



Published in final edited form as:

Clin Oral Investig. 2016 January ; 20(1): 91–99. doi:10.1007/s00784-015-1484-6.

A 7-day recall period for a clinical application of the Oral Health Impact Profile questionnaire

Niels Waller¹, Mike T. John², Leah Feuerstahler¹, Kazuyoshi Baba³, Pernilla Larsson⁴, Sanja Peršič⁵, Dóra Kende⁶, Daniel R. Reißmann⁷, and Ksenija Rener-Sitar^{8,9}

¹Department of Psychology, University of Minnesota, Minneapolis, USA

²Department of Diagnostic and Biological Sciences, School of Dentistry, University of Minnesota, Minneapolis, USA

³Department of Prosthodontics, Showa University, Tokyo, Japan

⁴Centre of Oral Rehabilitation, Prosthetic Dentistry, Norrköping, Sweden

⁵Department of Prosthodontics, School of Dentistry, University of Zagreb, Zagreb, Croatia

⁶Department of Prosthodontics, University of Pécs, Pécs, Hungary

⁷Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁸Department of Prosthodontics, Dental Division, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

⁹Department of Prosthodontics, University Dental Clinics, University Medical Center of Ljubljana, Slovenia

Abstract

Objectives—Aims were to investigate and compare the validity and reliability of Oral Health Impact Profile (OHIP) scores referencing 7-day and one-month recall periods in international prosthodontic patients.

Material and Methods—A sample of 267 patients (mean age = 54.0 years, SD = 17.2 years, 58% women) with stable oral health-related quality of life was recruited from prosthodontic treatment centers in Croatia, Germany, Hungary, Japan, Slovenia, and Sweden. These patients completed the OHIP on two occasions using a new 7-day recall period and the traditional one-month recall period. OHIP score validity and reliability were investigated with structural equation models (SEMs) that included OHIP_{past 7 days} and OHIP_{one-month} latent factors and single indicator measures of global oral health status. The SEMs assessed measurement invariance and the relative validities of the two OHIP latent factors (representing the two recall periods).

Corresponding author: Niels Waller, Professor, Director of Quantitative/Psychometric Methods, Department of Psychology, N657 Elliott Hall, University of Minnesota, 75 East River Road, Minneapolis Minnesota 55455, Phone: (612) 626-8729, nwaller@umn.edu.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Results—The SEMs provided cogent evidence for recall period measurement invariance for the two OHIP forms and equal validities ($r = .48$) with external measures of global oral health status.

Conclusion—When assessed in international prosthodontic patients, OHIP scores using the new 7-day recall period were as reliable and valid as the scores using the one-month recall period.

Clinical relevance—Conceptual advantages make a 7-day recall period a preferred frame of reference in clinical applications of the OHIP questionnaire.

Keywords

Oral Health; Quality of Life; Questionnaire; Multicenter Study; Validity; Reliability

Introduction

When measuring patient-reported outcomes, one of the most important elements in symptom and patient-perceived problem assessment is the recall period or time span that patients are asked to consider when responding to health-related questions. For example, whether the time span is limited to a one-day or to a one-year period may influence both the frequency and severity of reported symptoms. For the popular oral health-related quality of life (OHRQoL) questionnaires, such as the Oral Health Impact Profile (OHIP) and the Oral Impacts on Daily Performances, commonly applied recall periods are lifetime [1], twelve [2], six [3], three [4], and one [2] month(s). However, to capture rapid symptom relief after dental interventions, shorter recall periods, such as 7 days, might be necessary. A 7-day timeframe is commonly used in health-related quality of life assessment in medicine [5]. For example, PROMIS[®] (Patient Reported Outcomes Measurement Information System)—a system of highly reliable, precise measures of patient-reported health outcomes for physical, mental, and social well-being—frequently uses a 7-day recall period in its questionnaires or test-item banks [5]. Many PROMIS researchers contend that 7 days “is on the upper limits of ecological validity for specific events (especially for subjective symptoms), yet long enough to allow time for people to experience enough events” [5].

Although the original OHIP publication specified that all 49 items should refer to a fixed time period, it did not recommend a specific time period [6]. To our knowledge, a 7-day time period has never been used with the OHIP. To expand the applicability of the OHIP to periods of rapidly changing perceived oral health status, we argue that a 7-day recall period should supplement existing timeframes.

The aim of this study was to investigate and compare the relative validity and reliability of OHIP scores referencing 7-day and one-month recall periods in international prosthodontic patients.

Methods

Study setting, study design, and subjects

The study was an ancillary study initiated within the international Dimensions of Oral Health-Related Quality of Life (DOQ) Project [7]. The project analyzed 49-item OHIP [6] data from general population subjects and prosthodontic patients from six countries (Croatia,

Germany, Hungary, Japan, Slovenia, and Sweden) with validated language-specific OHIP instruments [8–13]. The international collaborators of the DOQ Project came from the Department of Prosthodontics, University of Zagreb, Zagreb, Croatia; the Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; the Department of Prosthodontics, University of Pécs, Pécs, Hungary; Department of Prosthodontics, Showa University, Tokyo, Japan; the Department of Prosthetic Dentistry, University of Ljubljana, Ljubljana, Slovenia; and the Centre of Oral Rehabilitation, Prosthetic Dentistry, Norrköping, Sweden. In each participating center, the authors received study approval from the institutional medical ethic committees and targeted a consecutive sample of prosthodontic patients. The intended sample size for each center was $N = 50$. This sample size was based on previous findings for OHIP-49 retest reliability coefficients [10]. These coefficients are commonly $> .75$, indicating that two OHIP scores with the same recall period are relatively stable and correlate highly. We used a coefficient of $.75$ for the predicted correlation between OHIP scores referencing different recall periods as our target for sample size calculation. Here, 50 subjects would allow us to determine $r = .75$ with a precision of $.63$ -to- $.87$ (95% confidence interval) in each country.

Patients were assessed on two occasions when their OHRQoL was assumed to be stable. Specifically, they were assessed either twice before the start of prosthodontic treatment or twice after the end of treatment. On average, two weeks elapsed between assessments. The order in which patients completed the two forms was determined by random assignment for each center using block-randomization performed by the statistical software STATA [15]. The study design is similar to test-retest studies that were performed for testing the psychometric properties for language-specific OHIP versions in these six countries [8–13]. However, instead of receiving two OHIPs with the same recall period, one OHIP form had a new, 7-day recall period and the other OHIP form had the commonly used one-month recall period. Not all subjects received questionnaires with the different recall periods and not all OHIP questionnaires were complete. We dropped seven subjects who completed OHIPs with the same recall period and eight subjects who provided insufficient OHRQoL information (6 or more missing items representing a threshold used earlier [14]). Missing values for OHIP questionnaires with five or fewer missing answers were imputed using a median imputation (within person and occasion response vectors) for each OHIP item. The final sample size was $N = 267$ with 59 patients coming from Croatia, 37 from Germany, 49 from Hungary, 50 from Japan, 50 from Slovenia, and 22 from Sweden. Data management was performed with Stata/IC 13.1 [15] and data analyses were performed in R [24].

Oral Health-Related Quality of Life assessment and global oral health status

For each of the 49 OHIP items, a subject rated how frequently he or she has experienced a certain impact on a 5-point scale (0 = ‘never’, 1 = ‘hardly ever’, 2 = ‘occasionally’, 3 = ‘fairly often’, 4 = ‘very often’). Whereas the original OHIP used a 12-month recall period, a one-month recall period has been used more frequently to capture recent oral health impacts. Additionally, following a recent suggestion [12], the word “jaw” was added to each OHIP item ending with the phrase “... because of problems with your teeth, mouth, and dentures” so that subjects referenced the entire stomatognathic system. On each occasion, subjects

were also asked to rate their global oral health status on a 5-point scale (0 = 'excellent', 1 = 'very good', 2 = 'good', 3 = 'fair', 4 = 'poor'). Due to miscommunication, Japanese subjects reported global oral health status on a 2-point scale (0 = 'good', 1 = 'poor').

Built on Locker's conceptual model of oral health [16], OHIP-49 items were initially grouped into seven domains: *Functional Limitation, Physical Pain, Psychological Discomfort, Physical Disability, Psychological Disability, Social Disability, and Handicap*. The Dimensions of OHRQoL Project [7] suggested, based on exploratory factor analytic results from 5,173 international participants and confirmatory factor analytic results from 5,022 participants, that *Oral Function, Orofacial Pain, Orofacial Appearance, and Psychosocial Impact* are the four major aspects of patient's self-perceived OHRQoL [16, 17] that are measured by the OHIP. However, the findings of the DOQ Project caution against the use of four dimension-scores due to the presence of a large general factor that accounts for the lion's share of reported symptom/problem comorbidity. Rather, the project authors recommend that OHRQoL measured with OHIP can be accurately described with a single summary score [17] that taps the general dimension. Therefore, in the present study, we used all items of the long OHIP (minus the 3 items that reference dentures) as our OHRQoL measure.

Data analysis

Reliability assessment—For each recall period and country, Cronbach's alpha [18] was calculated as a measure of the OHIP summary scores' internal consistency reliability. These reliability coefficients estimate the proportion of observed score variance that is due to true individual differences in OHIP summary scores. We computed 95% confidence intervals for the reliability coefficients using a method by Duhachek and Iacobucci [19].

Summary score analyses—OHIP summary scores were computed for each individual (at each occasion) as the sum of the 46 OHIP item scores (3 items referencing dentures were not used in the analyses). The means and standard deviations of these total scores were computed for each country by occasion (first or second) and recall period (7-day or one-month). In addition, we computed the means and standard deviations of the global oral health status indicator for each country, and we computed paired *t*-tests of mean differences between the two OHIP form recall periods and between the two sets of global oral health status indicators.

To assess the convergent validity of OHIP summary scores for the two recall periods, we correlated the recall period specific summary scores with the associated global oral health status scores. Confidence intervals for these Pearson correlations were constructed using Fisher's *r*-to-*z* transformation [31]. For each country, we tested whether these correlations were significantly different from each other. Because the two correlations are based on the same subjects, we used Steiger's method [20] for testing differences among dependent correlations.

Structural equation models—A series of structural equation models (SEMs) [21] was fit to the data to evaluate the measurement invariance and convergent validity of the two OHIP forms. In our first set of analyses, we tested the dimensional structure of each OHIP

form using separate confirmatory factor analytic (CFA) models (see Figure 1, Panel A). Next, we evaluated the OHIP form structural invariance to determine whether recall period choice influenced the relationships between the 46 manifest OHIP items and the underlying common factor of OHRQoL (see Figure 1, Panel B). Thirdly, we tested the convergent validity of each form by correlating the associated OHIP general factor with the global oral health status measure (see Figure 1, Panel C).

Prior to conducting the SEM analyses, we scaled the OHIP and global oral health status scores to remove country of origin mean-level effects from the data. Specifically, for each variable (i.e., item), we removed the country-specific item means to control for sample differences in perceived oral health (as shown in Table 1).

For the SEM analyses, all models were estimated using diagonally weighted least squares (DWLS) [22] estimation in the lavaan package [23] for the R software [24] programming environment. DWLS estimation has been shown to work well with ordinal data and to be robust to violations of multivariate normality [30].

Across the SEM analyses, model fit was evaluated using a standard collection of fit indices [21]. These indices included the log-likelihood chi-square test, the standardized root mean square residual (SRMR) [25], the root mean square error of approximation (RMSEA) [26], the comparative fit index (CFI) [27], the Tucker–Lewis index (TLI) [28], and the adjusted goodness-of-fit index (AGFI) [29]. To gauge the quality of our SEM results we consulted Nye and Drasgow [30] who recently investigated the performance of these fit indices using DWLS estimation under a variety of sample sizes and variable skewness conditions. Results for data sets most like the ones in our study suggested the following guidelines for adjudicating adequate model fit: RMSEA .02, SRMR .05, CFI .99, and TLI .99. A conservative cutoff of .95 was chosen for the AGFI.

Results

Characterization of prosthodontic patients from the six countries

The total sample included 267 adult prosthodontic patients from six international prosthodontics treatment centers. Summaries of the demographic variables, denture status, and proportion of follow-up assessments for these subjects are displayed in Table 1. In the aggregate data set, females represented fifty-eight percent of all respondents, and in each participating country the ratio of female to male prosthodontic patients was larger than one. The respondents had a mean age (SD) of 54.0 (17.2) years. Across countries, average subject ages varied from 40.0 (16.1) years for Slovenian patients to 68.6 (8.7) years for Japanese patients. Most (56.6%) prosthodontic patients had no removable dentures. The proportion of subjects with no removable dentures ranged from 18% in Japan to 96% in Slovenia. Across all countries, 51% of subjects completed the OHIP forms after treatment, though the proportion of post-treatment assessments ranged from 32% in Slovenia to 68% in Japan.

OHIP summary score analysis

As shown in Table 1, using data from all 267 patients, we found substantial differences in OHIP summary scores across countries. These average summary scores ranged from 16.2 in Sweden to 45.0 in Germany. Average OHIP summary scores were slightly higher when using the one-month recall period (34.9) than when using the 7-day recall period (32.1). However, *t*-tests showed significant mean score differences only for Croatia ($t(58) = 6.5, p < .001$) and Slovenia ($t(49) = 2.0, p = .047$). In contrast, none of the mean differences between the two sets of global oral health status scores (Table 2) reached statistical significance. For the 257 patients with complete OHIP and global oral health status data, summary scores from both OHIP forms were highly reliable in all (country specific) samples (Table 2). Cronbach's alpha for OHIP scores ranged from .93 to .98 for the 7-day recall period and from .95 to .98 for the one-month recall period. Regarding the correlational analyses, Table 2 reports (Pearson) correlations between OHIP summary scores and the global oral health status scores for each country. In most countries, these correlations were moderately high (median $r = .52$). For the Japanese subjects, due to the modified response format of the global oral health status scores, these correlations were slightly lower.

Finally, we evaluated differences in convergent validity associated with the two OHIP recall periods. Across the six (country-specific) samples, we found no significant differences between the across-form (i.e., recall period) OHIP-global oral health status correlations.

Structural equation models for Oral Health-Related Quality of Life

Previous findings within the DOQ Project [17] have demonstrated that a one-factor model (1FM) fits the 46-item OHIP reasonably well (using a one-month recall period). To corroborate this result in the current data, a 1FM confirmatory factor analysis was fit separately to each test form (see Figure 1, Panel A). Fit indices for these analyses are reported in the first two rows of Table 3. These findings suggest that the 1FM provides an accurate and parsimonious account of the latent structure of each OHIP test form.

Next, to investigate the effects of test form on the OHIP latent structure, we combined the two one-factor models into a two-factor (2FM) CFA with correlated factors (see Figure 1, Panel B). Because the two OHIP forms include the same 46 items, their item residual scores were allowed to covary across forms. To test across-form measurement invariance [32], we fit several models that varied the number of parameter equality constraints across test forms. In our first model, we allowed the factor loadings (Λ) and the residual variances (Θ) to vary across the two OHIP forms ($\lambda_i, \lambda_i', \theta_{ii}, \theta_{ii}'$ for $i = 1, \dots, 46$). As expected, this model fit well (see Row 3, Table 3). In our second model, to assess metric invariance [32], we constrained the corresponding factor loadings to be equal across the two test forms ($\lambda_i = \lambda_i'$ for $i = 1, \dots, 46$). As reported in Table 3, the fit statistics for this model indicated excellent model-data fit. Finally, to assess strict factorial invariance [32], we constrained both the factor loadings and the residual variances to be equal across the two OHIP forms ($\lambda_i = \lambda_i', \theta_{ii} = \theta_{ii}'$ for $i = 1, \dots, 46$). This model also fit the data well (see Table 3) and did not fit significantly worse than the metric invariance model ($\chi^2_{\text{dif}} = 10.4, df = 46, p = 1$). In the strict factorial invariance model, the two latent OHRQoL factors correlated .93, indicating that OHIP latent factor scores are highly correlated across recall periods.

Finally, we added the global oral health status indicators to the strict factorial invariance model (Panel C of Figure 1) and tested convergent validity invariance using two structural equation models. In the first, unconstrained model, we allowed the correlations between OHIP latent scores and the global oral health status scores to be estimated separately for the two recall periods ($\varphi \neq \varphi'$). In the second, constrained model, these correlations were required to be equal ($\varphi = \varphi'$). As shown in Table 3, both the unconstrained ($\varphi \neq \varphi'$) and the constrained ($\varphi = \varphi'$) models fit the data well. In the unconstrained model, the OHIP-global oral health status correlations were .50 (one-month) and .43 (7-day). In the constrained model, these correlations equaled .48. Constraining ($\varphi = \varphi'$) did not significantly worsen model fit ($\chi^2_{\text{dif}} = .10$, $df = 1$, $p = .76$). Thus, our data suggested that the convergent validity of the OHIP latent factors with our global measures of perceived oral health is not significantly affected by moving from the one-month to the 7-day recall period. For all models, the SRMR values were slightly higher than the previously described threshold value for indicating excellent model fit. Follow-up analyses of the residuals in our final model indicated that small amounts of item covariation remained in the data after fitting unidimensional OHIP latent factor models.¹

Discussion

In this study, analyzing data from 267 prosthodontic patients from six countries, we found that using a 7-day recall period instead of a one-month recall period did not impact test score reliability and validity of the long OHIP. In the country-specific analyses, OHIP summary scores had very similar reliability coefficients and slightly higher convergent validity coefficients with the 7-day recall period. Using an SEM approach, we found measurement invariance for OHIP item responses from the two recall periods. Furthermore, the correlations between self-reported global oral health and OHRQoL measured by the OHIP using either the 7-day or one-month recall periods were not significantly different.

When data from the two OHIP forms were compared, patients reported slightly more OHRQoL impairment (about 3 OHIP points) for the one-month recall period relative to the 7-day recall period. This difference did not reach 6 OHIP points, which represents the OHIP minimally important difference [33], in any of the six countries under study.

We assessed convergent validity of the OHIP summary scores by correlating data from both forms with a single-item measure of global oral health. These correlations were substantial for both recall periods, but the magnitudes of the correlations varied across countries. As expected, they were higher for the 7-day recall period in all countries except for Slovenia and the magnitudes of these convergent correlations were similar to those found in previous studies [34, 35]. All Cronbach's alpha reliability coefficients were very high and only small differences were observed across the two recall period test forms. Again, these reliability coefficients were similar to previously reported OHIP reliability coefficients [34, 35].

¹Our previous findings [16] indicated that in addition to a strong general factor of OHRQoL, the OHIP measures 4 weaker group factors that describe specific aspects of oral health (*Oral Function, Orofacial Pain, Orofacial Appearance, and Psychosocial Impact*). It is likely that inclusion of these weaker factors into our latent variable models would have improved the recovery of all item correlations. Unfortunately, our multi-site samples were not sufficiently large to enable us to rigorously evaluate these more complex latent variable models.

A small number of studies have previously considered the recall period for the OHIP. For example, a Finnish study [36] compared a 12-month with a one-month recall period for the 14-item OHIP. This study recruited adults from two sources; the first sample included patients awaiting orthognathic surgery ($N = 104$) and the second sample was a convenience sample of workers drawn from various workplaces in North Finland ($N = 111$). Similar to our study, the Finnish study did not find substantial differences between OHIP-14 summary scores referencing the two recall periods. Similarly, a German study [10] using the German language OHIP (OHIP-G) compared OHIP-G scores for lifetime, 12-month and one-month recall periods. These researchers found that a one-month recall period had the highest internal consistency reliability among the three test forms. No differences of clinically relevant score magnitude were observed among the three recall periods, but one statistically significant difference was detected when a one-month and a lifetime recall period were compared [10]. Considered in aggregate, the findings of these studies [10, 36] are in line with the results of our analyses. Specifically, all three studies found that modifying OHIP recall periods did not produce clinically significant differences in OHIP reliabilities and validities. Nevertheless, the interpretations of these studies have not been consistent. For instance, in the Finnish study, Sutinen et al. concluded that “although a standardized reference period of 12 months is recommended, in population surveys the use of a shorter (one-month) reference period does not appear to influence responses” [36]. In the German study [10], based on the expectation that memory is more accurate over shorter time periods, the authors recommended using a one-month recall period compared to longer reference periods.

Synthesizing the results of the two previous studies with those of the present study, we conclude there is no “correct” recall period for self-assessed oral health and that “[t]he recall period must correspond to the characteristics of the phenomenon of interest and the purpose of the assessment” [37]. More specifically, choice of recall period should be affected by the purpose and intended use of the OHIP scores, the patient population, the patients’ disease or condition, the treatment or device, and the study design [38]. Similarly, as noted by Norquist et al. “(1) recall depends on what the patient-reported outcome measure captures, its intended use, and attributes of the disease and study; (2) within the same disease area, recall can vary depending on the concept or phenomenon of interest; (3) recall must consider patient burden and their ability to easily and accurately recall the information requested; and (4) recall must be consistent with the duration of the trial and the scheduled clinic visits” [39]. While the choice of the OHIP recall period, such as the one-month or 7-day recall period, may not be substantially different for the patient’s burden, the measured phenomenon (e.g., current perceived oral health versus an average disease impact over a certain time) may be more accurately assessed using a particular recall period.

In measures of other health-related outcomes, researchers often prefer shorter recall periods to longer recall periods. For instance, Acaster et al. point out that “[i]n general, shorter recall periods (e.g., 24 hours, or 1 week at most) can be preferable to longer recall periods mainly because longer recall data can be heavily biased by current health and any significant events” [38]. In cancer patients, a study comparing a one-day with a 7-day recall period found comparable symptom reporting [40], and another study led to the recommendation of

a 7-day recall period [41]. In a study aimed to assess the accuracy of pain and fatigue items across different recall periods, recall periods of 3, 7, and 28 days generated similar ratings of pain and fatigue levels, suggesting that these recall periods may be exchangeable [42].

Strengths and limitations

This is the first international, multi-center study to assess the 7-day versus one-month recall period for the OHIP questionnaire, which is the most widely used OHRQoL instrument in dentistry. Because of our sample sizes, country estimates for the correlations of OHIP summary scores with global assessments of oral health had relatively wide confidence intervals. However, these coefficients were moderately high for all countries. Moreover, while correlations across test forms (i.e., recall periods) were high among OHIP summary scores, the absolute levels of OHRQoL measured with the two test forms were not necessarily similar.

Prosthetic patients from Croatia, Germany, Hungary, Japan, Slovenia, and Sweden represented quite different prosthetic populations in terms of their age, denture status, proportion of post-treatment (follow-up) OHRQoL assessment (as compared to pretreatment, baseline assessment), and OHRQoL impairment. However, samples did not vary substantially in their OHIP mean values (approximately 1 OHIP point) when the first assessment was compared to the second, providing evidence that average OHRQoL remained stable over the study period. Although the initially planned sample sizes were fifty patients per country, the exclusion of some questionnaires, due to missing data and a job position change for the Swedish collaborator, led to smaller samples for some countries.

Generalizability of results to OHRQoL dimensions, OHIP short forms, and other OHRQoL instruments

According to the DOQ Project [7], an individual's overall OHRQoL burden can be sufficiently summarized by a single, higher-order score despite the multidimensional nature of OHRQoL [16, 17]. Thus, we modeled the OHIP item responses using single latent factors. Nevertheless, further methodological work may provide more informative OHRQoL measures that better capture individual differences in the conceptually separable domains of oral health.

Many versions of the OHIP have been reported in the literature, including an abbreviated 14-item short form [43], a 5-item short form [44], and several condition-specific versions [45–47]. Summary scores from these alternate forms are known to correlate highly with summary scores from the long form OHIP [48–50]. These findings suggest that our results are likely to generalize to other OHIP versions. Finally, because the OHIP shares many similarities with other OHRQoL questionnaires, we believe that our recall period results are relevant for other OHRQoL measures.

Conclusion

The present study confirmed that recall periods do not have a large influence on OHIP scores or the correlations of scores with other global measures of perceived oral health. In

settings for which oral health changes quickly, we believe that the use of a 7-day recall period is a valuable option in OHRQoL measurement for two reasons:

- A 7-day recall period unifies the measurement timeframe that is used to assess other oral and medical conditions. This unification facilitates an integrated approach to the assessment of oral and general health.
- Short recall periods are conceptually appealing: All things considered, short recall periods should produce more valid and reliable results when health changes rapidly.

While we acknowledge that recall periods are situation-specific, we believe that, to achieve better global standardization, the 7-day timeframe should be OHIP's preferred recall period in clinical settings.

Acknowledgments

Research reported in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under Award Number R01DE022331.

References

1. López R, Baelum V. Oral health impact of periodontal diseases in adolescents. *J Dent Res.* 2007; 86:1105–1109. [PubMed: 17959905]
2. Slade, GD. The Oral Health Impact Profile. In: Slade, GD., editor. *Measuring Oral Health and Quality of Life.* University of North Carolina, Department of Dental Ecology; Chapel Hill: 1997. p. 93-104.
3. Adulyanon, S.; Sheiham, A. Oral Impacts on Daily Performances. In: Slade, GD., editor. *Measuring Oral Health and Quality of Life.* University of North Carolina. Department of Dental Ecology; Chapel Hill: 1997. p. 151-160.
4. Dolan, TA.; Gooch, BR. Dental Health Questions from the RAND Health Insurance Study. In: Slade, GD., editor. *Measuring Oral Health and Quality of Life.* University of North Carolina, Department of Dental Ecology; Chapel Hill: 1997. p. 65-70.
5. Schneider S, Choi SW, Junghaanel DU, Schwartz JE, Stone AA. Psychometric characteristics of daily diaries for the Patient-Reported Outcomes Measurement Information System (PROMIS): a preliminary investigation. *Qual Life Res.* 2013; 22:1859–1869. [PubMed: 23180166]
6. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. *Community Dent Health.* 1994; 11:3–11. [PubMed: 8193981]
7. John MT, Reissmann DR, Feuerstahler L, Waller N, Baba K, Larsson P, Celebic A, Szabo G, Renner-Sitar K. Factor analyses of the Oral Health Impact Profile - overview and studied population. *J Prosthodont Res.* 2014; 58:26–34. [PubMed: 24440578]
8. Szentpetery A, Szabo G, Marada G, Szanto I, John MT. The Hungarian version of the Oral Health Impact Profile. *Eur J Oral Sci.* 2006; 114:197–203. [PubMed: 16776768]
9. Petricevic N, Celebic A, Papic M, Renner-Sitar K. The Croatian version of the Oral Health Impact Profile Questionnaire. *Coll Antropol.* 2009; 33:841–847. [PubMed: 19860113]
10. John MT, Patrick DL, Slade GD. The German version of the Oral Health Impact Profile-- translation and psychometric properties. *Eur J Oral Sci.* 2002; 110:425–433. [PubMed: 12507215]
11. Renner-Sitar K, Celebic A, Petricevic N, Papic M, Sapundzhiev D, Kansky A, Marion L, Kopac I, Zaletel-Kragelj L. The Slovenian version of the Oral Health Impact Profile Questionnaire (OHIP-SVN): translation and psychometric properties. *Coll Antropol.* 2009; 33:1177–1183. [PubMed: 20102066]

12. Larsson P, List T, Lundstrom I, Marcusson A, Ohrbach R. Reliability and validity of a Swedish version of the Oral Health Impact Profile (OHIP-S). *Acta Odontol Scand.* 2004; 62:147–152. [PubMed: 15370634]
13. Yamazaki M, Inukai M, Baba K, John MT. Japanese version of the Oral Health Impact Profile (OHIP-J). *J Oral Rehabil.* 2007; 34:159–168. [PubMed: 17302943]
14. Locker D. Measuring oral health: A conceptual framework. *Community Dent Health.* 1988; 5:3–18. [PubMed: 3285972]
15. StataCorp. *Stata Statistical Software, Release 13.* StataCorp LP; College Station, TX: 2013.
16. John MT, Feuerstahler L, Waller N, Baba K, Larsson P, Celebic A, Kende D, Rener-Sitar K, Reißmann DR. Confirmatory factor analysis of the Oral Health Impact Profile. *J Oral Rehabil.* 2014; 41:644–652. [PubMed: 24909797]
17. John MT, Reißmann DR, Feuerstahler L, Waller N, Baba K, Larsson P, Celebic A, Szabo G, Rener-Sitar K. Exploratory factor analysis of the Oral Health Impact Profile. *J Oral Rehabil.* 2014; 41:635–643. [PubMed: 24909881]
18. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951; 16:297–334.
19. Duhachek A, Iacobucci D. Alpha's standard error (ASE): an accurate and precise confidence interval estimate. *J Appl Psychol.* 2004; 89:792–808. [PubMed: 15506861]
20. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull.* 1980; 87:245–251.
21. Kline, RB. *Principles and practice of structural equation modeling.* Guilford press; New York: 2011.
22. Jöreskog, KG.; Sörbom, D. *SPSS Inc. LISREL 8 user's reference guide.* Chicago, IL: Scientific Software International; 1996.
23. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw.* 2012; 48:1–36.
24. R Core Team. *R: A language and environment for statistical computing.* The R Foundation for Statistical Computing; Vienna, Austria: 2014. <http://www.r-project.org/> [Accessed 21 January 2015]
25. Bentler, PM.; Wu, EJC. *EQS for Windows user's guide.* Encino, CA: Multivariate Software; 1995.
26. Steiger, JH.; Lind, JM. Statistically-based tests for the number of common factors. Paper presented at the annual spring meeting of the Psychometric Society; Iowa City, IA. 1980.
27. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990; 107:238–246. [PubMed: 2320703]
28. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika.* 1973; 38:1–10.
29. Jöreskog, KG.; Sörbom, D. *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods.* 3. Mooresville, IN: Scientific Software; 1984.
30. Nye CD, Drasgow F. Assessing goodness of fit: Simple rules of thumb simply do not work. *Organ Res Methods.* 2010; 14:548–570.
31. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika.* 1915; 10:507–521.
32. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care.* 2006; 44(11 Suppl 3):S78–S94. [PubMed: 17060839]
33. John MT, Reissmann DR, Szentpétery A, Steele J. An approach to define clinical significance in prosthodontics. *J Prosthodont.* 2009; 18:455–60. [PubMed: 19374706]
34. van der Meulen MJ, John MT, Naeije M, Lobbezoo F. The Dutch version of the Oral Health Impact Profile (OHIP-NL): Translation, reliability and construct validity. *BMC Oral Health.* 2008; 8:11. [PubMed: 18405359]
35. Al-Jundi MA, Szentpétery A, John MT. An Arabic version of the Oral Health Impact Profile: translation and psychometric properties. *Int Dent J.* 2007; 57:84–92. [PubMed: 17506467]

36. Sutinen S, Lahti S, Nuttall NM, Sanders AE, Steele JG, Allen PF, Slade GD. Effect of a 1-month vs. a 12-month reference period on responses to the 14-item Oral Health Impact Profile. *Eur J Oral Sci.* 2007; 115:246–249. [PubMed: 17587301]
37. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Curr Med Res Opin.* 2009; 25:929–942. [PubMed: 19257798]
38. Acaster, S.; Cimms, T.; Lloyd, A. [Accessed 8 May 2014] Development of Methodological Standards Report: Topic # 3 - The Design and Selection of Patient-Reported Outcome Measures (PROMs) for Use in Patient Center Outcomes Research. Oxford Outcomes. 2012. <http://www.pcori.org/assets/The-Design-and-Selection-of-Patient-Reported-Outcomes-Measures-for-Use-in-Patient-Centered-Outcomes-Research1.pdf>
39. Norquist JM, Girman C, Fehnel S, DeMuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Qual Life Res.* 2012; 21:1013–1020. [PubMed: 21909804]
40. Shi Q, Trask PC, Wang XS, Mendoza TR, Apraku WA, Malekifar M, Cleeland CS. Does recall period have an effect on cancer patients' ratings of the severity of multiple symptoms? *J Pain Symptom Manage.* 2010; 40:191–199. [PubMed: 20579835]
41. Arnold BF, Galiani S, Ram PK, Hubbard AE, Briceno B, Gertler PJ, Colford JM Jr. Optimal recall period for caregiver-reported illness in risk factor and intervention studies: a multicountry study. *Am J Epidemiol.* 2013; 177:361–370. [PubMed: 23364878]
42. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain.* 2008; 139:146–157. [PubMed: 18455312]
43. Slade GD. Derivation and validation of a short-form oral health impact profile. *Community Dent Oral Epidemiol.* 1997; 25:284–290. [PubMed: 9332805]
44. John MT, Miglioretti DL, LeResche L, Koepsell TD, Hujoel PP, Micheelis W. German short forms of the Oral Health Impact Profile. *Community Dent Oral Epidemiol.* 2006; 34:277–288. [PubMed: 16856948]
45. Durham J, Steele JG, Wassell RW, Exley C, Meechan JG, Allen PF, Moufti MA. Creating a patient-based condition-specific outcome measure for Temporomandibular Disorders (TMDs): Oral Health Impact Profile for TMDs (OHIP-TMDs). *J Oral Rehabil.* 2011; 38:871–883. [PubMed: 21770999]
46. Allen F, Locker D. A modified short version of the oral health impact profile for assessing health-related quality of life in edentulous adults. *Int J Prosthodont.* 2002; 15:446–450. [PubMed: 12375458]
47. Wong AH, Cheung CS, McGrath C. Developing a short form of Oral Health Impact Profile (OHIP) for dental aesthetics: OHIP-aesthetic. *Community Dent Oral Epidemiol.* 2007; 35:64–72. [PubMed: 17244139]
48. Van Der Meulen MJ, John MT, Naeije M, Lobbezoo F. Developing abbreviated OHIP versions for use with TMD patients. *J Oral Rehabil.* 2011; 39:18–27. [PubMed: 21827525]
49. Baba K, Inukai M, John MT. Feasibility of oral health-related quality of life assessment in prosthodontic patients using abbreviated Oral Health Impact Profile questionnaires. *J Oral Rehabil.* 2008; 35:224–228. [PubMed: 18254801]
50. Larsson P, John MT, Hakeberg M, Nilner K, List T. General population norms of the Swedish short forms of oral health impact profile. *J Oral Rehabil.* 2014; 41:275–281. [PubMed: 24447237]

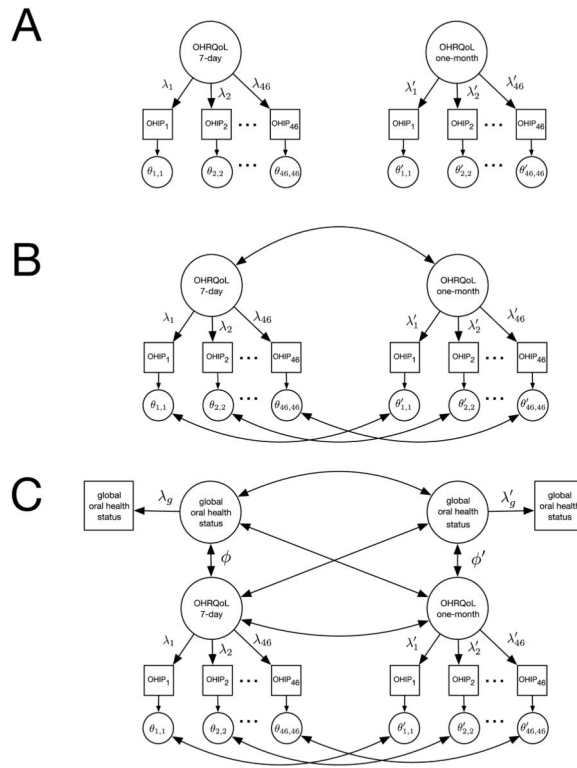


Fig. 1. Path diagrams for 1-factor (A), 2-factor (B), and convergent validity (C) structural equation models. Note. Rectangles denote manifest variables, circles denote latent variables, double-headed arrows denote correlations, and single-headed arrows denote regression pathways.

Sociodemographic characteristics, denture status, proportion of follow-up (post-treatment) assessments, and OHRQoL impairment in prosthodontic patients from the six countries of the Dimensions of OHRQoL Ancillary Project

Table 1

	All	Croatia	Germany	Hungary	Japan	Slovenia	Sweden
<i>N</i>	267	59	37	49	50	50	22
Females [%]	58.4	59.3	54.1	63.3	56.0	58.0	59.1
Age (SD)[years]	54.0 (17.2)	62.3 (13.2)	54.5 (13.8)	46.8 (15.4)	68.6 (8.7)	40.0 (16.1)	46.0 (15.4)
Removable denture status [% (N)]:							
<i>No removable</i>	56.6 (151)	47.5 (28)	48.6 (18)	67.3 (33)	18.0 (9)	96.0 (48)	68.2 (15)
<i>One removable</i>	35.2 (94)	37.3 (22)	40.5 (15)	26.5 (13)	80.0 (40)	4.0 (2)	9.1 (2)
<i>Two complete</i>	6.0 (16)	15.3 (9)	10.8 (4)	4.1 (2)	2.0 (1)	0.0 (0)	0.0 (0)
Unknown	2.2 (6)	0.0 (0)	0.0 (0)	2.0 (1)	0.0 (0)	0.0 (0)	22.7 (5)
Post-treatment assessments [%]	51.3	47.5	64.9	51.0	68.0	32.0	45.4
	Mean (SD)						
OHIP summary score at 1st occasion	34.0 (28.6)	38.5 (28.7)	45.9 (39.2)	31.7 (30.8)	36.0 (24.3)	27.5 (19.7)	17.3 (15.8)
OHIP summary score at 2nd occasion	32.9 (30.1)	37.0 (27.1)	44.0 (45.5)	31.3 (30.7)	34.9 (27.5)	27.4 (19.3)	15.0 (21.6)
OHIP summary score with a 7-day recall period	32.1 (29.4)	35.2 (27.7)	43.6 (42.8)	31.0 (31.0)	34.3 (28.4)	26.1 (17.9)	15.5 (15.9)
OHIP summary score with a one-month recall period	34.9 (29.2)	40.3 (27.8)	46.3 (42.1)	32.1 (30.6)	36.6 (23.3)	28.8 (20.9)	16.9 (21.6)

Table 2

OHIP summary score and global oral health status analyses by country

	Croatia (N = 59)	Germany (N = 34)	Hungary (N = 47)	Japan (N = 47)	Slovenia (N = 49)	Sweden (N = 21)
<i>Cronbach alpha (95% CI)</i>						
Alpha with a 7-day recall period	.97 (.95, .99)	.98 (.96, .99)	.96 (.93, .98)	.98 (.97, .99)	.94 (.90, .97)	.93 (.88, .99)
Alpha with a one-month recall period	.96 (.94, .99)	.98 (.96, .99)	.95 (.92, .98)	.97 (.95, .99)	.95 (.92, .98)	.97 (.93, .99)
<i>Mean (SD)</i>						
Global oral health status indicator administered with one-week OHIP	1.78 (.77)	2.59 (1.10)	2.34 (1.15)	0.45 (.50)	1.69 (.87)	1.62 (1.24)
Global oral health status indicator administered with one month OHIP	1.81 (.80)	2.65 (1.10)	2.36 (1.19)	0.45 (.50)	1.80 (.74)	1.43 (1.25)
<i>Convergent Validity Correlation (95% CI)</i>						
7-day recall OHIP and global oral health status indicator	.60 (.40, .74)	.63 (.37, .80)	.46 (.20, .66)	.40 (.13, .62)	.45 (.19, .65)	.61 (.24, .83)
One-month recall OHIP and global oral health status indicator	.60 (.41, .74)	.60 (.33, .78)	.45 (.19, .65)	.22 (-.07, .47)	.48 (.23, .67)	.56 (.16, .80)
Differences between convergent validity correlations	.00 (-.11, .10)	.03 (-.20, .30)	.01 (-.15, .18)	.18 (-.12, .48)	-.03 (-.27, .18)	.05 (-.40, .54)
<i>z (p-value)</i>						
Steiger's test for equal correlations	-.09 (.93)	.39 (.70)	.20 (.84)	1.25 (.21)	-.39 (.69)	.32 (.75)

Table 3

Fit statistics for structural equation models

Figure Panel	Model	χ^2	df	RMSEA	SRMR	CFI	TLI	AGFI
A.	7-day CFA IFM	697	989	<.001	.078	1.000	1.016	.968
A.	One-month CFA IFM	867	989	<.001	.082	1.000	1.007	.961
B.	2FM Unconstrained Λ, Θ	2913	4047	<.001	.077	1.000	1.016	.965
B.	2FM Constrained (metric invariance)	3195	4093	<.001	.081	1.000	1.012	.962
B.	2FM Constrained $\Lambda \Theta$ (strict factorial invariance)	3206	4139	<.001	.081	1.000	1.012	.962
C.	2FM Unconstrained (φ, ϕ), Constrained Λ, Θ	3450	4319	<.001	.081	1.000	1.011	.961
C.	2FM Constrained (φ, ϕ), Constrained Λ, Θ	3450	4320	<.001	.081	1.000	1.011	.961

Note: 1FM = one factor model; 2FM = two factor model.