

# Comparison of two methods in deriving a short version of oral health-related quality of life measure.

R. Saub<sup>1</sup>, D. Locker<sup>2</sup> and P. Allison<sup>3</sup>

<sup>1</sup>Community Dentistry, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia; <sup>2</sup>Community Dentistry, Faculty of Dentistry, University of Toronto, Ontario, Canada; <sup>3</sup>Faculty of Dentistry, McGill University, Montreal, Canada

**Objective** To compare two methods of developing short forms of the Malaysian Oral Health Impact Profile (OHIP-M) measure. **Method** Cross sectional data obtained using the long form of the OHIP-M was used to produce two types of OHIP-M short forms, derived using two different methods; namely regression and item frequency methods. The short version derived using a regression method is known as Reg-SOHIP(M) and that derived using a frequency method is known as Freq-SOHIP(M). Both short forms contained 14 items. These two forms were then compared in terms of their content, scores, reliability, validity and the ability to distinguish between groups. **Results** Out of 14 items, only four were in common. The form derived from the frequency method contained more high prevalence items and higher scores than the form derived from the regression method. Both methods produced a reliable and valid measure. However, the frequency method produced a measure, which was slightly better in terms of distinguishing between groups. **Conclusion** Regardless of the method used to produce the measures, both forms performed equally well when tested for their cross-sectional psychometric properties.

*Key words:* Oral health-related quality of life, psychometric methods, reliability, shortened version, validity.

## Introduction

There are a number of multi-items scales which have been shortened. The main reason of shortening a scale is to make it more practical and easy to use (Coste et al, 1997). Shortening an existing measure is a more tempting solution rather than developing a new one. This is because one can bypass the procedures of content identification and items pool composition (Coste et al, 1997). As such, it reduces the time to develop an efficient measure. Moreover, this approach produces a measure that is familiar to the user of the original form.

However, Coste et al (1997) found that there is no standardization on how a measure should be shortened. Locker and Allen (2002) suggested that there are three ways in which a measure can be shortened: 1) statistical approach (regression, internal consistency or factor analysis), 2) expert-based approach, and 3) item-impact method. The statistical method was the most frequently used in the process of shortening (Coste et al, 1997).

In dentistry, so far, only the Oral Health Impact Profile (OHIP) has been shortened from 49 items to 14 items. Slade (1997) shortened the OHIP using a controlled regression analysis while Locker and Allen (2002) used an item impact method. Both methods produced a different set of short forms. Despite having a different sets of items, both are valid and reliable.

Thus, the question as to which method is superior is a matter of debate. As Locker and Allen (2002) stated, “probably the method of developing a short form instrument is not as important as its content.” They suggested that different short forms may be needed according to the

purpose for which they are to be used. If the measure is to be used as a descriptive measure in a survey to document population oral health-related quality of life, then the aim should be to maximize the score. A measure consisting of low-prevalence items may fail to document the full extent of the oral health impact (Locker and Allen, 2002). In a case where a measure is to be used as a discriminative instrument, items affecting most patients may fail to distinguish between those who are and are not severely compromised (Locker and Allen, 2002). As such, the inclusion of items with low-frequency with relatively severe impact items will maximize the ability of the measure to discriminate between groups. On the other hand, if the measure is to be used as an evaluative measure, high frequency items which are more likely to demonstrate change as a result of health care interventions should be included in the instrument (Locker and Allen, 2002; Guyatt et al, 1986). Nevertheless, according to Juniper et al (1997), the ultimate approach is to test the properties of the instruments based on the purpose for which they were developed.

In a previous study, the 49 items of Oral Health Impact Profile was adapted for the Malaysian population as reported elsewhere (Saub et al., 2007). This adapted Malaysian version, known as L-OHIP(M), contained 45 items divided into seven domains, i.e functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap. The respondents are asked to answer on a five-point frequency Likert scale (very often, quite often, sometimes, once a while, and never). In order to provide an efficient way to collect data, the L-OHIP(M) was

shortened using a method known as “item-frequency” as described elsewhere (Saub et al, 2005). The psychometric properties of this version were tested separately and it was found to be reliable and valid for use in cross-sectional studies.

The aim of this paper is to report the result of the comparison made between two short versions of the Malaysian Oral Health Impact Profile, produced by using two different methods of shortening; regression and item-frequency method.

## Methods

### Data Source

The secondary data obtained from the reliability and validity study of L-OHIP(M) was used (Saub, 2004). A total of 220 respondents, a sub-sample of the Malaysian National Oral Health Survey of Adults (NOHSA 2000), completed the L-OHIP(M), where 149 completed the mail questionnaire and 71 completed the interviews. Six respondents were excluded from the analysis because they had more than nine items missing. Two method of scoring were computed: Additive score (ADD-score) and Simple count score (SC-score). ADD-scores were calculated by adding up the response codes and SC-score was calculated by summing the number of items reported as “very often” and “often”. A high score indicated poor OHRQoL.

### Shortening methods

Two methods of shortening were employed: item-frequency and controlled regression. The item-frequency method was based on the most commonly reported impact by 36 patients who were interviewed at the qualitative interview stage in the process of adapting the OHIP 49 items. The first part of the interview was a semi-structured, which consisted of open-ended questions and probes designed to elicit as much detail as possible from the respondent. The focus of the interview was on the ways in which their oral problems impact on daily life and psychosocial well-being. Following the responses to these questions, they were asked about areas of daily living that they had not mentioned spontaneously. After this component of the interview was completed the appropriate language version of the OHIP (translated into Malay language) was shown to each patient. Each statement was read to the patient, and the patient was asked to comment on the relevance of the statement (i.e. does the problem it describes apply to them). Content analysis was then performed. Two most commonly reported items from each subscale were chosen to form the short version (Saub et al, 2005). These items were extracted from the source data to form a 14 items which is known as Freq-SOHIP(M) version.

The controlled regression analysis, as described by Slade (1997), was performed on the source data. The total score of the 45 items was used as the dependent variable and all the 45 individual items as an independent variable. Stepwise regression was performed. The item that makes the least contribution to the R-squared and which is in a dimension that already has two items in the model was deleted. The model was then refitted. The procedure was continued until two items from each

dimension were selected. This form also contained 14 items and it was known as Reg-SOHIP(M).

### Analysis

The two forms produced were then compared in terms of content (common items, prevalence of individual items), the score, the reliability (Cronbach’s alpha and ICC), the validity, and the ability of the measures to distinguish between groups. For validity testing, the following hypotheses were tested: 1) there was a high correlation between ADD-score of the L-OHIP(M) with the ADD-score of the short form, 2) those who perceived that their oral health was very good or good would have a lower score than those who perceived as fair or poor, 3) those who perceived that they needed dental treatment would have a higher score than those who did not, 4) those who did not satisfy with their oral health would have a higher score than those who were satisfied with their oral health. Spearman’s correlation, Mann Whitney-U test, and Kruskal-Wallis tests were performed as required.

Two analyses were performed to test the ability to distinguish between groups: the differences in the mean ranks (DMR) between categories of the independent variables, obtained from the Mann-Whitney test and odds ratios based on the median splits (Allen and Locker, 1997).

## Results

Table 1 shows the respondents’ characteristics. The mean age was 42 year. The proportion of female respondents was slightly higher than the male respondents. Malays form half of the sample. All respondents had at least primary education.

Table 2 lists the 14 items for the regression and frequency forms of the S-OHIP(M). The two short forms had four items in common: “had painful ulcer”, “felt shy”, “had difficulty carrying out daily activities”, and “felt less confident of yourself”. The Reg-SOHIP(M) contained more low frequency items than the Freq-SOHIP(M). For the former, prevalence ranged from 6.1 to 30.4 percent while the latter ranged from 4.2 to 55.1 percent. The Freq-SOHIP(M) contained seven items of more than 20 percent compared to the Reg-SOHIP(M), which had only three items.

**Table 1.** Respondents’ characteristics

Sociodemographics	n (%)
Agegroup	
18-39	91 (42.5)
40-59	94 (43.9)
60+	29 (13.6)
Gender	
Male	96 (44.9)
Female	118 (55.1)
Ethnic	
Malay	123 (57.5)
Chinese	58 (27.1)
Indian	25 (11.7)
Other	8 (3.7)
Level of Education	
Primary and lower	63 (29.4)
Secondary	117 (54.7)
College and higher	34 (15.9)

**Table 2.** Item content and percent reported “very often”, “often” or “sometimes”

<i>Subscale</i>	<i>Reg-SOHIP(M)</i>		<i>Freq-SOHIP(M)</i>	
	<i>Item</i>	<i>%</i>	<i>Item</i>	<i>%</i>
Functional limitation	Trouble pronouncing words.	12.6	Difficult chewing any foods.	35.0
	Felt that foods you eat have not digested properly.	30.4	Had bad breath cause by dental problem	33.2
Physical pain	Had pain on the jaw.	16.4	Found it uncomfortable to eat any foods.	28.0
	Had painful ulcer in the mouth.	22.9	Had painful ulcer in the mouth.	22.9
Psychological discomfort	Felt shy.	24.3	Felt shy	24.3
	Felt stressed up	15.9	Felt discomfort due to food stuck.	55.1
Physical disability	Been unable to eat your favorite foods.	18.2	Had to avoid eating some foods.	28.5
	Had an interrupt meals.	18.2	Avoided smiling	13.6
Psychological disability	Been sad.	15.0	Your sleep been disturbed.	16.8
	Found it difficult to relax	13.6	Your concentration been affected.	13.6
Social disability	Been less tolerant of your spouse or family.	6.1	Avoided going out.	4.2
	Had difficulty carrying out daily activities.	7.9	Had difficulty carrying out daily activities	7.9
Handicap	Felt unwell.	12.6	Had to spend a lot of money.	13.6
	Felt less confident of yourself.	6.1	Felt less confident of yourself.	6.1

**Table 3.** Mean, median and range of ADD score and SC score

<i>Type of form</i>	<i>ADD score</i>	<i>SC2-score</i>	<i>SC3-score</i>
<i>Reg-SOHIP(M)</i>			
Mean(SD)	8.09 (7.44)	0.59 (1.40)	2.20 (2.83)
Median	6.0	0.0	1.0
Range	0 – 41.0	0 - 11	0 - 13
<i>Freq-SOHIP(M)</i>			
Mean(SD)	10.29 (7.55)	0.99 (1.56)	3.03 (2.99)
Median	9.0	0.0	2.0
Range	0 – 36.0	0 - 10	0 – 13
p value*	0.000	0.000	0.000

ADD score=Additive score, SC2 Score=number of items reported as “very often and often”, SC3-score=number of items reported as “very often, often and sometimes” \* Wilcoxon signed rank test

Table 3 shows the scores for both versions of the S-OHIP(M). The Freq-SOHIP(M) form had significantly higher scores than the Reg-SOHIP(M). The ADD score for Reg-SOHIP(M) and Freq-SOHIP(M) versions were 8.09 and 10.29, respectively.

The Cronbach’s alpha for Freq-SOHIP(M) and Reg-SOHIP(M) was 0.86 and 0.89 respectively. Both forms had a Cronbach’s alpha of more than 0.8, which indicates excellent consistency. For the ICC, the value for the Freq-SOHIP(M) was 0.77 and for the Reg-SOHIP(M) was 0.76. Both had similar ICC values of above 0.7, indicating that both were stable measures.

The scores of both of the short versions were highly correlated with the score of the L-OHIP(M) (Figure 1). This indicates that both of them were able to measure equally well the same construct as the long version. Construct validity was assessed for both of the S-OHIP(M) forms, showing that both of them followed the hypothesis as postulated (Table 4).

The differences in mean rank (DMR), obtained from the Mann-Whitney-U test, indicated that the Freq-SOHIP(M) was marginally better. Odds ratios (OR) based on the median splits also indicated that the Freq-SOHIP(M) was marginally better than the Reg-

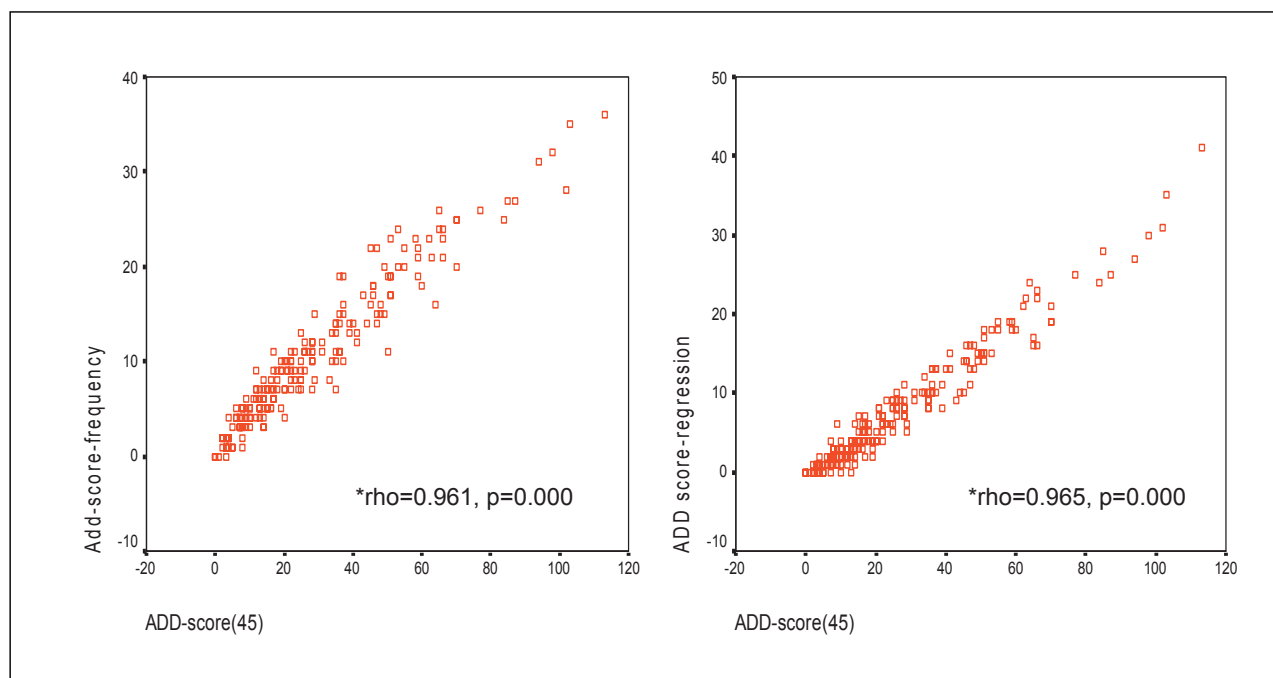


Figure 1. Correlation between short form score and long form score

Table 4. Comparison of the construct validity of the two forms of S-OHIP(M)

	<i>Reg-SOHIP(M)</i>	<i>Freq-SOHIP(M)</i>
Perceived oral health status		
Very good (n=2)	12.00 (16.97)	8.00 (11.31)
Good (n=103)	5.08 (5.58)	7.00 (6.08)
Fair (n=96)	10.46 (7.87)	13.21 (7.56)
Poor (n=6)	18.50 (4.59)	20.17 (4.96)
	<sup>2</sup> p=0.000	<sup>2</sup> p=0.000
DMR <sup>#</sup>	51	57
Odds ratio	6.0	5.1
Perceived dental treatment need		
Do not need treatment (n=75)	4.41 (5.22)	6.36 (5.59)
Need treatment (n=135)	10.06 (7.77)	12.44 (7.74)
	<sup>1</sup> p=0.000	<sup>1</sup> p=0.000
DMR	54	53
Odds ratio	3.5	4.1
Satisfaction with oral health		
Yes (n=126)	5.56 (5.95)	7.44 (5.92)
No (n=84)	11.65 (7.91)	14.38 (7.78)
	<sup>1</sup> p=0.000	<sup>1</sup> p=0.000
DMR	54	57
Odds ratio	4.6	5.6
Dental Status		
Dentate no denture	7.63 (6.95)	9.77 (7.10)
Dentate with denture	8.43 (7.82)	10.98 (7.89)
Edentate	11.60 (10.10)	13.20 (10.20)
	<sup>2</sup> p =0.270	<sup>2</sup> p=0.389
DMR*	10	11
Odds ratio	1.1 <sup>§</sup>	1.4 <sup>§</sup>

<sup>1</sup> Mann-Whitney test, <sup>2</sup>Kruskal-Wallis test

<sup>#</sup> Categories were pooled to very good/good vs fair/poor

\* Categories were pooled to Dentate no denture vs dentate with denture/edentate

<sup>§</sup> 95% confident interval includes 1

DMR-differences in mean rank.

SOHIP(M). This demonstrated that the item-frequency method produced a measure slightly better in distinguishing between groups (Table 4).

### Discussion

Ideally, to compare two different types of measure, these measures should be administered to the same sample independently. However, in this study, we utilized an existing long form of OHIP(M) data to create and compare these two measures. As such, this comparison has some limitations.

The original OHIP was shortened using the regression statistical approach (Slade, 1997), whereas the Malaysian short version was developed based on the most frequent items reported by patients, as discussed earlier. The question as to whether these two approaches yielded different sets of measures was investigated in this study. To make this comparison feasible, the same sample was used. However, it has to be noted that the comparison was not made with Slade's short form due to the changes that had occurred during the process of adaptation of the OHIP 49. Nevertheless, we have used the method used by Slade in developing the short form.

It was found that a different method of shortening produced a different set of items. Only four items were identical. Similar findings were also reported by Locker and Allen (2002) in their attempt to shorten the OHIP using an item-impact method. Even when using the same method in a different culture produced a different set of items (Wong et al 2002). This was found when producing the Chinese short version of OHIP, using the same method as the original short version. They found that only five items were identical in both the Chinese short form and the original short form. Although they have a different set of items, they were found to be reliable and valid. This may suggest that the methods used in producing a short version are not stable. It also suggests that different cultures perceive impact differently.

The frequency short form was found to have more items whose prevalence exceeded 20 percent and had higher scores compared to the regression short form, indicating that it identified more oral health impacts. Therefore, in a case where a measure is to be used as a descriptive measure, then Freq-SOHIP(M) would be the choice.

This study also found that both short forms showed excellent internal consistency reliability and good test-retest reliability when used for this population. A high correlation between scores from both short forms, and scores from the L-OHIP(M), indicated that both forms had good concurrent validity.

Both short forms performed equally well when testing their cross-sectional psychometric properties. However, the responsiveness of these measures was not tested in this study. Future research is indicated to determine which method will be superior in terms of detecting change.

### Acknowledgement

This study was supported by Vot F: Grant no: F0369/2001c, University of Malaya, and partial fellowship award, University of Toronto.

### References

- Allen, P.F. and Locker, D. (1997): Do item weights matter? An assessment using the oral health impact profile. *Community Dental Health* **14**, 133-138.
- Coste, J. Guillemin, F. Pouchot, J. Fermanian, J. (1997): Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology* **50**, 247-252.
- Guyatt, G.H. Bombardier, C. Tugwell, P.X. (1986): Measuring disease specific quality of life in clinical trials. *The Canadian Medical Association Journal* **134**, 889-895.
- Juniper, E.F. Guyatt, G.H. Streiner, D.L. King, D.R. (1997): Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology* **50**, 233-238.
- Locker, D. and Allen, P.F. (2002): Developing short-form measures of oral health-related quality of life. *Journal of Public Health Dentistry* **62**, 13-20.
- Saub R. (2004). Development of an oral health-related quality of life measure for the Malaysian adults population: Cross-cultural adaptation of the Oral Health Impact Profile. PhD thesis, University of Toronto Canada, 2004
- Saub, R. Locker, D. Allison, P. (2005): Derivation and validation of the short version of the Malaysian Oral Health Impact Profile. *Community Dentistry and Oral Epidemiology* **33**, 378-383.
- Saub, R. Locker, D. Allison, P. Disman M. (2007): Cross-cultural adaptation of the Oral Health Impact Profile (OHIP) for the Malaysian adult population. *Community Dental Health*. **24** (3), 166-175.
- Slade, G.D. (1997): Derivation and validation of a short-form oral health impact profile. *Community Dentistry and Oral Epidemiology* **25**, 284-296.
- Wong, M.C. Lo, E.C. McMillan, A.S. (2002): Validation of a Chinese version of the Oral Health Impact Profile (OHIP). *Community Dentistry and Oral Epidemiology* **30**, 423-430