Measuring Change in Quality of Life: Bias in Prospective and Retrospective Evaluation

Christine Blome, PhD*, Matthias Augustin

Institute for Health Services Research in Dermatology and Nursing, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

ABSTRACT

Objectives: Treatment effects on health-related quality of life (QOL) often differ depending on whether they are measured prospectively (before and after treatment) or retrospectively (after treatment only). These two approaches can be subject to different sorts of bias: Prospective evaluations may be biased by scale recalibration (a changed understanding of the response scale), and retrospective evaluations may be biased by recall bias (a wrong assessment of former QOL). **Methods:** On the basis of an analysis of the literature, we present an overview on possible biases in prospective and retrospective measurement of QOL and how these biases are named and defined in the literature. **Results:** The definitions of different biases are inconsistent. Many authors do not clearly distinguish measurement bias from true change. Furthermore, some consider only scale recalibration or only recall bias. **Conclusions:**

Much of the current discussion on bias in prospective and retrospective QOL measurement suffers from unclear definitions, especially of "response shift" and "recall bias," or from neglecting one of the possible biases. We suggest more elaborate definitions for different types of bias and recommend taking both kinds of bias into consideration when measuring change in QOL. The relevance of the different biases depends on the type of study, and so either prospective or retrospective assessment may be more appropriate.

Keywords: quality of life, recalibration, recall bias, response shift, thentest.

© 2015 Published by Elsevier Inc. on behalf of International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Introduction

In addition to prolonging life, the main purpose of medical treatment is to enhance the patients' health-related quality of life (QOL), which is usually considered a multidimensional construct including physical, social, emotional, and other influences of illness on subjective well-being [1,2]. Treatment benefit corresponds with how much QOL increases from before to after the intervention, which can be measured either prospectively or retrospectively, also called "indirect" versus "quasi-indirect" measurement of change [3]. Depending on how change in QOL is assessed—either prospectively or retrospectively—different kinds of bias can occur.

The purpose of this article was to show why the discussion on these biases could gain from a clearer definition of 1) what exactly we wish to measure and 2) what kinds of bias there are.

Validity of QOL Measurement

QOL is a psychological construct and thus an abstract concept that is not directly observable. Instead, it needs to be deduced from patients' responses to questions that are assumed to reflect aspects of their QOL [4]. One, however, can never be sure that the items actually do measure the intended construct. There is no gold standard to compare against—the standardized QOL questionnaires are the best instruments that are available for this purpose—and so convergent validity cannot be determined [5]. All the more important it is to collect data that support (instead of prove) an instrument's validity. This can be achieved either by testing hypotheses on correlations with other variables or on differences between groups of patients, or by asking medical experts and patients to rate the questionnaire's face validity [5]. Basing item development and selection on an explicit theoretical basis, that is, the respective QOL model to be reflected, can also support validity.

Further measurement difficulties arise from the fact that most QOL items have ordinal response scales, meaning that the distances between response options may not relate to equal distances on the underlying continuum of QOL. Transformations such as addition, subtraction, and averaging are inappropriate for ordinal scales. Nevertheless, averaging interval data is quite common in QOL research and is also accepted by many or most researchers in the field [4]. Alternatively, Rasch analysis can be applied during development to ensure that QOL instruments are one-dimensional, which is a prerequisite for building sum scores [6]. It has been debated for decades whether results will be distorted if Likert scales are treated as interval scales or not [7].

* *Address correspondence to:* Christine Blome, Institute for Health Services Research in Dermatology and Nursing, University Medical Center Hamburg-Eppendorf, Martinistraße 52, D-20246 Hamburg, Germany.

E-mail: c.blome@uke.de.

Some authors argue that sum scores over Likert items may be treated as interval data in the same way that sums over binary correct-incorrect items are interval [8,9].

In the light of these methodological difficulties and the nonprovability of construct validity in QOL instruments, it is important to take possible biases in QOL assessment into account to prevent (or at least detect) them. These biases depend in parts on whether QOL is measured prospectively or retrospectively.

Prospective versus Retrospective Measurement of Change

In a prospective measurement, patients rate their QOL both before and after treatment.

In a retrospective measurement, QOL data are collected after treatment only: Patients are asked to 1) rate their current, posttreatment QOL and 2) recall and rate their pretreatment QOL in a "thentest" (or "retrospective pretest") [10,11]. Alternatively, patients directly rate the extent to which their QOL has improved or worsened; this is called "transition rating" or "direct change measurement" [3].

In the so-called thentest method, prospective and retrospective measurement is combined: Patients rate their current QOL at time 1 ("pretest") and time 2 ("posttest"). At time 2, they also rate their former QOL at time 1 retrospectively (thentest). Prospective and retrospective treatment effect can differ widely—regarding not only effect size but also effect direction: In some studies, patients retrospectively rated their QOL better, and in others, worse than in the pretest [12].

On a global level, biases in QOL change measurement can be classified into 1) biases that can impair the validity of prospective change measurement only, 2) biases that can impair the validity of retrospective change measurement only, and 3) biases that can impair both measurements. In the following text, we give an overview over biases belonging to these categories.

Biases Specific to Prospective Measurement

Biases specific to prospective measurement are described under the terms "response shift," "recalibration response shift," or, less commonly, "scale recalibration" [13–15].

Response Shift

Response shift is an umbrella term for different phenomena, all of which imply that a QOL assessment can change although no change in objective circumstances has occurred. According to the well-established classification by Schwartz and Sprangers [10,13], response shift includes three different phenomena.

Recalibration response shift means that in the posttest, patients understand the response scale differently than in the pretest because their internal standards of interpretation have changed. An example: Before treatment, a patient answers the question "How intense is your pain?" with "very intense." After treatment, he still suffers as much pain as before. In the meantime, however, he has met other patients who have even stronger pain. Now he knows that in comparison to others, he has less pain. Accordingly, he now responds with "rather intense"—he chooses a different answer than before treatment although he knows very well that he still suffers as much pain as before. Thus, in the thentest, he rates his pretreatment pain as rather intense, too. Pretest and posttest of this patient are not comparable because his scale interpretation differs. Prospectively, a positive effect is being measured, although no change has

occurred; the retrospective measurement suggests no change and is thus more valid in this example.

Reprioritization response shift means that the values patients assign to different areas of life change between pretest and posttest: They give higher priority to some aspects of QOL whereas others become less important. An example: A patient finds much satisfaction in her work as a car mechanic, which plays a central role for her QOL. After she has been diagnosed with multiple sclerosis, she is not able to work full-time and at full power anymore. This is a great burden for her, and she rates her overall QOL in the pretest as 4 on a scale ranging from 0 (miserable) to 10 (perfect). In the course of her coming to terms with the chronic disease, she devotes herself more to other, less restricted areas of her life: She volunteers in a senior citizens' home and spends more time with her family. Over time, these areas become an ever more important aspect of her QOL appraisal. At posttest time, her handicaps in different areas of life have not subsided; her well-being, however, has greatly improved, and she rates her overall QOL as 7 because of the greater importance she attaches to the areas of volunteering and family.

Reconceptualization response shift means that patients change their definition of what QOL is. Here, not only is the relative importance of QOL components changing but the construct of QOL is also redefined altogether. An example: A young man defines QOL as a state of complete capability and independence. In consequence of a car accident, he nearly loses his eyesight, and so he often needs the help of others in everyday life. After a phase of despair and depression, he learns bit by bit to redefine QOL for himself. To him, QOL is now the ability to enjoy life in spite of physical restrictions, to appreciate the help of others, and to help others in return. At posttest time, his objective impairments in different areas of life have not changed. But yet, his well-being has improved and he rates his overall QOL better because he now bases his assessment on other factors than before.

Critique of Response Shift Definitions

This classification of response shift has been criticized [15,16]: The common term response shift suggests that all three mechanisms cause measurement error, when in fact only recalibration response shift does because here, a change in the patient's response behavior does not correspond with a real change in QOL. Ubel et al. [15] and Ubel and Smith [16] see the other two phenomena, reprioritization and reconceptualization, as adaptations to illness leading to a change in the true value, that is, the QOL actually experienced by a person.

Objective versus Subjective QOL

It could be argued that reprioritization and reconceptualization are also biases because QOL measurement improves although objective circumstances have not improved. We think that at this point, researchers should decide—and clearly state—how they wish to define QOL (the "true value" to be measured) and what exactly they want to measure: Is it objective QOL—objective impairments and circumstances that usually impair people's well-being? Or is it subjective QOL—the well-being actually perceived by a person [17]? In QOL questionnaires, both concepts are often mixed up. For example, in the EuroQol five-dimensional questionnaire (EQ-5D) [18], patients are asked for both how anxious or depressed they are (which is very close to well-being) and how much problems they have in washing or dressing themselves. (In our own works, we, too, have included items on both subjective and objective QOL in the same global score [19].)