ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

Journal of Clinical Epidemiology

BRIEF REPORT

Most domains of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 are reliable

Adam B. Smith[a,b,*], Kim Cocks[c], Matthew Taylor[a], David Parry[d]

[a]*York Health Economics Consortium, University of York, Market Square, York YO10 5NH, UK*
[b]*Research Innovation Office, University of York, York YO10 5DG, UK*
[c]*Department of Health Sciences, York Trials Unit, University of York, York YO10 5DD, UK*
[d]*AstraZeneca Pharmaceuticals Ltd, Alderley Park, Macclesfield SK10 4TF, UK*

Accepted 18 March 2014; Published online xxxx

Abstract

Objectives: The study's aim was to assess the internal reliability for the nine domains of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) to evaluate homogeneity across clinical studies and whether sample characteristics predict coefficient heterogeneity.

Study Design and Setting: A systematic literature review was undertaken. Internal reliability was assessed against Cronbach α coefficient >0.70. Reliability generalization was undertaken using fixed- and random-effects models. A weighted least squares regression model was applied to determine whether baseline sample characteristics (language, percentage of women, sample size, sample means and standard deviations, and cancer type) predicted variation in α coefficients.

Results: A total of 33 studies were identified. Eight domains demonstrated good internal reliability (unweighted/weighted by sample variance). One domain, Cognitive Functioning, consistently performed poorly. In terms of moderating variables, none of the sample characteristic variables explained sample variance for the Physical or Role Functioning domains. For the other domains, language, percentage of women, and sample means and variances accounted for some of the heterogeneity observed.

Conclusion: Most domains on the EORTC QLQ-C30 are reliable and may therefore be used to help inform decision-making processes, such as those involving individual patients.

Keywords: Reliability generalization; Internal reliability; Cronbach α; EORTC QLQ-C30; Patient-reported outcomes; Oncology

1. Introduction

Internal reliability, as measured by Cronbach α coefficient, is an important psychometric property of patient-reported outcome measures (PROMs). It is often erroneously assumed that the internal reliability of an instrument is stable across studies; however, reliability is not a constant property of an instrument and may vary across patient samples, administration method, and different languages.

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) [1] is one of the most widely used PROM in oncology clinical trials, research, and practice. The current version comprises 30 items, which are summed and transformed to create five functioning subscales (physical, role, social, emotional, and cognitive), a global quality of life subscale, three symptom subscales (fatigue [FA], nausea and vomiting [NV], and pain [PA]), and six individual items. The psychometric properties of the instrument have been assessed for a variety of cancers, clinical trials, and interventions.

A recent review of the psychometric properties of the EORTC QLQ-30 suggested that the internal reliability of the nine domains with multiple items taken together is modest at best [2]. These authors determined that the majority of studies (27 of 30, 90%) had found only limited or mixed evidence for the instrument's internal reliability (based on a criterion of at least 50% of subscales having a Cronbach α between 0.7 and 0.9). Strongly supportive evidence (Cronbach $\alpha > 0.7$ on all subscales) was found only in a single study (1 of 30, 3%). Although this study noted

Disclaimer: The views expressed herein are those of the authors and do not reflect the official policy or position of AstraZeneca Pharmaceuticals Ltd (UK) or York Health Economics Consortium.

Funding: A.B.S. was supported through a Knowledge Transfer Partnership Grant (KTP007957) funded jointly by the UK Technology Strategy Board and AstraZeneca Ltd.

* Corresponding author. Tel.: +44-(0)1904-324831; fax: +44-(0)1904-323628.

E-mail address: adam.smith@york.ac.uk (A.B. Smith).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2014.03.004

What is new?

Key findings

- This article presents the first reliability generalization of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30). Although there was a significant amount of heterogeneity between studies, most domains (eight of nine) of the EORTC QLQ-C30 demonstrated good internal reliability with the exception of Cognitive Functioning. A number of moderator variables such as percentage of women, language, and baseline mean scores and standard deviation were significant predictors for the heterogeneity in internal reliability.

What does this add to what was known?

- Previous research has failed to produce supportive evidence for the internal reliability of the EORTC QLQ-C30. The results of this study, however, suggest that most domains of the EORTC QLQ-C30 have good internal reliability.

What is the implication and what should change now?

- The need for reliable patient-reported outcome measures is of importance for both clinical trials and the treatment of individual patients. The EORTC QLQ-C30 is a reliable instrument and may therefore be used to aid decision-making processes in oncology clinical trials and clinical practice. Further work is required to aid the interpretation of scores on these instruments for individual patient care.

that the Cognitive Functioning domain performed poorly in general (as low as 0.19 in one case), they did not provide any further detailed breakdown of the properties of individual subscales of the EORTC QLQ-C30. These may vary considerably with some domains performing more optimally than others. Knowing which domains of the EORTC QLQ-C30 are reliable is key to any decision-making processes involving PROMs, particularly those affecting individual patients.

The aim of this study was to apply a statistically more sophisticated and robust method—a reliability generalization [3–5] analysis to the internal reliability coefficients of the EORTC QLQ-C30 (version 3.0) to (1) derive an overall average Cronbach α for each of the instrument's domains, (2) investigate whether there is heterogeneity in sampling schedules, (3) determine whether random-effects models are able to account for this variance, and (4) explore potential explanatory (moderating) variables that may account for any heterogeneity in sampling variance.

2. Method

2.1. Instrument

A reliability generalization of coefficient alphas was undertaken on the nine multi-item scales from the EORTC QLQ-C30, including the five functioning scales: physical functioning (PF, five items), role functioning (RF, two items), emotional functioning (EF, four items), cognitive functioning (CF, two items), and social functioning (SF, two items); global health scale (GHS, two items); and three symptom scales: FA (three items), PA (two items), and NV (two items).

2.2. Search strategy

A literature search was undertaken on PubMed on December 11, 2012, using the broad search terms: (''EORTC QLQ C30''[All Fields]) AND (''reliability''[All Fields] OR ''Cronbach''[All Fields]) with a limit of English language articles only. The data were extracted independently by the first author. As this was a single database extraction, there were no duplicates. Inclusion criteria were the use of the latest version of the EORTC QLQ-C30 (version 3) and studies involving cancer patients. Exclusion criteria included previous versions of the EORTC QLQ-C30, noncancer samples, small sample size ($N < 40$), unequal numbers of patients for each scale, and data not available for a majority of the scales (less than five of nine scales). The last two criteria were applied in to facilitate the calculations and comparisons across studies.

2.3. Data

The data collected from the articles included: Cronbach α (criterion: >0.70 and ≤ 0.90) [5], mean score and standard deviations, number of patients completing the instrument, primary diagnosis, proportion of women, average age, timing of assessment, administration mode, and language for baseline data only.

2.4. Reliability generalization

A previously published method was used (Appendix at www.jclinepi.com) to analyze the Cronbach α coefficients [6–9]. A weighted least squares regression (Enter method) was applied with transformed alphas as the dependent variable. The following were selected as potential moderating variables: diagnosis (five categories: breast, head and neck, mixed, lung, and other), language group (four categories: European, Asian, mixed, and Arabic), proportion of women, number of patients, and mean scale scores and standard deviations across studies. These variables were selected as they were available for most studies. The transformed score variance was included as a covariate. The analysis was undertaken