

File ID uvapub:102477
Filename Chapter 5: Instrument-order effects: using the Oral Health Impact Profile
 49 and the Short Form 12
Version unknown

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type PhD thesis
Title Some fundamental issues in oral health-related quality of life research
Author(s) J.M. Kieffer
Faculty ACTA
Year 2012

FULL BIBLIOGRAPHIC DETAILS:

<http://hdl.handle.net/11245/1.372260>

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, unless the work is under an open content licence (like Creative Commons).

CHAPTER 5
**Instrument-order effects: using the Oral Health Impact
Profile 49 and the Short Form 12**

INTRODUCTION

In oral health-related quality of life (OHRQoL) research, many studies use a combination of OHRQoL and health-related quality of life (HRQoL) questionnaires, and possibly other instruments (Allen et al. 1999, Broder et al. 2000, Kieffer et al. 2008b). To date, this has carried out without considering the possibility of instrument-order effects. While there is ample evidence in a wide range of disciplines that the ordering of items can influence research outcomes (Schuman et al. 1981, McClendon et al. 1988, Sudman et al. 1996, Weinberger et al. 2006, Kieffer et al. 2008a), there is very little literature that addresses instrument-order effects (McColl et al. 2003). Item-order effects occur because the context or conceptual framework surrounding the construct that is to be measured influences the responses. Therefore, in forming an opinion, people use a standard of comparison formed from information in the memory or from information given by a previous item (Sudman et al. 1996, Kieffer et al. 2008a). As such effects are well known within questionnaires (Schuman et al. 1981, McClendon et al. 1988, Sudman et al. 1996, Weinberger et al. 2006, Kieffer et al. 2008a), it is not unlikely that such effects may also be found between questionnaires, despite the fact that studies in HRQoL research did not find significant instrument-order effects (McColl et al. 2003, Cheung et al. 2004, Cheung et al. 2005, Rat et al. 2008). However, these studies dealt with generic versus disease-specific HRQoL instruments, and the subjects were characterized by specific diseases such as asthma or cancer. Thus subjects were probably already primed by their disease. Therefore, it is assumed that they will answer health questions, whatever the order, with their disease as their standard of comparison. Another study on norm data showed that scores on the Short-Form 36 (SF-36), which is a HRQoL instrument, were more favourable if a disease-specific instrument was administered first (Bowling et al. 1999).

In any case, such studies do not exist in OHRQoL research; furthermore, we could not find any HRQoL study, or OHRQoL study, for that matter, that dealt with instrument-order effects between two generic instruments. This is surprising considering the many OHRQoL studies that also assess HRQoL. Therefore, the aim of this study was to evaluate the effect of changing the administrative order of the Short Form 12 (SF-12) (an HRQoL instrument) and the Oral Health Impact Profile 49 (OHIP-49) (an OHRQoL instrument) which are often combined in OHRQoL studies, be it short form or original form questionnaires (Allen et al. 1999, Broder et al. 2000, Heydecke et al. 2003, McGrath et al. 2003, Kieffer et al. 2008b, Zimmer et al. 2009). As previously attained information influences one's judgment, we expect an effect to occur, implying that responses to later questions will tend to be more consistent with responses to earlier questions (Mason et al. 1994). That is, if the OHIP is administered first we anticipate higher responses (better HRQoL) on the SF-12, simply because judgments would be restrained to the oral impacts described by the OHIP-49 instead of the complete array of health.

MATERIALS & METHOD

Subjects (n=1622) took part in an epidemiological study carried out to evaluate the dental status of adults in the Netherlands and executed by The Netherlands Organisation for Applied Scientific Research (TNO) and the Academic Centre for Dentistry Amsterdam (ACTA). People living in and around 's-Hertogenbosch, the Netherlands, were selected

randomly to participate in the study (Schuller 2009). The sample was stratified on the basis of age (ranging between 25 and 75 yr).

The epidemiological study itself consisted of a survey, followed by a clinical examination of the mouth and a battery of questionnaires. A total of 1,018 dentate people participated in the clinical examination after which they were asked to complete the questionnaires at home and return them in a designated envelope.

The survey consisted, among other things, of questions regarding gender, education, and descent and of questions regarding the subject's opinion of their oral health and general health, which were both rated on a four-point scale from 'excellent' to 'poor'.

The battery of questionnaires consisted, among others, of the Dutch version of the OHIP-49 (Van der Meulen et al. 2008), which is an OHRQoL questionnaire, and the SF-12, which is a HRQoL questionnaire. In addition, an extra item was included in the SF-12 to measure general quality of life (QoL). The OHIP-49 contains 49 items representing seven subscales: Functional limitations, Physical pain, Psychological discomfort, Physical disability, Psychological disability, Social disability, and Handicap (Slade et al. 1994). For each item, respondents were asked how often, in the previous 4 wk, they had experienced a certain problem regarding their teeth, mouth or dentures. An example of an item of the OHIP-49 is: 'Have you had difficulty chewing any foods because of problems with your teeth, mouth or dentures?' They responded on a five-point scale, which was rated from 'very often' to 'never'. Thus, lower scores indicate a better OHRQoL. Scores for each subscale were calculated by summing the scores on the items. The SF-12 consists of 12 items that represent eight subscales: Physical function, Social functioning, Role physical, Role emotional, Mental health, Vitality, Bodily pain, and General health. The scores were summed and transformed according to the guidelines described by Ware et al. (Ware et al. 2002), and ranged from 0 to 100. A high score indicates a better HRQoL. The general QoL question asking subject: 'How would you rate your quality of life during the last 4 wk?', was rated on a five-point scale from 'excellent' to 'very poor'.

Two versions of the questionnaire battery were alternately allocated among the subjects. Eight out of nine subjects (Group 1) who participated in the clinical examination of the mouth received a version in which the OHIP-49 came first and was followed by the SF-12; and one out of nine subjects (Group 2) received a version in which the SF-12 came first and followed by the OHIP-49.

Statistical analyses

Subject characteristics in both groups, such as age, gender, education, and descent, but also subjects' opinions on their oral health and general health, which were assessed prior to the allocation to the two versions of the questionnaire battery, were compared using chi-square tests for categorical variables and independent-sample *t*-tests for continuous variables. The instrument-order effects were evaluated by comparing the responses of Group 1 with those of Group 2 on the SF-12 scores and the OHIP-49 scores by means of the non-parametric Mann–Whitney *U*-test. In addition, effect sizes (ES) were calculated. Because effect sizes are typically reliant on distribution, an alternative effect-size measure was used, which converts a *z*-score into an effect-size estimate by dividing the *z*-score by the square root of the total sample size (Eqn. 1) (Field 2005).

$$ES = z/\sqrt{N} \tag{1}$$

This estimate is, as with the distribution-based estimate, interpreted in terms of the ranges of magnitude of difference likely to be clinically meaningful: <0.2 (a small difference); 0.2-0.8 (a moderate difference); and > 0.8 (a large difference) (Cohen 1977).

RESULTS

Group 1 consisted of 686 subjects who completed and returned the first version of the questionnaire battery. No reminders were sent out. For 20 subjects their test number was missing or incorrect, and a further three subjects failed to fill in large parts of the questionnaire, leaving data from 663 subjects for analysis. Group 2 consisted of 83 participants who completed and returned the second version of the questionnaire battery. For two subjects the test number was missing or incorrect, leaving data from 81 subjects for analysis.

Table 1. Demographic characteristics of Group 1 and Group 2

Characteristics	Group 1	Group 2
Gender, Female	55.1 (633)	49.4 (79)
Education, High	50.8 (620)	48.7 (78)
Descent, Native	90.5 (631)	88.6 (78)
Age (yr)	50.6 (12.6) (663)	50.1 (12.0) (81)

Data are expressed as % (n) or as mean (standard deviation, SD) (n).

Group 1, subjects who received the OHIP-49 first, followed by the SF-12.

Group 2, subjects who received the SF-12 first, followed by the OHIP-49.

Table 1 shows the demographic characteristics of Group 1 and 2. Results showed no significant differences between groups with respect to gender, education, age, and descent. Neither did the results show differences on subjects' opinion on their oral health and general health between Group 1 and Group 2 (Table 2). We assume, therefore, that Group 1 and Group 2 are similar except for possible instrument order effects.

Table 2. Percentage distribution of subjects' ratings of their general health (GH) and oral health (OH)

Rating	OH		GH	
	Group 1(%)	Group 2(%)	Group 1(%)	Group 2(%)
Excellent	8.6	7.7	11.7	5.1
Good	66.0	65.4	75.6	75.9
Fair	18.9	20.5	10.0	15.2
Poor	6.5	6.4	2.7	3.8
Total	100	100	100	100

Group 1, subjects who received the OHIP-49 first, followed by the SF-12.

Group 2, subjects who received the SF-12 first, followed by the OHIP-49.

Because the assumption of equal variances was violated in several subscales, differences between the subscales were assessed using a nonparametric test (the Mann-Whitney *U*-test). Results are presented in Table 3. With regard to the two versions of the questionnaire battery, results showed a significant difference between Group 1 and Group 2 on the OHIP-49 subscale Psychological discomfort ($U = 22686.0$, $p=0.027$), where Group 1 indicated having more psychological discomfort than Group 2. The effect size indicated that the difference was small (Table 3). No significant differences were found between the

two groups with regard to the SF-12 subscale scores. However, results did show a significant difference between the two groups with regard to the additional QoL item ($U = 22956.0$, $p=0.045$), where Group 2 indicated having a poorer QoL than Group 1. However, the effect size also indicated a small difference ($ES=-0.07$).

Table 3. Mean subscale scores of Group 1 and 2 for the Oral health Impact profile (OHIP-49) and the Short Form (SF)-12 and mean differences between groups with confidence intervals (CI) and effect sizes.

	Group 1	Group 2	Mean diff.	95% CI ^b	Mann-Whitney test		
	Mean(SD) (n=663)	Mean(SD) (n=81)			z-score ^a	<i>p</i>	Effect size Z/\sqrt{n}
OHIP-49							
Functional limitations	11.9 (3.6)	12.0 (3.2)	-0.1	-0.95- 0.86	-0.062	0.951	0.00
Physical pain	11.7 (4.0)	11.2 (3.3)	0.5	-0.51- 1.46	-0.610	0.542	-0.02
Psychological discomfort	6.5 (3.0)	6.3 (3.2)	0.3	-0.45- 0.98	-2.210	0.027*	-0.08
Physical disability	10.0 (2.5)	9.7 (1.5)	0.3	-0.28- 0.93	-0.604	0.546	-0.02
Psychological disability	6.7 (2.2)	7.1 (3.7)	-0.4	-0.97- 0.14	-0.463	0.643	-0.02
Social disability	5.3 (1.2)	5.6 (2.5)	-0.3	-0.61- 0.06	-0.390	0.697	-0.01
Handicap	6.5 (1.7)	6.7 (2.7)	-0.2	-0.64- 0.21	-0.472	0.637	-0.02
SF-12							
Physical function	85.0 (25.4)	84.9 (23.9)	0.1	-5.70- 5.96	-0.280	0.779	-0.01
Social functioning	82.0 (21.9)	79.0 (21.7)	3.0	-2.08- 8.02	-1.426	0.154	-0.05
Role physical	75.1 (26.4)	72.8 (28.0)	2.3	-3.85- 8.44	-0.595	0.552	-0.02
Role emotional	80.6 (23.9)	77.0 (26.7)	3.6	-2.05- 9.21	-0.954	0.340	-0.04
Mental health	74.1 (18.0)	72.2 (20.0)	1.9	-2.36- 6.08	-0.637	0.524	-0.02
Vitality	65.9 (23.5)	66.4 (24.0)	-0.6	-6.02- 4.89	-0.038	0.970	0.00
Bodily pain	83.2 (23.4)	82.8 (22.4)	0.4	-5.02- 5.83	-0.310	0.757	-0.01
General health	67.1 (21.9)	62.5 (22.9)	4.6	-0.48- 9.68	-1.703	0.089	-0.06

^a Standardized score of the mean ranked difference

^b 95% Confidence intervals of the mean difference

* Ranked mean difference is significant at $p < 0.05$ (2-tailed).

Group 1, subjects who received the OHIP-49 first, followed by the SF-12

Group 2, subjects who received the SF-12 first, followed by the OHIP-49

SD, Standard Deviation

DISCUSSION

In contrast with our hypothesis, results did not show instrument-order effects on the SF-12 subscales. Order effects were detected with regard to one of the subscales of the OHIP-49 and general QoL, but in view of the small effect sizes these effects are negligible. The results seem reassuring and in line with other studies on generic versus disease-specific HRQoL instruments (McColl et al. 2003, Cheung et al. 2004, Cheung et al. 2005, Rat et al. 2008). However, while these previous studies concerned disease-specific samples, we chose a general population to ensure that our results would not be influenced by disease burden. We speculated that subjects characterized by specific oral conditions are likely to be primed by their disease, which could make their disease the standard of comparison

instead of the OHIP-49. We add, however, that we analysed a subsample of subjects who reported their oral health to be fair to poor, and a subsample who reported their oral health to be good to excellent, but also in these cases no instrument-order effects were found.

One possible explanation for the absence of order effects when administering the OHIP-49 first could be that subjects were exposed to a clinical examination of the mouth before filling out the battery of questionnaires. This could have primed them and is something that certainly needs to be taken into account. Furthermore, it has been shown that the SF-36 is quite insensitive when it comes to oral diseases, which could also be a reason for not detecting order effects on the SF-12 (Heydecke et al. 2003). While the two different versions of questionnaire battery were allocated by alternating the order as opposed to randomizing, because of its administrative simplicity, we have no reason to assume that bias could have risen from this allocation approach. We base this on the fact that the two groups were comparable with regard to demographic variables.

Therefore, the present results indicate that instrument-order effects do not take place with regard to the SF-12 and the OHIP-49. However, to be fully convinced we suggest studying instrument-order effects outside the context of a clinical examination in a dental setting to limit potential priming. In addition, we recommend studying order effects using other instruments alongside the OHIP-49, or with different OHRQoL instruments, to increase generalizability.

Acknowledgements- We would like to express our thanks to Dr. A. Schuller, of the department of Prevention and Care of the Netherlands Organization for Applied Scientific Research (TNO), and Dr. J.H.G. Poorterman, of the department of Social Dentistry and Behavioural Sciences of the Academic Centre for Dentistry Amsterdam (ACTA), who played a crucial role in the data collection for this study.