Assessing Levels of Agreement between Two Commonly Used Oral Health-Related Quality of Life Measures

Eduardo Bernabé, MSc; Cesar M. de Oliveira, PhD; Aubrey Sheiham, PhD; Georgios Tsakos, PhD

Abstract

Objective: This study aimed to assess the level of agreement between two commonly used oral health-related quality of life (OHRQoL) measures, the short form of the Oral Health Impact Profile (OHIP14) and the Oral Impacts on Daily Performances (OIDP). Methods: A sample of 1,675 15- to 16-year-old students attending all schools in Bauru (Sao Paulo, Brazil) was selected. The impact of oral conditions on quality of life in the last 6 months was reported using both OHIP14 and OIDP. To allow for comparison with the 100 percent OIDP score, OHIP14 scores were converted to percentages. Then, agreement between the two OHRQoL measures was analyzed using the Bland and Altman method. Results: The mean difference between OHIP14 and OIDP was 6.48 percent [confidence interval_{95%} (6.08; 6.89)], with higher scores reported for OHIP14 than for OIDP. Besides, 95 percent of the differences between the two OHRQoL measures were between -10.59 and 23.56 percent. Finally, differences between OHIP14 and OIDP increased significantly as the magnitude of their average increased (P < 0.001). Conclusion: There was a moderate level of agreement between OHIP14 and OIDP, which may be partly due to the fact that both OHRQoL measures assess different levels of oral impacts on quality of life in addition to having different scoring systems.

Key Words: oral health, quality of life, sociodental indicators, agreement, adolescents

Introduction

Although the impact of oral conditions on daily functioning can be assessed using a number of oral health-related quality of life (OHRQoL) measures (1), the short form of the Oral Health Impact Profile (OHIP14) (2) and the Oral Impacts on Daily Performances (OIDP) (3) are the two most widely used and have both gained some international recognition having been shown to be valid and reliable measures in different populations and cultural settings. Both measures are based on the same theoretical framework, namely, the WHO International Classification of the Impairments, Disabilities and Handicaps (4,5). That framework postulates that oral disease can lead to impairment, defined as any anatomic loss or

abnormality. Impairment may then lead to functional limitation. Another consequence of impairment could be pain and discomfort, either physical or psychological. Either of these consequences may lead to physical, psychological, and social disability. A final consequence is handicap or the experience of disadvantage (5). Whereas OHIP14 records quality of life impacts on the whole spectrum of the WHO model (2), OIDP only assesses impacts leading to disability and/or handicap, which have therefore been referred to as ultimate impacts (3).

Although some studies have psychometrically compared OHIP14 and OIDP (6-9), a proper evaluation of their criterion validity has not been carried out. Criterion validity refers to the agreement between a measure and a gold standard. However, in situations where there is no gold standard, as is the case for OHRQoL, well-established measures, such as OHIP14 and OIDP, can be used for comparison purposes (10-12). The question of interest is the extent to which different methods of rating produce comparable results for the intended purpose of measurement (13).

Despite previous studies reporting correlations from 0.42 to 0.78 between OHIP14 and OIDP (7,8), no study has specifically assessed their level of agreement. In fact, assessing criterion validity through correlation has been repeatedly criticized (14-17). Correlations are measures of association, but not of agreement, between a measure and its referent criterion. A high correlation does not expose systematic bias that can occur when one method provides consistently higher or lower scores than the other (14,15). Therefore, recommendations exist for more appropriate statistical evaluation methods (12, 16).

Checking the agreement between measures should be a mandatory step when investigating criterion validity (12,17). However, it is striking that no study on OHRQoL measures had previously addressed this issue. Therefore, the objective of this study was to assess the level of agreement between two commonly used OHRQoL measures, OHIP14 and OIDP.

Material and Methods

Stud	y Desig	gn. Through	а	two-
stage	cluster	sampling,		1,675

Send correspondence and reprint requests to Dr. Eduardo Bernabé, Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK. Tel.: +44 (0) 20-7679-5671; Fax: +44 (0) 20-7813-0280; e-mail: e.bernabe@ucl.ac.uk. Eduardo Bernabé, Cesar M. de Oliveira, Aubrey Sheiham, and Georgios Tsakos are with Department of Epidemiology and Public Health, University College London, London, UK. Manuscript received: 2/28/08; accepted for publication: 5/27/08.

© 2008, American Association of Public Health Dentistry DOI: 10.1111/j.1752-7325.2008.00112.x

adolescents were randomly selected from the 2,200 15- to 16-year-old students attending secondary schools in the city of Bauru, Sao Paulo, Brazil. First, a random sample of 15 schools was selected from a list of the 11 public and 10 private secondary schools in Bauru. The second stage was the random selection of an equal proportion of 15- to 16-year-old adolescents within each previously chosen school. That is, the proportion of adolescents sampled from each school was the same regardless of its size. In doing so, the probability that a participant was included in the sample was the same for all 15to 16-year-old adolescents in the population, and was a self-weighting sample (18,19).

Ethical approval was obtained from the Ethics Review Board of the Dental School at the University of Sao Paulo. Parents signed a consent letter agreeing for their children to participate in the study, and each adolescent gave verbal consent before the survey.

Data **Collection.** Information about the impact of oral conditions on quality of life in the last 6 months was collected through the OHIP14 and the OIDP. First, adolescents self-completed the OHIP14 in their classrooms. Then, students were individually interviewed with the OIDP in a private and quiet room by a single investigator (C.M.O.). The psychometric properties of both OHRQoL measures among these Brazilian adolescents and other Brazilian populations have been reported elsewhere (20-23).

OHIP14 assesses the frequency of problems associated with the mouth, teeth, or dentures on seven dimensions: functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability, and handicap. Adolescents were asked to rate each of the 14 items on a five-point ordinal scale coded 0 ("never"), 1 ("hardly ever"), 2 ("occasionally"), 3 ("fairly often"), and 4 ("very often"). A summary measure was obtained summing up all responses, which ranged between 0 and 56 points (2). To allow for comparison with the 100 percent scale of the OIDP score, the OHIP14 score for each adolescent was multiplied by 100 and divided by 56 to obtain a percentage.

OIDP assesses the ultimate oral impacts on eight daily performances, namely, eating, speaking, cleaning mouth, relaxing, smiling, studying, emotion, and social contact. For each daily performance, adolescents were first asked if they had experienced any impact. If no impact was reported, then a 0 score was assigned. If an impact was reported, its frequency (scale from 1 for "up to 5 days in total" or "less than once a month" to 5 for "more than 3 months in total" or "every or nearly every day"), and the severity of its effect on daily life (scale from 1 for "very minor effect" to 5 for "very severe effect") was scored. Then, a performance score was estimated by multiplying the corresponding frequency and severity scores for those affected

performances. The overall OIDP score was the sum of the eight performance scores (ranging from 0 to 200) multiplied by 100 and divided by 200 (3).

Statistical Analysis. The level of agreement between OHIP14 and OIDP was calculated using the method described by Bland and Altman (13). First, the unknown "true" values for OHRQoL were estimated by determining the average of both measures. Then, differences between OHIP14 and OIDP were calculated for each adolescent to the systematic estimate bias (12,13,16). The standard deviation of such differences was calculated to estimate the random error. Next, differences between the two measures were plotted against their average. Furthermore, the 95 percent confidence interval of the differences between the two measures was calculated to estimate the limits of agreement (12,13,16).





Results

The sample consisted of 951 female (56.8 percent) and 724 male (43.2 percent) participants, with a mean age of 15.83 ± 0.29 years. The adolescents were from high and low socioeconomic status (52.2 and 47.8 percent, respectively). The mean score was 8.83 ± 9.99 OHIP14 percent, with scores ranging from 0 to 76.79 and 13.8 percent of adolescents scoring zero, whereas the mean OIDP score was 2.35 ± 4.71 minimum percent with and maximum values of 0 and 43.50 percent, respectively, and 67.2 percent of adolescents scoring zero. Frequency distributions for OHIP14 and OIDP scores are presented in Figures 1 and 2, respectively.

Figure 3 shows the difference between OHIP14 and OIDP plotted against the average of the two measures. The mean difference between the two measures was 6.48 percent [CI_{95%} (6.08; 6.89)], indicating a consistent bias, with higher scores reported for OHIP14 than for OIDP. The standard deviation of the differences was 8.54 percent. Using this information, the lower and upper levels of agreement between OHIP14 and OIDP were -10.59 and 23.56 percent, respectively. Similar results were obtained when agreement between both OHRQoL measures was assessed by sex and socioeconomic status (data not shown).

The assumption of independence between the difference and the average of both measures could not be corroborated (Pitman test, P < 0.001). That is, there was an increase in the variability of the differences between OHIP14 and OIDP scores as the magnitude of their average increased. Therefore, limits of agreement were recalculated to control for this effect. Differences between OHIP14 and OIDP were linearly regressed on the range of their average values to plot the regressionbased expected value of the difference between both measures and the 95 percent limits of agreement (12,13). In doing so, differences between OHIP14 and OIDP were estimated using the equation $1.41 + 0.91 \times$ average. In addition, the lower and upper 95 percent limits of agreement were estimated by using the equations $-1.81 - 0.21 \times$ average and $4.63 + 1.01 \times$ average, respectively. Using this alternative method, the line of best agreement between OHIP14 and OIDP was estimated, taking into account the variation around that line with increasing levels of their average (Figure 4).

Discussion

This study found a moderate level of agreement between OHIP14 and OIDP because 95 percent of the differences between both OHRQoL measures lay within the values -10.59 and 23.56 percent. In addition, OHIP14 scores were around 6-7 percent higher than the corresponding OIDP scores. According to Bland and Altman (16), the decision about how far apart different measures can be without leading to problems is a question of clinical judgment. Ideally, this should be defined in advance to help in the interpretation of results. Researchers can use both measures interchangeably if the difference between them is not enough to cause difficulties in clinical interpretation (12). Although several studies have used OHIP14 and OIDP in different contexts, no study has concluded how big a difference between simultaneous evaluations for OHIP14 or OIDP should be in order to be clinically relevant rather than only statistically significant (24). A five-point difference has been reported as the minimal important difference for OHIP14 (25), which can be transformed to a 9 percent difference on the 100 percent scale used here. However, there is little evidence that OHRQoL score differences reflect changes of clinical relevance (24). Based on the current state of knowledge, we consider that



the level of agreement between OHIP14 and OIDP was moderate because further analysis showed that 49.8 percent of their differences lay within the range of \pm 5 percent, 60.5 percent lay within the range of \pm 7 percent, and 74.7 percent lay within the range of \pm 9 percent. Nevertheless, more research is needed to support the assumption that differences of up to 9 percent in any of the two OHRQoL measures are not of clinical relevance.

There might be two main explanations for this moderate agreement between OHIP14 and OIDP. The first explanation is related to the different scope of the two OHRQoL measures. Although both measures were derived from the same theoretical framework (4,5), OHIP14 assesses oral impacts on all levels of this model (2), whereas OIDP only assesses oral impacts on the disability and handicap level (3). In this regard, it has been claimed that OIDP was designed to assess only ultimate oral impacts because evaluating impacts on the whole spectrum of the theoretical framework may be redundant and cause double-scoring of the same impacts at different levels, thereby overestimating the real magnitude of impacts of oral conditions on quality of life (3,26). This may be especially important, considering that one of the primary intended uses of OIDP is to facilitate oral health needs assessment. Therefore, depending on the application of the OHRQoL measure, scoring at different levels may be seen as redundant especially in cases not focusing on providing a detailed profile of impacts but rather requiring specific information on impacts to facilitate needs assessment.

The second but not less important explanation is related to the different scoring systems used in the two OHRQoL measures. Whereas OHIP14 only assesses the frequency of oral impacts on quality of life, OIDP assesses both the frequency and the severity of such oral impacts. The OIDP severity ratings allow subjects to provide a personal score reflecting the relative importance of the experienced oral impact on their daily life (3). This individually sensitive weighting system gives prominence and increased validity to the views of respondents (27), thereby complying with significant conceptual and technical criteria suggested in the relevant literature on HRQoL (28,29) and OHRQoL (1).

An interesting additional finding of this study was that the variability in the differences between OHIP14 and OIDP increased proportionally to the size of the "true" value for OHRQoL. That is, differences between both measures were higher at higher values of their average (Figure 3). Because 12.7 percent of adolescents reported a zero score in both measures (i.e., floor effect), this indicates that the agreement between OHIP14 and OIDP was higher among adolescents with lower levels of oral impacts on quality of life than among those with higher levels, which can be explained by the frequency distribution of OHIP14 and OIDP scores among these Brazilian adolescents. Bland and Altman (12) have reported this effect as the most common departure from the assumptions for their method. Researchers could ignore this effect if they keep in mind that limits of agreement would be wider for small "true" values and narrower for large "true" values (12,16). However, an alternative method, based on linear regression, has been proposed to take into account this variation in the aforementioned relationship (16). Using the linear regression method, we estimated the 95 percent limits of agreement between OHIP14 and OIDP, which permitted the prediction of the differences between both measures and their limits of agreement according to the magnitude of the "true" value of OHRQoL (Figure 4).

Finally, the order of administration of both OHRQoL measures may



have had an effect on the present findings because all adolescents received both measures in the same order. Some psychological studies have previously demonstrated that changing the sequential order of administration of questionnaires can alter the pattern of responses (30,31). Two reasons are commonly cited to explain why order effects could appear. First, respondents may become fatigued or lose concentration during the second administration and therefore tend to misunderstand or omit more items. Second, respondents may become familiar with a given topic and/or desensitized by previous questions on the same topic, thus producing a different pattern of answers (31). However, recent studies have reported that the order in which different HRQoL measures were administered in cancer patients did

not have major effects on the number of missing values, mean scores and variability, validity and internal consistency (32-34). A suggested explanation of these apparently contradictory findings is that questions on HRQoL are less stigmatizing and less threatening than questions about psychological problems (33,34). Because these studies were based on patient-based rather population-based samples, than better studies assessing order effects on the performance of OHIP and OIDP are needed to elucidate this point.

Conclusions

A moderate level of agreement between OHIP14 and OIDP was found in this sample of Brazilian adolescents. OHIP14 scores were 6-7 percent higher than the corresponding OIDP scores, with 95 percent of the differences between both OHRQoL measures being between –10.59 and 23.56 percent. The moderate agreement between the two measures may be partly due to the difference in their scopes – OHIP14 assesses oral impacts on all levels of the model of oral health, whereas OIDP assesses those oral impacts on the disability and handicap levels only – and their different scoring systems – OHIP14 assesses the frequency of oral impacts, whereas OIDP assesses the frequency and severity of oral impacts.

Acknowledgments

Eduardo Bernabé was supported by the Programme Alβan, the European Union Programme of High Level Scholarships for Latin America, Scholarship N° E06D1000352PE.

References

- Locker D, Allen F. What do measures of "oral health-related quality of life" measure? Community Dent Oral Epidemiol. 2007;35:401-11.
- Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol. 1997; 25:284-90.
- Adulyanon S, Sheiham A. Oral impacts on daily performances. In: Slade GD, editor. Measuring oral health and quality of life. Chapel Hill: University of North Carolina; 1997. pp. 151-60.
- WHO. International classification of impairments, disabilities and handicaps. Geneva: World Health Organization; 1980.
- Locker D. Measuring oral health: a conceptual framework. Community Dent Health. 1988;5:3-18.
- Robinson PG, Gibson B, Khan FA, Birnbaum W. A comparison of OHIP 14 and OIDP as interviews and questionnaires. Community Dent Health. 2001; 18:144-49.
- 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.
- Soe KK, Gelbier S, Robinson PG. Reliability and validity of two oral health related quality of life measures in Myanmaradolescents. Community Dent Health. 2004;21:306-11.
- Baker SR, Pankhurst CL, Robinson PG. Utility of two oral health-related qualityof-life measures in patients with xerostomia. Community Dent Oral Epidemiol. 2006;34:351-62.

- Anastasi A, Urbina S. Psychological testing. 5th ed. New Jersey: Prentice-Hall; 1997.
- Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 3rd ed. New York: Oxford University Press; 2003.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135-60.
- Bland JM, Altman DG. Measurement in medicine: the analysis of method comparison studies. Statistician. 1983;32:307-17.
- Hebert JR, Miller DR. The inappropriateness of conventional use of the correlation coefficient in assessing validity and reliability of dietary assessment methods. Eur J Epidemiol. 1991;7:339-43.
- Delcourt C, Cubeau J, Balkau B, Papoz L. Limitations of the correlation coefficient in the validation of diet assessment methods. CODIAB-INSERM-ZENECA Pharma Study Group. Epidemiology. 1994;5:518-24.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307-10.
- Schmidt ME, Steindorf K. Statistical methods for the validation of questionnaires – discrepancy between theory and practice. Methods Inf Med. 2006;45:409-13.
- Scheaffer RL, Mendelhall W, Ott L. Elementary survey sampling. 5th ed. London: Duxbury Press; 1996.
- Korn EL, Graubard BI. Analysis of health surveys. 1st ed. New York: John Willey & Sons; 1999.

- de Oliveira CM, Sheiham A. The relationship between normative orthodontic treatment need and oral health-related quality of life. Community Dent Oral Epidemiol. 2003;31:426-36.
- de Oliveira CM, Sheiham A. Orthodontic treatment and its impact on oral healthrelated quality of life in Brazilian adolescents. J Orthod. 2004;31:20-7.
- Ferreira CA, Loureiro CA, Araujo VE. Psychometrics properties of subjective indicator in children. Rev Saude Publica. 2004;38:445-52.
- Oliveira BH, Nadanovsky P. Psychometric properties of the Brazilian version of the Oral Health Impact Profile-Short Form. Community Dent Oral Epidemiol. 2005;33:307-14.
- Brondani MA, MacEntee MI. The concept of validity in sociodental indicators and oral health-related quality-of-life measures. Community Dent Oral Epidemiol. 2007;35:472-8.
- Locker D, Jokovic A, Clarke M. Assessing the responsiveness of measures of oral health-related quality of life. Community Dent Oral Epidemiol. 2004;32:10-8.
- 26. Sheiham A, Steele JG, Marcenes W, Tsakos G, Finch S, Walls AW. Prevalence of impacts of dental and oral disorders and their effects on eating among older people; a national survey in Great Britain. Community Dent Oral Epidemiol. 2001;29:195-203.
- Tsakos G, Marcenes W, Sheiham A. Evaluation of a modified version of the index of Oral Impacts On Daily Performances (OIDP) in elderly populations in two European countries. Gerodontology. 2001;18:121-30.

- Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. JAMA. 1994;272:619-26.
- 29. Guyatt GH, Cook DJ. Health-status, quality-of-life, and the individual. JAMA. 1994;272:630-1.
- Lucas CP. The order effect: reflections on the validity of multiple test presentations. Psychol Med. 1992;22:197-202.
- Jensen PS, Watanabe HK, Richters JE. Who's up first? Testing for order effects in structured interviews using a counterbalanced experimental design. J Abnorm Child Psychol. 1999;27:439-45.
- 32. Kemmler G, Holzner B, Kopp M, Dunser M, Margreiter R, Greil R, Sperner-Unterweger B. Comparison of two quality-of-life instruments for cancer patients: the functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30. J Clin Oncol. 1999;17:2932-40.
- 33. Cheung YB, Wong LC, Tay MH, Toh CK, Koo WH, Epstein R, Goh C. Order effects in the assessment of quality of life in cancer patients. Qual Life Res. 2004;13: 1217-23.
- Cheung YB, Lim C, Goh C, Thumboo J, Wee J. Order effects: a randomised study of three major cancer-specific quality of life instruments. Health Qual Life Outcomes. 2005;3:37.

Copyright of Journal of Public Health Dentistry is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.