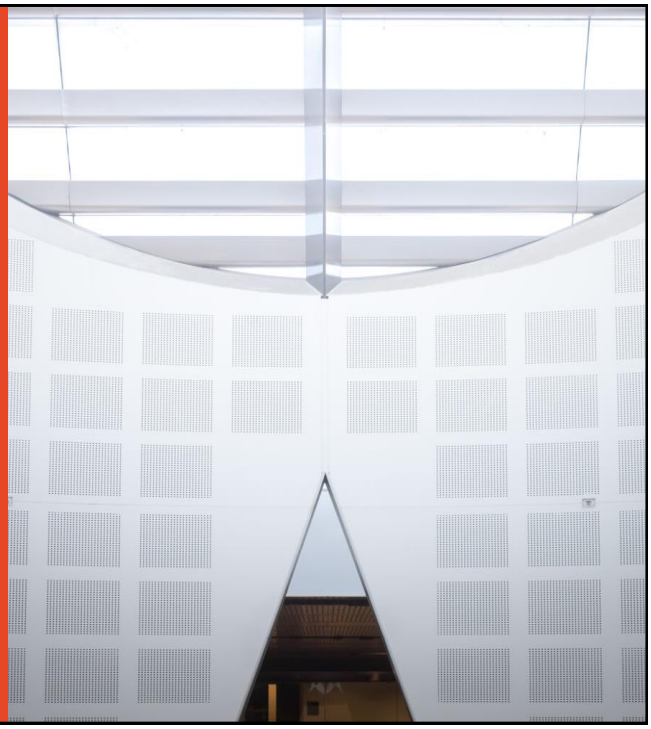
Patient-reported outcome measures: Key design principles to generate high-quality PRO evidence

Dr Claudia Rutherford
Deputy Director, Quality of Life Office,
University of Sydney
claudia.rutherford@sydney.edu.au

THE UNIVERSITY OF SYDNEY



Overview

- Common terminology
- Conceptual framework quality of life (QOL) / patient-reported outcome (PRO) in cancer
- Why should we measure PROs
- How to measure PROs in clinical research
  - What methods are available
  - How to select the most appropriate measure(s)
- Design principles for incorporating PROs in research

Common terminology

Terminology & common usage

Broad umbrella: 'Quality of Life'

A PRO is 'a measurement of any aspect of a patient's health status that comes directly from the patient, without interpretation of the patient's responses by a clinician or research associate'

FDA Guidance (2006)

'Patient-reported outcomes PRO'

Example PROs (subjective/unobserved): HRQOL; Function; Symptom (pain, fatigue), Sexual function, Body image

Patient-reported experiences (PREs)

Patient Reported Outcome Measures

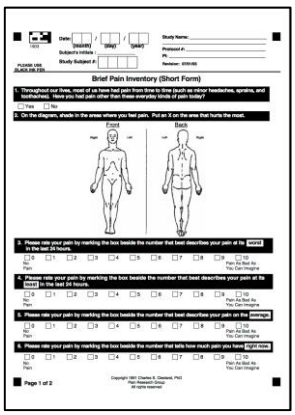
*Capture the patient's perspectives about how illness or care impacts on their **health and well-being***

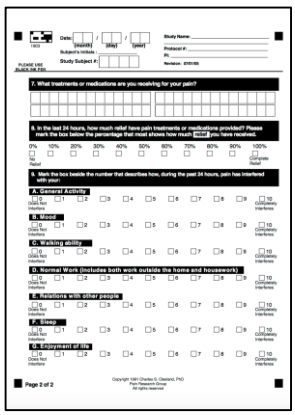
Patient Reported Experience Measures

*Capture the patient's perception of their **experience with health care or services***

PRO/E vs PRM

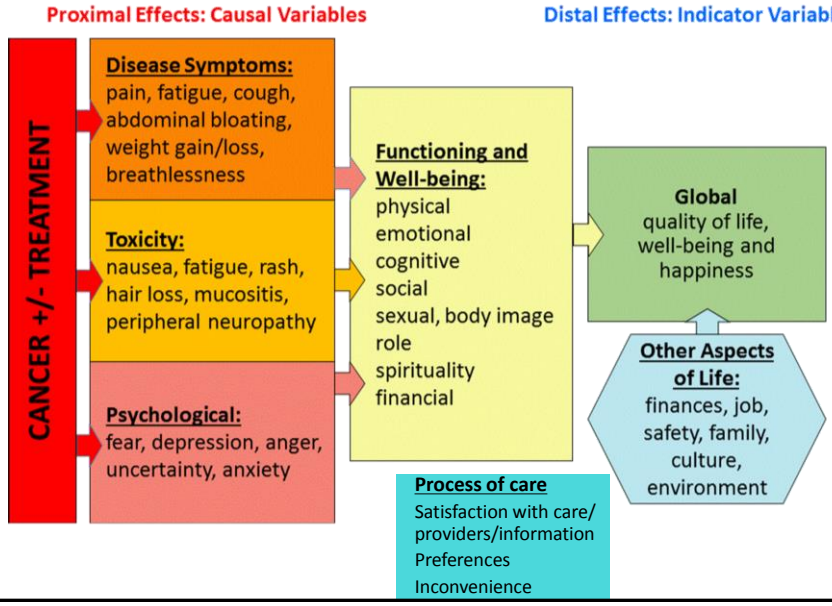
- PRO/E = the patient-reported outcome/experience, e.g. pain / experience of care in relation to provision of analgesics – timeliness, information, communication with care providers
- PRM = the measurement tool used to assess the PRO/PRE, e.g. BPI (Brief Pain Inventory)


Brief Pain Inventory (Short Form)



What are the key QOL / PRO issues in cancer?

How does cancer and its treatment(s) affect a person?



Why assess PROs in clinical research?

Why PROs matter in healthcare?

- Survival almost always paramount - Keep people alive with best possible QOL
- QOL/PROs important consideration in most chronic diseases
  - Understand palliative benefits vs toxic side-effects of treatment
  - Short & long term treatment effects
  - How best to provide supportive care over & above therapeutic care, when and for how long
  - Identify what patients' need and are we helping
- Understanding 'impact' of disease
  - Symptoms and Functional outcomes
  - Affects during acute treatment phase, ongoing care, 'survivorship'
  - Individuals at risk of poor psychological outcomes

WHEN do PROs value-add?

- When response and/or survival differences are likely to be similar
  - non-inferiority trials
- When comparing treatments with quite different toxicity profiles, complexity, or methods of delivery
- When symptom improvement is an endpoint of the trial
  - advanced disease or supportive care trials
- When survival is not the primary study objective

How to measure key PROs in clinical research?

Types of PRO measures

- PRMs often consist of multiple items (questions) that score together to give a score on a particular outcome such as pain, anxiety and fatigue.
- Allow standardised assessment; that is, they use the same standard set of questions, response options and scoring
  - useful for assessing the same patient or sample repeatedly on the same outcomes over time

1. **Generic measures**
2. **Disease-specific measures**
3. **Single domain/symptom-specific**

Generic PRMs

- Designed to measure range of constructs.
- Applied across multiple diseases, treatments/healthcare programmes, and populations (not disease-specific).
- Applicable across diseases and healthy populations.
- Useful for broad comparisons of the relative impact of healthcare programs across diseases.
- Population normative data (Aus, USA, UK, etc).
- Multi-dimensional.

Example: e.g. SF-36, WHO-QOL, PROMIS-10

- 10-items scored into 2 DOMAINS: **Global Physical Health** component and **Global Mental Health** component

Patrick & Deyo, 1989

Disease/condition-specific PRMs

- Tailored for specific diseases or conditions
  - goal to detect minimally important effects in individuals
- Tend to be more responsive to subtle changes in patient's condition, so better suited to measuring outcomes at the individual level.
- Not suitable for comparisons with other conditions.
  - Example: E.g. EORTC QLQ-C30, FACT-G – cancer-specific QOL
  - Tailored for specific cancers e.g. brain: QLQ-BN20, FACT-Br
    - Scores for multiple outcomes:
      - **Symptoms**
      - **Function**
      - **Global QOL**
- Disease-specific PRMs sometimes cover generic content.

Domain/symptom/treatment-specific measures

- Measure one construct (e.g. anxiety) or symptom (e.g. pain) or treatment (e.g. chemotherapy).
- Appropriate if:
  - symptom-targeted intervention,
  - expected differences in symptoms between treatment arms
- Often not disease-specific
  - Pain - Brief Pain Inventory,
  - Anxiety & depression - DASS, HADS
- Some are disease-specific (EORTC and FACIT)
  - Cancer-related fatigue, diarrhea, cancer cachexia, lymphedema
  - Cancer-treatment: Bone marrow transplant, cystectomy, + more…

Utility measures

- Used for economic evaluation to evaluate cost-effectiveness of interventions.
- Economic evaluation has some special requirements for HRQOL measurements.
  - Single metric (utility) is used to quality-adjust survival
  - Values of 1 (full health) to 0 (dead); negative values allowed (worse than dead)
- Range of HRQOL domains but combined into single score (rather than domain-specific scores like profile measures).
- Commonly used multi-attribute utility instruments (MAUIs): EQ-5D, SF-6D, Health Utilities Index (HUI2, HUI3), AQOL.

HRQOL measured in 2 fundamentally different ways

HRQOL measured with a 'profile measure'

1. Subjective – how the respondent (patient) feels
2. Self-report about own health
3. **Multidimensional**
   - **Physical**
   - **Emotional**
   - **Social**
4. **Profile (multiple scales)**
5. Ordinal scales with clinical interpretation

Useful to identify which aspects of a person's health is impacted.

Utility weights (hence QALYs) must be measured with a 'preference-based measure'

1. Subjective – assesses community preferences for QOL vs survival
2. Population preferences about a series of health states, by surveying the a community sample using a preference-based method
3. **Multi-dimensional**
4. **Index (single scale)**
5. Cardinal scale (interval, anchored at 0 and 1) with survival-QOL trade-off interpretation

Useful as an estimate of disease burden.

Adapted from CREST

Proxy assessment

When/why used

- If patient cannot self-report (dementia, children, brain injury)

The problem with proxies

- Proxies report on 'health status' vs patients report of their HRQOL – the latter is subjective Cross-informant variance, particularly for non-observable domains (emotional vs physical function outcomes).

Strategies to overcome problems with proxies

- Collect patient AND proxy reports, if possible, for as long as possible
- Use proxies consistently in longitudinal studies
- Restrict to observable domains (e.g. physical function, "pain actions")

Standardised measurement

E.g. QLQ-C30 Physical Functioning items & scale

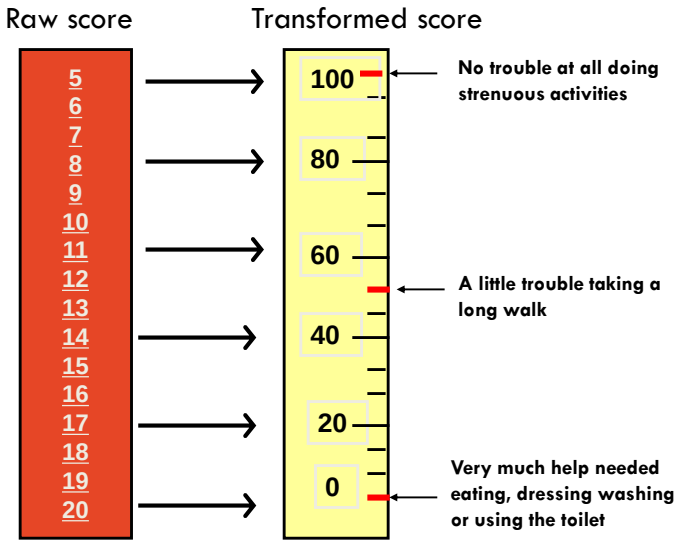
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?
2. Do you have any trouble taking a long walk?
3. Do you have any trouble taking a short walk outside of the house?
4. Do you need to stay in bed or a chair during the day?
5. Do you need help with eating, dressing, washing yourself or using the toilet?

Response scale for each item

Not at All	A Little	Quite a Bit	Very Much
1	2	3	4

Sum to give total scale score – Possible score range 5 - 20

Physical Functioning score or "Ruler"



The right PROM(s) for your study is one that:

- Covers all relevant outcomes (e.g. symptoms and aspects of QOL) that are expected to be affected by the disease and/or intervention/treatment
  - especially those where treatment differential (between trial arms) is expected
- Review the content of the questionnaires
  - What issues covered by items (questions) and how are they worded?
  - How are items combine into scales? Domain vs Total score
  - Select the PRM(s) that **best match target PROs/PREs**
- Has evidence of careful development / demonstrated validity, reliability, and responsiveness
  - Preferably in population(s) same/similar to your target population and similar context

Snyder 2007. Value in Health

Other considerations for selecting PRMs

- **Static questionnaires** (standard set of items)
  - Work on paper and on computer
  - Unless brief with simple scoring, can be a burden to administer and score.
- **Dynamic questionnaires** require computer-based assessment and access to validated item banks and computer-adaptive test (CAT) software.
  - CAT measures are more efficient and allow more domains to be assessed (more precise measurement based on fewer items).
  - CAT requires computer assessment (select items based on respondent answers to previous items).

Other considerations for selecting PRMs continued

Attributes of the questionnaires

- Design, layout and instructions,
- Framing of questions,
- Response format and Recall period

Cultural appropriateness

- Important consideration but difficult as not all PRMs have cultural validation studies.

Language availability

- Not just language translation but also cultural appropriateness.

PRO Design principles and other considerations

Justify inclusion of PROs to be measured

- What is the **rationale** for measuring the PRO(s) in the study – think about:
  - What's known / not known (evidence from literature)
  - Why it matters – e.g. clinical importance (clinical expertise)
  - Which PROs relevant / expected to be affected by intervention/treatment
    - acute symptoms or side effects OR late effects (e.g. cumulative tmt burden or persistent problems (e.g. fatigue, pain, anxiety, fear recurrence)
  - Define your study PROs – if you're measuring symptoms, which specific symptoms? If measuring QOL, are all QOL domains relevant? Explicit = appropriate PROM

Should PROs be included in every trial?

- Include if HRQOL/PRO data will inform future decision making
- QOL assessment requires time and effort
  - Researchers: planning, data entry, analysis, reporting;
  - Patients: completing;
  - Staff at recruiting sites: ensuring PRO assessments are completed or record reasons for non-completion;
  - Cost: Q'aire licence & scoring manual, staff, admin costs.
- Needs careful thought and planning:
  - what aspects of QOL really matter to patients and their managing clinicians within the context of a particular trial
  - If not done well, results can be misleading

Who will be included in PRO sub-study?

- Ideally, the same participants as those evaluated for all study endpoints
  - Scientific – credibility, generalisabiliy and interpretation of results
- In practice, some limitations that make self-assessment infeasible
  - Language availability
  - Literacy level
  - Cognitive impairment
  - Physical impairment

Patient burden: How many items is too many items?

Questionnaire length

- Balance between gathering enough information to determine functionality vs limiting patient burden.
  - No longer than 20 mins at baseline, shorter thereafter (Basch 2012, JCO)
- Repetition across questionnaires
- Number of assessments

Where will PROs be assessed?

- In clinic is convenient but may not be the most informative time
  - E.g. Cyclic, acute toxic effects like nausea may be missed if PRO consistently assessed just before a dose of chemotherapy
- Online PRO assessment allows completion:
  - at more informative time points feasible
  - non-clinic-based populations, e.g. survivorship cohorts

How will PROs be assessed?

- Modes of Administration (MOA)
  - The PLACE of completion + HOW the patient completes the questionnaire
  - Paper (hard-copy) vs electronic (online) vs telephone-assisted
    - Pros and cons of each MOA
    - Logistics and costs need to be considered – set-up, collect data, enter/check data, follow-up/remind non-responders
  - Results from MOA meta-analysis support the use of mixed MOA with a study
    - a useful strategy for reducing missing PRO data

Rutherford 2016 Qual Life Res

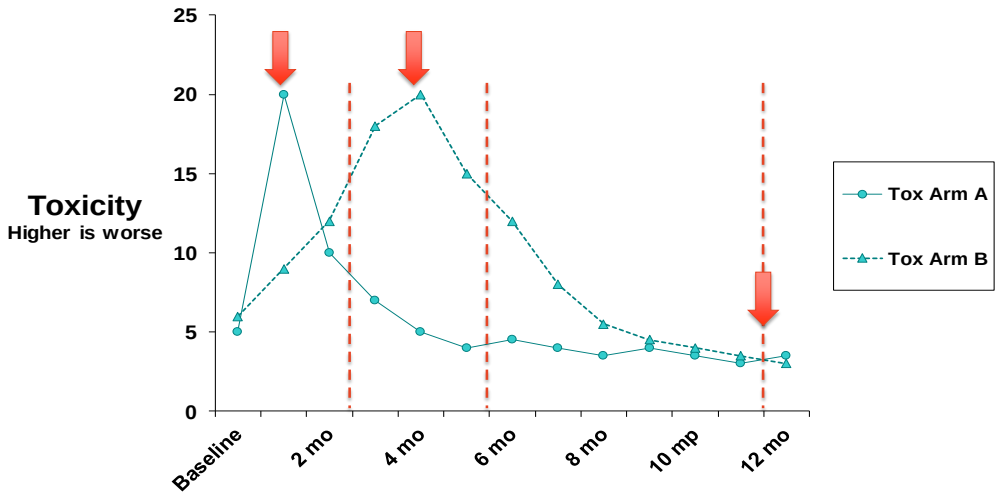
Timing of PRO assessments - Baseline

- Getting the timing right matters!
- Baseline – always!
  - Intervention studies = Before intervention starts
  - Survivorship studies = at recruitment
- Time period around the target event (e.g. surgery, end of treatment) within which effect of interest will be observed and not diluted.

Timing of PRO assessments – Follow-up

- Follow-up time points, think about:
  - what are the acute and late effects
  - what are the expected trajectory of treatment effects over time
- Time-points for PROs should coincide with maximum expected treatment impact
  - E.g. chemo/radio-therapy, end treatment best time to pick up cumulative toxicity
  - Different aspects of impact, e.g. short-term toxicity v long-term benefit
- Timing of maximum treatment impact may differ between arms
  - E.g. long vs short regimens– measure both arms at both times
- How long to continue follow-up

Example: When would you measure?



The missing PRO data problem

- Missing PRO data can cause loss of power and bias
- Seriously affect the external validity (generalizability) of the results
- Think about PRO missing data at all stages of research
- Minimise PRO missing data where possible

Mercieca-Bebber R, Palmer MJ, Brundage M et al. Design, implementation & reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. BMJ Open 2016; 6.

Example: Ovarian cancer cohort study

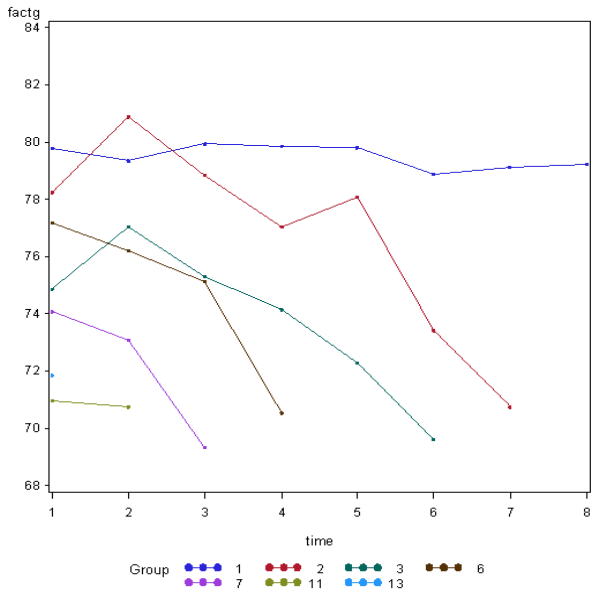
Y axis: QOL (FACT-G), higher score = better QOL

X-axis: 8 QOL assessment time-points

Graph is stratified by number of assessments completed

Those who drop out early start out worse and have steeper declines

Mercieca-Bebber RL, et al. Asia-Pacific Journal of Clinical Oncology, Accepted June 2016.



SPIRIT-PRO – Guidelines for PRO inclusion in clinical trials

Special Communication

February 6, 2018

Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols

The SPIRIT-PRO Extension

Melanie Calvert, PhD[1]; Derek Kyte, PhD[1]; Rebecca Mercieca-Bebber, PhD[2]; et al

» Author Affiliations

JAMA. 2018;319(5):483-494. doi:10.1001/jama.2017.21903

CONSORT-PRO recommendations for reporting PRO endpoints

Reporting of Patient-Reported Outcomes in Randomized Trials

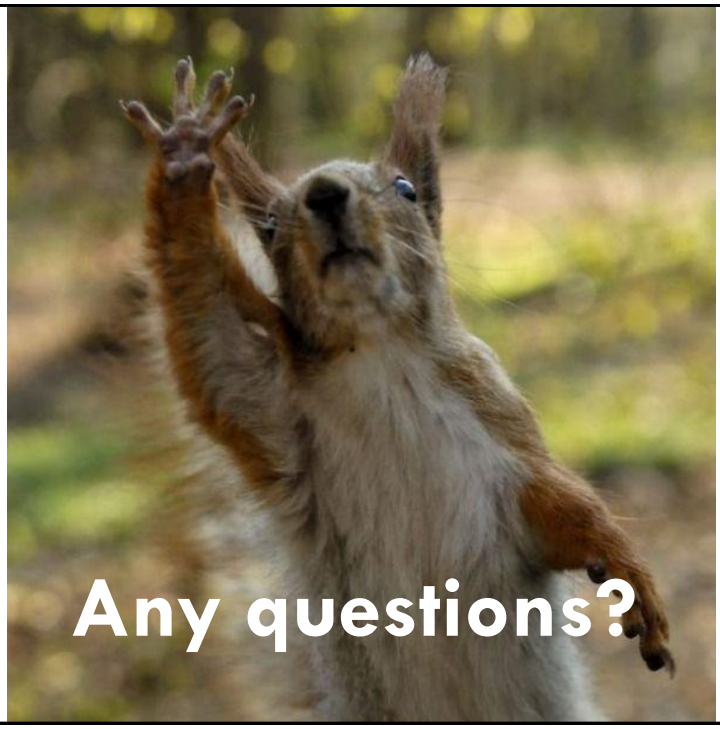
The CONSORT PRO Extension

Melanie Calvert, PhD
Jane Blazeby, MD
Douglas G. Altman, DSc
Dennis A. Revicki, PhD
David Moher, PhD
Michael D. Brundage, MD
for the CONSORT PRO Group

The CONSORT (Consolidated Standards of Reporting Trials) Statement aims to improve the reporting of randomized controlled trials (RCTs); however, it lacks guidance on the reporting of patient-reported outcomes (PROs), which are often inadequately reported in trials, thus limiting the value of these data. In this article, we describe the development of the CONSORT PRO extension based on the methodological framework for guideline development proposed by the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network. Five CONSORT PRO checklist items are recommended for RCTs in which PROs are primary or important secondary end points. These recommendations urge that the PROs be identified as a primary or secondary outcome in the abstract, that a description of the hypothesis of the PROs and relevant domains be provided (ie, if a multidimensional PRO tool has been used), that evidence of the PRO instrument's validity and reliability be provided or cited, that the statistical approaches for dealing with missing data be explicitly stated, and that PRO–specific limitations of study findings and generalizability of results to other populations and clinical practice be discussed. Examples and an updated CONSORT flow diagram with PRO items are provided. It is recommended that the CONSORT PRO guidance supplement the standard CONSORT guidelines for reporting RCTs with PROs as primary or secondary outcomes. Improved reporting of PRO data should facilitate robust interpretation of the results from RCTs and inform patient care.

JAMA. 2013;309(8):814-822 www.jama.com

THE CONSORT (Consolidated Standards of Reporting Trials) Statement, first published in 1996 and most recently revised in 2010,[1,2] provides evidence-based recommendations to improve the completeness of reporting of randomized controlled trials (RCTs). The statement focuses on parallel-group trials, but a number of extensions for reporting other trial designs (cluster, noninferiority, and equivalence), interventions (nonpharmacologic and herbal therapies), and for specific data, such as harms have been developed.[3] The CONSORT Statement is endorsed by major journals and


Any questions?