BACKGROUND

Sample size affects the robustness of psychometric results. Why is this a problem? Clinical trials are powered for the statistical methods used to evaluate the primary endpoint, which may or may not include a patient-reported outcome (PRO) measure. Currently, there is little guidance for psychometric sample size requirements for the development of PRO measures used in clinical trials. What are the consequences? Sample sizes that are too large or too small may influence conclusions regarding the reliability and validity of the PRO measures used in clinical trials. With sample sizes that are too large, the evaluation may be compromised for the PRO measures, resulting in more rigorous methods than needed. With sample sizes that are too small, the applicable analyses methods may be limited, models may not converge, and results lack statistical power to achieve the desired outcomes. First step to change the problem forward: A more rigorous approach to PRO sample size practices is to describe current sample size practices in the literature.

OBJECTIVE

To describe the statistical methods and corresponding sample size practices in published psychometric evaluation studies of PRO measures likely developed for use in clinical trials over the past decade.

METHODS

Systematic Literature Review and Data Extraction

- Data source: PubMed
- Eligibility criteria: Article published in peer-reviewed journals

Validation or psychometric evaluation abstracts and corresponding full-text articles that mention one of four concept blocks (i.e., construct validity, content validity, response shift, and test-retest reliability) were included. They required a unique set of evaluation methods. PRO measures likely developed for use in clinical trials:

- No review articles, opinion pieces, duplicate articles, or gray literature
- Articles written in English
- Articles published on or after January 1, 2004

Literature review and data-extraction process:

- The search strategy included a comprehensive list of four concept blocks designed to capture mentions of (1) PROs and (2) clinical trials that were reported and published as full development/validation studies in PubMed. The search was performed from May 1, 2004, to May 1, 2014.
- The primary reviewer (TC) trained a team of four reviewers (WHC, LN, VW, LM) and provided a set of instructions and a template for extracting data.
- Five reviewers assigned titles and abstracts for review.
- Reviewers independently read titles and abstracts for inclusion and exclusion.

Discrepancies were resolved based on consensus between reviewers.

Overall study sample sizes ranged from 39 to 14,038. The overall mean sample size was 577.3 (standard deviation [SD], 1241.38; median, 249.0; minimum, 39; maximum, 14,038).

Within each study, sample sizes were reported by method. Results: 7 excluded abstracts were reviewed and 18 included abstracts (of which 12 full-text articles were reviewed for inclusion). Given the low error rate, further verification was not warranted.

Descriptive Studies

- Twenty-six countries were represented in the studies included in these analyses. Studies were most often based in the United States (36.0%), United Kingdom (13.3%), and Germany (6.1%). Approximately 9% of studies were conducted in multiple countries.

- Among the 181 studies included in the final analyses, the majority (75%) included the development of at least one PRO measure: gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gait (11.0%), gai...