Community Dent Oral Epidemiol 2006; 34: 53–62

Community Dentistry and Oral Epidemiology

Assessing oral health-related quality of life in general dental practice in Scotland: validation of the OHIP-14

Marcelo José Fernandes[1], Danny Adolph Ruta[2], Graham Richard Ogden[3], Nigel Berry Pitts[4] and Simon Alexander Ogston[1]

[1]Section of Public Health, Health Informatics Centre, University of Dundee, Dundee, [2]Public Health Research Group, School of Population and Health Sciences, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne, [3]Unit of Oral Surgery and Medicine, University of Dundee Dental School, Dundee, [4]Dental Health Services Research Unit, Health Informatics Centre, University of Dundee, Dundee, UK

Fernandes MJ, Ruta DA, Ogden GR, Pitts NB, Ogston SA. Assessing oral health-related quality of life in general dental practice in Scotland: validation of the OHIP-14. Community Dent Oral Epidemiol 2006; 34: 53–62. © Blackwell Munksgaard, 2006

Abstract – *Objectives:* To validate the Oral Health Impact Profile (OHIP)-14 in a sample of patients attending general dental practice. *Methods:* Patients with pathology-free impacted wisdom teeth were recruited from six general dental practices in Tayside, Scotland, and followed for a year to assess the development of problems related to impaction. The OHIP-14 was completed at baseline and at 1-year follow-up, and analysed using three different scoring methods: a summary score, a weighted and standardized score and the total number of problems reported. Instrument reliability was measured by assessing internal consistency and test–retest reliability. Construct validity was assessed using a number of variables. Linear regression was then used to model the relationship between OHIP-14 and all significantly correlated variables. Responsiveness was measured using the standardized response mean (SRM). Adjusted R^2s and SRMs were calculated for each of the three scoring methods. Estimates for the differences between adjusted R^2s and the differences between SRMs were obtained with 95% confidence intervals. *Results:* A total of 278 and 169 patients completed the questionnaire at baseline and follow-up, respectively. Reliability – Cronbach's α coefficients ranged from 0.30 to 0.75. Alpha coefficients for all 14 items were 0.88 and 0.87 for baseline and follow-up, respectively. Test–retest coefficients ranged from 0.72 to 0.78. Validity – OHIP-14 scores were significantly correlated with number of teeth, education, main activity, the use of mouthwash, frequency of seeing a dentist, the reason for the last dental appointment, smoking, alcohol intake, pain and symptoms. Adjusted R^2s ranged from 0.123 to 0.202 and there were no statistically significant differences between those for the three different scoring methods. Responsiveness – The SRMs ranged from 0.37 to 0.56 and there was a statistically significant difference between the summary scores method and the total number of problems method for symptomatic patients. *Conclusions:* The OHIP-14 is a valid and reliable measure of oral health-related quality of life in general dental practice and is responsive to third molar clinical change. The summary score method demonstrated performance as good as, or better than, the other methods studied.

Key words: oral health; primary health care; quality of life

Marcelo Fernandes, DHSRU, Health Informatics Centre, Mackenzie Building, Kirsty Semple Way, Dundee DD2 4BF, UK
Tel: +44 1382 420107
Fax: +44 1382 420101
e-mail: m.j.fernandes@dundee.ac.uk (or) mfernandes10@hotmail.com

Submitted 01 June 2004;
accepted 27 June 2005

There is an increasing recognition that oral disorders can have a significant impact on physical, social and psychological well-being. This has resulted in a greater clinical focus on quality of life improvement as a major, if not a primary outcome of dental care, and has led to the development of a number of instruments that aim to measure dental outcomes in terms of the

impact of changes in oral health on quality of life.

Among these, the short form of the Oral Health Impact Profile (OHIP), is emerging as a powerful tool in the assessment of Oral Health-Related Quality of Life (OHRQoL). The short form version (OHIP-14) consists of 14 items organized in seven sub-scales, which address aspects of oral health that may compromise someone's physical, psychological and social well-being (1).

The original 49-item OHIP was developed by Locker and Slade (2) and based on Locker's conceptual model of oral health (3) and includes seven domains namely: functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap.

Since its development, the OHIP-14 has been preferred to the OHIP-49 by a number of researchers due to its practicality. A considerable body of evidence now exists on the validity and reliability of the OHIP-14 in a number of hospital settings and dental conditions, including surgical removal of impacted molars (4), elderly partially edentulous and complete edentulous patients seeking dental rehabilitation (5), and oral lichen planus (6). Three different scoring methods have been reported in studies using the OHIP-14: a summary OHIP-14 score (which is expressed as the sum of the seven raw sub-scale scores on a scale from 0 to 4 where a high score signifies worse OHRQoL); a weighted and standardized summary score (where weights are attributed to every question within the domain); and the total number of problems reported (i.e. occasionally, often, or very often with a possible range of 0–14 problems) (1, 7). Only one study has specifically assessed the validity of the OHIP-14 in a UK context (8). Robinson et al. found that the OHIP-14 (using the summary score) demonstrated superior validity to another oral health-related quality of life measure, the Oral Impacts on Daily Performance (OIDP), when used in a dental hospital setting. However, the authors did not assess instrument reliability and responsiveness.

There have not been, to date, reports on the validation of the OHIP-14 on a UK-based sample of patients attending general dental practice. The OHIP-14 was used in the 1998 UK Adult Dental Health Survey, which sampled over 5000 members of the general UK population (9). This study provided good indirect evidence for the validity of the OHIP-14 in general dental practice, but general populations are not necessarily representative of general dental practice attenders. In addition, no rigorous formal validation of the OHIP-14 was reported in the UK Adult Dental Health Survey, and only one scoring method (the total number of reported problems) was used. Convincing evidence for the validity and reliability of the OHIP-14 in general dental patients is essential if the instrument is to be used to measure OHRQoL in primary care research and practice in the UK.

Responsiveness, or the ability of a health status measurement tool to detect clinically important changes over time is a critical requirement of an outcome measure (10) (and indeed, the OHIP-14 was originally intended to assess long-term effects on OHRQoL). To date, little evidence exists for the responsiveness of OHIP-14 to clinical change over time. Recently Locker et al. (11) assessed the responsiveness of the OHIP-14 over a 1-month period in a dental care programme for the elderly using several formal statistical methods. They concluded that in this context the OHIP-14 was able to detect modest change in OHRQoL and that relatively large samples would be required to detect minimally important clinical differences (defined as five-point scale).

The Scottish Molar Actuarial Life Table Project (The Scottish MALT Project) is a longitudinal observational study looking at the natural history of lower impacted wisdom teeth. Patients with pathology-free impacted third molars were recruited from general dental practices in Tayside, Scotland and were followed for a year to assess clinical pathological changes and development of symptoms associated with impaction. Despite the publication of several studies and guidelines (12–14) on the management of impacted wisdom teeth, controversy still exists over whether retained asymptomatic impacted lower wisdom teeth will remain symptom-free. Also, these guidelines state that not all previously asymptomatic teeth that develop symptoms should be removed, but no assessment has been reported of the impact that the retention of impacted third molars will have on OHRQOL.

The design of the Scottish MALT study provided the perfect opportunity to assess for the first time not only the validity and reliability of the OHIP-14 in a primary care setting but also whether or not this measure is responsive to changes in the clinical

status of impacted wisdom teeth over a 1-year period.

Methods

Identification and recruitment of patients

The study received approval from the Tayside Medical Ethics Committee. All general dental practices with panoramic radiography facilities in Tayside were invited to participate in the study. The researcher, a qualified dentist, visited practices and reviewed panoramic radiographs taken in those practices between 1995 and 2000.

These radiographs had been taken for routine check-ups and specific pathological investigations unrelated to the wisdom teeth considered in the study. Case notes were also reviewed in order to generate a population of general dental practice attenders with at least one pathology- and symptom-free impacted lower third molar. All patients aged from 18 to 70 years meeting the inclusion criteria received an information leaflet giving a brief description of the study and were invited to participate by a letter with an appointment time. They were also given a phone number with the option to change or to cancel the appointment.

Patients attended a clinical assessment in their local dental surgery where the research dentist confirmed the presence of impaction and excluded patients with third molar-related pathology and/or symptoms. The clinical variables included a basic periodontal examination where the Community Periodontal Index (CPI) score was obtained for every patient. The angulation of the impaction was classified according to Winter (15) and the number of teeth present was also noted. Patients were also assessed for temporomandibular joint (TMJ) dysfunction as these conditions are often mistaken for wisdom teeth problems. After the collection of clinical variables patients completed a self-administered sociodemographic questionnaire containing lifestyle questions including dental behaviour questions and also the OHIP-14. The OHIP-14 asked about problems patients might have encountered with their teeth, mouth or dentures during the previous year. Patients who failed the appointment were given two more opportunities to attend.

Responding patients were re-assessed a year later by the same research dentist. In the assessment the presence or absence of the wisdom teeth considered at baseline was noted and patients were asked the following questions: 'Have you had any pain coming from your wisdom tooth in the past year?' 'Have you had any discomfort coming from your wisdom tooth in the past year?' 'Did you come to see your dentist in the past year due to problems with your wisdom tooth?' 'Has your dentist prescribed any antibiotics to you because of problems with your wisdom tooth in the past year?' 'How many times did you have this problem in the past year?' Patients reporting pain were asked to rate the pain in a 0–10 scale. Symptoms were defined as mild pain (scale 1–5), severe pain (scale 6–10), infection (pain + prescription of antibiotics), discomfort or irritation and food stagnation. Information was cross-checked with general dental practice's notes and in those few cases where there was a record of symptoms patients were reminded of particular episodes. On this occasion patients were asked to complete the OHIP-14 again.

Assessing reliability

Instrument reliability was measured by assessing internal consistency and test–retest reliability. Internal consistency was assessed using Cronbach's α (which measures the correlation between items, i.e. questions), for each of the seven health domains, and for all 14 items. Alpha values were calculated at baseline and at follow-up. To assess test–retest reliability, we calculated the intraclass correlation coefficient (ICC) based on the repeated administration of the questionnaire to the same patients 1 year later. Only patients reporting no symptoms in the previous year were included in the test–retest analysis. In order to compare the performance characteristics of the three scoring methods, correlation coefficients are reported for each method. Cronbach's α values and test–retest ICCs above 0.5–0.7 (16–18) are generally considered to indicate sufficient reliability for an instrument or scale to be used to make group comparisons; instruments or scales with coefficients above 0.85 are considered reliable enough for individual patient comparisons.

Assessing validity

We assessed construct validity (i.e. the extent to which OHIP-14 scores are related to specified variables in accordance with an established theory or 'hypothetical construct') by testing a number of hypotheses. The following hypotheses were tested: patients with fewer than 25 teeth have higher OHIP-14 scores at baseline than patients with 25 teeth or more; patients reporting symptoms

including pain at 1-year follow-up have higher OHIP-14 scores at 1 year than asymptomatic patients; that OHIP-14 scores are related to clinically detectable pathology (measured by presence or absence of gingival bleeding and CPI score); that patients citing oral problems as the reason for visiting the dentist have higher OHIP-14 scores than patients attending for a routine check-up or follow-up; that patients who attend the dentist regularly have lower OHIP-14 scores when compared with those who attend less often; that the OHIP-14 score is related to oral hygiene (as measured by time since last visit, frequency of brushing, use of mouthwash and dental flossing); and that OHIP-14 scores are higher in patients of lower socioeconomic status [as measured by employment status, years in education and Carstairs and Morris postcode based deprivation categories (19)]. Additional variables included in this analysis were age, sex, and smoking history and alcohol consumption. Analysis of variance and *t*-tests were used to test for differences between these groups of patients. SPSS regression (SPSS v. 11.5.0; SPSS Inc., Chicago, IL, USA) was used with specified independent variables ('enter' method) to model the relationship between OHIP-14 and all significantly correlated clinical and sociodemographic variables identified above. We tested each variable individually and those found not to be significant were left out from subsequent analysis. All significant variables were entered into the regression model and adjusted R^2s were calculated. This process was repeated for each of the three OHIP-14 scoring methods. Adjusted R^2s were generated for each of the scoring methods at baseline and at follow-up. The values of adjusted R^2 were compared against each other by a bootstrap method in which we repeatedly selected new random samples from a population generated from the original data and refitted the same models and recalculated the R^2 values in each bootstrap sample. This enabled the calculation of confidence intervals for the differences between R^2s.

Assessing responsiveness

Responsiveness, or sensitivity to clinical change, was measured using the standardized response mean (SRM). The SRM is equal to the mean change in score divided by the standard deviation of individuals' changes in scores. An SRM of 0.2 indicates a small effect or clinical change, 0.5 moderate, and 0.8 or greater a large effect (16–18). SRMs were calculated for patients reporting symptoms associated with third molar impaction (i.e. clinical change) within the 1-year follow-up and also for those not reporting symptoms (i.e. no clinical change). These calculations were repeated for each of the three scoring methods. The SRMs obtained were then analysed against each other by using the bootstrap method described before.

Results

Identification and recruitment of patients

Six of seven general dental practices invited agreed to participate. The total number of panoramic radiographs screened was 4299 and these generated 714 potentially eligible patients. A total of 164 patients had moved away from their dental surgery and could not be contacted leaving a total of 550 remaining patients. Of these, 40 were no longer registered with any of the general dental practices involved in the study at the time of recruitment. Thus 510 eligible patients were invited to participate in the study. In total 278 patients attended the baseline appointment (54.5%), 76 (15.0%) refused to participate, and the remaining 156 failed to attend (30.5%). Failures were due to a variety of reasons including lack of time to attend the appointment and dental anxiety.

The baseline appointments generated 278 completed OHIP-14 questionnaires. A comparison between attenders and nonattenders showed that attenders were significantly older than nonattenders (mean ages 32.46 and 30.04, respectively; $P < 0.05$). Nonattenders also included a significantly larger percentage of patients from lower socioeconomic groups (70% compared with 59% from those who attended the appointment; $P = 0.03$).

At the baseline appointment nine patients did not fit our inclusion criteria and were excluded from further analysis. All remaining patients followed their routine dental treatment and were re-invited for another appointment a year later. Fifteen patients had moved away from their dental surgery and one was no longer registered with any of the general dental practices involved and therefore could not be contacted, leaving a total of 253 remaining patients. From those, two patients decided to drop out the study (0.8%).

In total 178 patients attended the follow-up appointment (70.4%). Despite the fact that patients were again given three opportunities to attend,

failures were due mainly to a lack of convenient time to attend the appointment. The follow-up appointments generated 169 completed OHIP-14 questionnaires (61% of those completing a baseline assessment). Nine questionnaires were not completed due to lack of time or because the patient had not brought reading glasses.

Patient characteristics

Of the 278 patients attending a baseline appointment and completing the OHIP-14 107 (38.5%) were males and 171 (61.5%) female. This distribution is consistent with the fact that more females normally attend general dental practice, this has been well described in the literature in the past. The mean age for males was 34.4 years (range 18.65–69.96 years) and for females, 30.84 years (range 19.02–67.50 years). The clinical characteristics are summarized in Table 1. In total, 297 impacted lower third molars were analysed at the follow-up appointments – the majority of those were mesially impacted (43.4%) followed by vertical impaction (26.9%), distal impaction (17.2%), and horizontally impacted teeth (12.5%). The proportion of partially erupted and unerupted teeth were 58.8% and 41.2%, respectively.

At baseline the study population had a mean OHIP-14 summary score of 6.52 (SD 7.92; range 0–50), a mean weighted and standardized score of −0.12 (SD 5.97; range −5.03–30.1) and a mean number of problems of 2.25 (SD 2.74; range 0–13). Patients who attended the 1-year follow-up appointment showed a mean OHIP-14 summary score of 6.70 (SD 7.20; range 0–50) and a mean weighted and standardized score of −0.02 (SD 5.34; range −4.66–31.4). The mean total number of problems for this population at follow-up was 2.07 (SD 2.57; range 0–13).

Assessing reliability

Cronbach's α coefficients for each of the seven health domains ranged from 0.49 (functional limitation) to 0.74 (physical discomfort) at baseline and 0.30 (functional limitation) to 0.75 (social disability) at 1-year follow-up. Only the functional limitation domain fell below 0.5 either at baseline or follow-up. Alpha coefficients for all 14 items combined were 0.88 and 0.87 for baseline and follow-up, respectively.

Test–retest reliability

Intraclass correlation coefficients for patients reporting no symptoms between baseline and follow-up were 0.76, 0.72 and 0.77 using the total OHIP-14, weighted/standardized and number of problems scoring methods, respectively.

Assessing validity

Tables 2 and 3 show the results of the various tests of construct validity for baseline and follow-up, respectively. Patients reporting higher OHIP-14 scores using all three scoring methods (i.e. poorer OHRQoL) at baseline were significantly more likely (at the 1% level) to: have fewer than 25 teeth; use occasional or daily mouthwash; have reported trouble with their teeth at their last dental visit; only visit the dentist when having trouble; and be a smoker. Patients with higher OHIP-14 scores were also significantly more likely not to drink alcohol (significant at 5% level).

At follow-up patients reporting high OHIP-14 scores on all three scoring methods were significantly more likely to report pain (with those reporting severe pain having significantly higher scores than those reporting mild/moderate pain),

Table 1. Clinical characteristics by sex and three age groups at baseline for subjects attending both appointments

	Males			Females		
Clinical characteristics	Age group 1[a]	Age group 2[b]	Age group 3[c]	Age group 1[a]	Age group 2[b]	Age group 3[c]
n	43	12	13	78	28	5
Mean number of teeth	26.24	26.17	19.77	26.22	25.78	21.80
Highest CPI score (%)						
0–2	68	74	30	61	30	0
3–4	32	26	70	39	70	100
Gingival bleeding (as in number of sextants) (%)						
≤2	54	56	73	68	54	20
>2	46	44	27	32	46	80

[a]18–34.9 years of age.
[b]35–49.9 years of age.
[c]50–70 years of age.

Table 2. Validity: OHIP-14 scores by sex, age, clinical and sociodemographic characteristics and dental behaviour at baseline

Variable	Summary scores		Weighted/standardized scores, score (SD)	Number of problems, score (SD)
	n	Score (SD)		
Sex				
Male	107	6.62 (7.96)	−0.03 (6.04)	2.38 (2.81)
Female	171	6.36 (7.89)	−0.25 (5.89)	2.04 (2.62)
Age group (years)				
18–34.99	193	5.94 (7.13)	−0.51 (5.23)	2.16 (2.59)
35+	85	7.82 (9.39)	0.71 (7.25)	2.44 (3.03)
Postcode deprivation category				
Least deprived (deprivation categories I, II and III)	115	6.22 (7.72)	−0.62 (5.90)	1.97 (2.65)
Most deprived (deprivation categories IV, V and VI)	162	6.77 (8.08)	0.31 (6.02)	2.58 (2.92)
Highest CPI				
Less severity (scores 0 –2)	152	6.10 (6.57)	−0.76 (4.99)	1.95 (2.42)
More severe (scores 3–4)	87	7.48 (7.47)	0.12 (5.37)	2.10 (2.53)
Bleeding on probe				
Limited to two sextants	126	6.02 (6.14)	−0.81 (4.75)	1.94 (2.44)
Bleeding in more than two sextants	110	7.54 (7.67)	0.03 (5.51)	2.31 (2.62)
Number of teeth				
Fewer than 25	63	9.00** (9.43)	1.43** (6.99)	2.92** (3.46)
25 or more	177	5.77** (5.56)	−1.10** (4.10)	1.82** (2.04)
Mouthwash				
Never	87	4.55** (5.10)	−2.00** (3.80)	1.30** (1.91)
Occasionally or daily	152	8.18** (8.23)	0.62** (6.05)	2.64** (2.85)
Brushing				
Twice a day (gold standard)	139	6.06* (6.10)	−0.85 (4.73)	1.93 (2.26)
Less than twice a day or more than twice a day	100	7.97* (8.90)	−0.40 (6.33)	1.93 (2.26)
Flossing				
Regular flossing	114	6.25 (5.40)	−0.73 (4.09)	1.98 (2.16)
No flossing	124	7.54 (9.18)	0.11 (6.68)	2.35 (3.05)
Employment				
Unemployed	17	9.47 (8.19)	1.46 (5.88)	3.00 (3.32)
Employed	218	6.70 (7.42)	−0.43 (5.47)	2.10 (2.57)
Education				
No further education	87	8.22* (9.49)	0.66* (6.88)	2.60 (3.24)
Further education	149	6.09* (5.91)	−0.86* (4.45)	1.90 (2.17)
Time since last dental appointment				
Up to 2 years ago	223	6.73* (7.04)	−0.39* (5.22)	2.12 (2.58)
More than 2 years ago	8	13.5* (15.10)	3.72* (10.85)	3.88 (3.83)
Reason for last dental visit				
Trouble with teeth	71	10.61** (10.26)	2.33** (7.54)	3.49** (3.39)
Check-up	165	5.22** (5.10)	−1.48** (3.81)	1.56** (1.97)
Frequency of seeing a dentist				
Just when in trouble	31	12.5** (12.5)	3.94** (9.39)	4.13** (3.94)
Regular	208	6.01** (5.95)	−0.97** (4.31)	1.85** (2.22)
Smoking				
Smoker	135	8.36** (7.99)	0.65** (5.91)	2.65** (2.80)
Nonsmoker	104	4.91** (6.19)	−1.60** (4.59)	1.50** (2.22)
Alcohol				
Nondrinker	39	9.33* (11.41)	1.41* (7.96)	3.00* (3.64)
Drinker	193	6.30* (6.27)	−0.73* (4.75)	−0.37* (5.46)

Significant differences between groups: *$P < 0.05$; **$P < 0.01$.

symptoms and to have fewer than 25 teeth (significant at the 1% level).

When these variables were entered into a stepwise linear regression model at baseline, three significant variables – use of mouthwash, reason for last dental visit and frequency of seeing a dentist – entered the model for summary scores and weighted and standardized scoring methods.

Table 3. Validity: OHIP-14 scores by sex, age and clinical characteristics at follow-up

Variables	Summary scores		Weighted/standardized scores, score (SD)	Number of problems, score (SD)
	n	Score (SD)		
Sex				
Male	65	7.31 (8.89)	0.57 (6.75)	2.22 (3.14)
Female	104	6.32 (5.92)	−0.40 (4.23)	1.98 (2.15)
Age group (years)				
18–34.99	112	6.63 (7.15)	−0.13 (5.23)	2.01 (2.50)
35+	57	6.82 (7.35)	0.17 (5.60)	2.19 (2.74)
Number of teeth				
Fewer than 25	48	8.92** (9.52)	1.76** (7.11)	2.96** (3.20)
25 or more	120	5.71** (5.75)	−0.82** (4.21)	1.67** (2.13)
Symptoms/signs				
Symptoms/signs present	41	10.2** (7.4)	2.60** (5.69)	3.37** (3.08)
No symptoms or signs	121	5.33** (6.57)	−1.05** (4.77)	1.59** (2.2)
Pain				
No pain	142	5.96* (6.86)	−0.60* (5.00)	1.81* (2.41)
Pain present	20	10.85* (7.52)	3.18* (5.88)	3.65* (3.11)

Significant differences between groups: *$P < 0.05$; **$P < 0.01$.

The use of mouthwash and the number of teeth entered the model for the total number of problems method. Together those variables were able to explain 19.8%, 18.9% and 20.2% or the variation in OHIP-14 scores for the summary scores, weighted and standardized, and number of problems, respectively These results are summarized in Table 4. At follow-up, three significant variables entered the regression model. Symptoms and number of teeth entered the model for summary scores and weighted and standardized scoring methods, and pain and number of teeth for the number of problems scoring method. These variables were able to explain 12.3%, 13.5% and 14.5% of the variation in OHIP-14 scores for the summary scores, weighted and standardized, and number of problems methods, respectively. These results are summarized in Table 5.

When the adjusted R^2s obtained from the regression models for each of the three different scoring methods were compared against each other at baseline and at follow-up, all the 95% confidence intervals for the difference between any two of those methods included 0 indicating that these differences were not statistically significant.

Assessing responsiveness

The SRMs for patients reporting symptoms within the 1-year study period were 0.55, 0.56 and 0.37 for the summary scores method, the weighted and

Table 4. Validity: stepwise regression of OHIP-14 scores by three different scoring methods and condition-specific variables for 278 patients at baseline

Scoring method	Regression coefficient	Standard error	*t*	Significance level
Summary scores ($R^2_{adj} = 0.198$[a])				
Use of mouthwash	3.276	0.872	3.75	<0.001
Reason for last appointment	−3.083	0.092	−3.10	0.020
Frequency of seeing a dentist	−3.174	1.399	−2.27	0.024
Constant	14.79	3.27	4.51	<0.001
Weighted and standardized scores ($R^2_{adj} = 0.189$)				
Use of mouthwash	2.411	0.649	3.71	<0.001
Reason for last appointment	−2.083	0.738	−2.82	0.005
Frequency of seeing a dentist	−2.666	1.042	−2.55	0.011
Constant	5.824	2.439	2.38	0.018
Total number of problems ($R^2_{adj} = 0.202$)				
Use of mouthwash	1.196	0.389	3.07	0.003
Number of teeth	0.948	0.432	2.19	0.030
Constant	3.124	1.611	1.94	0.050

[a]Amount of variation these variables are able to account for OHIP-14 scores in the sample (i.e. 19.8%).
Reference categories: use of mouthwash: 0 = use, 1 = no use; reason for last appointment: 0 = no trouble, 1 = trouble; frequency of seeing a dentist: 0 = regularly, 1 = when in trouble; number of teeth: 0 = fewer than 25, 1 = 25 or more.

Table 5. Validity: stepwise regression of OHIP-14 scores by three different scoring methods and condition-specific variables for 169 patients at 1-year follow-up

Scoring method	Regression coefficient	Standard error	*t*	Significance level
Summary scores ($R^2_{adj} = 0.123$)				
Symptoms	3.425	0.786	2.53	0.012
Number of teeth	3.558	1.183	3.12	0.002
Constant	0.655	1.616	0.40	0.686
Weighted and standardized scores ($R^2_{adj} = 0.135$)				
Symptoms	2.937	1.172	2.50	0.013
Number of teeth	2.842	0.838	3.39	0.001
Constant	−4.791	1.186	−4.04	<0.001
Total number of problems ($R^2_{adj} = 0.145$)				
Measure of pain	1.492	0.318	4.696	<0.001
Number of teeth	1.521	0.408	3.70	<0.001
Constant	−0.379	0.572	−0.66	0.509

Reference categories: symptoms: 0 = no symptoms, 1 = symptoms; number of teeth: 0 = fewer than 25, 1 = 25 or more; measure of pain: 0 = no pain, 1 = pain.

standardized method and the total number of problems, respectively. The SRMs for asymptomatic patients were, however, much lower: 0.08, 0.01 and 0.13 for the summary scores method, the weighted and standardized method and the total number of problems, respectively. When the SRMs obtained for each of the three different scoring methods were compared against each other, the 95% confidence intervals for the difference between the summary scores method and the total number of problems method for patients reporting symptoms indicated significance (0.06–0.37). All other comparisons included 0 indicating lack of statistical significance.

Discussion

The aim of our study was to validate the OHIP-14 in a population of general dental practice attenders. Our sample included all patients registered with participating practices in whom a panoramic radiograph was taken between 1995 and 2000. It is likely that this sample will have included irregular attenders – i.e. patients who only register when attending for an immediate problem. If these irregular attenders were less likely to participate in a research study this may well explain the response rate and the observed differences in age and socioeconomic status between respondents and nonrespondents at baseline. However, we believe the results of the follow-up analysis, which included patients who developed problems during the year, and had a low drop-out rate (<30%), allow us to draw conclusions about the validity of the OHIP-14 in a wider general practice population of patients with and without current dental problems.

One of the main inclusion criteria in our study was for patients to have a symptom-free impacted lower third molar. Selecting patients by a higher or lower initial level of some variable (in this case the absence of symptoms or pathology) can lead to a regression to the mean bias, due to the degree of biological variability of the data or error. However, the purpose of our study is to validate an OHRQoL instrument in a primary care context and to correlate changes identified by the instrument with changes in oral health reported by patients overtime. Whether those changes are a result of regression to the mean will not affect these observed correlations or the conclusions drawn.

No consensus exists about which criteria should be used to assess reliability, validity and responsiveness. However, the baseline and follow-up α coefficients and test–retest coefficients observed in this study for the 14-item scale combined would generally be considered to indicate that the OHIP-14 is a reliable instrument for the purposes of group comparisons. Coefficients for the seven separate domains would similarly be considered to indicate scales sufficiently reliable for group comparisons, with the exception of the functional limitation scale. The same level of reliability was observed for all three scoring methods.

None of the socioeconomic variables analysed was found to be statistically significantly correlated with OHIP-14 scores using the three scoring methods. Among clinical variables only the number of teeth was considered to be significantly correlated with OHIP-14 scores for all three scoring methods. This was one of the clinical variables identified by

Slade in the regression model used in the original validation of the OHIP-14 (1). The observed relationship between OHIP-14 scores and the use of mouthwash appears to conflict with our prior hypothesis that poor oral hygiene should be associated with high OHIP-14 scores. In our study patients who used mouthwash had significantly higher OHIP-14 scores than those who did not use mouthwash, this contrasted with the fact that other indicators of poor oral hygiene (i.e. only visiting the dentist when having trouble, brushing once a day or less or being a smoker, for example) showed the expected relationship to OHIP-14 scores. This has caused us to speculate that patients who use more mouthwash may in fact have worse oral health (and consequently worse OHRQoL as measured by the OHIP-14). The use of mouthwash may rather reflect an attempt by a significant proportion of patients to mask oral problems.

In our follow-up assessments the onset of symptoms, number of teeth and pain were strongly correlated with higher OHIP-14 scores, for the three methods studied. The measure of pain was not only strongly correlated with OHIP-14 scores at follow-up but was also related to severity of pain. These clinical variables were also able to explain a substantial amount of variation in OHIP-14 scores in the regression analysis for a study of this kind. However, regardless of which scoring method was used, over 80% of the variation in OHIP-14 scores remained unexplained. It is possible that other nonclinical patient characteristics which were not measured in our study may account for a good part of the unexplained variation in OHIP-14 scores. For example in a study with older men where three different oral quality of life assessment tools were compared: the Oral Health Quality of Life Measure (OHQOL), the OHIP-49 and the Geriatric Oral Health Assessment Instrument (GOHAI) showed that individuals with a higher 'negative affect' had higher OHIP-49 scores when compared with people with a lower negative affect (20). In fact, in that study, from the three tools analysed, the OHIP showed the strongest correlation with negative affect. Negative affect was defined as a general disposition to experience distress, including aversive mood states such as anger, disgust, scorn, guilt, fearfulness and depression. The theory that best explains this phenomenon is known as 'symptom perception hypothesis', where individuals with a particular personality trait (i.e. negative affect) are more likely to perceive and complain about health problems (21, 22). A better understanding of the psychological factors influencing OHRQoL may have implications for the way in which the need for dental care is assessed and the type interventions offered.

When comparing the overall psychometric properties using the three scoring methods there is little to separate them on reliability and validity. However, the number of problems method does appear to be less responsive than the two other methods. The simplicity of the summary method over the weighted and standardized method – the latter requires statistical manipulation – combined with the fact that its ordinal scale and range of scores is more familiar and intuitive to clinicians, would lead us to recommend it as the scoring method of choice when using the OHIP-14 in this patient population. This recommendation is supported by Allen and Locker, who also concluded that simple scoring methods were as good as weighted and standardized methods (7).

In summary, the OHIP-14 demonstrates good levels of reliability, validity and is responsive to third molar clinical change when used to assess OHRQoL in general dental practice. Among the three methods studied the summary scores method performed as well or better and was simpler to use. Variations in OHRQoL in general dental patients are largely unexplained by clinical variables however, and further research is required to explore the effect of personality and other psychological variables on OHRQoL.

Acknowledgements

The development of this study was supported by The Wellcome Trust (061636/HS/SH/MW/sf). The project also had collaboration from EastRen (East of Scotland Research Network) who facilitated access to dental practices in Tayside (project no. 524-01). The authors thank all dental practices involved and also Mrs Gloria Burns and Dr Kim Stringer. We are also grateful to two anonymous referees for their helpful comments on an early draft of this paper.

Professor Pitts acknowledges support from the Chief Scientist Office who core funds the Dental Health Services Research Unit. The views expressed in this paper are those of the authors and may not be shared by the funding bodies.

References

1. Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol 1997;25:284–90.

2. Locker D, Slade G. Oral Health and quality of life among older adults: The Oral Health Impact Profile. J Can Dent Assoc 1993;59:830–3, 837–8, 844.
3. Locker D. Measuring oral health: a conceptual framework. Community Dent Health 1988;5:3–18.
4. McGrath C, Comfort M, Lo C, Luo Y. Patient-centred outcome measures in oral surgery: validity and sensitivity. Br J Oral Maxillofac Surg 2003;41:43–7.
5. Locker D, Matear D, Stephens M, Lawrence H, Payne B. Comparison of the GOHAI and the OHIP-14 as measures of the oral health-related quality of life of the elderly. Community Dent Oral Epidemiol 2001;29:373–81.
6. Hegarty AM, McGrath C, Hodgston TA, Porter SR. Patient-centred outcome measures in oral medicine: are they valid and reliable? Int J Oral Maxillofac Surg 2002;31:670–4.
7. Allen PF, Locker D. Do weights really matter? An assessment using the oral health impact profile? Community Dental Health 1997;14:133–8.
8. Robinson PG, Gibson B, Khan FA, Birnbaum W. Validity of two oral health-related quality of life measures. Community Dent Oral Epidemiol 2003;31:90–9.
9. Office for National Statistics. Adult Dental Health Survey – oral health in the United Kingdom 1998. London: Office for National Statistics; 2000.
10. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology; the essentials, 3rd edn. Baltimore: Williams & Wilkins; 1996. p. 25.
11. Locker D, Jokovic A, Clarke M. Assessing the responsiveness of measures or oral health-related quality of life. Community Dent Oral Epidemiol 2004;32:10–8.
12. Scottish Intercollegiate Guidelines Network. Management of unerupted and impacted third molar teeth. Edinburgh: SIGN; 2000.
13. Royal College of Surgeons of England Faculty of Dental Surgery. The management of patients with third molar teeth: report of a working party convened by the Faculty of Dental Surgery, The Royal College of Surgeons of England. London: Faculty of Dental Surgery RCS (Eng); 1997.
14. Song FLD, Glenny AM, Sheldon T., Prophylactic removal of impacted third molars. An assessment of published reviews. York: NHS Centre for Reviews and Dissemination; 1996.
15. Winter GB. Impacted mandibular third molars. St Louis, American Medical Book Co.; 1926.
16. Bowling A. Measuring disease. Milton Keynes: The Open University Press; 1995.
17. McDowell I, Newell C. Measuring health. A guide to rating scales and questionnaires. Oxford: Oxford University Press; 1996.
18. Bowling A. Measuring health. A review of quality of life measurement scales. Milton Keynes: The Open University Press; 1997.
19. Carstairs V, Morris R. Deprivation and health in Scotland. Aberdeen: Aberdeen University Press; 1991.
20. Kressin NR, Reisine S, Spiro III A, Jones JA. Is negative affectivity associated with oral quality of life? Community Dent Oral Epidemiol 2001;29:412–23.
21. Watson D, Clark LA. Negative affectivity: the disposition to experience aversive emotional states. Psychol Bull 1984;96:453–90.
22. Friedman HS, Booth-Kewley S. The ''disease-prone personality''. Am Psychol 1987;42:539–55.