

The value of item banks and PROMIS for patient-reported outcome measurement

Dr. Caroline Terwee

President PROMIS Health Organization, coordinator Dutch-Flemish PROMIS National Center
VU University Medical Center
Department of Epidemiology and Biostatistics

Recent developments in the measurement of Patient-Reported Outcomes

- Challenges with traditional Patient-Reported Outcome Measures (PROMs)
- The value of Item Response Theory-based item banks and Computerized Adaptive Testing (CAT)
- PROMIS

Challenges with PROMs



- Questionnaires are often considered too long
- All patients need to complete the same questions at all times, but not all questions are relevant for all patients (or not at all times)
- Scores are difficult to interpret because of the ordinal nature of most scales
- Scores are incomparable across different PROMs because each PROMS has its own scale (metric)
- Quality of PROMs is not always good (enough) or this is not known

- Item banks, based on Item Response Theory (IRT) methods, preferably used as Computerized Adaptive Test (CAT)
- With CAT the computer selects questions based on answers to previous questions.
- Patients get more relevant questions and they need to complete less questions
- CATs have better measurement properties than traditional PROMs

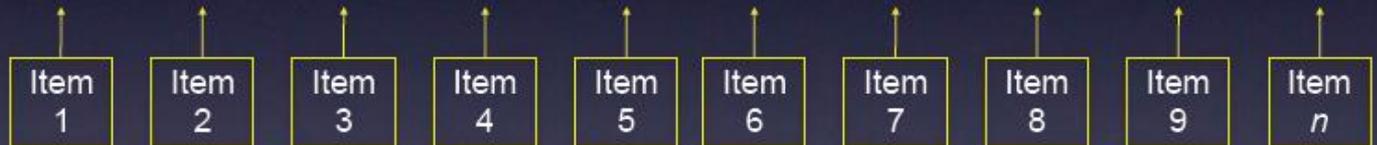
CATs are the gold standard instruments of the future



Example IRT item bank



Physical Functioning Item Bank



Are you able to get in and out of bed?

Are you able to stand without losing your balance for 1 minute?

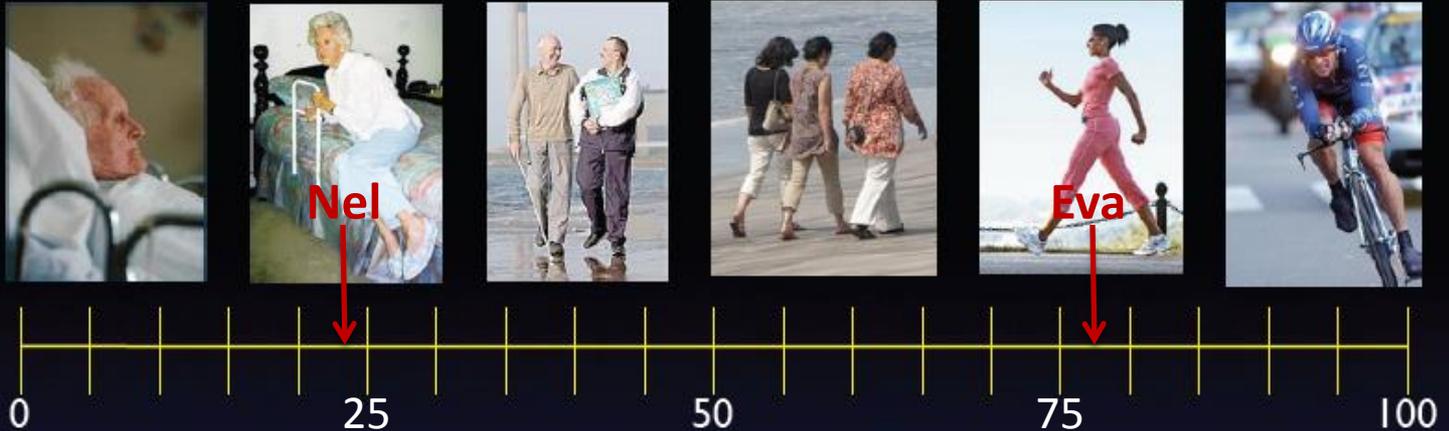
Are you able to walk from one room to another?

Are you able to walk a block on flat ground?

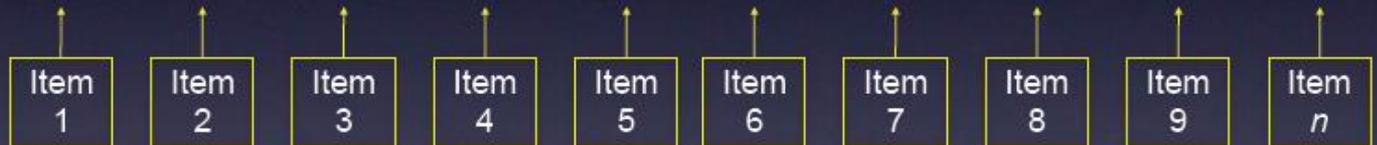
Are you able to run or jog for two miles?

Are you able to run five miles?

Example IRT item bank



Physical Functioning Item Bank



Are you able to get in and out of bed?

Are you able to stand without losing your balance for 1 minute?

Are you able to walk from one room to another?

Are you able to walk a block on flat ground?

Are you able to run or jog for two miles?

Are you able to run five miles?

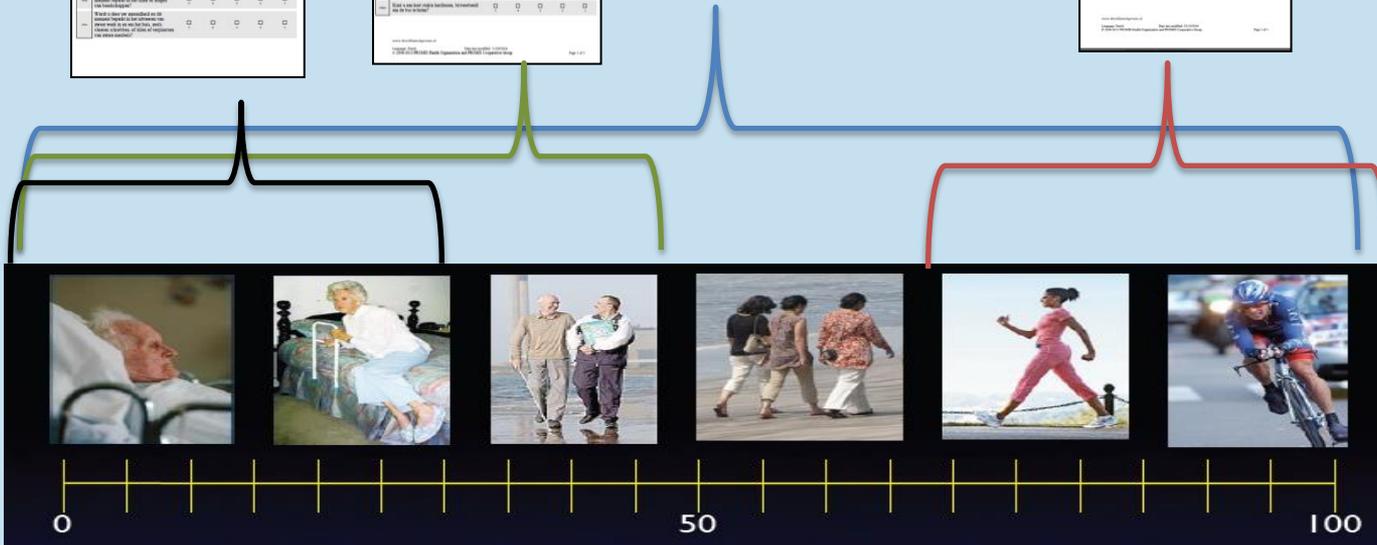
population-specific short forms

Three sample PROMIS short forms are displayed, each with a title in Dutch and English, and a table of items with response options.

Form 1 (Left): "Middel- en Hoge Functione: Lichaamlijk functioneren - Short form 16". It contains 16 items related to physical functioning, such as "Ik kan mijn gewicht dragen" and "Ik kan mijn gewicht dragen".

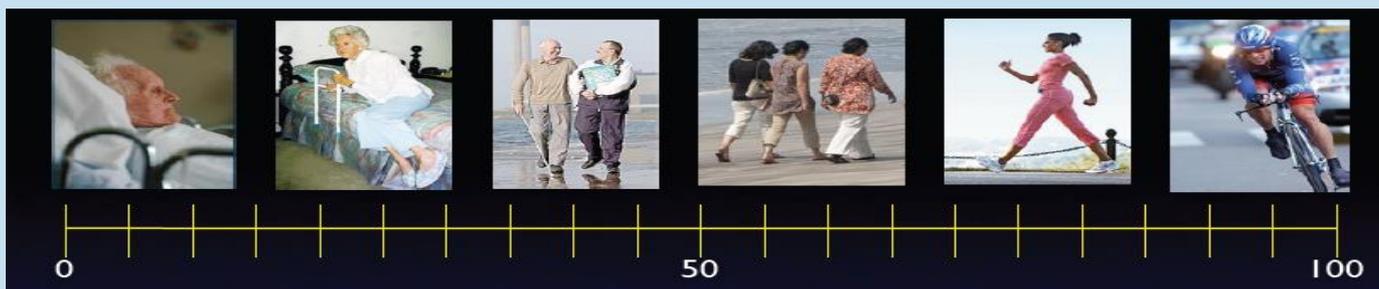
Form 2 (Middle): "Middel- en Hoge Functione: Lichaamlijk functioneren - Short form 16". It contains 16 items related to physical functioning, such as "Ik kan mijn gewicht dragen" and "Ik kan mijn gewicht dragen".

Form 3 (Right): "Middel- en Hoge Functione: Lichaamlijk functioneren - Short form 16". It contains 16 items related to physical functioning, such as "Ik kan mijn gewicht dragen" and "Ik kan mijn gewicht dragen".



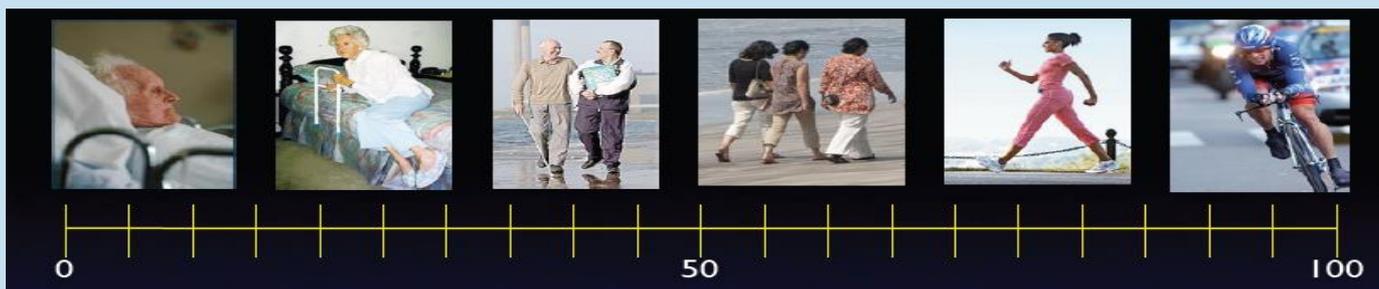
Computerized Adaptive Testing (CAT)

Are you able to walk at a normal speed?



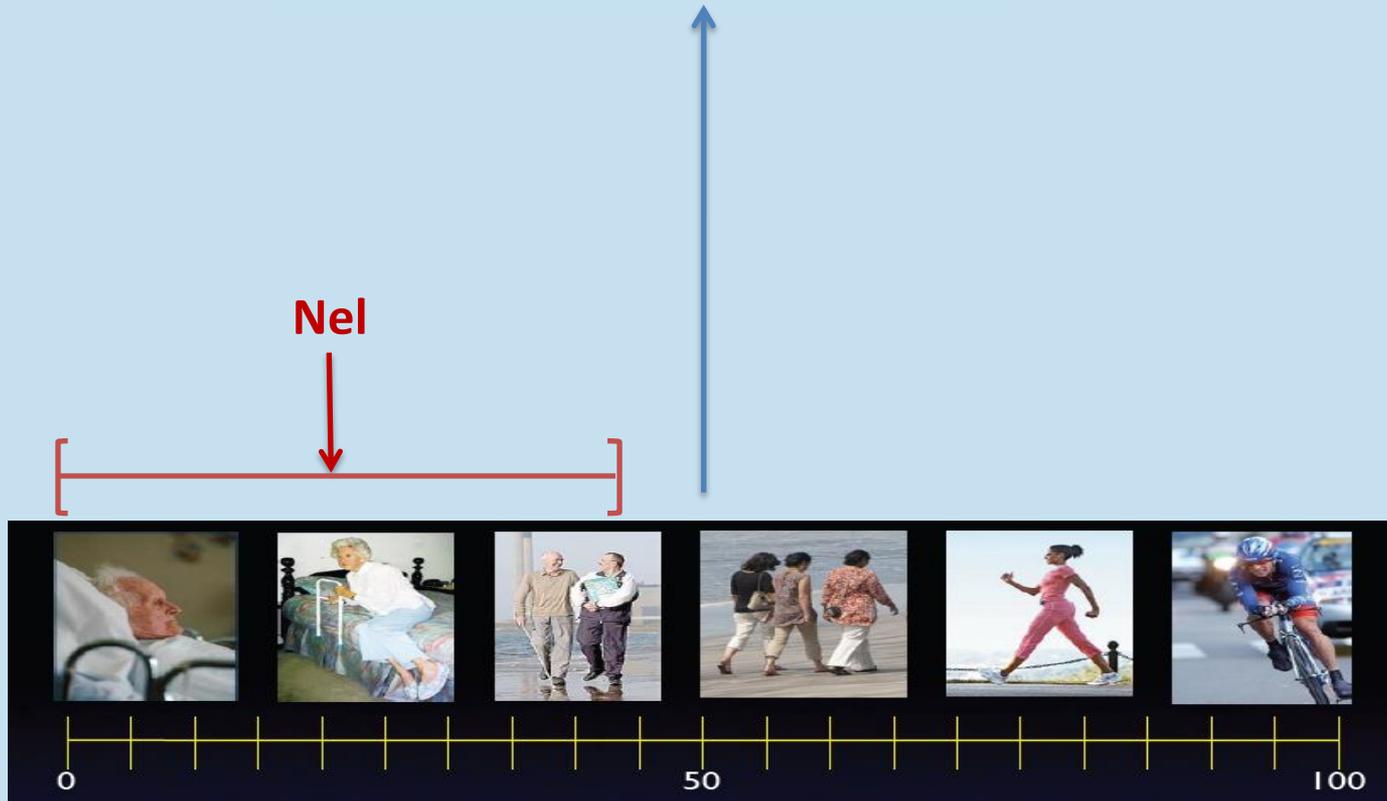
Computerized Adaptive Testing (CAT)

Are you able to walk at a normal speed? NO



Computerized Adaptive Testing (CAT)

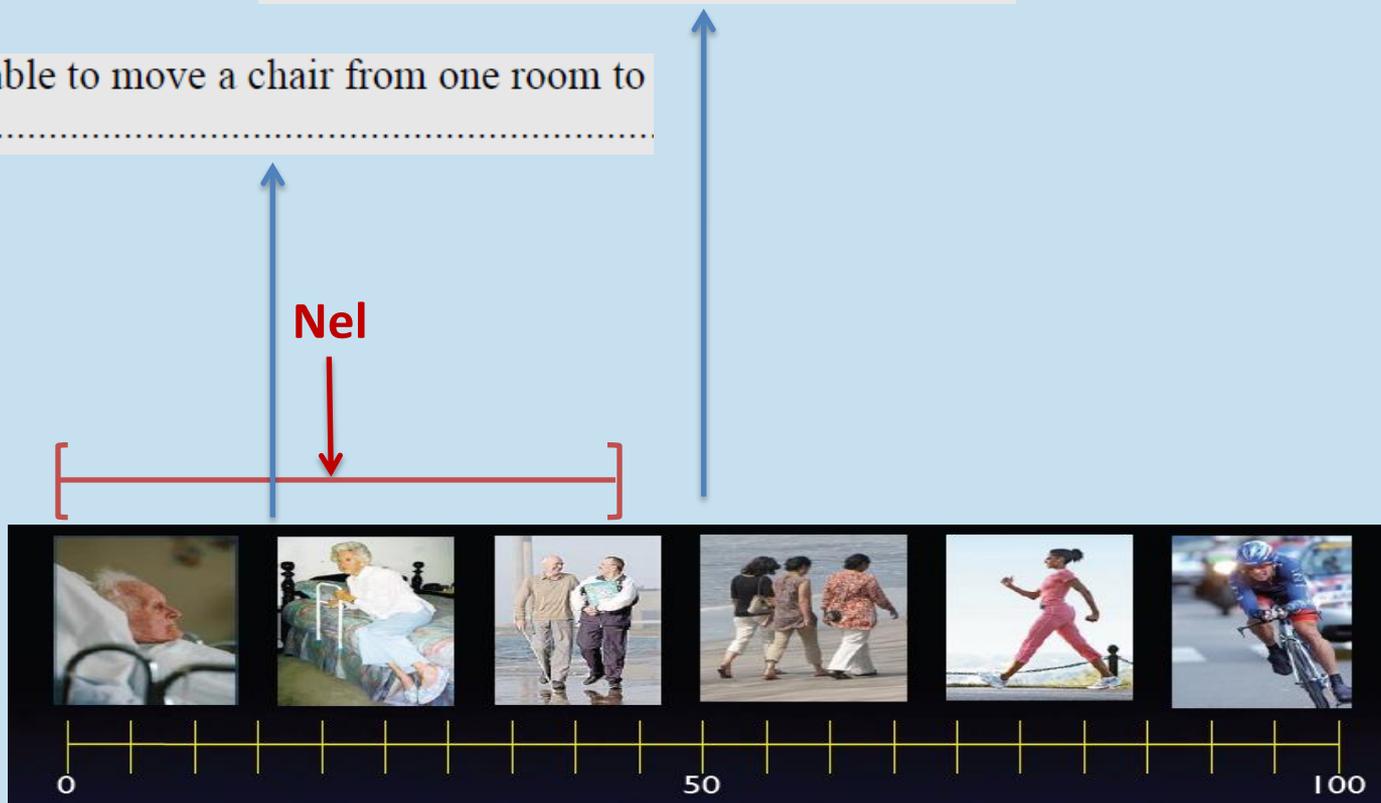
Are you able to walk at a normal speed? NO



Computerized Adaptive Testing (CAT)

Are you able to walk at a normal speed? NO

Are you able to move a chair from one room to another?.....



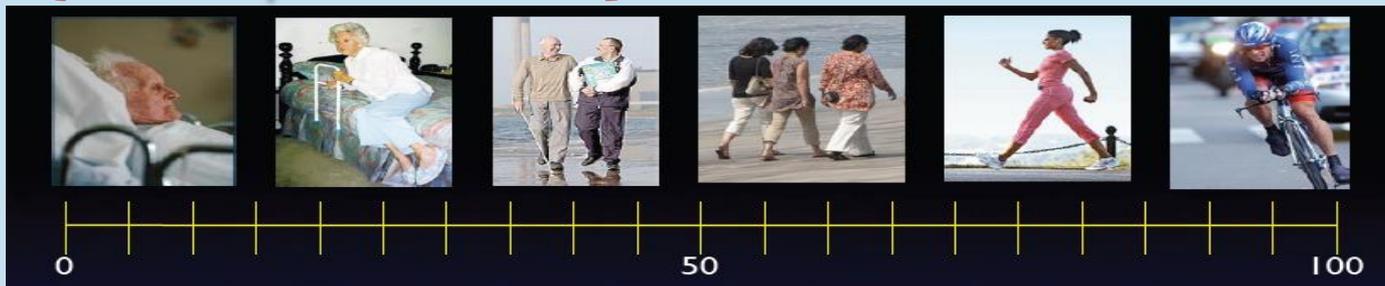
Computerized Adaptive Testing (CAT)

Are you able to walk at a normal speed? NO

Are you able to move a chair from one room to another?.....

YES

Nel



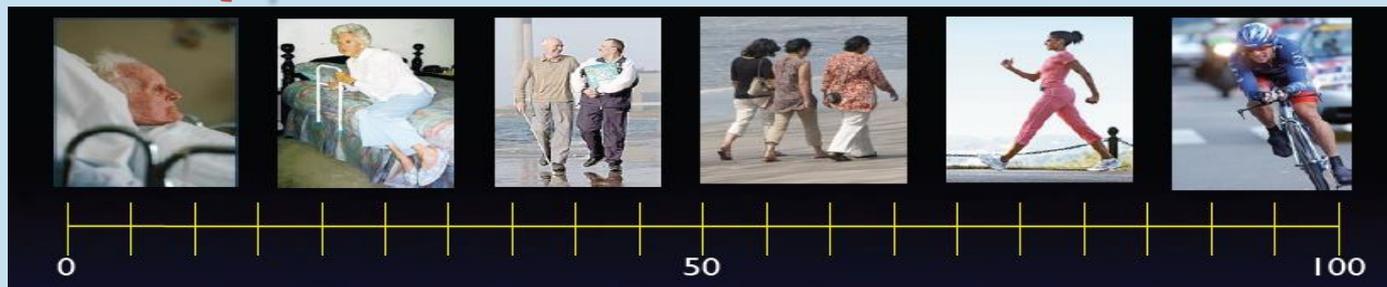
Computerized Adaptive Testing (CAT)

Are you able to walk at a normal speed? NO

Are you able to move a chair from one room to another?.....

YES

Nel



Computerized Adaptive Testing (CAT)

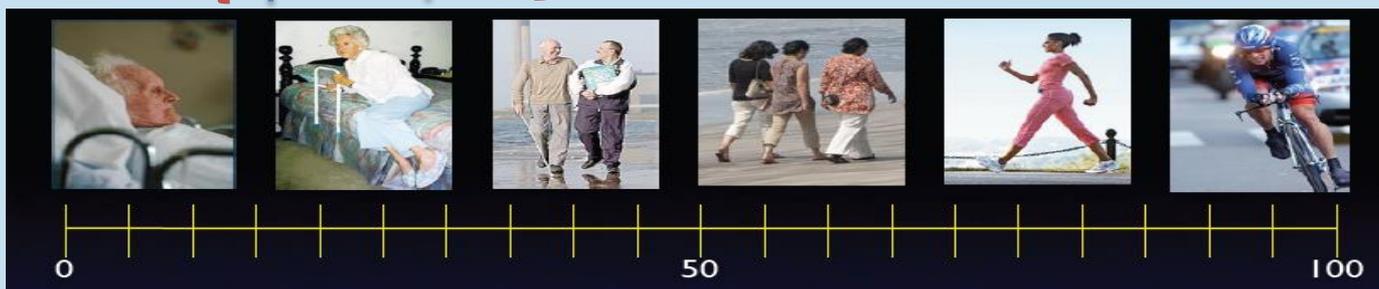
Are you able to walk at a normal speed? NO

Are you able to move a chair from one room to another?.....

YES

Are you able to stand up on tiptoes?.

Nel



Computerized Adaptive Testing (CAT)

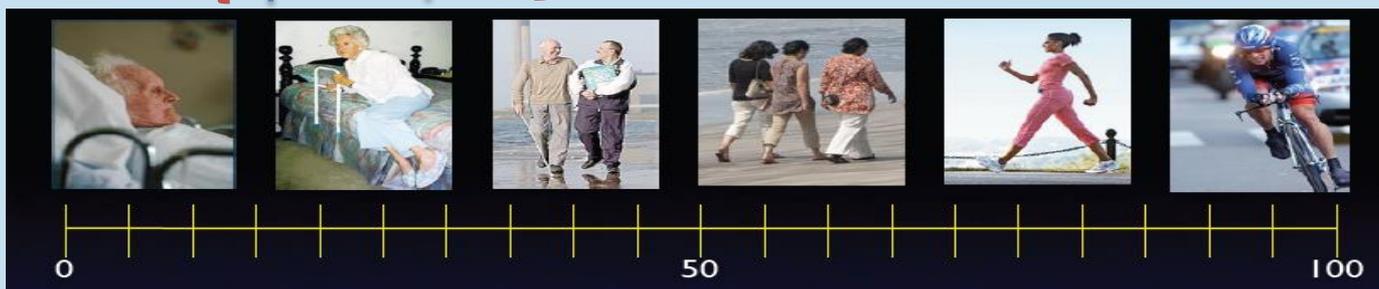
Are you able to walk at a normal speed? NO

Are you able to move a chair from one room to another?.....

YES

Are you able to stand up on tiptoes? NO

Nel



Computerized Adaptive Testing (CAT)

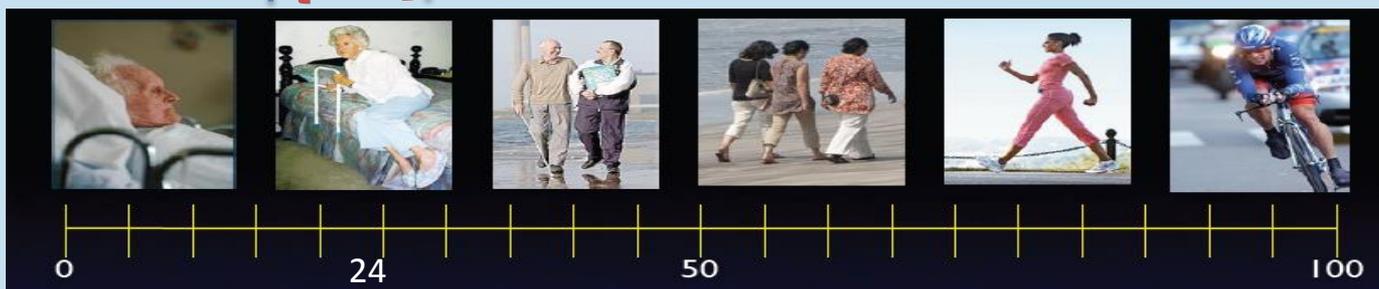
Are you able to walk at a normal speed? NO

Are you able to move a chair from one room to another?.....

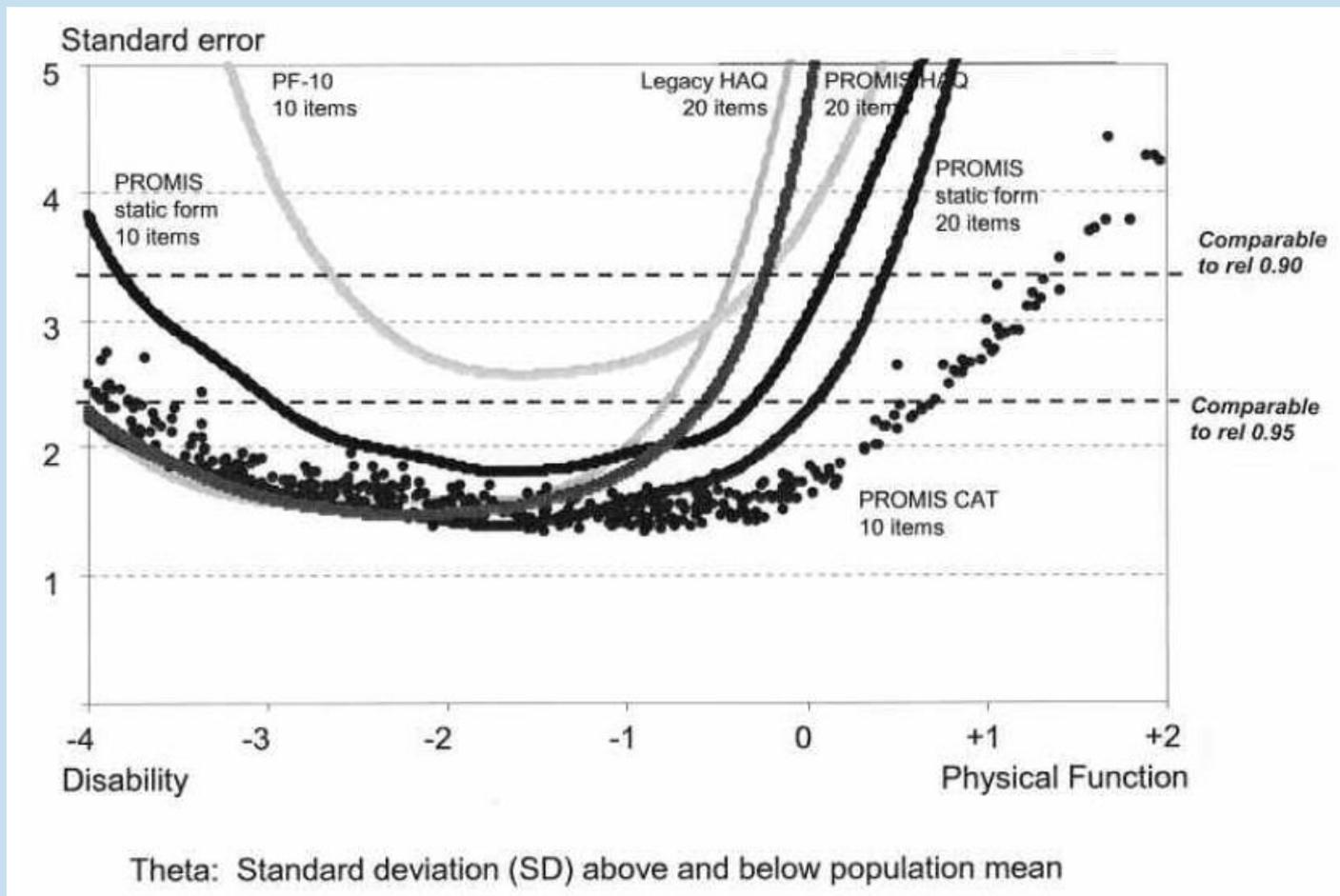
YES

Are you able to stand up on tiptoes? NO

Nel



Reliability of CATs



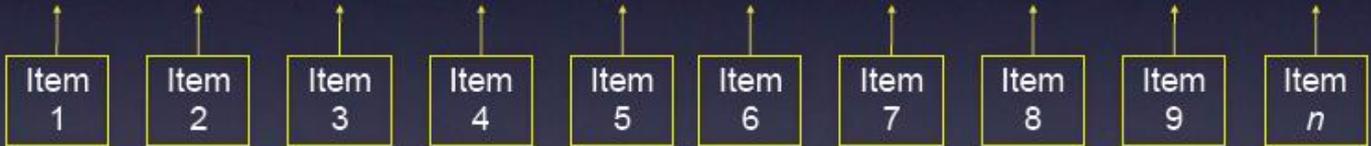
Item bank development

- Large item pools are created based on all existing instruments
- Patients are involved in focus groups or interviews to ensure relevance and comprehensiveness (content validity)
- Items are rewritten into a consistent format
- Statistical analyses are performed on large datasets (>1000) to ensure unidimensionality and to fit IRT item parameters (item location and discriminative ability).
- Differential Item Functioning analyses are performed to check if the item parameters are consistent across subgroups or populations
- CAT software is developed

Item banks are dynamic



Physical Functioning Item Bank

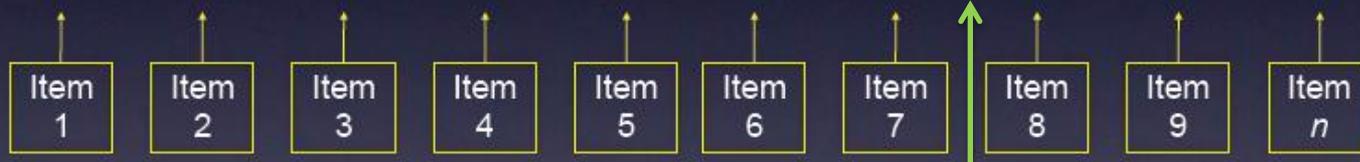


- Are you able to get in and out of bed?
- Are you able to stand without losing your balance for 1 minute?
- Are you able to walk from one room to another?
- Are you able to walk a block on flat ground?
- Are you able to run or jog for two miles?
- Are you able to run five miles?

Item banks are sustainable

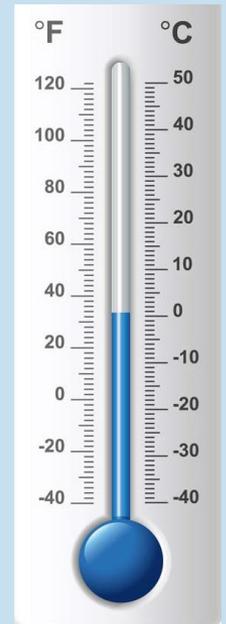
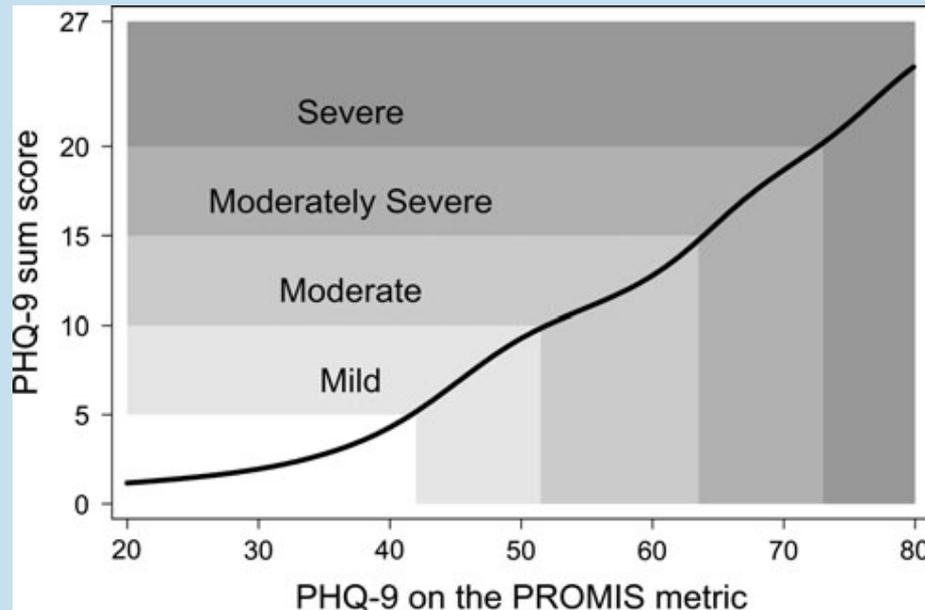


Physical Functioning Item Bank



- Are you able to get in and out of bed?
- Are you able to stand without losing your balance for 1 minute?
- Are you able to walk from one room to another?
- Are you able to walk a block on flat ground?
- Are you able to run or jog for two miles?
- Are you able to run five miles?
- Are you able to use a smartphone?

Other PROMs can be scored on the IRT metric



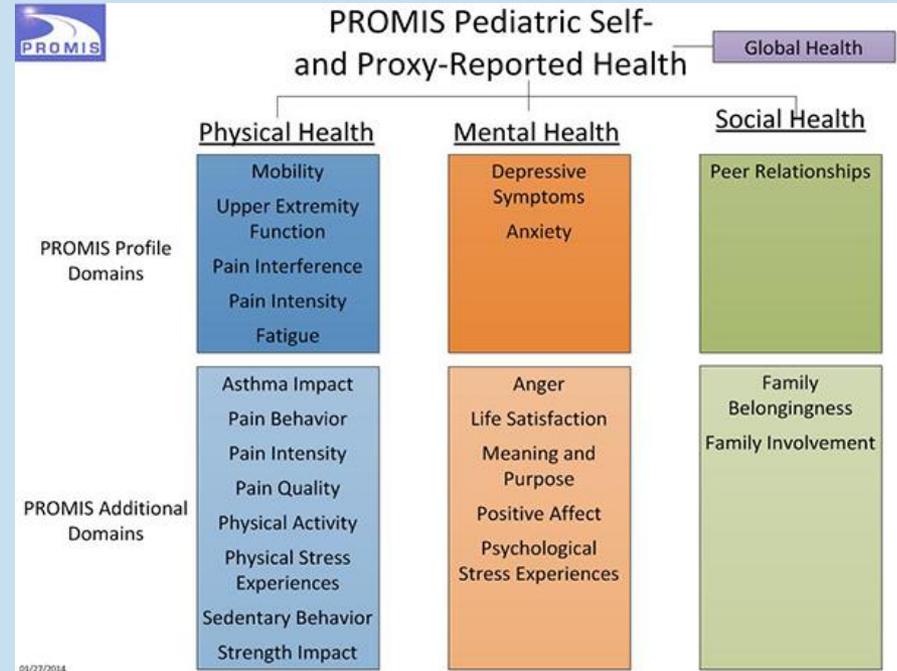
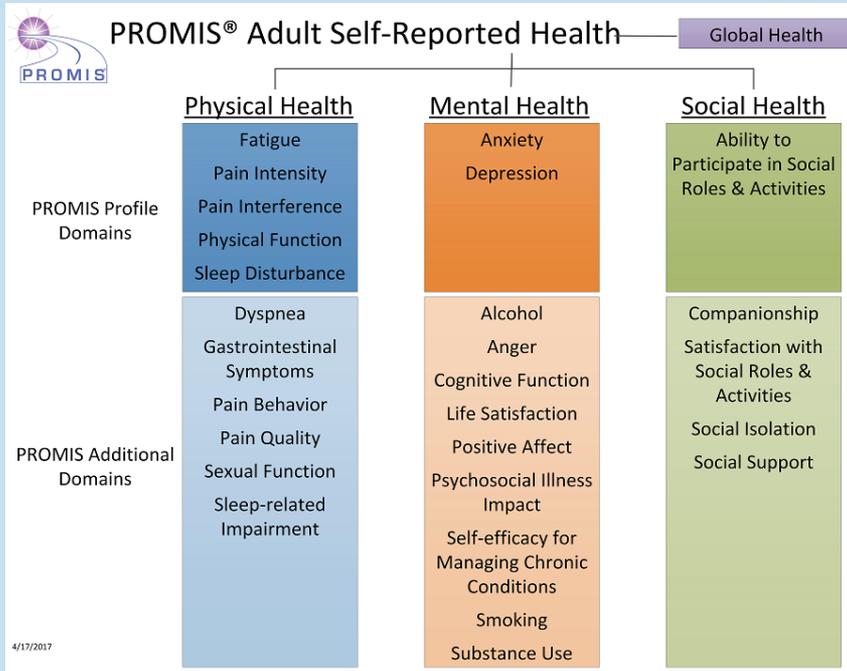
[Gibbons et al. Qual Life Res 2011;20:1349-1357.](#)

www.prosettastone.org

www.common-metrics.org

This helps interpreting scores when changing from traditional PROMs to PROMIS. It makes scores of different PROMs comparable, e.g. for meta-analyses

Patient-Reported Outcomes Measurement Information System (PROMIS)



- Fixed short forms are available (e.g. 4,6, or 8 items).
- Custom short forms can be created from item banks
- All item banks can be used as CAT

- Short forms and whole item banks are being translated in many languages

- CAT software is becoming available in different countries

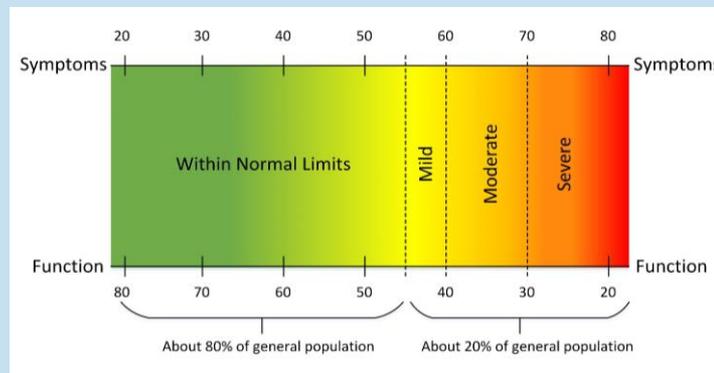
Example custom short form

- An RCT (not published yet) evaluated the effect of a personalized e-health care program on return to normal activities after surgery
- The primary outcome was time elapsing between surgery and return to “normal activities”
- Normal activities was measured by asking participants to select at baseline eight activities from the PROMIS Physical Function item bank. These 8 activities were scored on the IRT metric (personalized PF score)
- Time to return to these eight activities after surgery was measured for each patient as the primary outcome, as well as change in personalized PF score on the IRT metric
- The intervention was effective, in contrast to previous interventions



PROMIS is generic

- PROMIS item banks are generic: they can universally be applied across patient populations.
- Scores of PROMIS instruments are comparable within and between populations
- All scores are expressed as T-scores with a score of 50 representing the mean of the general population (SD 10)



Generic instruments are becoming important because of multimorbidity

- Patients increasingly have multiple conditions

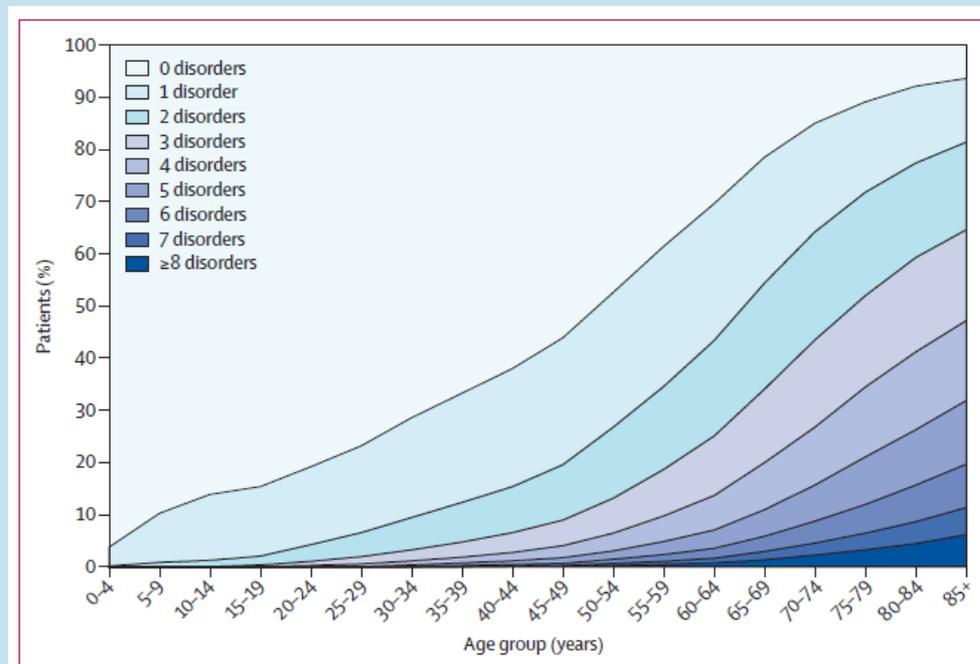
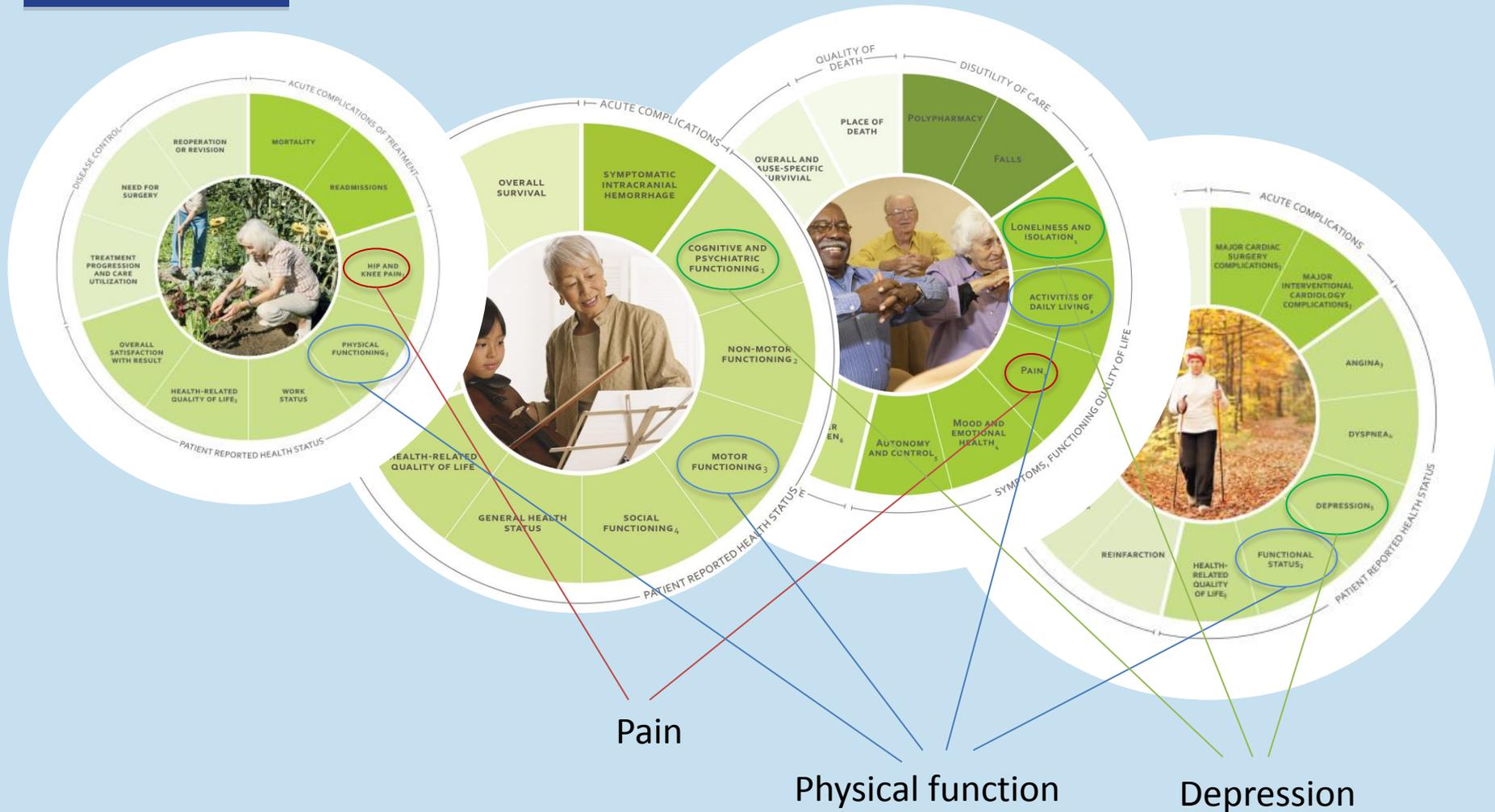


Figure 1: Number of chronic disorders by age-group

Barnett K et al. Lancet 2012;380:37-43.

- Multimorbidity influences the effects of treatment and health care decision making
- Examples:
 - *patients may not benefit optimally from treatment if they are depressed or have cognitive problems*
 - *treatment may need to be postponed or not given at all because of other (more urgent) diseases*
- It increasingly makes sense to measure how a patient feels and functions with or without (multiple) disease(s) (also in evaluating the effects of interventions in clinical trials).

Relevante outcomes are universal



Examples of commonly relevant outcomes:

Pain

Fatigue

Anxiety

Depression

Physical Function

Ability to participate in social roles and activities

These outcomes can be measured with PROMIS, in all patients

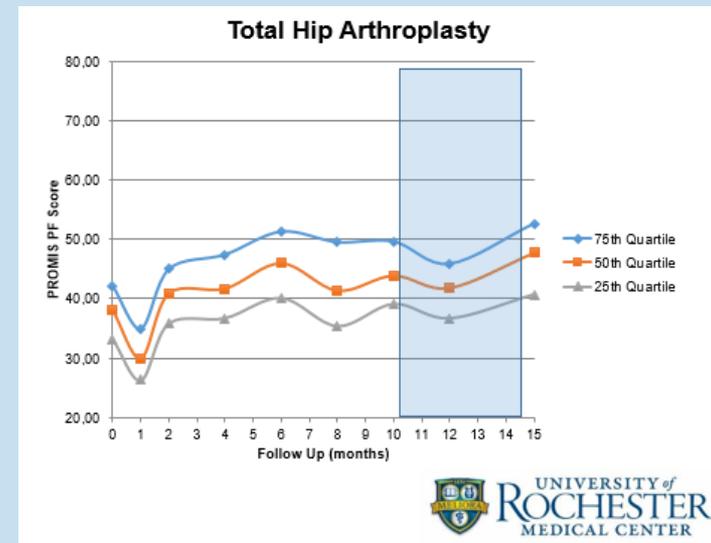
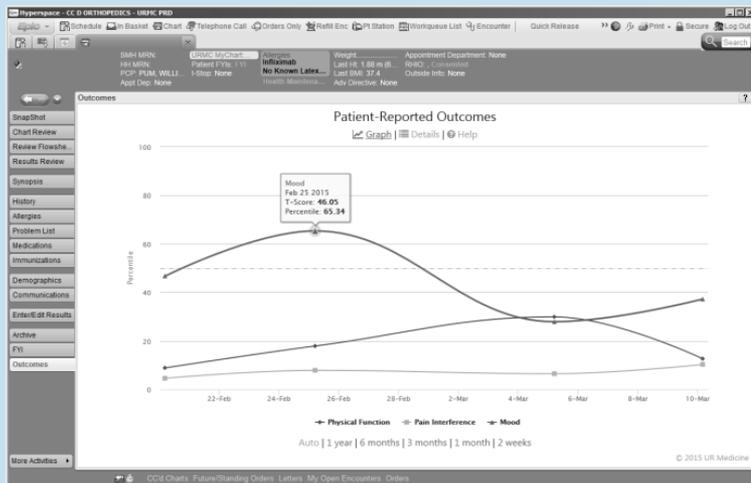
Use of PROMIS in Rochester Medical Center

Large-scale implementation in Rochester Medical Center

Every patient, every clinic, every visit

CAT Physical Function, Pain Interference, Depression

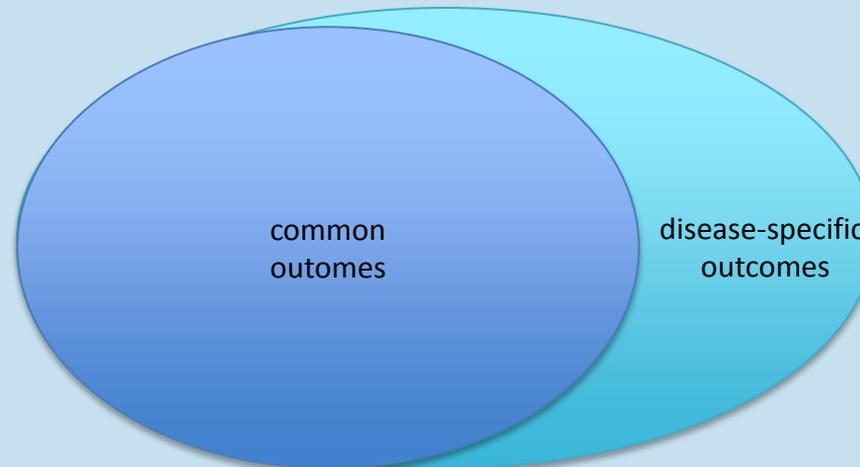
Completion time: 2.4 minutes



N Engl J Med 2017;377:6-9

Common outcomes could be measured in all patients using generic IRT-based item banks, such as PROMIS

In addition, disease-specific outcomes could be added for each specific disease



Generic IRT-based item banks are responsive

- Generic IRT-based item banks are more responsive than traditional generic instruments (such as SF-36) because they measure more precisely
- Therefore they can replace disease-specific instruments

Methods: We studied two Legacy (original) Physical Function/Disability instruments (HAQ, PF-10), their item-improved derivatives (Item-Improved HAQ and PF-10), and the IRT-based PROMIS Physical Function 10- (PROMIS PF 10) and 20-item (PROMIS PF 20) instruments. We compared sensitivity to detect 12-month changes in physical function in 451 rheumatoid arthritis (RA) patients and assessed relative responsiveness using *P*-values, effect sizes (ES), and sample size requirements.

Results: The study sample was 81% female, 87% Caucasian, 65 years of age, had 14 years of education, and had moderate baseline disability. All instruments were sensitive to detecting change (< 0.05) in physical function over one year. The most responsive instruments in these patients were the Item-Improved HAQ and the PROMIS PF 20. IRT-improved instruments could detect a 1.2% difference with 80% power, while reference instruments could detect only a 2.3% difference ($P < 0.01$). The best IRT-based instruments required only one-quarter of the sample sizes of the Legacy (PF-10) comparator (95 versus 427). The HAQ outperformed the PF-10 in more impaired populations; the reverse was true in more normal populations. Considering especially the range of severity measured, the PROMIS PF 20 appears the most responsive instrument.

Conclusions: Physical Function scales using item improved or IRT-based items can result in greater responsiveness and precision across a broader range of physical function. This can reduce sample size requirements and thus study costs.

Fries et al. Arthritis Res Ther 2011;13:R1-47.

Validation while using

- Generic item banks do not need to be validated in each and every population before they can be used
- The construct is broadly covered by the large number of items
- Content validity is further ensured through extensive patient involvement in the development of the item bank
- Item parameters are being established in large patient samples
- Differential Item Functioning (DIF) is evaluated across populations
- DIF can be assessed in each population in which the item bank is used. If large DIF is found, the item parameters can be adjusted for the specific population, without losing data and without losing comparability of scores because the metric stays the same.

- IRT-based item banks and CAT offer a solution to the challenges of traditional PROMs
- With CAT patients get more relevant questions and they need to complete less questions
- Item banks offer a common metric on which other PROMs can also be scored for meta-analyses
- Item banks offer a sustainable measurement system for the future
- PROMIS is the largest system of IRT-based item banks for PROs, that is being translated, validated and implemented across the world
- The generic nature of PROMIS item banks makes them universally applicable across patient populations, without the need for validation in each and every population.
- Using PROMIS can make outcome measurement more efficient and more comparable.



www.healthmeasures.net/promis

www.dutchflemishpromis.nl

cb.terwee@vumc.nl

[PHO Conference - October 28 & 29,](#)

[2018 in Dublin, Ireland](#)

www.promishealth.com