# Saliva-based colorimetric test as an index of gingival inflammation in epidemiologic studies

BRIAN A. BURT, DAVID M. RODER, JAMES C. CECIL AND STEPHEN A. EKLUND

*Program in Dental Public Health, School of Public Health, The University of Michigan, Ann Arbor, Michigan, U.S.A.*

Abstract – There is a pressing need for a reliable, low-cost method of assessing the gingival and periodontal status of large population groups. Existing indexes, despite their value in dental public health, are still too subject to examiner variability for use by uncalibrated examiners. This study describes an evaluation of a quick, inexpensive, extraoral colorimetric test for gingival inflammation, based on a reaction between saliva and the test material. It could probably be applied by non-professional personnel. In this study, the test was applied to a population of elementary schoolchildren, dental hygiene students and faculty, and adult inmates of two correctional institutions. These populations were chosen on the assumption that they would exhibit varying intensities of gingival inflammation. Values obtained from the colorimetric test carried out by one researcher were compared with Gingival Index (GI) scores observed by a different examiner. Results suggest that the colorimetric test may be a valid, reliable means of detecting major differences in the prevalence of gingival inflammation in *most adult populations*, although having little, if any, useful application among children at the mixed dentition stage.

Monitoring the gingival and periodontal status of large population groups has long been a major problem (3, 16), for no index yet developed possesses the desirable combination of simplicity, validity, reliability and low cost of application. Although established indexes such as RUSSELL's PI (12), RAMFJORD's PDI (11), and LÖE's GI (9) have led to major increases in knowledge of the epidemiologic aspects of gingival and periodontal disease, all require trained examiners and calibration if variability within and between examiners is to be controlled (2, 4–7, 13, 15). After a long history of addressing inherent problems in periodontal indexes, the World Health Organization produced its first

"Basic methods" manual (16), which suggested a simple index for the collection of comparable data from a variety of environments where examiner training was impractical. Despite the simplicity of the index, however, experience from over 30 health administrations showed that reliability remained a big problem (6, 17). A number of revisions, suggested as a result of this testing, have been included in the second edition of the "Basic methods" manual (17).

Clearly, there is a need for simple, low-cost methods of determining the gingival and periodontal status of large populations with reasonable reliability. Given the inherent difficulties with any index

requiring the clinical assessment of periodontal tissues, efforts to develop valid and reliable extraoral tests appear worthwhile. It is perhaps more likely that universal measures of gingival and periodontal disease, intended for the purposes of broad global epidemiologic comparisons as described by the World Health Organization (16, 17), will come from extraoral tests rather than from any type of direct clinical evaluation.

The purpose of this paper is to describe the evaluation of an extraoral, colorimetric test for gingival inflammation, marketed under the brand name Gindex® (Janar Company, P. O. Box 1845, Grand Rapids, Michigan 49501, U.S.A.). The test is intended to measure the extent of gingival inflammation by recording colorimetrically the level of hemoglobin in a sample of saliva (about 5 ml), obtained by having the subject swish the saliva around the mouth for 1 min and then spitting it into a beaker. The validity of the test is based on several assumptions: (a) that the hemoglobin in the exudate coming from the crevices of inflamed gingival units is dispersed evenly in the saliva, (b) that the hemoglobin is the only constituent of the saliva which reacts with the test material, and (c) that the hemoglobin level in the saliva sample relates directly to the extent and severity of gingival inflammation in the mouth.

When added to hemoglobin, the hydrogen peroxide in the Gindex test liquid releases oxygen, which can oxidize orthotolidine dihydrochloride to a blue derivative. The intensity of this color change is matched against a color chart and given a value from 25 to 100 (100 represents maximum color change) to denote the level of hemoglobin in the saliva sample, and presumably therefore the extent of the inflammation in the mouth. This scale appears to be ordinal, and the numbers used (25–100) are chosen arbitrarily to describe increasing intensity of color change.

Gindex has been previously tested, under clinical conditions, on 81 subjects aged 15–60 years, by comparing Gindex values with scores from Löe's Gingival Index (1). A statistically significant positive correlation (r = 0.57) was reported. These promising results appeared to be worth testing in a field situation. Since (a) the retail cost of the Gindex material was only about US $ .50 per test, (b) each test would only take about 3–4 min per subject, and (c) the test could possibly be performed by non-professional persons, the potential of the technique for epidemiologic purposes seemed considerable.

## MATERIAL AND METHODS

Three populations in southeastern Michigan were purposively chosen for the study. They were:
A. 136 children in grades 3 and 4 at an elementary school, comprising 67 females and 69 males. One child was 7 years old, 63 were 8 years old, 60 were 9 years old, and 12 subjects were aged 10 years.
B. 52 adults from a school of dental hygiene, of whom 46 were female students aged 18–23 years, two were female hygienists aged 26–27 years, and four were male dentists aged 24–32 years.
C. 65 adults, comprising 63 males from two correctional institutions plus two low-income females. The ages ranged from 18 to 66 years with 30 subjects under 25 years of age, 27 aged 25–34 years, and 8 aged 35 years or more.

These groups were specifically chosen in order to have contrasting levels of disease to test. It was presumed that the children would have considerable levels of gingivitis resulting from dental plaque, the eruption of teeth, and exfoliation. Adults from correctional institutions were considered likely to have substantial levels of gingivitis, whereas subjects from the school of dental hygiene were thought likely to have very low levels of gingival inflammation.

All subjects were seen by the one examiner (B.A.B.) with mouth mirrors and explorers. A portable dental chair and fiberoptic light were used when surveying the children, whereas fixed dental operatory facilities were available for examining the adult populations.

Gingival Index (GI) values were obtained using the criteria and scoring method described by Löe (9). To save time, a part-mouth approach was used (9); examinations were restricted to gingival tissues adjacent to the following permanent teeth: (a) the maxillary right and mandibular left central incisors; (b) the maxillary right and mandibular left first molars; and (c) the maxillary left and mandibular right first premolars. The premolars were often unerupted in the children; in these instances, the corresponding lateral incisors were substituted for examination. Only teeth considered to be fully erupted were scored.

A total of 89 children were examined twice for the GI, 2–7 days apart, to assess intraexaminer variability. The examiner was unaware of the subjects' initial scores when reexamining, though he did know when the repeat examinations were in progress. It was not practical for repeat examinations to be carried out for the adult populations.

A second researcher (D.M.R.) independently obtained a saliva sample from each subject. The sample was obtained by having the subject first rinse the mouth with water, then swish subsequently produced unstimulated saliva around the mouth for 1 min, then spit the saliva into a beaker. This routine is that described by the manufac-

turer. It is presumed that the salivary swishing picks up hemoglobin in the exudute from areas of gingival inflammation, though it should be emphasized here that the test uses saliva, not crevicular fluid.

Subjects provided saliva samples concurrently in groups of four, thereby reducing the time for the Gindex test to between 3 and 4 min per person.

The extent of blue color change achieved in the test was matched against the manufacturer's color chart by the same person who obtained the saliva, and assigned a value of 25, 40, 55, 70, 85, or 100. Occasionally color changes fell between optional shades on the color chart; in these instances, a midpoint value was assigned.

Two reliability checks of the Gindex test were carried out. In the first, to assess intraexaminer consistency at reading the color values, two saliva samples from 81 children, taken 2–7 days apart, were tested independently by the one examiner (D.M.R.). Reliability of the test itself was checked by having two portions of the same saliva sample from 73 children tested by D.M.R. and a third examiner who took no further part in the study. No attempt was made to standardize these two Gindex examiners.

The correspondence between Gindex and GI values was determined using Spearman's coefficient of rank correlation (8), as were the intraexaminer variability checks of both GI and Gindex and the reliability check of the Gindex. During the course of the study, it became apparent that a marked differentiation in the intensity of color change existed between Gindex values of 70 or more and lower values. Because it seemed that this differentiation might have enhanced examiner consistency, Gindex values were also classified on a dichotomous scale divided at this level.

## RESULTS

### ASSOCIATIONS BETWEEN GINDEX AND GI

For the 136 schoolchildren, there was no direct correlation between the Gindex results and GI scores ($r_s = 0.13$; $P > 0.05$). However, when data were separated out for the 73 children for whom two Gindex tests were carried out to assess reliability of the test, it was found that their mean Gindex score was considerably higher than that for other children (Table 1). There was no apparent reason for this finding, presumably it was a chance occurrence. However, it was thought advisable to keep data for these two groups of children separate, as a major purpose of the study was to seek intergroup associations between Gindex test results and GI scores. Accordingly, mean Gindex and GI scores, as well as proportions of subjects with Gindex scores of at least 70, were calculated for the following four groups: (a) the 73 children reexamined to evaluate the reliability of the Gindex test; (b) the remaining

Table 1. Mean Gingival Index (GI) scores, mean Gindex scores, and percentages with Gindex scores of 70 or more, by study population

| Study population | $n$ | Mean GI ($\pm$ s.e.) | Mean Gindex ($\pm$ s.e.) | Present with Gindex $\geq 70$ |
|---|---|---|---|---|
| Dental hygiene group | 52 | $0.20 \pm 0.02$ | $44.76 \pm 3.33$ | 30.8 |
| Schoolchildren | 63 | $0.54 \pm 0.03$ | $48.76 \pm 2.86$ | 33.3 |
| Schoolchildren* | 73 | $0.68 \pm 0.03$ | $52.49 \pm 3.24$ | 42.5 |
| Institutionalized group | 65 | $0.95 \pm 0.05$ | $58.46 \pm 3.50$ | 49.2 |

* The 73 schoolchildren were those who received a Gindex test from each of two examiners, to check the reliability of the test.

63 children; (c) the dental hygiene group; and (d) the institutionalized group. Results (Table 1) demonstrate quite marked intergroup differences in GI scores, though intergroup differences in Gindex values are less clearcut. Irrespective of the index used (mean GI value, mean Gindex value, or percentage of subjects with Gindex values of 70 or more), however, a similar direction of differences is evident between the four population groups, though again it is possible that this finding could be due to chance.

No intragroup correlation between GI scores and Gindex values was apparent for the child populations or the dental hygiene group, though a direct correlation was evident for the institutionalized group ($r_s = 0.41$; $P < 0.01$). When Gindex values

Table 2. Cross-tabulation of dichotomous Gindex values with Gingival Index (GI) scores, for three populations

| Population | GI scores | Gindex values | |
|---|---|---|---|
| | | $<70$ | $\geq 70$ |
| 136 schoolchildren | 0 −0.6 | 50 | 27 (35.1%) |
| | 0.61 −1.2 | 30 | 21 (41.2%) |
| | 1.21 −1.8 | 4 | 4 (50.0%) |
| 65 inmates of correctional institutions | 0 −0.6 | 19 | 6 (24.0%) |
| | 0.61 −1.2 | 11 | 20 (64.5%) |
| | 1.21 −2.0 | 3 | 6 (66.7%) |
| 52 dental hygiene students and faculty | 0 −0.2 | 18 | 10 (35.7%) |
| | 0.21 −0.4 | 13 | 5 (27.8%) |
| | 0.41 −0.6 | 5 | 1 (16.7%) |

were dichotomized (scores of 70 or more, below 70), it seemed appropriate to relate them to GI values by regression analysis, in which (a) the Gindex value was the indicator variable; (b) age and sex were covariables, and (c) the GI value was the dependent variable (10). The regression coefficient thus derived was 0.18, which while statistically significant ($P < 0.05$), was not large.

Table 2 shows cross-tabulations between dichotomous GI data and GI scores for the child population and the two adult groups; again differences are sharpest in the institutionalized group.

EXAMINER CONSISTENCY

A total of 89 children were examined clinically for GI scores on two occasions by the same examiner. The mean GI value was $0.57 \pm 0.03$ (s.e.) for the first examinations, compared with $0.62 \pm 0.03$ (s.e.) for repeat examinations, an increase of 8.8 %. The correlation coefficient ($r_s$) for the two sets of values was 0.76 ($P < 0.01$).

A total of 81 children provided two saliva samples, 2–7 days apart, both read by the same examiner to check intraexaminer variability. The mean Gindex value for the first tests was $50.44 \pm 2.84$ (s.e.), and $50.59 \pm 2.84$ (s.e.) for repeat tests. However, the correlation coefficient for the two sets of values was only 0.25 ($P < 0.05$). When the percentages of scores that were 70 or more were compared, results again were similar; 37.0 % had scores of at least 70 in the first test, and 35.8 % were scored at this level in the second test. However, only 13 (43.3 %) of the 30 subjects with an initial rating of 70 or more also received a rating of this level in the second test, suggesting considerable intraexaminer variability. Among the 51 subjects who received a score of less than 70 in the first test, 16 (31.4 %) were assigned a value of 70 or more in the subsequent test. Therefore, although the examiner appeared highly consistent when assessing the group's collective Gindex values, the agreement between successive findings for individual children was extremely low. It cannot be stated with certainty that this latter finding was solely a result of true intraexaminer variability, for it could have at least partly represented some inherent variability within the product.

To assess the reliability of the Gindex test, two examiners independently tested the same saliva sample from each of 73 children. The mean values

obtained were $52.49 \pm 3.24$ (s.e.) and $57.77 \pm 2.85$ (s.e.), the second mean being 10.1 % higher than the first. The correlation coefficient ($r_s$) for the two sets of values was 0.60 ($P < 0.01$). One examiner found 31 children (42.5 %) to have Gindex values of 70 or more, and the corresponding value for the second researcher was 32 (43.8 %). Of the 31 children who were assigned a value of 70 or more by the first examiner, 24 received a similar score from the second – a number significantly higher than would occur by chance in 1 % of occurrences ($\chi^2 = 22.37$; 1 df).

No check was made on whether differences in color perception existed between these two Gindex examiners, but it is possible that interexaminer differences were at least partly a result of this factor.

DISCUSSION

In the present study, overall Gindex data for the child populations were similar, irrespective of whether they were obtained by two different examiners or one examiner at different times. The similarity was most apparent when Gindex data were classified on a dichotomous scale. From this limited evidence, it seems that the Gindex test might be useful for broad comparisons in situations where a calibration of researchers is impractical.

To be valid, the Gindex values must relate to the condition under study. Therefore, it was encouraging to observe in the present investigation that intergroup differences in gingival inflammation were similar, irrespective of whether they were expressed by Gindex or GI values. Nevertheless, the non-significant associations between Gindex and GI values among children and dental hygiene subjects, together with the low intraexaminer and interexaminer consistency of Gindex values for individual children, indicate that the Gindex test has little, if any, intragroup discriminatory power within these populations.

For subjects in the correctional institutions, however, there was a significant direct association between Gindex and GI values, indicating that the Gindex test could discriminate between different levels of gingival inflammation within this population group. The finding is consistent with ABBOTT's observation of a similar correlation for 81 subjects aged 15–60 years (1). It seems evident that AB-BOTT's study population and the present institution-

alized subjects had similar, moderate levels of gingival disease, as indicated by GI values of about 1.0.

The lack of a significant association between Gindex and GI values in children may at least partly be a result of the eruptive and exfoliative gingivitis occurring in this age group. We believe that inflammation of this origin frequently was not reflected in the GI values, particularly as values were obtained through partial recording. Moreover, it was observed that children often had difficulty following the manufacturer's instruction to swish saliva between their teeth before expectorating, being easily distracted by their classmates when asked to provide samples concurrently in small groups. Perhaps the saliva samples obtained from children were unsatisfactory for this reason.

A number of false positives are apparent in Table 2; some high Gindex scores were obtained for the dental hygiene group, even though very little gingivitis could be found in any of these subjects during the clinical examinations. Several of the dental hygiene subjects who registered high Gindex scores commented that they were under "examination stress," and another was known to be diabetic. What effect stress, diabetes, and other oral environmental factors may have on the colorimetric reaction is unknown, but they are perhaps worthy of further investigation. It could be suggested, too, that the manufacturer look into the possibility of incorporating a colorimeter with the Gindex kits, so as to reduce variability in results attributable to human differences in color perception.

It is possible that the correlation coefficients between GI and Gindex values of 0.57 (1) and 0.41 (present study) may underestimate the association of Gindex values with *actual* disease levels. Even the level of intraexaminer variability demonstrated for GI values in the present study ($^r$s = 0.76; $P <$ 0.01) could have led to an underestimation of the real association. In any event, it is worth noting that even the GI and PI only correlated against each other at the 0.7 level when applied to an adult population by the researchers who originated the indexes (14). Moreover, weaker associations have been detected between other widely accepted indexes (14). In this context, the correlation coefficient obtained between Gindex and GI values for inmates of correctional institutions in the present study seems reasonable.

Also, we must state that we are not necessarily accepting the GI as an absolute measure of the quantity and quality of gingival disease. In our opinion, the Gingival Index should not be looked upon as a fine discriminatory measure, especially when partial recording is used. As a consequence, there may be little clinical difference in disease quality and quantity associated with the three levels of GI scores presented for the dental hygiene group in Table 2. In that event, the lack of a correlation between Gindex and Gingival Index scores would be understandable. When broader differences in GI scores exist, as with the institutionalized group, the discriminatory power of the Gindex test becomes more apparent.

It must be emphasized again that the Gindex test, in its present form, has no place in clinical trials, nor in any situation where highly sensitive measures are called for. However, for broad epidemiologic studies with adult groups it may have considerable utility as a quick and inexpensive index which does not require trained examiners.

# REFERENCES

1. ABBOTT, B. H.: *The reliability of the Gindex™ test in determining gingival inflammation.* Thesis. University of Michigan, Ann Arbor 1977.
2. ALEXANDER, A. G., LEON, A. R., RIBBONS, J. W. & MORGANSTEIN, S. I.: An assessment of the inter- and intra-examiner agreement in scoring gingivitis clinically. *J. Periodontal Res.* 1971: *6:* 146–151.
3. BARMES, D. E.: What a public health planner wants from a periodontal assessment. *Int. Dent. J.* 1976: *26:* 429–434.
4. DAVIES, G. N.: The different requirements of periodontal indices for prevalence studies and clinical trials. *Int. Dent. J.* 1968: *18:* 560–569.
5. DAVIES, G. N. & BARMES, D. E.: An evaluation of proposed revisions to the WHO manual "Oral health surveys – Basic methods." *Community Dent. Oral Epidemiol.* 1976: *4:* 55–65.
6. DAVIES, G. N., HOROWITZ, H. S. & WADA, W.: The assessment of periodontal disease for public health purposes. *J. Periodontal Res.* 1974: *9:* 62–70.
7. DAVIES, G. N., KRUGER, B. J. & HOMAN, B. T.: An epidemiological training course. *Aust. Dent. J.* 1967: *7:* 17–28.
8. GIBBONS, J. D.: *Nonparametric statistical inference.* McGraw-Hill, New York 1971.
9. LÖE, H.: The gingival index, the plaque index, and

the retention index systems. *J. Periodontol.* 1967: *38:* Suppl.: 610–616.

10. NETER, J. & WASSERMAN, W.: *Applied linear statistical models; regression, analysis of variance, and experimental designs.* Irwin, Homewood, IL 1974.

11. RAMFJORD, S. P.: The Periodontal Disease Index (PDI). *J. Periodontol.* 1967: *38:* Suppl.: 602–610.

12. RUSSELL, A. L.: A system of classification and scoring for prevalence surveys of periodontal disease. *J. Dent. Res.* 1956: *35:* 350–359.

13. SHAW, L. & MURRAY, J. J.: Diagnostic reproducibility of periodontal indices. *J. Periodontal Res.* 1977: *12:* 141–147.

14. SHIP, I. I., COHEN, D. W. & LASTER, L.: A study of gingival, periodontal, and oral hygiene examination methods in a single population. *J. Periodontol.* 1967: 38: Suppl.: 66–76.

15. SMITH, L. W., SUOMI, J. D., GREENE, J. C. & BARBANO, J. P.: A study of the intra-examiner variation in scoring oral hygiene status, gingival inflammation and epithelial attachment level. *J. Periodontol.* 1970: *41:* 671–674.

16. World Health Organization: *Oral health surveys – Basic methods.* 1st ed. WHO, Geneva 1971.

17. World Health Organization. *Oral health surveys – Basic methods.* 2nd ed. WHO, Geneva 1977.