Journal of Clinical Epidemiology 69 (2016) 79–88

Journal of Clinical Epidemiology

Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust

Johannes M. Giesinger[a,1], Jacobien M. Kieffer[a,1], Peter M. Fayers[b,c], Mogens Groenvold[d,e], Morten Aa. Petersen[d], Neil W. Scott[b], Mirjam A.G. Sprangers[f], Galina Velikova[g], Neil K. Aaronson[a,*], on behalf of the EORTC Quality of Life Group

[a]*Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*
[b]*Institute of Applied Health Sciences, University of Aberdeen, Foresterhill Road, AB25 2ZD Aberdeen, UK*
[c]*Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Postboke 8905, N-7491 Trondheim, Norway*
[d]*The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, Bispebjerg Bakke 23, 2400 Copenhagen, Denmark*
[e]*Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen, Denmark*
[f]*Department of Medical Psychology, Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands*
[g]*Leeds Institute of Cancer and Pathology, St James's Institute of Oncology, University of Leeds, Beckett Street, LS9 7TF Leeds, UK*

Accepted 21 August 2015; Published online 28 September 2015

Abstract

Objective: To further evaluate the higher order measurement structure of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30), with the aim of generating a summary score.

Study Design and Setting: Using pretreatment QLQ-C30 data (N = 3,282), we conducted confirmatory factor analyses to test seven previously evaluated higher order models. We compared the summary score(s) derived from the best performing higher order model with the original QLQ-C30 scale scores, using tumor stage, performance status, and change over time (N = 244) as grouping variables.

Results: Although all models showed acceptable fit, we continued in the interest of parsimony with known-groups validity and responsiveness analyses using a summary score derived from the single higher order factor model. The validity and responsiveness of this QLQ-C30 summary score was equal to, and in many cases superior to the original, underlying QLQ-C30 scale scores.

Conclusion: Our results provide empirical support for a measurement model for the QLQ-C30 yielding a single summary score. The availability of this summary score can avoid problems with potential type I errors that arise because of multiple testing when making comparisons based on the 15 outcomes generated by this questionnaire and may reduce sample size requirements for health-related quality of life studies using the QLQ-C30 questionnaire when an overall summary score is a relevant primary outcome.

Keywords: Health-related quality of life; Questionnaires; EORTC QLQ-C30; Measurement model; Higher order factor scores; Confirmatory factor analysis

1. Introduction

Patient-reported outcome measures (PROMs) are currently seen as important outcomes in both observational studies and clinical trials. They represent the patients' voice in determining the burden of disease and its treatment. One of the most widely used PROMs in oncology is the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30) [1]. The QLQ-C30 is a multidimensional health-related quality of life (HRQOL) questionnaire composed of six functional scales, three symptom scales, and a number of additional single item scales (15 outcomes, in total).

Although the QLQ-C30 provides a wealth of information about the HRQOL of patients, it also presents an analytic challenge because of the multiple outcomes it generates, and the concomitant risk of committing a type I error due to multiple testing [1,2]. In some studies, it is

Funding: The work of J.M.G. was supported by a grant from the Austrian Science Fund (FWF J3353).

[1] Shared first authorship.

* Corresponding author. Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Tel.: +31-20-5122481; fax: +31-20-5122322.

E-mail address: n.aaronson@nki.nl (N.K. Aaronson).

http://dx.doi.org/10.1016/j.jclinepi.2015.08.007

What is new?

Key findings

- We found a robust single higher order factor model to be the best performing measurement model for the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30).
- The resulting QLQ-C30 summary score exhibits equal or superior known-groups validity and responsiveness to change over time as compared to the individual QLQ-C30 scales.

What this adds to what was known?

- The results support the robustness of a single-factor higher order measurement model for the QLQ-C30.
- The validity and responsiveness of the QLQ-C30 summary score is equal to, and in many cases superior to the original, underlying QLQ-C30 scale scores.

What is the implication and what should change now?

- The EORTC Quality of Life Group recommends using the QLQ-C30 summary score to supplement the 15-outcome profile generated by the QLQ-C30.
- The availability of a summary score can facilitate more reliable hypothesis testing analyzing QLQ-C30 data.
- If the QLQ-C30 summary score is chosen as the primary focus of a study, then its use can reduce the risk of type I errors that can occur when making comparisons based on the original 15 outcomes generated by the QLQ-C30. Thus, it may be possible to reduce sample size requirements for health-related quality of life studies using the QLQ-C30 questionnaire.

possible to reduce the number of statistical tests performed by defining a limited set of QLQ-C30 scales that are of primary interest. Preferably this is done on an a priori basis to avoid selective, post-hoc reporting of results [3,4]. However, in many studies, it may be difficult to prespecify which QLQ-C30 scales are of most interest. In such cases, investigators frequently rely on the two-item scale assessing overall quality of life [5–8].

The disadvantage of this very brief two-item overall quality of life scale is that it may have less measurement precision than is desired for detecting group differences over time. In addition, it may not be a conceptually appropriate summary of the QLQ-C30, which contains a relatively large number of symptom scales and items [9,10].

On the basis of such considerations, Hinz et al. and Nordin et al. introduced and investigated summary scores for the QLQ-C30. Hinz et al. [11] used a total score derived from summing up all 30 items of the questionnaire and two separate summary scores based on the sum of all items of the functioning domains and of the symptom domains, respectively. Nordin et al. [12] investigated the known-groups validity of the two-item global quality of life scale and three alternative scoring algorithms for the QLQ-C30 based on (1) the 15 QLQ-C30 scale means; (2) the sum of all individual QLQ-C30 items (except for the item on financial problems); and (3) the sum of the scales assessing physical function, emotional function, quality of life, fatigue, nausea/vomiting, pain, appetite, and diarrhea. For all proposed summary measures, change was categorized in one way or the other into improved, unchanged, and worse. The three alternative scoring approaches performed considerably better than the original, two-item quality of life scale. Although this study documented that the QLQ-C30 global quality of life scale may not be particularly well suited for detecting changes between patient groups and/or changes over time, the alternative summary scoring algorithms proposed were generated in an ad hoc manner, without rigorous empirical testing of hypothesized measurement models.

Cognizant of the need to have a solid empirical basis for any proposed higher summary score for the QLQ-C30, Gundy et al. [10] used structural equation modeling to test seven alternative higher order measurement models for the QLQ-C30. All models exhibited a moderate-to-good model-data fit. The model that showed the best statistical fit (slightly better that the other models) was a two-factor model of physical and mental health. This is conceptually similar to the SF-36 Health Survey component scores, and the factor structure of the PROMIS domain mapping project [10,13–16] Although appealing conceptually, Gundy et al. questioned if this advantage outweighs the model's relatively complex nature, and if perhaps a more parsimonious and simpler model would be more suitable.

The aims of the present study were to (1) identify the best performing higher order model among those suggested by Gundy et al. [10]; (2) test the validity and responsiveness of the best performing higher order factor score(s) as compared to that of the underlying individual scales of the QLQ-C30; and (3) develop an additional scoring algorithm for summary score(s) for the QLQ-C30 on the basis of a higher order measurement model.

2. Methods

2.1. Data source

The QLQ-C30 data used for these analyses were collected originally for the EORTC Quality of Life Cross-