Community Dental Health (2016) **33,** 107–115
Received 6 July 2015; Accepted 12 November 2015

doi:10.1922/CDH_3605Robinson09

Choosing a measure of Health Related Quality of Life

P.G. Robinson

School of Clinical Dentistry, University of Sheffield

This paper provides practical advice on the choice of health related quality of life measures. It starts by making explicit a series of underlying assumptions and then advises on selecting a measure as a trade-off between three sets of conditions:

- The purpose of collecting HQoL information, which considers the objectives of the study, the level of analysis, the population to be studied and the audience to whom the data will be presented.
- The qualities of the measure, including the need for a strong conceptual basis, pragmatic considerations, face and content validity, adequate psychometric properties and for the measure to be acceptable to the people participating in the study.
- The use of the measure, including the mode of administration and resource requirements.

An earlier version of this paper was published as Robinson, P.G. (2016): Wahl der messinstrumente zur ermittlung der gesundheitsbezogenen lebensqualitat. In: Kovacs, L., Kipke, R., Lutz, R. (eds) Lebensqualitat in der medizin. Wiesbaden: Springer VS, pp201-222.

Key words: *health related quality of life, methodology, psychometrics, methods, scales, investigative techniques*

Introduction

The last few decades have seen tremendous increases in life expectancy in many parts of the world. Consequently there has been a greater focus on the quality as well as the quantity of life. In addition, health has been increasingly deprofessionalised, which has led to greater lay participation in health care. This deprofessionalisation is manifest in the very concepts of health that we use. We have moved from the purely biomedical to incorporate the psychological and social causes and consequences of health conditions. Cumulatively, these changes result in the need to measure those aspects of quality of life that are related to health. Similar trends echo in oral health where ideas about health have expanded from the biomedical counting of cavities and teeth to incorporate assessments of the impacts of oral conditions on everyday life.

The measurement of health-related quality of life (HQoL) is not easy. The concept is a relatively novel, purely academic and somewhat arbitrary construction with no lay meaning (Tsakos *et al*., 2012). As a consequence it is poorly and inconsistently defined and there can be no gold standards of HQoL against which measures can be validated or calibrated. It may therefore be the case that some measures are not fit for purpose. Nevertheless, health related quality of life is now used for a wide variety of purposes (Table 1), and its measurement is an important part of general and oral health research.

This paper provides practical advice on the choice of health related quality of life measures. It will start making explicit a series of underlying assumptions, which will include defining the terms to be used.

Table 1. Potential uses of HRQoL data (Robinson *et al*., 2002)

Political	• Planning public health policy • Planning resource allocation
Clinical	• Communication tools • Commissioning programs of care • Evaluating • Assessing outcomes of new treatment • Aid understanding of patient perspective • Screening, identifying and prioritising patient problems and preferences • Involving patients in decision making and self-care • Monitoring & evaluating individual patient care • Identifying which patients may benefit from the treatment • Predicting outcomes to provide appropriate care • Audit
Research	• Evaluating interventions • Elucidating the relationships between different aspects of health
Public Health	• Describing and monitoring illness in population • Planning, monitoring and evaluating services • Needs assessment and prioritization • Encouraging greater lay participation in health care
Theoretical	• Exploring models of health • Describing factors influential to health

Correspondence to: Professor Peter G Robinson, School of Oral and Dental Sciences, University of Bristol, Lower Maudlin Street, Bristol, BS1 2LY Email: peter.g.robinson@bristol.ac.uk

The advice on choice of measure will be presented as a trade-off between three sets of conditions:

1. The purpose of collecting HQoL information, which considers the objectives of the study, the level of analysis, the population to be studied and the audience to whom the data will be presented.
2. The qualities of the measure, including the need for a strong conceptual basis, pragmatic considerations, face and content validity, adequate psychometric properties and for the measure to be acceptable to the people participating in the study.
3. The use of the measure, including the mode of administration and resource requirements.

Definitions and Assumptions

For the purposes of this paper Quality of Life is considered to be *an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns* (WHO, 1995).

The *position in life* might include many domains such as personal safety, financial security and emotional needs. Clearly some of these domains are outwith the purview of health care and so it is necessary to distinguish between these non-medical factors and *health*-related quality of life, which is the topic of this paper. For our purposes health-related quality of life can be regarded as *the functional effect of a medical condition and/or its consequent therapy upon a person*. (Adapted from ISOQoL, 2015).

Given the scope of this journal, the focus of this paper is a further subset of HQoL: Oral Health related Quality of Life (OHQoL), which is *the impacts of oral disorders on everyday life that are important to people and of sufficient magnitude to affect perception of their life overall* (Locker and Allen, 2007).

HQoL measures tend to take the form of items addressing different domains, with the items answered on ordinal scales. Scores may be recorded as the prevalence of impact (the proportion of participants who have scored any item above a given threshold), extent of impact (the proportion of items scored above a threshold by each participant) or as a total score (the sum of the item codes). As illustration, the Oral Health Impact Profile (OHIP) and Oral Impacts on Daily Performance (OIDP) are commonly used to measure OHQoL.

OHIP enquires about functional limitations, pain, psychological discomfort and disability, physical disability, social disability and handicap arising from oral conditions. The original version comprised 49 items that enquire about the frequency of impacts on five point Likert Scales (Slade and Spencer, 1994). Subsequent versions have fewer items (Slade 1997) or have been adapted to consider the consequences of specific oral conditions, such as missing teeth (Allen and Locker, 2002). OHIP is well evaluated and has been widely tested in hundreds of studies.

OIDP is concerned with the disabling and handicapping consequences of oral conditions on the ability to eat, speak, clean one's teeth, sleep, smile, enjoy contact with others, work and maintain one's emotional state. It considers both the frequency and severity of impacts on these daily performances (Adulyanon and Sheiham, 1997).

It is not the purpose of this paper to provide a rationale for assessing the subjective impacts of oral conditions, however, doing so necessarily emphasises the perspective of the person experiencing them. This person-centredness carries key implications. The major implication arises from the first phrase of the WHO definition above relating to the *individual's perception*. We are concerned with the individual's subjective feelings of personal well-being in relation to health. Therefore HQoL data are most valid if they are self-assessed. An indirect consequence of this is that the measures we use should ideally be derived with lay input so that they enquire about the things of relevance to lay people and use the language that they use.

It is also helpful to distinguish this personal feeling of well-being from functioning (what people can do) and from utility, where participants assign values to states, which form the basis of health economic measures, neither of which are considered in this paper.

Your Purpose

The starting point when selecting an HQoL measure is to think about why you are measuring it in the first place. The objectives of any study will include both explicit and implicit assumptions about the analytic objectives, the level of analysis, the population being studied and the audience for the findings, all of which will inform the choice of measure.

Analytic purpose and level of analysis

Analytic objectives can be categorised as descriptive (where the intention is to describe the level of impact in some way), discriminative (which compare levels of impact in different groups or conditions) and evaluative (where the intention is to assess responses to health care, either in individual patients or to contribute to the evidence-base for an intervention) (Hyland, 2003).

The level of analysis refers to whether entire populations are studied (as in public health or to prioritise care according to need), samples (as is usually the case in research) or in individuals (such as in clinical practice, or when screening). Measures used with individuals must be very accurate whereas the limitations of modest accuracy can sometimes be overcome statistically when assessing HQoL in groups or samples of people.

The interactions between analytic purpose and level of analysis can be arranged in a matrix to guide the selection of an HQoL measure (Table 2).

Evaluative measures for use in groups or samples (cell C2 in Table 2) should focus on the specific aspects of health that will respond (either positively or negatively) to the intervention so that they exclude the random variation ('background noise') in scores arising from the irrelevant items (Hyland, 2003). In order to detect changes it is essential that evaluative measures are stable in the absence of true change. Ideally they will be precise to detect small gradations of change, which may necessitate the use of more points on the measurement scales (such as seven point Likert scales).

Table 2. Interactions between analytic objective and level of analysis

	Population	*Sample*	*Individual*
Describe	A1	A2	A3
Discriminate	B1	B2	B3
Evaluate	C1	C2	C3

Evaluative measures should also avoid floor and ceiling effects (where many participants report the lowest or highest possible scores) so that most items score in the middle of their range.

Discriminative measures used in samples or groups (B2 in Table 2) are required to detect wide ranges of different levels of severity of impact (known as 'content validity') so that the scores have good spread for correlational analysis (Hyland, 2003). It follows from this that some of the items might have floor or ceiling effects that award higher or lower scores only to the most severely or mildly affected participants. Longer measures with more items may be necessary in order to detect differences between groups although fewer response categories may be needed, so that participants might only respond to each item with a yes/no answer.

Our recent development of an OHQoL measure to detect the specific impacts of dentine hypersensitivity illustrates these points very well (Boiko *et al*., 2010). Dentine hypersensitivity can cause short sharp dental pain when eating or drinking something cold, such as ice-cream. We had been commissioned to develop the measure for a manufacturer to help them evaluate toothpastes used to manage this problem.

In fact dental sensitivity to ice-cream is very common. This meant that there was a ceiling effect in the items related to ice-cream (many participants reported the impact) and so the ice-cream related items were useful for discriminating between healthy and mildly affected people (eg in screening) (Robinson *et al*., 2014). However, the pain from such a cold stimulus is rather unresponsive to treatment, meaning that the ice-cream related items were not useful in evaluations of treatment.

Measures used with individuals (typically patients) are often being used with very specific purposes and so require fewer items (Hyland, 2003). Discriminative uses with individuals (B3 in Table 2) might identify specific impacts to inform clinical decisions. Hence the items need only be relevant to that particular patient group or condition. This direct relevance to a particular condition is a form of content validity and should be evident as a high internal reliability of the measure (that is all the items relate to the same underlying construct). Internal reliability is evaluated using Cronbach's alpha, and alphas greater than 0.9 are recommended for measures to be used in individuals (Nunally, 1978).

Evaluative measures for use in individuals (C3 in Table 2) should focus only on the individual aspects of the patient's health that will change. For example, a sleeping pill should help the patient sleep and reduce daytime tiredness. It therefore needs very few items but those items should be very responsive, which may require small gradations (i.e. greater precision) to detect small changes. However, this narrow focus may exclude unanticipated or broader subjective changes, so that sharply focussed measures might be supplemented with generic or global measures.

By contrast measures to be used in entire populations (A1, B1, C1) will naturally involve many participants and so the minimisation of participant burden and administrative cost may be paramount. Such measures may contain very few items and be generic in scope, capturing the impacts of a large range of conditions.

The population under study

Health can be regarded as the manner in which one interacts with one's environment (Dubos, 1959). Culture forms part of that environment and also forms the lens through which one view's the world (Helman, 2003). Therefore health and culture are intricately linked, and of course culture varies between and within populations. It follows then that the items within HQoL measures must be relevant to the culture in which they are being used. Our item on difficulties eating ice-cream would not yield useful information if used in a population that rarely ate ice-cream for instance.

Consequently measures should be studied carefully to ascertain whether they appear to be measuring HQoL and locally relevant aspects of it (this is face and content validity, and a recurring theme in this paper). In addition, elaborate processes exist to translate measures between languages to ensure both linguistic and cultural equivalence. Translated measures also require revalidation in their new language (Herdman *et al*., 1998; John *et al*., 2002).

Careful translation of this type is necessary and there are often small but important differences between the different forms that result. Interestingly, the domains of HQoL appear to be relatively consistent across cultures, but the emphasis placed upon them may vary.

The cognitive ability of participants will also influence the selection of measure. Children and people with cognitive impairments will require measures constructed in a manner appropriate to their capacity. This may involve more straightforward wording and the use of pictures that may well restrict the scope of the measure. It should be remembered that cognitive ability is dynamic, so that some impairments are progressive, as is the case with Alzheimer's disease. Conversely, as children age their reading and intellectual abilities develop.

A further consideration when investigating children and young people's HQoL is that different aspects of health become more or less important to them over time. For example, we might expect adolescents to become more concerned about their appearance and embarrassment than younger children. For these reasons some HQoL measures have analogous forms specific to different age groups (Jokovic *et al*., 2002; 2004).

Audience

The final aspect of the purpose of the study to consider is the intended audience for the results. Possible recipients of research findings might include lay people, patients, planners, health care workers or academics, all of whom will have different needs for the data, areas of interest and levels of interpretation. One can easily imagine that a politician would *pretend* to understand DMFT scores (the mean number of decayed, missing and filled teeth) but would be genuinely interested in the proportion of children in their constituency (and their parents, who are registered voters) kept awake by toothache. Data collection should therefore be planned with this in mind so that the results are adequate for their purpose, both providing the necessary information and not collecting data that will not be used.

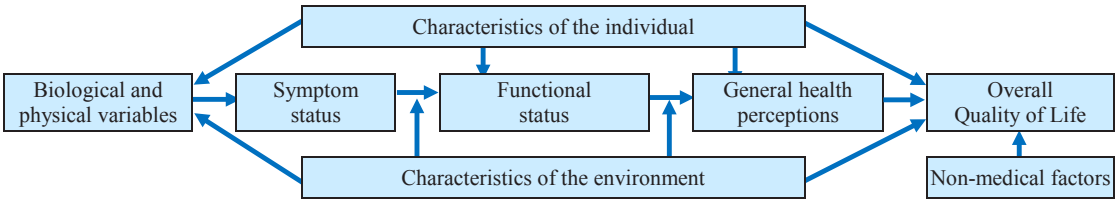

Figure 1. *The Wilson and Cleary model (1995) linking clinical variables with Health-Related Quality of Life: a conceptual model of patient outcomes*

The qualities of the measure

Measures of HQoL need a firm conceptual basis and should fulfil certain pragmatic considerations. They must also have face and content validity, adequate psychometric properties and be acceptable to the participants providing the information.

Conceptual clarity

As has already been noted, the concept of HQoL is purely academic with limited lay meaning. Because of this softness of the concept it is essential that each measure is based on an explicit definition of HQoL. Furthermore, the measure should relate to an underlying theoretical model or construct. For example, both OHIP and OIDP are founded on Locker's conceptual model of oral health (Locker, 1988; Slade and Spencer, 1994; Adulyanon and Sheiham, 1997) and our Dentine Hypersensitivity Experience Questionnaire (DHEQ) was based on Wilson and Cleary's (1995) model linking clinical variables to quality of life (Figure 1) (Boiko *et al*., 2010).

This conceptual clarity serves a number of purposes. It makes the assumptions underlying the measure explicit and also helps to ensure that the measure is compatible with current understanding. In addition, the underlying framework guides the validation of the measure by as we can determine whether the scores of the measure correlate with the framework. This 'construct validity' has two components; we would expect the HQoL scores to be related to other domains in the construct (convergent validity) and would similarly expect the scores to be unrelated to domains that are not present in the construct (divergent validity). Although this paper is more concerned with the selection of existing measures rather than developing new ones, it is essential to choose a model that suits one's purpose, rather than adhering to one with which you are familiar, but which does not include the factors relevant to your current study.

Pragmatic considerations

The development of an HQoL measure is time consuming and costly and there will not be a body of comparator data for a new measure. It is therefore immensely preferable to use an existing measure whenever possible. As we have already seen, the measure selected must be appropriate to the participants under investigation.

MacEntee and Brondani's (2015) comprehensive review indicates that OHIP is a widely used family of measures of different length that may be administered in a variety of ways and with numerous published data for guidance and clarification. The widespread use of OHIP does not preclude all of the considerations raised in this paper; for instance it may not be sensitive to the impacts of some dental conditions and its factor structure may require revision (Bekes *et al*., 2009; Baker *et al*., 2008). Nevertheless, OHIP can be used as a reasonable starting point in the selection of an OHQoL measure.

It's important that the measure must guide knowledge more than simply providing a classification, which is why it is necessary to consider the purpose of the investigation. The investigators can then clearly articulate the use of the measure at the relevant level and analytic purpose so that they can search for an evidence-base for its use in similar circumstances.

Pragmatic considerations also require the results to be readily interpreted and applicable to policy where necessary.

Face and content validity

Face validity refers to whether a measure looks as though it will measure the things of interest. Content validity refers to the relevance and coverage of questions; that is whether the measure enquires about all the different aspects of the condition. As we have seen, content validity may vary between discriminative and evaluative measures, so that evaluative measures must focus on aspects of the condition that may change as participants' health improves or deteriorates. In this case content validity involves being responsive to different levels of severity.

Face validity and content validity are assessed by studying the items and relating them to the experiences of people with the condition. Validation is typically carried out by academics with expertise in the condition, but may be supplemented by the insights of patients or other lay people with first-hand experience of it. Face and content validity are absolutely crucial in the selection of the measure. Although they require no mathematical tests they are indicated by floor and ceiling effects. They are also deceptively difficult to assess, hence they can be under-valued by researchers and not given sufficient attention.

As an example of less than ideal content validity, an early evaluation of OHQoL measures found profound floor effects. One quarter of patients attending a dental emergency department reporting no impacts on their quality of life using one measure. The measure in question had been designed to capture only the most disabling and handicapping consequences of oral conditions and also had a long reference period, whereas emergencies are acute by definition (Robinson *et al*., 2003).

The reference period is the window of time for which participants are required to report their experiences. So, for example, OIDP enquires about the frequency and severity of impacts on daily life over the previous six months (Adulyanon and Sheiham, 1997). Reference periods are especially important in evaluative studies. A longer period is useful to detect differences in infrequent impacts to allow treatment effects to emerge. Conversely if it is too long the evaluation is delayed and there is greater risk of recall bias. Longer reference periods are also relatively insensitive to short term impacts, such as the transient discomfort after treatment (Reissmann *et al*., 2015).

However, the reference periods of measures can be adjusted to some extent, although this prevents comparison of results between studies using different periods.

In another example the content validity of an OHQoL measure produced conflicting results when evaluating a reservoir bite guard (Robinson *et al*., 2005). The bite guard was a soft plastic gum shield that leaked artificial saliva into the mouths of people with dry mouth during the night. Our randomised controlled trial of the bite guard used both OIDP and OHIP14 but only OIDP detected a treatment effect. The purpose of the bite guard was to help people sleep, but sleeping is not considered in OHIP14! (Slade, 1997).

Specific approaches to assessing face and content validity involve checking that the measure was developed using qualitative data from a similar population, by piloting it and by using condition specific measures. Content validity can also be assessed by checking that the items are relevant to the known impacts of condition and that they detect variations in its extent and severity.

The contrast between generic and condition-specific measures is helpful. Generic measures detect the impacts of a variety of conditions. They are therefore useful for comparing the impacts of different conditions but less sensitive or responsive to the specific impacts of a given condition. Heydecke and colleagues (2003) showed that generic measures could not distinguish between the benefits of different forms of dental treatment. Nevertheless they may be of value in evaluative studies as they may detect unanticipated benefits or side effects of treatment. By contrast specific measures focus on the known impacts of that condition, organ or intervention and so are more sensitive to those impacts and responsive to changes in them, but less useful for comparing the impacts of different diseases (Brazier and Fitzpatrick, 2002).

Care is required when applying these generalisations. For example, OHQoL measures are organ specific but not specific to any particular oral condition, as has been illustrated by two studies of OHQoL in people with dentine hypersensitivity. Bekes and colleagues (2009) found that the scores for a generic measure (OHIP14) varied by less than 10% between people with and without sensitive teeth. On the other hand, a condition specific measure (DHEQ) was able to discriminate between the impacts of mild and severe sensitivity as experienced by people with the condition in the general population and those taking part in a trial of toothpastes for it (Boiko *et al*., 2010).

It is often the case that a single questionnaire is insufficient or ideal in a study and it may be better to deploy a battery of measures. A generic measure can be used to compare the experience of participants with those with other conditions and to identify any unexpected effects. This can be complemented by a condition specific measure that will be sensitive to anticipated impacts and will respond to changes. These difficulties have been anticipated in the development of some measures, which have been constructed in a modular format. The EORTC has both generic core modules as well as bolt-on condition specific modules (EORTC, 2015).

It is often valuable to incorporate a small number of global questions that can serve as anchors for construct validation and when assessing responsiveness. For example, a global health rating simply asks participants to rate the health of their mouth as excellent, very good, good, fair or poor (Dolan, 1998). Locker and Allen (2007) also recommended including items that place the impact of the condition in the context of life overall (that is, in relation to non-medical aspects of quality of life). Such broader questions prevent data collection being too specific and so may capture any unanticipated effects.

Psychometric properties

As already described, most HQoL measures are scales combining loosely related items. The softness of the concept of HQoL often prevents any validation against a gold standard, because there is none. Test theory attempts to overcome these difficulties of validation by stressing reliability, and we are left with a battery of tests beloved of PhD students that appear to acquire gravitas because they are numeric and seemingly objective. In practice, many HQoL measures meet basic psychometric standards, but this is not a substitute for face and content validity (For example, see Luckett *et al*., 2011).

Psychometrics are covered very well elsewhere (Streiner and Norman, 2003), so will be summarised here in three broad categories; reliability, validity and precision, only as far as they pertain to the selection of a measure.

There are two components of reliability: internal reliability, which is the extent to which the items in a measure all consider the same underlying construct and test-retest reliability, which is the stability of the scores from repeated testing in participants whose quality of life is stable.

Internal reliability is evaluated using tests such as Cronbach's alpha, alpha if item deleted and split half reliability tests. Most carefully designed measures have alphas in excess of 0.7, which is considered to be the threshold acceptable for use in groups. Alphas greater than 0.9 are required for measures used with individuals (Nunally, 1978). It should be noted that there is a mathematical artefact by which reliability increases when there are more items in a measure, so that internal reliability is rarely the sole criterion for selecting a measure.

Test-retest reliability is evaluated using the Intra-Class Correlation (ICC) co-efficient. It is especially important in evaluative studies as small treatment effects might be masked by random variation in unstable measures. Figure 2 depicts the results of a randomised controlled trial of school-based health promotion in which low scores indicate good quality of life. As can be seen, there was a small but significant difference between the groups at follow up. This difference was small in part because quality of life appeared to improve in both groups (scores in HQoL measures often appear to improve with repeated administration). However, this difference would not have been significant had there been large variations in each group.

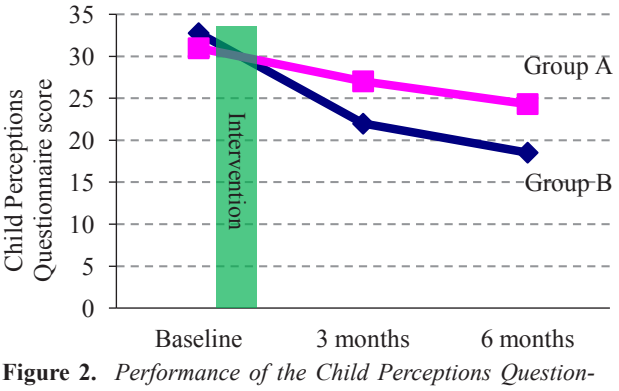

Figure 2. *Performance of the Child Perceptions Questionnaire in a randomised controlled trial of a health promotion intervention*

Validity refers to whether a measure is measuring what it is supposed to measure and can be considered in three broad categories: criterion validity, construct validity and responsiveness.

Criterion validity should be assessed by comparing the scores from the measure against a gold standard, but as we have seen, this is rarely present. Construct validity considers the extent to which the scores from the measure correlate to the underlying construct (convergent validity) and do not correlate to unrelated ideas (divergent validity). Both forms are assessed using appropriate hypothesis tests. Figure 3 shows a clear correlation between Dentine Hypersensitivity Experience Questionnaires scores and the effect of the condition on life overall (Machuca *et al*., 2013). We would not expect such a correlation with eye colour for instance. Convergent validity should involve significant but only mild to moderate associations, as very strong correlations might suggest that the new measure is offering no additional insights.

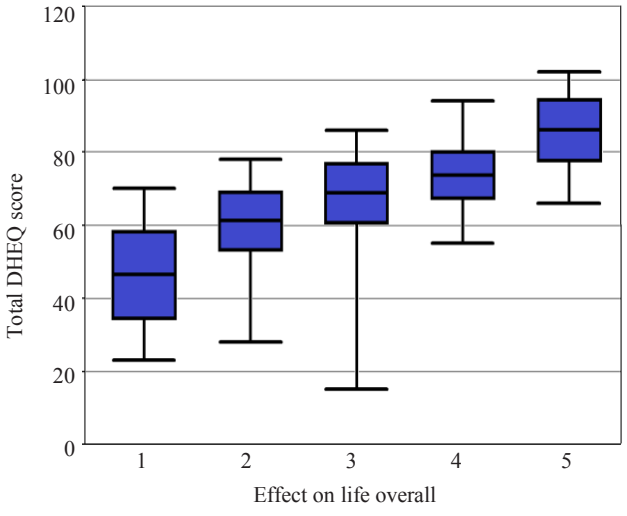

Figure 3. *Correlation between Dentine Hypersensitivity Experience Questionnaire (DHEQ) scores and Effect on Life Overall among people with dentine hypersensitivity*

A final aspect of construct validity is structural validity, which refers to the different dimensions of a scale. For example, OHIP (Slade and Spencer, 1994) was designed to detect impacts on seven dimensions of Locker's (1988) model of oral health: functional limitation, pain, psychological discomfort, physical disability, psychological disability, social disability and handicap. Factor analysis is often used to identify or confirm the domains or dimensions within a multifactor scale. Ideally there should be strong correlations between the items on each dimension and individual items should relate to ("Load on to") only one dimension. Baker and colleagues (2008) also used structural equation modelling to assess the relationships between the dimensions of OHIP. Exhaustive exploratory and confirmatory factor analyses of OHIP provide strong evidence that OHIP actually contains four highly correlated dimensions, which John and colleagues (2014a;b) termed oral function, oro-facial pain, oro-facial appearance and psychosocial impact, however they also felt that OHQoL could be summarised with a single score.

The presentation and interpretation of OHQoL data are discussed later in this manuscript.

Responsiveness is the ability of a measure to detect changes over time, and is therefore very important in evaluative studies. It may be assessed using one or more of three broad strategies (Baker *et al*., 2014; Revicki *et al*., 2008). Within individuals the effect size is calculated as the change over time expressed as a proportion of the baseline standard deviation. Effect sizes of greater than 0.5 are regarded as moderate and greater than 0.8 as large (Cohen, 1998). It should be noted that the effect size is a function of the effectiveness of the intervention and the spread in the data as well as the ability of the measure to respond to changes. The second approach uses an external referent, such as a global change rating. Change indices such as Cohen's effect size or standardised response means may then be calculated for each change category. Alternatively a minimum important difference can be calculated as the mean HQoL change in participants who reported any improvement in life overall. The final approach calibrates responsiveness against treatments of differing efficacy and may be expressed as an observed treatment effect, a relative percent improvement or standardised effect size by trial arm

The precision of a measure is its ability to recognise fine distinctions between different levels of impact and again it is especially important for evaluative measures, although less so for discrimination.

As might be imagined, no single measure will perform very well across all of these properties. Stability is in almost direct tension with both responsiveness and precision because a measure that is stable is less likely to respond to change. Similarly, one way to increase the reliability of a measure is to increase the size of the gradations (reduce precision). Researchers must trade off all the different properties according to the purpose of their study.

Acceptability

Our participants face time, emotional and cognitive challenges when they give us personal information. Burdening them with measures that take a long time to complete is unfair in its own right. In addition, it may result in participants failing to complete the study, omitting questionnaire items or falling into response sets, where they lose concentration and score every item in the same way. To a certain extent some of these difficulties can be overcome by making measures user-friendly. However, the notion of respecting our participants is reasonable. Some researchers have suggested maximum numbers of items in questionnaires or interview schedules for these reasons. These upper limits depend on the mode of administration (telephone, 20 items; mailed questionnaires, 60 items; face to face interview 80 items) and are provided here only as a guide (Furlong *et al*., 2005).

These numbers can be exceeded if the completion of questionnaires is supervised (for example, if children complete them at school, supervised by a teacher or researcher) or by the use of inducements (prizes, honoraria, etc.) to recognise or reward participants for giving their time. Both of these approaches bring their own challenges. There may be concerns about the independence of data collected under supervision, although we have used this method many times to gain seemingly valid data (Baker *et al*., 2010; Gururatana *et al*., 2014; Nammontri *et al*., 2013).

Inducements carry both practical and moral implications that might best be discussed with local ethics committees during research protocol development.

One other aspect of acceptability that requires consideration is the threats to participants' self-esteem when they answer questions about personal and sensitive topics (Waters *et al*., 2009). Once again this is an area where it may be invaluable to involve lay people affected by the condition throughout the research process, so that they can advise on strategies to overcome these challenges. At the lesser end of the scale we observed that whilst answering questions about dentine hypersensitivity sensitised some people to the condition, for others it helped them see their pain in context, as a relatively minor problem (Krasuska, 2014).

The advice on the features of the measures in this section can be supplemented by using checklists or tools specifically designed to appraise health-related quality of life measures. Those produced by the Scientific Advisory Committee of the Medical Outcomes Trust (2002) and by Terwee and colleagues (2007) consider not only the psychometric properties but also the approaches taken in their development and the breadth of their evaluation.

Using the measure

We have seen that measures may be administered as paper questionnaires, as interview schedules or via electronic means. Whilst pragmatic considerations often mean the mode of administration is determined by the resources available in relation to the time, staff and equipment required, some modes are better suited to particular participant types or study objectives. For example, Porritt and colleagues (2014) used smart phones to collect daily assessments of the impact of sensitive teeth. To further complicate matters, some measures are better suited to different modes of administration. For instance, complex measures with contingent questions are completed more comprehensively when administered in interviews or using automated electronic data collection (Robinson *et al*., 2003).

As the mode of administration may influence both the level of completion and the scores derived, the mode should be consistent throughout a study. Furthermore, results collected using different modes should be compared only with great care.

The interpretation of HQoL data is beyond the scope of this paper, but will influence the selection of a measure. A key issue pertains to the lack of intuitive meaning in HQoL data, which can manifest in many ways. For example, a single score aggregates variations across different dimensions and so may mask important distinctions in impact. The extraction of a painful tooth will relieve pain but this impact may be replaced by the psychological discomfort of an unsightly gap. We know that some scores are worse than others, but how much worse? At what point does an impact become clinically meaningful? Researchers are advised to think very carefully about the kinds of impacts they anticipate in their studies and to check that the measure they select will allow them to detect and interpret these impacts. Tsakos and colleagues (2012) provide a good review of the interpretation of OHQoL data.

A central tenet of this paper is that HQoL data should be self-reported whenever possible. Unfortunately, this is not possible in young people (the age of seven years seems to be a lower threshold) or in those with other cognitive impairments (Waters *et al*., 2009, Feeny *et al*., 1998). In these cases proxies in the form of relatives, carers or health care workers can provide data. In general the use of proxies yields biased data that suggest worse HQoL than those provided by the persons themselves (Schölzel-Dorenbos *et al*., 2007). Proxies may be able to assess functional impacts relatively accurately, but may omit emotional and spiritual aspects of life (Waters *et al*., 2009).

Conclusions

Health related quality of life is now an outcome whose measurement is essential in health research. This paper has overviewed some complexities of measuring HQoL and the consequent difficulties of choosing a measure to do so. The diverse requirements outlined here mean that, despite the vast numbers of measures already available, there is rarely a perfect fit for any situation. Instead researchers should select HQoL measures by making trade-offs between the purpose of collecting the data, the inherent qualities of the measures and the practicalities of their use to find the optimal fit for their situation. The creation of a new measure is a considerable task in its own right and should only be undertaken if there is sufficient time for its development and validation before it is used substantively.

In summary, the steps to select an HQoL measure are to:

1. Make an explicit statement of your analytic purpose, level of analysis and intended audience as you would do in any research protocol
2. Scrutinise the existing measures according to the requirements of step 1 above
3. Determine whether they measure what you are interested in. Do they enquire about the relevant aspects of life? If you need them to, will the items discriminate between people or respond to variations in health?
4. Will it be feasible to use the existing measures in terms of the cognitive abilities of your participants, the burden to them and the administrative costs?
5. Are the psychometrics acceptable?
6. Use existing measures if possible.

Acknowledgment

The author of this paper and the editor of this journal are indebted to László Kovács, Roland Kipke, and Ralf Lutz, editors of the original publication, Lebensqualität in der Medizin, for their kind permission to publish this translation.

References

Adulyanon, S. and Sheiham, A. (1997): Oral Impacts on Daily Performances. In, Slade GD (Ed), *Measuring Oral Health and Quality of Life*. University of North Carolina, Chapel Hill, 151-160.

Allen, F. and Locker, D. (2002): A modified short version of the oral health impact profile for assessing health-related quality of life in edentulous adults. *International Journal of Prosthodontics* **15**, 446-450.

Baker, S.R, Gibson, B., and Locker, D. (2008): Is the oral health impact profile measuring up? Investigating the scale's construct validity using structural equation modelling. *Community Dentistry and Oral Epidemiology* **36**, 532-541.

Baker, S.R., Mat, A. and Robinson, P.G. (2010): What psychosocial factors influence adolescents' oral health? *Journal of Dental Research* **89**, 1230-1235.

Baker, S.R., Gibson, B.J., Sufi, F., Barlow, A. and Robinson, P.G. (2014): The Dentine Hypersensitivity Experience Questionnaire: a longitudinal validation study. *Journal of Clinical Periodontology* **41**, 52–59.

Bekes, K., John, M.T., Schaller, H-G. and Hirsch, C. (2009): Oral health-related quality of life in patients seeking care for dentin hypersensitivity. *Journal of Oral Rehabilitation* **36**, 45–51.

Boiko. O.V., Baker, S.R., Gibson, B.J., Locker, D, Sufi, F., Barlow, A.P.S. and Robinson, P.G. (2010): Construction and validation of the Quality of Life Measure for dentine hypersensitivity (DHEQ): *Journal of Clinical Periodontology* **37**, 973–980.

Brazier, J. and Fitzpatrick, R. (2002): Measures of health-related quality of life in an imperfect world: a comment on Dowie. *Health Economics* **11**, 17-19.

Cohen, J. (1988): *Statistical power analysis for the behavioural sciences* 2nd edn. New York: Lawrence Erlbaum, 273-288.

Dolan, T.A., Peek, C.W., Stuck, A.E. and Beck, J.C. (1998): Three-year changes in global oral health rating by elderly dentate adults. *Community Dentistry and Oral Epidemiology* **26**, 62-69

Dubos R (1959): *The mirage of health: utopia, progress and biological change*. London: Harper Collins.

EORTC (2015): *Quality of Life*. http://groups.eortc.be/qol/eortc-qlq-c30

Feeny, D., Juniper, E.F., Ferrie, P.J., Griffith, L.E. and Guyatt, G. (1998): Why not just ask the kids? Health-related quality of life in children with asthma. In: Drotar, D. and Mahwah, N.J. (Eds) *Measuring Health-Related Quality of Life in children and adolescents implications for research and practice*. Lawrence Erlbaum Associates Publishers, pp171-185.

Furlong, W., Barr, R.D., Feeny, D. and Yandow, S. (2005): Patient-focused measures of functional health status and health-related quality of life in pediatric orthopedics: A case study in measurement selection. *Health and Quality of Life Outcomes* **3**, 3.

Gururatana, O., Baker, S.R. and Robinson, P.G. (2014): Determinants of children's oral-health related quality of life over time. *Community Dentistry and Oral Epidemiology* **42**, 206–215.

Helman, C.G. (2003): *Culture Health and Illness* (5th Ed). Hodder Arnold, London

Herdman, M., Fox-Rushby, J. and Badia, X. (1998): A model of equivalence in the cultural adaption of HRQoL instruments: the universalist approach. *Quality of Life Research* **7**, 323-335

Heydecke, G., Boudrias, P., Awad, M.A., De Albuquerque, R.F., Lund, J.P., Feine, J.S. (2003): Within-subject comparisons of maxillary fixed and removable implant prostheses: Patient satisfaction and choice of prosthesis *Clinical Oral Implants Research* **14**, 125-30.

Hyland, M.E. (2003): A brief guide to the selection of quality of life instruments. *Health and Quality of Life Outcomes* **1**, 24

ISOQoL (2015): *What is Health-Related Quality of Life Research?* www.isoqol.org/about-isoqol/what-is-health-related-quality-of-life-research.

Jokovic, A., Locker, D., Stephenson, M., Kenny, D., Tompson, B. and Guyatt, G.H. (2002): Validity and reliability of a questionnaire for measuring child oral health related quality of life. *Journal of Dental Research* **81**, 459-463.

Jokovic, A., Locker, D., Tompson, B. and Guyatt, G.H. (2004): Questionnaire for measuring oral health related quality of life in eight- to ten-year-old children. *Pediatric Dentistry* **26**, 512-518.

John, M.T., Patrick, D.L. and Slade, G.D. (2002): The German version of the Oral Health Impact Profile-translation and psychometric properties. *European Journal of Oral Science* **110**, 425-433.

John, M.T., Reissmann, D.R., Feuerstahler, L., Waller, N., Baba, K., Larsson, P., Celebić, A., Szabo, G., Rener-Sitar, K. (2014a): Exploratory factor analysis of the Oral Health Impact Profile. *Journal of Oral Rehabilitation* **41,** 635-643.

John, M.T., Feuerstahler, L., Waller, N., Baba, K., Larsson, P., Celebić, A., Kende, D., Rener-Sitar, K. and Reissmann, D.R. (2014b) Confirmatory factor analysis of the Oral Health Impact Profile. *Journal of Oral Rehabilitation* **41**, 644-652.

Krasuska, M. (2014): *Response shift and oral health quality of life in dentine hypersensitivity*. Thesis, University of Sheffield. http://etheses.whiterose.ac.uk/8310

Locker, D. (1988): Measuring oral health: A conceptual framework. *Community Dental Health* **5**, 3-18.

Locker, D. and Allen, F. (2007): What do measures of 'oral health-related quality of life' measure? *Community Dentistry and Oral Epidemiology* **35**, 401-411.

Luckett, T., King, M.T., Butow, P.N., Oguchi, M., Rankin, N., Price, M.A., Hackl, N.A. and Heading, G. (2011): Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. *Annals of Oncology* **22**, 2179-2190.

MacEntee, M.I. and Brondani, M. (2015): Cross-cultural equivalence in translations of the oral health impact profile. *Community Dentistry and Oral Epidemiology* In Press.

Machuca, C., Baker, S.R., Sufi, F., Mason, S., Barlow, A. and Robinson, P.G. (2014): Derivation of a short form of the Dentine Hypersensitivity Experience Questionnaire. *Journal of Clinical Periodontology* **41**, 46-51.

Nammontri, O., Robinson, P.G. and Baker, S.R. (2013): Enhancing oral health via sense of coherence: a cluster randomized trial. *Journal of Dental Research* **92**, 26-31.

Nunally, J.C. (1978): *Psychometric Theory* (2nd Edn). New York: McGraw-Hill.

Porritt, J.M., Sufi, F., Barlow, A. and Baker, S.R. (2014): The role of illness beliefs and coping in the adjustment to dentine hypersensitivity. *Journal of Clinical Periodontology* **41**, 60-69.

Reissmann, D.R., Poulopoulos, G. and Durham, J. (2016): Patient perceived burden of implant placement compared to surgical tooth removal and apicectomy. *Journal of Dentistry* **43**, 1456-1461.

Revicki, D., Hays, R.D., Cella, D. and Sloan, J. (2008): Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* **61**, 102–109.

Robinson, P.G., Baker, S.R. and Gibson, B.J. (2014): Ice cream-related quality of life: constructing a questionnaire to capture changes in the impacts of dentine hypersensitivity. In: *Dentine Hypersensitivity: Developing a person-centred approach to oral health*. Robinson, P.G. (Ed). London: Elsevier.

Robinson, P.G., Carr, A.J. and Higginson, I. (2002): Choosing a measure of health related quality of life. In: Carr, A.J., Higginson, I., Robinson, P.G. (Eds) *Quality of Life*. London: BMJ Publishing.

Robinson, P.G., Gibson B., Khanm F.A. and Birnbaum, W. (2003): Validity of two Oral Health Related Quality of Life measures in a UK setting. *Community Dentistry and Oral Epidemiology* **31**, 90-99.

Robinson, P.G., Pankhurst, C.L. and Garrett, E.J. (2005): Randomised controlled trial: Effect of a reservoir biteguard on quality of life in xerostomia. *Journal of Oral Pathology and Medicine* **34**, 193-197.

Schölzel-Dorenbos, C.J.M., Ettema, T.P., Bos, J., Boelens-van der Knoop, E., Gerritsen, D.L., Hoogeveen, F., de Lange, J., Meihuizen, L. and Dröes, R-M. (2007): Evaluating the outcome of interventions on quality of life in dementia: Selection of the appropriate scale. *International Journal of Geriatric Psychiatry* **22**, 511-519.

Scientific Advisory Committee of the Medical Outcomes Trust (2002): Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research* **11**, 193-205.

Slade, G.D. (1997): Derivation and validation of a short-form oral health impact profile. *Community Dentistry and Oral Epidemiology* **25**, 284-290.

Slade, G.D. and Spencer, A.J. (1994): Development and evaluation of the Oral Health Impact Profile. *Community Dental Health* **11**, 3-11.

Streiner, D.L. and Norman, G.R. (2003): *Health measurement scales* 3rd edn. Oxford: Oxford University Press.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A., Knol, D.L. and Dekker, J. (2007): Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* **60**, 34-42.

Tsakos, G., Allen, P.F., Steele, J.G. and Locker, D. (2012): Interpreting oral health-related quality of life data. *Community Dentistry and Oral Epidemiology* **40**, 193-200.

Waters, E., Davis, E., Ronen, G.M., Rosenbaum, P., Livingson, M. and Saigal, S. (2009): Quality of life instruments for children and adolescents with neurodisabilities: how to choose the appropriate instrument. *Developmental Medicine & Child Neurology* **51**, 660–669.

WHO (1995): The world health organization quality of life assessment (WHOQOL): Position paper from the world health organization. *Social Science and Medicine* **41**, 1403-1409.

Wilson, I.B. and Cleary, P.D. (1995): Linking clinical variables with health-related quality of life: conceptual model of patient outcomes. *Journal of the American Medical Association* **273**, 59-65.