Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://dare.uva.nl/document/353104

File ID 353104
Filename Chapter 9: Summary and general discussion

SOURCE (OR PART OF THE FOLLOWING SOURCE):
Type Dissertation
Title Some fundamental issues in oral health-related quality of life research
Author J.M. Kieffer
Faculty Faculty of Dentistry
Year 2012
Pages 97
ISBN 978-94-6108-247-3

FULL BIBLIOGRAPHIC DETAILS:
http://dare.uva.nl/record/408435

Copyright

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.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

CHAPTER 9
Summary and general discussion

SUMMARY

Primarily this thesis has been focused on the structural and psychometric properties of the Oral Health Impact Profile 49 (OHIP-49), and secondly on the conceptual foundation of oral health-related quality of life (OHRQoL), health-related quality of life (HRQoL) and quality of life (QoL).

The OHIP-49 distinguishes seven dimensions of oral health and is based on a conceptual model of oral health. The dimensions are hierarchically ordered so that the impacts described by the dimensions are gradually more disruptive to one's life. In the study presented in Chapter 2, it was argued that items in later subscales are more general than items in earlier subscales because of this hierarchy. This could potentially create a situation wherein order effects might occur when changing the sequence of the items. Indeed, when administering two versions of the OHIP-49, one upholding the original hierarchical structure, and an alternative in which subscales containing more general items were presented first, results show higher scores on the more general subscales (Psychological disability, Social disability, Handicap) of the alternative version. This indicates that the lack of influence of the more specific items in earlier subscales changes one's conceptual framework surrounding the content of the items in latter subscales. Although effect sizes were generally small, we argued that these order effects were nonetheless noteworthy, because subscale inter-correlations suggested different factor structures for both versions. Thus the OHIP-49 does seem to be susceptible to item-order effects.

Expanding on the fact that the dimensions of the OHIP-49 are based on a hierarchical model, in Chapter it is argued that events described by dimensions higher in the hierarchy would be judged as having more impact on quality of life, than events described by dimensions lower in the hierarchy. To evaluate this, the relative severity was assessed of the impact on daily life with which each dimension of the OHIP-49 is judged relative to all other dimensions. Subjects' judgments were assessed using the method of direct ranking and the method of paired comparison. Results of paired comparison showed the Handicap and the Psychological disability dimensions (dimensions highest in the hierarchy) to be regarded as having the most severe impact on daily life. Furthermore, results of the direct ranking method, wherein subjects ranked the dimensions according to severity, showed that the ordering approximated the hierarchical model on which the OHIP-49 is based. Because differences were found between the severity of impacts described by the dimensions, it was proposed to either include subscale-weights, or to at least increase the number of items in subscales where the impact on QoL is judged as being more severe.

Chapter 4 describes a study that examined the internal responsiveness of the short Oral Health Impact Profile (OHIP-14) and its ability to differentiate between patients with and without pre- and postoperative complaints resulting from third-molar surgery. Results showed that the OHIP-14 was able to differentiate between the first preoperative day and all days within the postoperative week. Furthermore, one month postoperatively, mean OHIP-14 scores were reduced to the preoperative level. In addition, differences could be shown between patients with and without pre- and postoperative complaints, partial and complete mucosa coverage and the level of impaction of the third molar. The OHIP-14 can

be considered internally responsive to changes of oral conditions as a result of surgical third molar removal.

Chapter 2 focuses on item-order effects. Chapter 5 examines order effects between questionnaires and evaluates the effect of changing the administrative order of the SF-12 and the OHIP-49. It was argued that when administering the OHIP-49 first, the judgments on the SF-12 would be restrained to the oral impacts described by the OHIP-49 instead of the complete array of health. Thus making the responses to later questions on the SF-12 more consistent with responses to the earlier questions on the OHIP-49. Results showed no major instrument-order effects. However, this study was conducted in a dental setting, which may have influenced the outcomes. Moreover, to the best of our knowledge it is the only study within oral-health research dealing with this topic. Therefore, further research is needed to rule out instrument-order effects.

The rationale behind the study presented in Chapter 6 was, that before the emergence of the concept QoL, medical and dental research were mainly focused on objective outcomes. However, the patients' view and thus the use of subjective measures have become increasingly important when it comes to general health and oral health outcomes. Therefore, the concepts oral health, general health and QoL have become an integral part of each other and the question arises as to how these concepts are connected. The study described on Chapter 6 aimed at assessing this relationship. Results showed an association between self-reported general health and, with the exception of the Role-emotional and Mental-health subscales, all subscales of the RAND-36 (a HRQoL questionnaire). Results also showed an association between self-reported oral health and the subscales of the OHIP-49 (except for the Psychological disability subscale). However, no association was found between self-reported general health and the OHIP-49, or between self-reported oral health and the RAND-36. Findings suggest that oral health, general health and QoL have different determinants. Furthermore, oral health and general health appear to be mostly unrelated in this seemingly healthy population. It is proposed that if no apparent disease is present, oral and general health must be regarded as separate constructs.

In research, a theory gives an explanation as to how concepts, such as OH and QoL are related. Assumptions about these relationships guide the selection of the measurement model underlying the methods and techniques to be used for developing and analyzing questionnaires. Chapter 7 focuses on the presumed measurement model underlying OHRQoL questionnaires and outlines statistical complications due to model misspecification. Methods for developing and analyzing questionnaires, such as Factor Analysis and Cronbach's alpha, presume psychological constructs to be latent, inferring a reflective-measurement model with the underlying assumption of local independence. Local independence implies that the latent variable explains why the variables observed are related. Many OHRQoL questionnaires are analysed as if they were based on a reflective-measurement model; local independence is thus assumed. This assumption requires these questionnaires to consist solely of items that reflect instead of determine OHRQoL. It is argued that this assumption is not tenable, because many OHRQoL questionnaires contain a mix of both a formative and a reflective measurement model, thus violating the assumption of local independence. It is furthermore suggested, that the model-

misspecification is due to a lack of theory of OHRQoL, and thus the importance of a theoretical understanding of the construct OHRQoL is stressed.

Chapter 8 implements the idea of Chapter 7 to the OHIP-49, for this questionnaire is also often analysed as if it were based on a reflective measurement model. This assumption is untenable however, because the OHIP contains items that describe symptoms as well. Clearly, symptoms are determinants of QoL and underpin a formative-measurement model. As has been said, many traditional statistical methods are not applicable to instruments containing formative items. Therefore, the aim of the study presented in Chapter 8 was to detect formative items in the OHIP-49 using an approach based on frequency analysis using statistics analogous to chi-square statistics. The rationale behind this approach is that contrary to reflective items, formative items have an asymmetric relationship with QoL; one symptom is sufficient to cause a decline in QoL. Furthermore, a decline in QoL does not necessarily mean that a person suffers from all symptoms. However, there was cause to assume that the approach used did not represent a valid and complete method to identify formative items. Putting the items of the OHIP under further scrutiny, it was argued that perhaps all items in the OHIP are formative due to the fact that they are attributed to the oral cavity.

GENERAL DISCUSSION AND CONCLUSION

When assessing QoL, like with many psychosocial constructs, we typically use questionnaires. Within dentistry we want to assess the effect of oral health, or better yet, the effects of oral disease on one's QoL. Thus we want to assess 'oral health-related quality of life' (OHRQoL). As has been said throughout this thesis, one of the most frequently used instruments for assessing OHRQoL is the OHIP-49 and its shorter adaptation, the OHIP-14. Both have been based on Locker's model of oral health, which perpetuates a hierarchical order of dimensions of oral health (Locker 1988). In the following paragraphs I will formulate three standing points. Firstly, understanding the underlying model of a questionnaire is essential to the interpretation of the responses given on a questionnaire. Secondly, the OHIP measures the status of oral health, rather than the impact of oral health on QoL. And thirdly, poor understanding of the conceptual model underlying a questionnaire, or more importantly underlying a construct, leads to model misspecification.

Interpretation of responses

It is imperative for the researcher to understand the subjects' responses to a questionnaire. If not fully understood, then ascertaining a valid assessment of (OHR) QoL becomes a problem. The fact that the OHIP is based on a hierarchical model has, for one, an effect on how the questions are answered. Both Chapter 2 and 3 concentrate on the OHIP-49 being based on Locker's hierarchical model of oral health (Locker 1988), and in both chapters the idea is elaborated that people define a standard of comparison to measure their opinion against. However, Chapter 2 focuses on the item level of the OHIP-49, and on the standard of comparison formed by external stimuli. On the other hand, Chapter 3 focuses on the subscale level of the OHIP-49 and on the standard of comparison being formed from information existing in the memory. In Chapter 2, it has been shown that neglecting the underlying hierarchical order of the dimensions could potentially influence the factor structure of the OHIP. Furthermore, the proposed hierarchy influences the way in which

the dimensions are valued. Indeed, dimensions higher in the hierarchy were judged as having a more severe impact and demonstrated the tenability of the model underlying the OHIP. Moreover, while differences in severity between dimensions are clear, this is not reflected by the instrument itself. No weights are given to the subscales when computing subscale scores. Hence, it could even be said that the dimensions higher in the hierarchy are in fact even more underestimated because they are reflected by the smallest number of items. Thus understanding the underlying model of a questionnaire is essential to the interpretation of the responses given to the questions.

However, knowing the underlying model of a questionnaire is not only essential to the interpretation of the responses given to the questions, but also to the interpretation of the outcomes of a questionnaire. The following paragraph will go into what we are actually measuring with the OHIP.

Oral Health vs. Oral Health-Related Quality of Life

Chapter 4, while presenting a straightforward responsiveness study, also outlines an interesting discussion on the variability of oral complaints across time. It even goes as far as suggesting that the responses given on the OHIP are influenced by that point in time where problems and/or pain are felt the most, and are not a summation of the pain felt over a certain time period, questioning as to whether the OHIP measures a state or a trait. Locker's hierarchical model of oral health (Locker 1988) is the underlying model of the OHIP-49. Therefore, it seems reasonable to assume that the OHIP-49 in fact measures oral health status rather than the impact of oral health on QoL as suggested in Chapter 8. Moreover, this was the intent of the developers of this instrument (Slade et al. 1994). Thus the OHIP was originally developed to measure social impacts of oral disease, with some items reflecting symptoms and some reflecting social impacts, but has gradually become an OHRQoL questionnaire. Unfortunately, this renders the interpretation of the results in for instance Chapter 6 questionable, because at that point we intended to measure OHRQoL. However, if the OHIP, which we used to measure OHRQoL, actually measures oral health, then part of the purpose of Chapter 6 is defied; that is, getting more insight into the relationship between OH, GH and QoL. At best, the results show a relationship between self-reported oral health and most of the dimensions of oral health. However, one could argue that oral health or disease (such as toothache) affects one's QoL, and therefore that the OHIP measures causal factors of QoL, suggesting that the underlying measurement model of the OHIP is formative, as argued in Chapter 8.

The following paragraph refers to the consequences of model misspecification and what these consequences, in a practical sense, would mean for the OHIP.

Misspecification of the measurement model

The first chapters of this thesis have typically been focused on measuring OHRQoL on item, subscale and instrument level, using the OHIP. However, the last two chapters are more fundamental and focus on differences between reflective and formative measurement models and on issues pertaining to model-misspecification. MacKenzie et al. (MacKenzie et al. 2005) empirically tested the effects of model misspecification and found that it can lead to type I or type II errors of inference in hypothesis testing. Additionally, also goodness-of-fit indices are influenced by model misspecification (Hu et al. 1998, Fan et al.

2005, MacKenzie et al. 2005). Therefore, misspecification of the measurement model underlying a questionnaire may result in severe biases in the structural parameters of the theoretical model. Furthermore, it may lead to incorrect interpretations of effect sizes or of significance testing regarding the relationship between concepts (Jarvis et al. 2003). Hence, differentiating between reflective- and formative-measurement models is of fundamental importance to the understanding of the conceptual or theoretical basis of QoL (Kieffer et al. 2009). Nevertheless, in research with the OHIP this differentiation is mostly omitted.

In Chapter 8 we tried to implement an approach for differentiating between reflective items and formative items (Fayers et al. 1997b) in the OHIP-49. The rationale behind this approach is that, contrary to reflective items, formative items have an asymmetric relationship with Quality-of-Life (QoL); for determinants characterize causes that are sufficient to change QoL, but are not necessary to change QoL. However, the approach used did not seem to adequately identify formative items, perhaps because it works under the assumption that the symptoms, if present, actually affect QoL. Furthermore, many participants reported not having any symptoms or side effects, making the data quite skewed. This has also been expressed in Chapter 6, where in retrospect the fact that we were dealing with healthy individuals perhaps defeated the purpose of the study. Because if illness does not occur, we cannot measure it, and hence we cannot show a relationship, let alone conclude there is no relationship. Furthermore, as has been said in Chapter 8, it was suggested that many, if not all items in the OHIP are in fact formative of oral health due to the make-up of the items themselves; that is, one part of the items ask for the frequency of a symptom or side effect, while the other part is attributed to the oral cavity. Therefore, many items in the OHIP-49 ask for the frequency of symptoms attributed to the oral cavity and <u>not</u> for the impact of symptoms on one's daily life, leaving out the most important part; namely that what we wish to measure: the impact of oral health on one's QoL. This means that the OHIP only measures OHRQoL under the assumption that oral symptoms, if present, actually affect QoL.

It is important to note that concerns about formative measurement have been expressed (Borsboom et al. 2003, Howell et al. 2007, Edwards 2010). Throughout chapters 7 and 8 we have argued that reflective measures only co-vary because these measures all have a relationship with the common factor. The measures are homogeneous and tap into a unidimensional construct. This means that if one measure should be omitted, it would not change the meaning of the construct. Thus the measures are interchangeable (Bollen 1984, Bollen et al. 1991, Fayers et al. 1997b, Edwards et al. 2000, Atkinson et al. 2006, Kieffer et al. 2009, Edwards 2010). However, with formative measurement the construct is multidimensional and the measures are heterogeneous. As such, formative measures affect the construct, and the meaning of a construct becomes dependent on the measures that have been introduced into the model (Howell et al. 2007, Edwards 2010). Moreover, because we are dealing with heterogeneous measures that if correlated, do so for other reasons than the construct itself, the measures in a formative measurement model do not belong to the same nomological network and do not show internal consistency. As such, a formative model technically does not explain the correlation between its measures which is necessary to attain a predictable model (Edwards et al. 2000, Edwards 2010). Hence, problems with formative measurement are far-reaching and extend to causality, dimensionality, parameter

identification, internal consistency, construct validity and measurement error (Edwards 2010).

Seen from a philosophical angle, reflective measurement is characterized by a critical realist perspective, which establishes that constructs exist in the real world, which can only be measured to the best of our ability, but can never be measured perfectly, hence the error terms included in reflective measurement models (Borsboom et al. 2003, Edwards 2010). On the other hand, formative measurement chiefly comes from an operationalist perspective. From this perspective constructs are seen as composite variables that are a linear combination of formative measures, and do not exist independently from its measures (Borsboom et al. 2003, Edwards 2010). Therefore, the meaning of a construct is in fact a product of its measures and the construct itself does not exist in the real world. This makes formative measurement quite problematic, and preference is given to reflective measurement (Borsboom et al. 2003, Howell et al. 2007, Edwards 2010).

In sum, the OHIP mostly seems to be a formative measure of oral health, which could perhaps measure OHRQoL under the assumption that oral symptoms affect one's QoL. However, the problem with conceptualizing QoL in terms of oral health states is that the OHIP was not originally developed to measure OHRQoL. In the following paragraph I will discuss the lack of consensus regarding the concept of QoL.

Conceptual framework

The poet's eye, in a frenzy rolling,
doth glance from heaven to earth, from earth to heaven
And as imagination bodies forth
the forms of things unknown, the poet's pen
turns them into shapes, and gives to airy nothing
a local habitation and a name.
(By Shakespeare)

This passage implies that the imaginary and a simple stroke of a pen can give meaning and authenticity to indistinct objects and feelings. A phenomenon well received in poetry, however, for the social sciences it is a whole different ballgame.

Quality of life (QoL) has become a fundamental part of many disciplines and one of the most popular patient-reported outcomes in oral health and health-care research. But up till now the only rationalization for its existence is that it gives a voice to the patient in an otherwise dominating biological and medical world (Armstrong et al. 2004). As with this thesis, much research concerns the measurement of QoL, be it health related or oral-health related, which confines the discussion of the concept to reliability and validity questions of various instruments that are said to measure QoL, and to debates as to whether or not a newly developed instrument falls within the QoL concept. Yet, it seems as if the concept itself has been taken for granted (Rosenberg 1992). However, there exists a lack of consensus with regard to the conceptualization of QoL. The evidence for the diversity of approaches to measuring QoL is compelling. So what is QoL? Everyone has an idea, but, for that matter, we cannot seem to agree on a theoretical model or on a definition. There are many articles dealing with the array of definitions and conceptualizations of QoL. (Meeberg 1993, Farquhar 1995, Dijkers 1999, Haas 1999, Brown et al. 2004, Moons et al.

2006). A major reason underlying this diversity is the multidisciplinary use of the concept (Farquhar 1995). Researchers all too often stay within the confinements of there own expertise, which ultimately leads to diversity within disciplines as well. This can also be seen within dental- and health-care research, where QoL has more interpretations than one cares for (Locker et al. 2007b); perhaps a reason for some researchers to either measure other concepts or quit defining the concept all together (Farquhar 1995). There are other reasons for the lack of consensus with regard to the conceptualization of QoL:

1. QoL is an exchangeable concept, in that many conceptualizations of QoL actually seem conceptualizations of other concepts, such as happiness, satisfaction or well-being, implying that QoL can either be measured by symptoms of depression and mood states (happiness) or by the extent to which a person positively appraises one's personal life conditions (satisfaction) or by both (well-being) (Haas 1999, Brown et al. 2004, Moons et al. 2006).
2. As has been noted before, QoL is used as an umbrella term for patient-reported outcomes. This reason is related to the foregoing in that QoL is used interchangeably with (oral-) health status and functional status. The indiscriminate use of the term QoL has reinstated many instruments measuring (oral) health status as (oral) HRQoL instruments (Dijkers 2007, Locker et al. 2007b). Hence, the most frequently used instruments for measuring (oral) HRQoL capture the frequency of impacts that stem from (oral) diseases which clinicians, researchers and research subjects find important. However, they do not capture the meaning of these impacts to a person's QoL (Locker et al. 2007b). Consequently, it becomes a 'definition by default' (Edwards 1985) in which it is the operational definition that has assigned meaning to the concept, and therefore, without a conceptual definition, QoL is then defined as it were by the items that have been selected on any grounds whatsoever (Farquhar 1995).
3. Already addressed, misspecification of the measurement model is another reason for the lack of consensus. If we cannot distinguish between cause and effect, outcomes can be seriously misinterpreted.
4. Last but not least, QoL is in the eye of the beholder. However, as said, many instruments include determinants, and thus operate under the assumption that these determinants, such as oral diseases, if present, actually affect one's QoL. But do they really? Although people within cultures do value comparable sets of attributes, on an individual level they vary in relative importance (Brown et al. 2004). Indeed, younger people may vary from older people in what they find important in life. It has also been shown that against all odds, physically disabled people actually report a good QoL: this phenomenon is called the disability paradox (Moons et al. 2006). This supports the idea that importance of determinants differ across individuals. Weighting the dimensions of QoL could perhaps be the solution to the individuality of QoL, because, if weights are not used then one is in fact weighting by selection or rejection, as has been discussed in Chapter 3. However, who decides what weights should be given to a certain dimension? Even more importantly, who defines these dimensions? And this question brings us back to square one; what defines QoL? (Farquhar 1995).

Perhaps we need to go back to the drawing table and start developing a new theory of QoL. Edwards proposes an alternative to the formative approach (Edwards 2010); choosing reflective constructs that cause QoL. That means that not the formative measures, but the reflective constructs give meaning to QoL (Edwards et al. 2000, Edwards 2010). However, there are reasons to be sceptical about this approach, because as described in Edwards (Edwards 2010) and in Borsboom (Borsboom et al. 2003) the construct remains a product of its measures, making the construct nothing more than a label.

To conclude

In the chapters, I have reported some of the limitations of the studies described. However, the most obvious limitation regarding this thesis is that much insight into the conceptualization of QoL, HRQoL and OHRQoL came at the end of this thesis. That is, earlier chapters report outcomes of OHRQoL, whilst in later chapters the notion of the measuring OHRQoL is challenged. Thus the OHIP is first referred to as an instrument measuring OHRQoL, which is later disputed. Moreover, in earlier chapters, outcomes such as Cronbach's alpha are reported for the OHIP, even though this is later contested on the grounds of the underlying measurement model of the OHIP. A clear case of progressive insight. It goes without saying that the concept QoL has added significantly to the importance of the patients' view in many research efforts. However, I would like to argue that without the benefit of a specific theoretical or conceptual model of (OHR) QoL, understanding, and in particular measuring the construct is not feasible. This makes the problems with the concept QoL fundamental in nature, bringing us to the core business of science; developing a theory. Attempts at providing and testing conceptual models of QoL have been made in different research fields, e.g. (Feeny et al. 1995, Wilson et al. 1995, Ferrans 1996, Awad et al. 1997, Baker 2007, Sprangers et al. 2008). Being fully aware that this is a scientific field in progress, we can only hope that such efforts will continue, perhaps even join forces to give us broader insight into the conceptualization of QoL. That is why I want to conclude by saying that the key to a concept's definition is understanding the causal relationship between the concept and its measures, and moreover, that it should not be an accumulation of existing and already defined concepts, for that can only lead to erroneous conclusions and misguided policies (Bradley 2001, Taillefer et al. 2003).