The reproducibility of the Denplan Oral Health Score (OHS®) in general dental practitioners

S. Delargy[1], M. Busby[2], S. McHugh[3], R. Matthews[2] and F.J.T. Burke[1]

[1]*Primary Dental Care Research Group, The University of Birmingham School of Dentistry, St Chad's Queensway, Birmingham B4 6NN, UK;*
[2]*Denplan (UK), Victoria Road, Winchester SO23 7RG, UK;* [3]*University of Glasgow Dental School, Glasgow G2 3JZ, Scotland, UK.*

Objective: To assess the reproducibility of the Denplan Oral Health Score (OHS®) and to examine whether general dental practitioners are more reproducible at measuring oral health when they use the OHS than when they use their usual method of examining patients. ***Design***: This was a single centre study designed to compare dentists' reproducibility using the Oral Health Score with their reproducibility when using the method of examination that they normally employ in general practice. ***Setting:*** The study was carried out at The University of Birmingham School of Dentistry during 2001. ***Participants:*** Ten local general dental practitioners were recruited by advertising. The patients were selected from various clinics in the University of Birmingham School of Dentistry to provide a broad spectrum of oral health. ***Main Outcome Measures:*** The dentists scored oral health by two methods; firstly on a 10 cm linear scale after using their everyday method of examination and secondly using the structured format of the OHS examination, where the Oral Health Scores were expressed as a percentage. ***Results:*** Overall, the majority of the participating dentists demonstrated better intra-examiner reproducibility with the OHS than with their own method of examination. There was also higher inter-examiner reproducibility with the OHS than with the dentists' own method.

Key words: Oral health index, oral health score, reproducibility.

Introduction

Burke and Wilson (1995) suggested that the availability of an accurate, universally accepted method of measuring oral health would create numerous opportunities to advance oral health and dentistry worldwide, by providing the necessary feedback to general dental practitioners and third party insurers in relation to the outcome of treatment and the state of oral health of patients within their care. They further suggested that, internationally, a universal measure of oral health would allow comparison among different communities, along with an understanding and appreciation of the problems associated with them.

Whilst indices have been developed to measure specific dental diseases and conditions, it has been suggested that an overall index of oral health encompassing the different conditions may be useful in comparing population health status, inequality and in deciding on resource allocation (Marcus et al 1983).

Many composite measures of oral health have been described, including a comprehensive multidimensional Index of Treatment Need (IDN) (Lambert and Freeman, 1967), the Oral Health Grading (Bulman et al., 1968), the Oral Health Status Index (Marcus et al., 1983) and a modified version of this, the Childrens' Oral Health Status Index (COHSI) (Koch et al., 1985) and, more recently, the Clinical Minimum Data Set (Ireland et al 2001a). However, none of these has found acceptance in terms of widespread use by general dental practitioners .

Details of an Index of Oral Health (OHX) have been presented by Burke and Wilson (1995). Use of the index involves assessing patient comfort and satisfaction in addition to the assessment of caries, periodontal disease, tooth wear, mucosa, occlusion and, where appropriate, dentures. Previously accepted standards and/or indices were adapted to comprise the various elements of the index. Among these are the Adult Dental Health Survey criteria for caries (Todd and Lader, 1991), the clinical acceptability of restorations (Ryge, 1980) and the Faculty of General Dental Practitioners' self assessment manual (1991), tooth wear (Smith and Knight ,1984) and periodontal status as measured by CPITN (Ainamo et al, 1982).

The OHX has been subjected to two pilot studies on reliability, with the results of a study comparing two general dental practitioners indicating an acceptable value (Burke and Wilson 1995), and results of a study comparing four dentists (two GDPs and two hospital-based dentists) indicating correlation coefficients between the examiners of 0.94 to 0.99, demonstrating good correlation between the examiners (Burke et al., 1994). Further studies of the reproducibility of the Oral Health Index are ongoing.

The OHX was modified by a focus group at Denplan, a private, UK-based dental capitation company, to produce the Oral Health Score (OHS®). The components of the original OHX were not changed, nor were the scientifically based measurement protocols. The principal difference in the two scoring systems was the calculation of the overall score, so that the OHS was expressed as a percentage rather than a fraction of the maximum achievable score. If good reproducibility could be demonstrated among general dental practitioners, the

OHS could have many applications including: measuring predictability of treatment; quality assurance; peer review and patient motivation.

In order for the OHS to be widely accepted it must be valid, easy to use, involve a minimal amount of equipment and time and be reproducible. Ease of use and validity of the OHS have recently been assessed by Burke and co-workers (2003) with the results indicating that the OHS is an easy to use and valid measure of a patient's oral health. The aim of this investigation was to determine the inter-examiner and intra-examiner reproducibility of the OHS and to compare the reproducibility of the OHS with that of the dentists' usual method of examining patients (the Dentists' Own Score: DOS).

Method

This was a quantitative data study conducted at The University of Birmingham School of Dentistry. The fifteen patients involved were selected by the principal author from clinics in the University of Birmingham School of Dentistry in order to attempt to ensure a broad spectrum of oral health in both groups. The only inclusion criteria were that the patients should be over 18 years of age and available to attend on the selected dates for the study. Ten local general dental practitioners were recruited by local advertising and from BRIDGE (**B**irmingham **R**esearch **i**n **D**ental **G**eneral Practic**E**, a group of West Midlands dentists interested in practice based research) and asked to make themselves available for four sessions, which were scheduled at weekly intervals (Streiner and Norman,1995) to minimize a learning effect on behalf of the dentists. At the first session each dentist was asked to examine 10 patients using the method of examination they employed in everyday practice. They were then asked to indicate on a 10 cm line (where one end represented poor oral health and the other end good oral health) an oral health assessment for each patient. The following week the same group of patients was examined in the same way and a second score awarded on the linear scale. At the third session the dentists were given a presentation on the rationale and use of the OHS. At this session the dentists examined a different group of 10 patients using the OHS proforma . At the final session this second group of patients were examined again using the OHS examination, and a further OHS calculated. Five of the ten patients examined using the DOS also participated in the OHS sessions, therefore these patients' scores could be compared across the four visits.

The DOS were expressed as percentages in order to facilitate comparison with the OHS. Estimates of inter and intra-examiner reproducibility were calculated using the statistical technique of components of variance, with these components being estimated by means of repeated measures of analysis of variance (Dunn, 1989). The repeated measures statistical analyses also examined for a "Visit" effect for each dentist (i.e. a systematic difference between the mean Visit 1 and mean Visit 2 scores for a given dentist, which would indicate 'poor' reproducibility and /or learning effects on the part of the dentists).

Results

Table 1(a) presents the summary statistics of the DOS for each dentist for the two visits. At visit 1 the mean DOS ranged from 50% to 75%, with a similar range (48% to 78%) at visit 2. Table 1(b) presents the summary statistics of the OHS for each dentist for the two visits. On visit 1, the mean Oral Health Scores ranged from 71% to 80%, with a similar range (72% to 79%) at visit 2. Given such a small range of scores across the dentists, it is not surprising that there is little difference between the average OHS at the two visits. The range of scores in Tables 1(a) and 1(b) indicates more variability with the DOS than for the OHS. Table 1(a) demonstrates that the highest mean DOS was 78% while the lowest mean score was 48%. Table 1(b) demonstrates that the highest mean OHS was 80% while the lowest mean score was 71%.

Figure 1 illustrates that the five patients scored by both methods were scored higher with the OHS than the DOS. This figure also illustrates that there is a greater range of scores used for these patients with the DOS than the OHS.

The greater variability of the Dentists' Own Scores than the OHS is further illustrated in Figures 2(a) and (b), highlighted by the fact that there is more clustering of the scores around the line of equality with the OHS. It is also evident from the range of scores that the DOS uses more of the scale.

Figure 2(b) also shows that, whilst for many of the patients there would seem to be agreement between dentists on the OHS, it is not for all patients and there is a suggestion of better 'agreement' between dentists for patients with 'good' oral health than for patients with 'poorer' oral health. That is, for patients with "poorer" oral health, at the bottom left hand of the plot, there is an indication of more variability between dentists.

It is, however, possible to achieve a low OHS and Table 1(b) demonstrates that the lowest OHS recorded was 35 compared with the lowest DOS of 15.
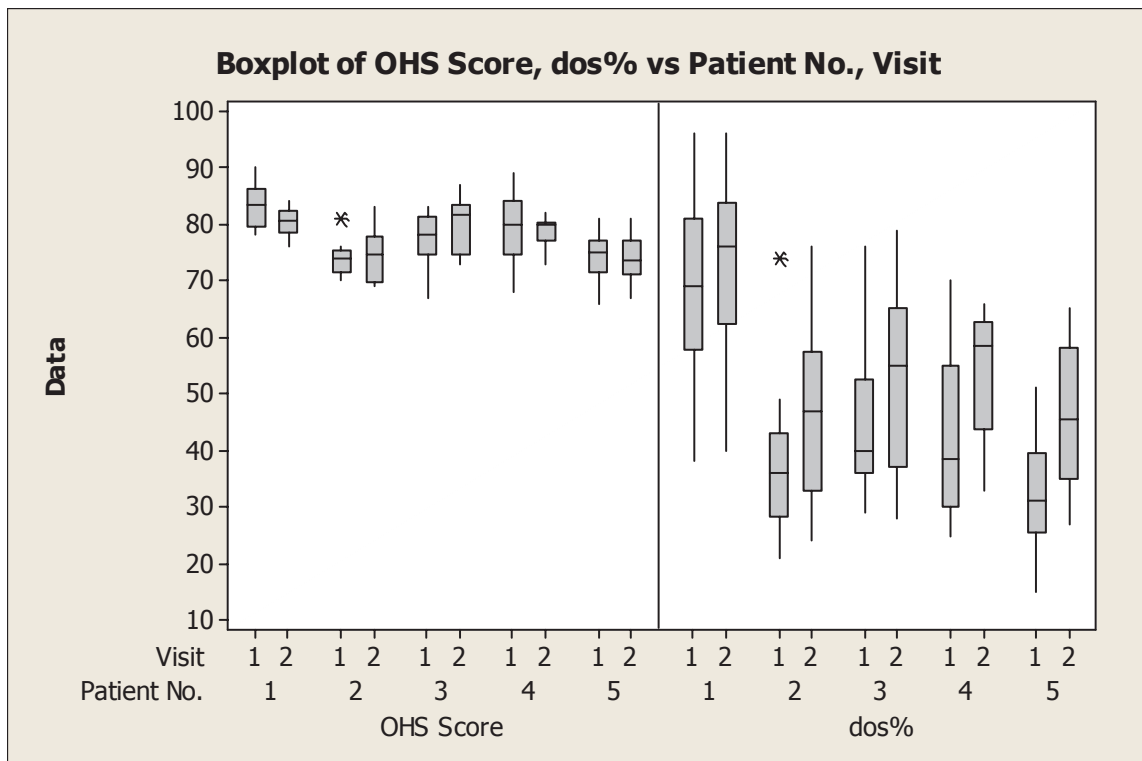
Inter-Examiner Reproducibility

Table 2 contains the estimates and 95% confidence intervals for the inter-examiner reproducibility of the DOS and the OHS. From these estimates, there is a suggestion of higher inter-examiner reproducibility when the structured format of the OHS is used rather than the DOS. The point estimate of inter-examiner reproducibility for the OHS is 82%, compared with 61% for the DOS. From the 95% confidence interval, "at worst" the inter-examiner reproducibility for the OHS could be 70% and "at best" 95%. In the case of the DOS, from the 95% confidence interval, "at worst" the inter-examiner reproducibility could be 48% and "at best" 88%. Shrout (1998), in a revision of the often quoted reproducibility standards provided by Landis and Koch (1977), suggests that an inter-examiner reproducibility estimate of 82% may be interpreted as substantial, whereas an inter-examiner reproducibility estimate of 61% may be interpreted as moderate. According to Shrout, inter-examiner reproducibility estimates in the range 81% to 100% may described as substantial, 61% to 80% as moderate and 41% to 60% as fair. The significant 'visit effect' for the DOS

Table 1(a). Summary Statistics of Dentists' Own Scores for Each Dentist

Dentist	Visit 1					Visit 2				
	Mean %	StDev	Min	Max	Range	**Mean%**	StDev	Min	Max	Range
1	**59**	24	26	96	71	**68**	21	41	96	55
2	**51**	15	32	69	36	**61**	4	55	67	12
3	**75**	11	50	89	39	**78**	9	64	87	24
4	**59**	23	27	94	68	**64**	21	27	90	64
5	**60**	25	21	91	71	**73**	21	33	94	61
6	**55**	25	29	96	67	**54**	25	23	91	69
7	**55**	32	15	96	81	**68**	20	38	97	59
8	**58**	29	28	99	72	**71**	16	57	96	39
9	**68**	26	31	98	67	**66**	27	24	98	74
10	**50**	19	28	84	56	**48**	22	28	83	56

Table 1(b). Summary Statistics of Oral Health Scores for Each Dentist

Dentist	Visit 1					Visit 2				
	Mean %	StDev	Min	Max	Range	**Mean%**	StDev	Min	Max	Range
1	**75**	11	53	92	39	**72**	14	40	86	46
2	**77**	15	39	92	53	**72**	16	35	92	57
3	**80**	10	55	96	41	**79**	7	67	92	25
4	**75**	10	53	92	39	**76**	10	52	88	36
5	**76**	12	48	92	44	**78**	8	63	92	29
6	**73**	13	43	92	49	**72**	12	44	84	40
7	**76**	9	53	92	39	**75**	12	52	92	40
8	**78**	10	58	92	34	**75**	12	50	92	42
9	**71**	13	41	90	49	**73**	12	45	94	49
10	**74**	11	52	90	38	**73**	12	47	90	43

Figure 1. *Boxplots of 5 Patients' Scores by Both Methods, across all 10 Dentists*
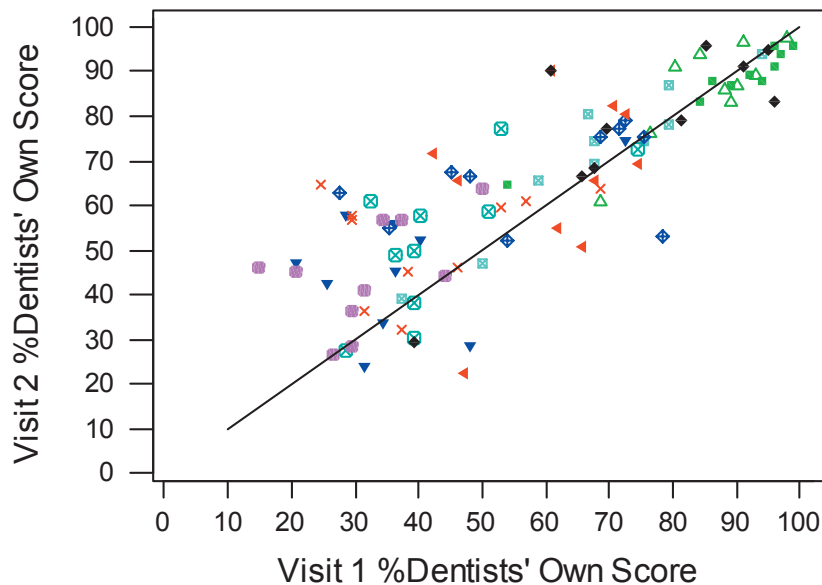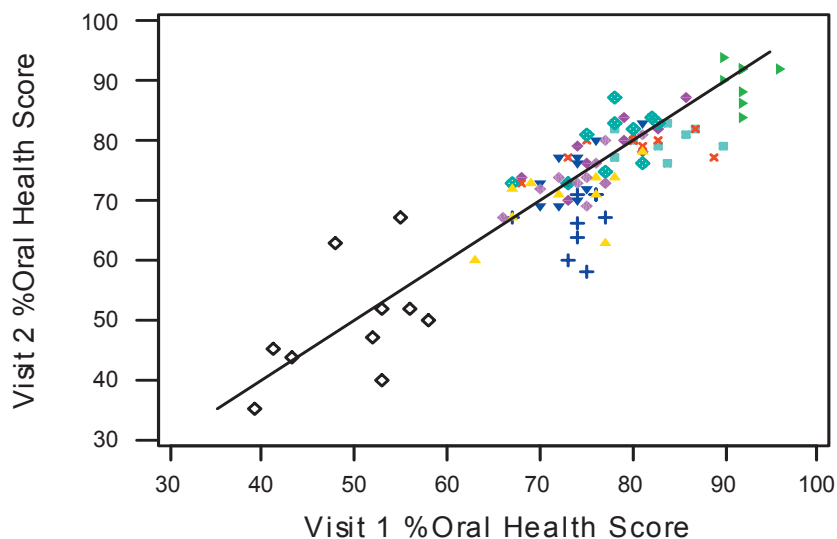
Figure 2(a). *Plot of Dentists' Own Scores*

Figure 2(b). *Plot of Oral Health Scores*

Table 2. Inter-Examiner Reproducibility

Estimate of Inter-Examiner Reproducibility (95% CI)	
Dentists' Own Score	*OHS*
61%[1]	82%
(48, 88)%	(70, 95)%

[1] Significant visit effect

was due to five of the dentists (dentists 1, 2, 5, 7 and 8) scoring patients higher on the second visit than the first, introducing a 'systematic bias' towards visit 2.

Intra-Examiner Reproducibility

Table 3 presents point estimates of intra-examiner reproducibility of the OHS. Good agreement was found between the two visits for most of the dentists, with the intra- examiner reproducibility estimates ranging from 81% to 97%. Table 3 also presents the point estimates of reproducibility for the DOS, which range from 41%

to 98% and demonstrates that 7 out of 10 dentists show greater reproducibility when using the OHS than the DOS. Significant visit effects are highlighted for dentists 1,2,5,7 and 8 with the DOS indicating poor reproducibility. The summary statistics for the DOS in Table 1(a) show that these dentists scored patients' oral health as being higher on visit 2.

Table 3 also demonstrates that there is a significant visit effect for dentists 2 and 8 when they used the OHS. These dentists scored oral health higher on visit 1 than on visit 2 and this bias towards visit 1 can be seen in Table 1(b).

Table 3. Inter-Examiner Reproducibility

	Estimate of Intra-Examiner Reproducibility	
	Dentists' Own Score	OHS
Dentist 1	97%[1]	87%
Dentist 2	41%[1]	92%[1]
Dentist 3	80%	81%
Dentist 4	76%	92%
Dentist 5	85%[1]	82%
Dentist 6	92%	92%
Dentist 7	77%[1]	89%
Dentist 8	79%[1]	93%[1]
Dentist 9	98%	92%
Dentist 10	87%	97%

[1] Significant Visit Effect
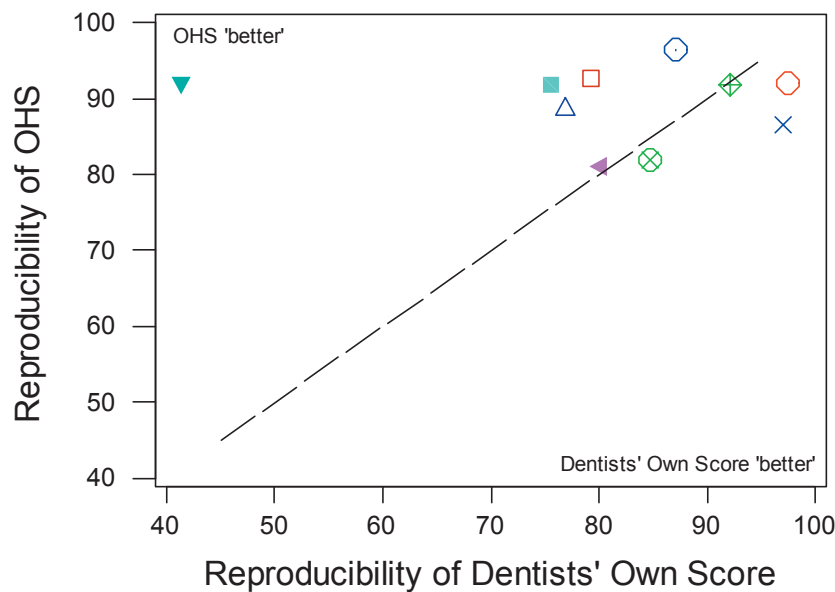
Figure 3. *Plot of Intra-Examiner Reproducibility Estimates for Each Dentist*

Figure 3 compares of the estimates of intra-examiner reproducibility of the OHS and DOS and illustrates that six of the dentists had greater reproducibility using the OHS than the DOS.

Discussion

Adoption of an oral health scoring method appropriate to general practice is of more relevance today than when the original paper on OHX was published in 1995, given the increasing impetus of clinical governance. In order for the OHS to be adopted for widespread use in practice it must be shown to be reproducible. Whilst a number of other composite measures of oral health exist, none of them have been adopted for widespread use in clinical practice, so it was therefore considered inappropriate to compare the OHS with these other indices. It was felt that the reproducibility of the OHS should be compared with the methods which general dental practitioners currently adopt to assess a patient's oral health. Whilst the authors accept that getting dentists to 'score' is not part of their usual examination practice, it was required in order to compare the two methods of examining patients. Indeed, the participating dentists had no objections to the methodology used and stated that it made them "think" more about their perceptions of patients' oral health.

The "visit effect", in which five dentists systematically scored patients' oral health as being higher on visit 2 when using the DOS, demonstrates the poorer reproducibility of this method since care was taken to ensure that the patients did not receive any dental treatment between visits.

The majority of dentists involved in this study were more reproducible within themselves when they used the OHS rather than their own method of scoring patients. In imposing weightings for each component of the OHS, the variation both between and within examiners should be limited, leading to improved reproducibility. The structured format of the OHS should ensure that all the components of oral health are examined and may be considered thereby to minimize operator error. In this respect, Ireland and co workers (2001 b) demonstrated that the quality of clinical record keeping of a sample of 50 dentists, piloting the Denplan Excel programme, improved on their adoption of the structured examination.

Patients tend to have higher scores when the OHS is used than when the DOS is used as demonstrated in Figure 1. The participating dentists did not tend to score

the OHS at less than 60%, with the exception of one patient, who scored 35%. Seven of the patients scored lower than 60% when the dentists used their own scoring method, the lowest score awarded being 15%. This may be due to the fact that, when examining patients using their own scoring system, dentists score patients with several missing teeth, dentures or heavily restored teeth as having relatively poor oral health, whereas such a patient can achieve a high OHS as this method distinguishes between dental status and oral health.

The intra-examiner reproducibility of the OHS is substantial (Shrout, 1998) and the OHS therefore has potential applications in practice such as patient motivation, measuring effectiveness of treatment and measuring oral health over time. The inter-examiner reproducibility is also good, suggesting that it could be adopted for epidemiology or comparing the performances of different practices. In general, there was better agreement when the OHS was used for patients with "good" oral health. This is to be expected as more variables are introduced as oral health deteriorates. Further work needs to be done on investigating the reproducibility of the OHS in patients with poor oral health. In this respect, it is suggested that the reproducibility of the individual components of the OHS should be assessed to identify which components are less reproducible before measures can be taken to improve the reproducibility of the overall OHS.

Finally, the results of this study suggest that there are likely benefits to patients for whom examinations are completed using the OHS, given that a standardized, structured examination routine may produce a more robust assessment of a patient's oral health.

Acknowledgements

References

Advisory Board in General Dental Practice. Self-assessment manual and standards. (1991): London, Royal College of Surgeons of England.

Ainamo J., Barnes D., Beagrie G., Cutress T., Martin J. and Sardo-Infirri J. (1982): Development of the World Health Organisation (WHO) Community Periodontal Index of Treatment Needs (CPITN). *International Dental Journal* **32,** 281-291.

Bulman J.S., Richards N. D., Slack G.L. and Willcocks A.J. (1968): Demand and need for dental care: a sociodental study. Oxford University press, London, pp97-103.

Burke F.J.T., Greene P.R., Roberts C. (1994): Reproducibility studies on a newly designed index of oral health. *Journal of Dental Research* **73,** 449.

Burke FJT, Wilson NHF. (1995): Measuring oral health: an historical view and details of a contemporary oral health index (OHX). *International Dental Journal* **45,** 358–370.

Burke FJT, Busby M., McHugh S, Delargy S., Mullins A., Matthews R.. (2003): Evaluation of an oral health scoring system by dentists in general dental practice. *British Dental Journal* **194,** 215-218.

Dunn G. (1989). Design and analysis of reliability studies. Oxford University Press, New York.

Ireland R.S., Jenner A.M., Williams M.J., and Tickle M. (2001a): A Clinical Minimum Data Set for Primary Dental Care. *British Dental Journal* **190**: 663-667.

Ireland R.S., Harris RV, Pealing R. (2001b): Clinical record keeping by general dental practitioners piloting the Denplan " Excel" accreditation programme. *British Dental Journal* **191,** 260-263.

Koch A.L, Gershen, J.A and Marcus, M. A (1985) Children's oral health status index based on dentists' judgment. *Journal of the American Dental Association* **110**: 36-42.

Landis J., Koch G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**:159-174.

Lambert C. Jr. and Freeman E. (1967) The clinic habit. College and University Press. Newhaven, Connecticut.

Marcus M., Koch A.L. and Gershen, J.A. (1983) Construction of a population index of adult oral health status derived from dentists' preferences. *Journal of Public Health Dentistry*, **43(4)**: 284-294.

Ryge, G. (1980): Clinical criteria. *International Dental Journal* **30,** 347-357.

Shrout, P.E. (1998): Measurement reliability and agreement in psychiatry. *Statistical Methods for Medical Research* **7,** 301-317.

Smith, B.G.N. and Knight, J.K. (1984): An index for measuring the wear of teeth. *British Dental Journal* **156,** 435-438.

Streiner, D.L. Norman G.R. (1995) Health Measurement Scales: A Practical Guide to their Development and use. 2nd edn. Oxford University Press, Oxford.

Todd, J .E. and Lader, D. (1991) Adult Dental Health 1988. United Kingdom. London, HMSO.