Acta Odontologica Scandinavica, 2008; 66: 135–138

informa
healthcare

ORIGINAL ARTICLE

No significant retest effects in oral health-related quality of life assessment using the Oral Health Impact Profile

MIKE T. JOHN[1], DANIEL R. REIßMANN[2], OLIVER SCHIERZ[3] & FINBARR ALLEN[4]

[1]*Department of Diagnostic and Biological Sciences, University of Minnesota, Minneapolis, USA and* [2]*Department of Prosthetic Dentistry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany,* [3]*Department of Prosthodontics and Materials Science, University of Leipzig, Leipzig, Germany, and* [4]*Department of Restorative Dentistry, University Dental School and Hospital, Wilton, Cork, Republic of Ireland*

Abstract

Objective. To investigate retest effects in assessment of the Oral Health Impact Profile (OHIP) in patients requiring prosthodontic treatment. ***Material and methods.*** The German 49-item OHIP was administered in 21 patients requiring prosthodontic treatment on three pretreatment occasions: twice (t_1 and t_2) on day 1, 1–2 h apart, and on day 2 (t_3), 3–78 days later. Differences in reliability coefficients, OHIP summary scores, and item responses, and the influence of time on score differences were analyzed. ***Results.*** Some evidence of retest effects in oral health-related quality of life (OHRQoL) was found (the difference in reliability coefficients between t_1–t_2 and t_2–t_3 was 0.08; $p = 0.02$). ***Conclusions.*** The findings support repeated OHRQoL assessment over short periods of time when perceived oral health is rapidly changing. This is important because many oral conditions have acute symptoms and treatments have immediate OHRQoL effects that need to be measured.

Key Words: *OHIP, quality of life, temporal stability, test-retest reliability*

Introduction

Although patient-rated oral health-related quality of life (OHRQoL) has become a core outcome measure in dental research, measurement of the construct is still challenging. A key measurement property of any OHRQoL measure is its ability to capture the impact of an intervention by measuring pre/post-treatment differences. This is usually undertaken by administering questionnaires pre- and post-intervention, and measuring the difference between the two scores. Responses given in the second administration of a measure may be influenced by those given in the first administration (i.e. retest effects). Treatment effects may then be impossible to determine, because true post-treatment changes would be confounded with retest effects. In particular, this would affect assessment of conditions with acute signs and symptoms. At the moment, evidence supporting or refuting retest effects in OHRQoL is lacking.

The aim of this study was to investigate OHRQoL retest effects when using the Oral Health Impact Profile (OHIP), a widely used OHRQoL measure, in a clinical setting. Our research hypotheses were that retest effects would be present for OHRQoL measurements over short periods of time and absent between repeated assessments over longer time periods.

Material and methods

Subjects (n = 21, mean (SD) age 54.2 (14.8) years, 52% women) were a convenience sample of patients seeking prosthodontic treatment in the Department of Prosthodontics and Materials Science, University of Leipzig, Germany. They had completed an OHRQoL questionnaire on three occasions prior to active treatment procedures or any change in clinical oral health status (i.e. oral health was considered to be stable across the time-frame of questionnaire administration). In addition, evidence

Correspondence: Mike T. John, Department of Diagnostic and Biological Sciences, University of Minnesota, 515 Delaware Street SE, Minneapolis, MN 55455, USA. Tel: +1 612 624 3130. Fax: +1 612 626 0138. E-mail: mtjohn@umn.edu

(Received 12 August 2007; accepted 25 March 2008)

ISSN 0001-6357 print/ISSN 1502-3850 online
DOI: 10.1080/00016350802077835

for a substantial change in perceived oral health was absent from patient histories. Questionnaire data were collected at three time-points, namely:

- t_1: at the beginning of the first consultation when treatment was planned, day 1.
- t_2: after the first consultation was finished, about 1–2 h after t_1 on day 1.
- t_3: between 3 and 78 days later, i.e. the day when treatment started but before treatment, day 2.

OHRQoL was measured using OHIP-G, the German version [1] of the Oral Health Impact Profile [2]. The format for the questions was "In the past month, have you had ... because of problems with your teeth, mouth, or dentures?" and responses were given on a Likert scale ranging from 0 ("never") to 4 ("very often"). Summary scores (OHIP-G49) were derived by summing the response codes across all 49 items (possible range 0–196), with higher scores indicating poorer OHRQoL. Each time (t_1, t_2, t_3) the subjects answered the OHIP questions without having access to their previous responses, and they always used the same 1-month recall period.

It was hypothesized that retest effects would be present for OHRQoL measurements between t_1 and t_2, because the assessments took place on the same day, and absent between t_2 and t_3, because the two occasions were at least 3 days apart (mean time between administrations: 29 days, max: 78 days; number of patients >30 days: 9).

Several findings would favor the presence of retest effects at day 1. First, percent agreement, i.e. the proportion of identical OHIP responses for all subjects for all questions comparing t_1–t_2 and t_2–t_3, should be larger for day 1 compared with agreement on different days. Second, the correlated Pearson coefficients comparing OHIP summary scores on different occasions ($r_{t1–t2}$ versus $r_{t2–t3}$) should be different, because scores from the same day should correlate more highly than scores from different days. Small time-dependent test-retest differences would favor the absence of retest effects between day 1 and day 2. Differences <6.8 OHIP units would be considered as clinically not relevant based on results for baseline follow-up differences comparing patients treated with removable partial dentures to patients treated with fixed partial dentures in our patient population [3]. A regression analysis was performed to determine the influence of time between t_2 and t_3 on test-retest differences.

In addition to test-retest differences and Pearson correlation coefficients, intraclass correlation coefficients (ICCs) based on a one-way repeated measures ANOVA [4] were calculated; these are commonly applied for test-retest analyses. Guidelines were applied to judge the magnitude of reliability coefficients; coefficients >0.75 are considered excellent reliability [5]. Finally, in addition to temporal stability, internal consistency was measured using Cronbach's alpha [6]. All analyses were performed using the statistical software package STATA, Release 9 (StataCorp. 2005, Stata Statistical Software, College Station, Tx), with the probability of a type I error set at the 0.05 level.

The institutional review board of the School of Medicine, University of Leipzig, approved the study.

Results

OHIP-G49 summary scores on all three occasions presented similar internal consistency. Cronbach's alpha >0.95 and lower bounds of all confidence intervals (CIs) >0.92 were observed. No substantial differences were evident between occasions. Mean (SD) summary score differences between t_1 and t_2 were 2.0 (7.1) OHIP units and mean (SD) differences between t_2 and t_3 were -0.8 (3.5) OHIP units. Both mean differences were small and not statistically significant. These results indicated the absence of a change in OHRQoL over time. ICCs for assessment of test-retest differences were 0.97 (95% CI: 0.94–0.99) for t_1–t_2 and 0.89 (95% CI: 0.80–0.98) for t_2–t_3.

The two indications for retest effects on day 1 were present. First, percent agreement for OHIP responses for t_1–t_2 (79%) was larger than that for t_2–t_3 (77%). Second, testing the correlated Pearson correlation coefficients (which were identical to the test-retest intraclass correlation coefficients) for t_1–t_2 against t_2–t_3 presented statistically significant results. The difference between reliability coefficients was 0.08 ($p=0.02$) when comparing the t_1–t_2 with the t_2–t_3 interval.

A small and statistically non-significant coefficient for the time influence on test-retest differences was observed (regression coefficient for 1 day: -0.04, 95% CI: -0.29–0.21, R-squared and adjusted R-squared <0.01).

Discussion

The observed pattern of findings in this study fit a priori hypotheses about the presence of test-retest effects. Retest effects were present when they were expected to exist, because the test-retest correlation coefficient for the one-day reassessment was statistically significantly larger than the test-retest correlation coefficient for the reassessment that took place over several days. However, although the difference in coefficients was statistically significant, it was small at 0.08. Both correlation coefficients were still excellent according to guidelines for the interpretation of reliability coefficients [5]. Therefore, even though test-retest effects were found, they were small. Test-retest effects were absent when they

were expected to be absent, because the time influence on test-retest difference was small and statistically not significant when reassessment time intervals between 3 and 78 days were investigated. Based on previous results where we found 6.8 OHIP units as an average OHIP change difference over two follow-ups when comparing treatment-related change in clinically distinct groups of prosthodontic patients [3], we consider the test-retest difference change over the time period of 78 days as not clinically significant. This pattern of test-retest effects showed that retest effects (also carryover, practice, memory, learning, and fatigue effects) for the 49-item Oral Health Impact Profile did not undermine its reliability over short time periods.

To assess the temporal stability of questionnaire scores, any chosen interval between administrations is a compromise. If the time is too short, the initial response may be remembered and memory may contaminate test-retest reliability [7]. If the time is too long, true changes in the construct may be reflected in the scores, bearing in mind that quality of life is known to be a dynamic construct [8]. When a time interval of 1–2 weeks was used for OHIP, excellent reliability coefficients were observed [9–11]. For a time interval of 3 months between administrations, coefficients were lower, although Cronbach's alpha values characterizing the internal homogeneity of the scores were excellent [2,12]. These data may indicate a time influence on reliability, which is in line with our study results. However, when our findings are compared with retest reliability assessment for very short time intervals (1–3 days, ICC: 0.87 [13]), our results are slightly higher. Even though our sample size was not large, the confidence intervals indicated sufficient precision for most results. For example, both lower limits of the 95% confidence interval of the reliability coefficients were above the 0.75 value that is proposed as the threshold for excellent reliability [5]. Therefore, the existence of test-retest effects, as demonstrated by a statistically significant difference in reliability coefficients, did not change the categorized interpretation of the level of reliability. When we looked at the confidence interval limits of the time influence regression coefficient over longer time periods than about 1 month, our study did not have enough statistical power to exclude a substantial time influence on test-retest differences. For example, whereas the regression coefficient of -1.2 OHIP units for a 1-month time period (30 days times -0.04 OHIP units per day) indicated small test-retest differences, application of the lower limit of the 95% confidence interval of the regression coefficient resulted in -8.7 OHIP units (30 days times -0.29 OHIP units per day). This value of test-retest differences exceeded the 6.8-unit criterion that we set as the threshold for clinical significance.

In general, OHRQoL retest assessment is similar to health-related quality of life (HRQoL) instruments that have used intervals between 10 min and 1 month, with most studies using time intervals between 2 days and 2 weeks [14]. The presence of retest effects for HRQoL questionnaires is often not clear, and several explanations (such as social desirability and legitimation) have been offered [15,16].

The fact that retest (memory) effects were present in our study was not too surprising, given the very short interval between administrations. However, we suggest that the length of the instrument (i.e. the considerable number of questions) decreases the likelihood that individuals try to remember the answers given earlier when asked a second time. Our results may therefore not necessarily generalize to short OHIP versions. Support for this hypothesis comes from the Sickness Impact Profile, a generic HRQoL instrument with 235 items (long form) or 146 items (short form) where items are grouped within 14 categories [17]. Reliability coefficients for test-retest data with a mean of 24 h between administrations were high (0.96), but percent agreement between answers reached only 71%, a value similar to what we observed. The substantial variation in item answers over a very short time interval supports the absence of substantial memory effects in long questionnaires. However, this variation does not necessarily translate into insufficient reproducibility of the instrument summary score. The item's random error often cancels out, resulting in favorable summary score reliability.

The absence of substantial retest effects of OHIP scores is a prerequisite for assessment of OHRQoL changes over short time periods. This is important for assessment of perceived oral health, because many oral conditions have acute symptoms and treatments have immediate OHRQoL effects that need to be measured.

Declaration of interest: The authors reports no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

[1] John MT, Patrick DL, Slade GD. The German version of the Oral Health Impact Profile: translation and psychometric properties. Eur J Oral Sci 2002;110:425–33.

[2] Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. Community Dent Health 1994; 11:3–11.

[3] John MT, Slade G, Szentpetery A, Setz J. Oral health-related quality of life in patients treated with fixed, removable and complete dentures 1 and 6–12 months after treatment. Int J Prosthodont 2004;17:503–11.

[4] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

[5] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions, 3rd edn. Hoboken, NJ: John Wiley; 2003.

[6] Cronbach LJ. Coefficient alpha and the internal reliability of tests. Psychometrika 1951;16:297–334.
[7] McKelvie SJ. Does memory contaminate test-retest reliability? J Gen Psychol 1992;119:59–72.
[8] Allison PJ, Locker D, Feine JS. Quality of life: a dynamic construct. Soc Sci Med 1997;45:221–30.
[9] Saub R, Locker D, Allison P. Derivation and validation of the short version of the Malaysian Oral Health Impact Profile. Community Dent Oral Epidemiol 2005;33:378–83.
[10] 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.
[11] Yamazaki M, Inukai M, Baba K, John MT. Japanese version of the Oral Health Impact Profile (OHIP-J). J Oral Rehabil 2007;34:159–68.
[12] Bae KH, Kim HD, Jung SH, Park DY, Kim JB, Paik DI, et al. Validation of the Korean version of the oral health impact profile among the Korean elderly. Community Dent Oral Epidemiol 2007;35:73–9.
[13] de 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.
[14] Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003;56:730–5.
[15] Ormel J, Koeter MW, van den Brink W. Measuring change with the General Health Questionnaire (GHQ). The problem of retest effects. Soc Psychiatry Psychiatr Epidemiol 1989;24:227–32.
[16] Pevalin DJ. Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest effects. Soc Psychiatry Psychiatr Epidemiol 2000;35:508–12.
[17] Badia X, Alonso J. Validity and reproducibility of the Spanish Version of the Sickness Impact Profile. J Clin Epidemiol 1996;49:359–65.