Unbiased measurement of health-related quality-of-life

King, B.L.

Citation for published version (APA):
King, B. L. (2011). Unbiased measurement of health-related quality-of-life

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
In the introduction, **Chapter 1**, we presented the background and aims of the thesis. The general consensus among psychometricians and applied researchers is that measurement invariance is prerequisite for comparisons of HRQoL over time. When measurement invariance is violated and response shift and/or measurement bias is present, the assessment of change is compromised. Despite this general consensus, we do not know the best way to test measurement invariance, which factors will affect measurement bias and response shift and under what circumstances these factors play a role. Hence, the general objectives of this thesis were to investigate measurement invariance in existing sets of HRQoL data in diverse patient populations, to account for any measurement biases or response shifts identified, and assess true effects on unbiased HRQoL. To achieve this, structural equation modeling (SEM) was used. In applying the procedure outlined by Oort [1] to the data used in this thesis we anticipated that methodological problems would arise. As a result, our secondary aim was to address some of these problems in two methodological papers that used empirical examples to illustrate the problem.

In Chapters 2 – 5 we applied SEM to investigate measurement bias and response shift in diverse patient populations. In **Chapter 2** we illustrated two perspectives of measurement bias and response shift that were proposed by Oort et al [2], the measurement and conceptual perspectives. From the measurement perspective, the focus is on bias in the measurement of HRQoL, and from the conceptual perspective the focus is on bias in the explanation of HRQoL. We illustrate both perspectives using HRQoL data collected from patients with breast, lung, pancreatic and esophageal cancer before and after surgery. Using SEM we were able to investigate the presence of measurement bias, response shift in measurement, explanation bias, and response shift in explanation. Additional exogenous variables that we considered included: cancer site, health status, sex, age, optimism, and social comparison. We identified six measurement biases. Five of these were considered response shift, with two reconceptualizations of the domains bodily pain and general health, and three direct effects between sex, optimism, upward comparison and general health. The direct effect between age and physical functioning was constant over time, and therefore not considered
response shift, but bias. Both general health and physical functioning should measure HRQoL exclusively, but actually were indicative of the above mentioned patient characteristics. Two explanation biases were found that suggested optimism and upward comparison directly affect the latent variable Mental HRQoL, independent of patients’ actual health conditions. These effects were stable over time. Using the procedure outlined by Oort [1] and extending it to include additional exogenous variables [2], we were able to model both the measurement and conceptual perspectives.

For the remaining chapters we only focused on the measurement perspective of testing measurement invariance. In Chapter 3 we illustrated the flexibility of SEM. Two different questionnaires, one related to disability (the Disability Scales, Study 1) and the other to HRQoL (the SF-12, Study 2), were measured over three measurement occasions. Patients were newly diagnosed multiple sclerosis patients who were enrolled in a large registry (North American Research Committee on Multiple Sclerosis). In Study 1, the measurement occasions were random and were accounted for by introducing time since diagnosis as an exogenous variable. Health state was operationalized as fixed group membership and thus multiple covariance matrices were included for each health state group. In Study 2, measurement occasions were fixed. Health state was operationalized as time-varying covariates, accounted for by exogenous variables relapse (yes/no) and symptoms (worse/same/better). Overall, in Study 1, progressive and relapsing patients reported greater disability than stable patients but there was little longitudinal change. We did find recalibration response shift in the sensory disability item of the Performance Scales, indicating that stable and relapsing patients initially overestimated their sensory disability. In Study 2, relapse and symptoms were associated with overall HRQoL, a response shift was present, though difficult to interpret, and mental health and vitality were biased with respect to age. No true change in HRQoL was found.

In Chapter 4 we tested measurement invariance in a cross sectional sample of cancer patients who were about to undergo radiotherapy for the first time. Our aim was to identify whether valid comparisons for HRQoL could be
made with respect to the patient characteristics of age, sex, previous treatment for cancer, and information regarding treatment preferences using the EORTC QLQ C30. We believed that valid comparisons may not be possible due to the heterogeneity of the sample, making it highly likely that HRQoL scores would be biased. As the sample was small, we tested invariance using restricted factor analysis rather than multiple-group SEM and included additional variables as exogenous variables. The results indicated that there were two violations of invariance, these violations were associated with age and previous treatment. Older patients reported worse physical functioning, and patients who had received treatment prior to radiotherapy reported worse emotional functioning given their overall latent Functioning of HRQoL. Prior to accounting for bias, the model was not interpretable due to poor fit; the inclusion of the direct effects led to a substantial improvement in overall model fit and a model that could be interpreted.

The final paper in the applied section is Chapter 5. In this chapter our focus was on perceived stigma in primary care patients with depression/anxiety symptoms and/or at-risk drinking behaviors. Patients were randomized into two models of care, enhanced referral or integrated care. Stigma was measured using the SAMSHA Mental Health and Alcohol Abuse Stigma Assessment over three measurement occasions. We used SEM to conduct a longitudinal multi-group (treatment groups) analysis. We also included additional exogenous variables that allowed us to investigate bias in relation to the wording of items and a number of patient characteristics. The results suggested that a two-factor measurement model (Perceived Stigma and Comfort Level) fit the data. One item (item 4) measuring whether it would be difficult for the respondent to start treatment if others knew they were to have treatment was identified as exhibiting recalibration response shift. Considering the impact of response shift and true change on observed change for this item, we see that the observed change is almost fully attributable to response shift. Additionally, four biases were found regarding Item 3 which assessed whether the respondents believe people would think differently of them if they received treatment. This indicates that Item 3 is perhaps too general and certain respondent characteristics had undue influence.
on their responses to this item. Finally, bias was found on Item 6, which assessed respondents comfort in talking to a mental health provider. People with higher mental HRQoL felt more comfortable even when Comfort Level was held constant. While these findings provided insight into the psychometric qualities of the Stigma Assessment, accounting for these biases had little impact on change in Perceived Stigma and Comfort Level over time. Perceived Stigma decreased in both groups over time at a comparable rate. Comfort Level did not significantly change over time for either treatment group.

In Chapters 6 and 7 methodological issues were considered and explored. In **Chapter 6** we reconsidered Oort’s [1] original four step procedure, and proposed a three step procedure to investigate measurement bias and response shift, using SEM. These are: (1) Establishing a measurement model, (2) Testing measurement invariance across measurement occasions, and (3) Testing measurement invariance with respect to exogenous variables. We not only focused on providing a detailed description of these three steps, but in proposing these steps, we also attempted to guard against chance findings and to circumvent the constraint interaction. If either of these problems are ignored, incorrect conclusions regarding bias and change in the common factor means may be drawn. To illustrate these three steps we investigated measurement invariance in a sample of HIV patients, where HRQoL was measured over four biannual measurement occasions. In this illustrative example we identified two instances of response shift in Step 2: biased factor loadings and intercepts of both ‘Health Distress’ and “Energy and Fatigue”, which were primarily due to recalibration. In Step 3 we found one instance of response shift and one of bias: the direct effects between CD4-cell count and Role Functioning were considered to be response shift, whereas the direct effects between CD-4 cell count and Emotional Functioning were stable over time and not considered response shift. In both steps the response shifts identified were difficult to interpret. However, test taking itself can have an effect on response behavior, which may change with time. The respondents may have become more accustomed to both their disease and taking the test, which perhaps led to response shift. Regardless, we believe
that this new procedure does help to safeguard against chance findings and the constraint interaction was successfully avoided.

In **Chapter 7** we investigated two different procedures that can be used to test measurement invariance. Both procedures were argued for on theoretical grounds and then applied to an illustrative example of longitudinal data from lung cancer patients. The first procedure extended the traditional use of modification by incorporating the suggestions of Saris, Satorra and Van der Veld [3]. They suggest including the expected parameter change and the power of the modification indices to assess bias. The second procedure presented is an extension of our three step procedure presented in Chapter 6. We extended the procedure by calculating standardized observed parameter changes as an additional safeguard against chance findings. The standardized observed parameter change values provide information regarding the substantive change in parameters between the fully constrained parameters and the parameters that have been freed to be estimated. The results from the illustrative example differed between procedures with regard to the parameters associated with response shift and the conclusions about change in the common factor means. We concluded that these differences were primarily associated with the two major differences between procedures. The first difference concerns testing parameters as single or multiple parameters. The other difference is related to relying on expected or on observed parameter change.

In the last chapter, **Chapter 8**, we provide a general discussion of the findings of this thesis. In addition to this, we discuss limitations of the studies, and what is still unknown that could lead to future research.
References

