The Effect of Varying Diagnostic Thresholds upon Clinical Caries Data for a Low Prevalence Group

N.B. Pitts and H.E. Fyffe

J DENT RES 1988 67: 592
DOI: 10.1177/00220345880670031401

The online version of this article can be found at:
http://jdr.sagepub.com/content/67/3/592
The Effect of Varying Diagnostic Thresholds upon Clinical Caries Data for a Low Prevalence Group

N. B. PITTS and H. E. FYFFE

Dental Health Services Research Unit, Department of Periodontology and Community Dentistry, Dental School, Park Place, Dundee DD1 4HR, Scotland

This study investigated the impact of employing differing diagnostic thresholds on clinical caries data in studies of groups with low caries prevalence. Data from clinical examinations of 287 Hong Kong dental students were analyzed by means of the CARIES microcomputer software package. This software allows for re-calculation of raw data according to three different diagnostic thresholds (D3, D2, and D1). When “enamel” and “initial” lesions (as defined by WHO criteria) were included in the calculation of DMFT, its value increased from 3.0 (D3) to 5.9 (D1), while the percentage of individuals considered “caries-free” decreased from 28.2% to 7.0%. Little change was found in the magnitude of the intra-examiner reproducibility, when calculated at each threshold, for a random 10% of the subjects. It was unfortunately not possible to calculate inter-examiner reproducibility in this study. The use of criteria which might be misinterpreted as being similar, but which use differing effective diagnostic thresholds, can dramatically influence the reported level of dental caries. In view of these findings, it may be necessary for the question of diagnostic thresholds to be re-examined and to receive greater emphasis in future studies.


Introduction.

Most epidemiological caries studies employ criteria in which a tooth, or tooth surface, is only recorded as decayed when cavitation is obvious, and which ignore all caries present at a less severe level (Klein et al., 1938; Todd et al., 1982; Downey and Teagle, 1979). In 1981, the Federation Dentaire Internationale (FDI) adopted a number of “global goals for oral health” which proposed that by the year 2000 the global average for Decayed, Missing, and Filled teeth (DMFT) in 12-year-olds should be no more than 3 (FDI, 1982). Here, clinical caries was to be diagnosed at the level at which clinical intervention was required, which is commonly taken to mean when clinical cavitation has occurred. The workshop on Epidemiology of Dental Caries set up by the British Association for the Study of Community Dentistry in 1982 also recommended one, relatively gross, level of diagnosis: “caries at the cavitation level, i.e., involving dentine”, for prevalence studies, since this would conform to the criteria in current use and, they hoped, allow for a higher degree of reproducibility between and among examiners (Palmer et al., 1984). Few studies have considered the extent to which inclusion or exclusion of pre-cavitation lesions (“initial” and “enamel” lesions, to use the World Health Organization terminology [WHO, 1979]) may affect DMF counts, particularly after the recent decline in caries prevalence.

In cases where the presence of “early” lesions as well as cavities has been scored and reported, it has been shown that large differences may occur between DMFT calculated inclu-
lumination. A plane mouth mirror and an Ash sickle probe No. 54 were used by the examiner. Where necessary, the probe was used to clean gross debris from fissures.

During the examinations, all lesions (including "enamel" and "initial" caries) were recorded. The diagnostic criteria employed are listed in Table 1. The only modification made to the WHO criteria, after a pilot study, was to allow slight staining in an otherwise sound-looking fissure to be scored as sound, since this type of slight staining was found to be exceedingly common and was not felt by the examiners to represent initial lesions. The WHO criteria applied to the other categories for permanent teeth were as follows: "extracted due to caries" and "lost due to other reasons", "filled", "filled with primary caries", "filled with secondary decay", "un-erupted", and "excluded" teeth. In line with WHO recommendations, only teeth extracted due to caries were counted as missing.

The examinations were conducted by three dentists from the Faculty of Dentistry of the University of Hong Kong. One dentist carried out the examinations of the 1980 and 1981 intakes, while the other two dentists each examined just one annual intake. Each had previously studied the criteria, and two examiners were trained and calibrated using volunteer patients by the third, one of the authors. A scribe recorded the scores for each tooth on a prepared form (since the software to allow for direct input was not available at the start of the study), and the examinations were tape-recorded so that checks for recording errors could be made on completion of each session. Approximately 10% of the subjects were randomly selected to undergo a repeat examination (by the examiner responsible for that particular student intake) to allow for assessment of intra-examiner reproducibility (Horowitz, 1972). Unfortunately, due to the timing of the departure of the examiners from the University staff and from Hong Kong, it was not feasible to calculate inter-examiner reproducibility using the same subjects.

**Caries software package.**—The data were analyzed by use of part of the suite of programs contained in the Clinical and Radiographic Information from Epidemiological Surveys (CARIES) software package. This software, which was written with the assistance of the Dental Data Processing Unit of the University of Hong Kong, is written in Microsoft BASIC, runs on IBM PC-compatible microcomputers, and requires that WHO diagnostic criteria and scoring codes be used for the clinical examination. The package includes programs to validate input data (to reject impossible or inappropriate values) and to calculate DMFT and DMFS and their components for every subject in the examination and for the group as a whole. The software can also re-calculate the DMF indices and their components according to three different diagnostic thresholds, D1, D2, and D3 (as detailed in Table 2), thus allowing for inclusion or exclusion of "enamel" and "initial" lesions in the computations. Although radiographic data can be utilized by the package (either in conjunction with corresponding clinical data or on its own), for this study the package was employed solely to process clinical data.

**Intra-examiner reproducibility.**—To calculate intra-examiner reproducibility, we used the following methods: Dice’s coincidence index for both ‘sound to sound’ and ‘decay to decay’ (Dice, 1945); kappa (Cohen, 1960); the modified percentage reproducibility index (Shaw and Murray, 1975); and finally Pearson’s correlation coefficient (Rugg-Gunn et al., 1976).

**Results.**

**Mean DMF.**—The average time for an examination (all of which were carried out at the D1 threshold) to be carried out was approximately three minutes, with a range of 2–5 minutes. Table 3 shows the mean decayed, missing, and filled teeth and surface counts from the same data (for all of the 287 students) calculated at each of the three diagnostic thresholds. For both DMFT and DMFS, it is evident that, as the sensitivity of the diagnostic threshold was increased from D3 (least sensitive) to D1 (most sensitive), the value of the calculated index almost doubled.

When the separate components of the indices are considered, it becomes obvious that, as expected, the decay component is the one most affected by the imposed thresholds. An increased sensitivity of diagnostic threshold had no effect on the M component, which remained at 0.29 missing teeth (or 0.86 missing surfaces) per person. Similarly, filled surfaces remained constant at 2.97 per person regardless of threshold. There was a

**TABLE 1**

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Surface Sound</td>
<td>No evidence of treated or untreated clinical caries (slight staining allowed in an otherwise sound fissure).</td>
</tr>
<tr>
<td>1</td>
<td>Initial Caries</td>
<td>No clinically detectable loss of substance. For pits and fissures, there may be significant staining, discoloration, or rough spots in the enamel that do not catch the explorer, but where loss of substance cannot be positively diagnosed. For smooth surfaces, these may be white, opaque areas with loss of lustre.</td>
</tr>
<tr>
<td>2</td>
<td>Enamel Caries</td>
<td>Demonstrable loss of tooth substance in pits, fissures, or on smooth surfaces, but no softened floor or wall undermined enamel. The texture of the material within the cavity may be chalky or crumbly, but there is no evidence that cavitation has penetrated the dentin.</td>
</tr>
<tr>
<td>3</td>
<td>Caries of Dentin</td>
<td>Detectably softened floor, undermined enamel, or a softened wall, or the tooth has a temporary filling. On approximal surfaces, the explorer point must enter a lesion with certainty.</td>
</tr>
<tr>
<td>4</td>
<td>Pulpal Involvement*</td>
<td>Deep cavity with probable pulpal involvement. Pulp should not be probed.</td>
</tr>
</tbody>
</table>

* Included with D3 in computations.

**TABLE 2**

<table>
<thead>
<tr>
<th>Threshold</th>
<th>DMFT Indices Applying to the Data Via the Caries Software Package</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3</td>
<td>Excludes (W.H.O.) &quot;enamel&quot; and &quot;initial&quot; lesions</td>
</tr>
<tr>
<td>D2</td>
<td>Excludes (W.H.O.) &quot;initial&quot; lesions</td>
</tr>
<tr>
<td>D1</td>
<td>Includes all clinically detected lesions</td>
</tr>
</tbody>
</table>

**TABLE 3**

<table>
<thead>
<tr>
<th>Diagnostic Threshold</th>
<th>Mean DMFT (SEM)*</th>
<th>Mean DMFS (SEM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3</td>
<td>3.01 (0.19)</td>
<td>4.75 (0.32)</td>
</tr>
<tr>
<td>D2</td>
<td>4.16 (0.21)</td>
<td>6.12 (0.33)</td>
</tr>
<tr>
<td>D1</td>
<td>5.91 (0.23)</td>
<td>8.40 (0.37)</td>
</tr>
</tbody>
</table>

*SEM = Standard error of the mean.
slight decrease in the filled tooth count, from 1.97 (D3) to 1.78 (D1). This was the result of five subjects in whom a tooth had both a restoration and a code 1 or 2 lesion. When this occurred, the tooth was counted as filled until the lesion was recognized by the diagnostic threshold, at which point the tooth was classed as decayed.

The mean DT component itself increased from 0.75 at threshold D3, to 1.96 at D2, and to 3.84 at threshold D1. Similarly, for mean DS there was a marked increase from 0.92 at threshold D3, to 2.29 at D2, and to 4.58 at threshold D1.

**Discussion.**

Microcomputers with the appropriate software enjoy a number of features which are theoretically desirable for the handling of epidemiological caries data. Although the software was not finalized in time to allow for chair-side input in this study, many powerful microcomputers are now available as portable versions, which facilitate immediate entry of data during a clinical examination. This reduces the need to fill out separate record forms or cards for each individual, reduces problems of record storage, and also reduces the risk of error during transcription of data into computer files. The automatic checking facility which ensures that only appropriate codes can be entered by the scribe at the time of examination should also minimize errors. Although data could be uploaded directly onto a mainframe computer, the attraction of using a microcomputer (particularly one which is IBM PC-compatible) is that workers who have no ready access to comprehensive computing facilities may well be able to acquire or borrow an ‘office’ desktop computer.

Another benefit of the CARIES software package is the centralized storage of data in a systematic manner, which allows for efficient access and rapid re-calculation of raw results according to different thresholds. This makes it feasible to select appropriate sets of results from a study to deal with different situations. For example, the appropriate diagnostic threshold must be employed when comparisons are made with the results of other studies if valid conclusions are to be drawn. It may, however, also be useful to examine a different facet of the caries pattern by the additional reporting of results computed with a different threshold.

As caries prevalence decreases, it is becoming increasingly difficult to demonstrate differential effects between different formulations of caries-preventive agents in clinical trials. Some now suggest that the sensitivity of the diagnostic criteria used should be increased (Glass et al., 1983), since much of the beneficial effect of fluoride in toothpaste, for example, is in preventing the progression of pre-cavitation lesions to cavitated ones and ‘reversing’ incipient lesions. A proportion of this cariostatic action may be missed if pre-cavitation lesions are ignored by the diagnostic threshold employed (Pitts, 1983). Similarly, in epidemiologic surveys the relative contribution made by pre-cavitation and initial lesions to the total caries experience (which may change with falling prevalence) will be unknown unless sensitive thresholds are employed.

When compared with data available for children, there is a scarcity of corresponding information for young adults with which to compare these Hong Kong DMF figures, although the DMF results appear to be low in comparison with those recently reported for Denmark, Finland, Israel, Japan, and the U.S.A. by Gordon and Newbrun (1986). In a national survey in Britain, a DMFT of 14.9 for the 16-24-year age group, based on the D3 threshold, was reported (Todd et al., 1982). This is approximately five times that of the Hong Kong group. More recently, however, a national survey of UK children reported (with the D3 threshold) a mean DMFT of 3.1 and 5.9 for 12- and 15-year-olds, respectively (Todd and Dodd, 1985). The Hong Kong data (which used the same diagnostic threshold) showed a mean DMFT of 3.01, indicating that the caries prevalence in this group of 20-year-olds was, in the early 1980’s, on a par with that of British 12-year-olds and considerably lower than that of British 15-year-olds.

Many workers have refrained from including pre-cavitation lesions in DMF counts, since it has been suggested that the benefits derived from the extra information may be outweighed by an increased examiner and method error. The results of a number of studies do not, however, support this contention (Backer-Dirks, 1964; Haugejorden and Slack, 1975; Espelid and Tveit, 1986). While the logistics of this study precluded an assessment of inter-examiner reproducibility, it was possible to record intra-examiner reproducibility for each examiner.

**TABLE 4**

PERCENTAGES OF A GROUP OF 287 HONG KONG DENTAL STUDENTS CONSIDERED CURRENTLY FREE OF DECAY (DECAYED TEETH = 0), OR FREE OF DECAY EXPERIENCE (DMF TEETH = 0), AT EACH OF THE THREE DIAGNOSTIC THRESHOLDS

<table>
<thead>
<tr>
<th>Diagnostic Threshold</th>
<th>Decayed Teeth = 0</th>
<th>DMF Teeth = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3</td>
<td>61.0%</td>
<td>28.2%</td>
</tr>
<tr>
<td>D2</td>
<td>33.5%</td>
<td>15.0%</td>
</tr>
<tr>
<td>D1</td>
<td>14.3%</td>
<td>7.0%</td>
</tr>
</tbody>
</table>

**TABLE 5**

OVERALL INTRA-EXAMINER REPRODUCIBILITY FOR ALL THREE EXAMINERS, AT THE THREE DIAGNOSTIC THRESHOLDS, FOR A RANDOMLY SELECTED 10% OF SUBJECTS

<table>
<thead>
<tr>
<th>Reproducibility Index</th>
<th>Diagnostic Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D3</td>
</tr>
<tr>
<td>Coincidence Index (sound to sound)</td>
<td>0.999</td>
</tr>
<tr>
<td>Coincidence Index (decay to decay)</td>
<td>0.841</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.789</td>
</tr>
<tr>
<td>Modified percentage reproducibility</td>
<td>99.7%</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.98</td>
</tr>
</tbody>
</table>
There are many suggested methods of calculating reproducibility (Shaw and Murray, 1975), the choice of the most appropriate method(s) being to an extent governed by survey study design and objectives, as well as by personal preference. While the use of Pearson's correlation coefficient is well-established, it was considered that the Kappa statistic and Dice's coincidence index also have a number of advantages (Nuttall and Paul, 1984). Many methods give few guidelines for interpretation of values. Landis and Koch (1977), however, proposed a scale for interpreting Kappa results. At each of the three diagnostic thresholds employed here, the mean Kappa values lie at the top of the "substantial agreement" range. When the individual examiners are considered, it is evident that each Kappa was at the borderline between "moderate" and "substantial" agreement (with the D1 agreement being higher), while examiners B and C's values were either at the borderline of, or within, the "almost perfect" agreement range. The important point to be made, however, is that although the individual examiners showed differences in intra-examiner reproducibility according to the index calculated, little difference was found in their ability to diagnose with a similar level of reproducibility at each diagnostic threshold. Fears of loss of intra-examiner reproducibility when the more sensitive thresholds are used might, therefore, have been overstated. Further studies of inter-examiner reproducibility with low-prevalence groups are, however, still required.

Although the lower caries levels encountered in this group may be of a lower reproducibility achieved, it might be postulated that the presence of fewer, less severe lesions would result in a greater number of more difficult, borderline decisions at times of higher caries activity. There was a slight deterioration in reproducibility from D3 to D1, but the magnitude of this, as demonstrated in the small changes in the various reproducibility indices and the relatively small increase in the standard error of the mean of the DMF values (Table 3), is, we would submit, insufficient to warrant exclusion of pre-cavitation lesions from DMF counts in all situations. It is possible that the pre-examination self-cleaning and good examining conditions (dental light and availability of compressed air) contributed to the reproducibility observed. These conditions can be provided quite readily — even, where necessary, with portable equipment.

Although it is obvious that increasing the sensitivity of any diagnostic threshold will give a different interpretation to the results of an examination, the aim of this study was to assess the magnitude of this effect when applied to epidemiological caries data. The results show that the use of such thresholds can have a dramatic effect on the reported level of dental caries. DMFT and DMFS indices were almost doubled by inclusion of "enamel" and "initial" lesions, while the "caries-free" percentage was decreased to approximately one-quarter of its D3 value by inclusion of these lesions (Tables 3 and 4). In view of these findings, it may be necessary for the question of choosing a diagnostic threshold which is appropriate to the survey objective(s) to be re-examined and to receive greater emphasis in future studies. There is a chance that the apparent underestimation of disease prevalence which arises when insensitive thresholds are used may be misinterpreted by research workers, dentists, and health planners.

From the results of this study, it would seem that it is feasible and potentially advantageous to use a microcomputer with software suitable for storage of raw epidemiological caries data recorded at the D1 threshold to calculate DMFT and DMFS at differing diagnostic thresholds. It is evident that the use of criteria which might be misinterpreted as being similar, when in fact they use differing effective diagnostic thresholds, can dramatically influence the reported level of dental caries, and considerable care should therefore be taken in use of the term "caries-free" when diagnostic thresholds which ignore pre-cavitation lesions have been employed. Finally, calculation of various reproducibility indices suggests that there may be only minimal loss of intra-examiner reproducibility when the more sensitive diagnostic criteria are employed on low caries prevalence groups. Further studies in this area, including an assessment of inter-examiner reproducibility, are required and planned.

Acknowledgments.

We are indebted to the dental students of Hong Kong University (HKU) for agreeing to participate, to all the various members of the Faculty of Dentistry, HKU, for their cooperation in mounting the caries examinations during the busy Induction and Orientation weeks, and in particular to Dr. C. Taylor and Dr. W. Yung (formerly HKU) for acting as examiners. We are also grateful to Mr. D. Carthy, Head of HKU's Dental Data Processing Unit (DDPU), and to Mr. L. Pong (DDPU) for the assistance in producing the CARIES package.

REFERENCES


PALMER, J.D.; ANDERSON, R.J.; and DOWNER, M.C. (1984):


