3A, 3B, and 4A
Continued Documentation of Relevance of Item Content and Generalizability as needed	Stages 1 and 2
Continued Documentation of Relevance of Item Content and Generalizability	Stages 4B and 5

External Psychometric Criteria

Criteria	Applies To
IRT: DIF analyses by different disease conditions and relevant population characteristics (e.g. age, sex, etc.)	Stages 3 and 4
CTT: Evidence supporting responsiveness and interpretation guidelines (MID, responder criteria)	Stages 3 and 4
POPULATION: General population and multiple disease conditions; Relevant language translations	Stages 3 and 4
FORMAT: CAT, short form, and study tailored forms	Stages 3 and 4
MODE: Evidence supporting multiple modes of administration (CAT, paper, IVRS, computer)	Stages 3 and 4