Validity of the Community Periodontal Index of Treatment Needs’ (CPITN) for population periodontitis screening

Validação do Community Periodontal Index of Treatment Needs’ (CPITN) para identificação de periodontite em populações

Abstract

The aim of the present study was to validate two versions of CPITN for periodontitis diagnosis. A sample of 400 individuals underwent full mouth periodontal examination including Clinical Attachment Loss, Periodontal Pocket Depth, and Sub-gingival Calculus. Full and partial CPITN versions were derived from this exam (gold standard). Contingency tables were constructed and operational characteristics obtained, as well as ROC curves. The results show 58% sensitivity for full CPITN and 80.6% specificity. Positive and negative predictive values were 87% and 46.3%, respectively. According to the test, estimated periodontitis prevalence was 46%, while the figure obtained with the gold standard was 69%. The partial version of the CPITN showed 50% sensitivity and 87.1% specificity. Positive and negative predictive values were 89.6% and 43.9%, respectively. Estimated periodontitis prevalence, through partial CPITN, was 30.5%. Adjusted global agreement (kappa) for partial and full CPITN was 0.32 and 0.29, respectively. Both CPITN versions disagreed significantly with gold standard results (chi-square p < 0.001). As a conclusion, both total and partial CPITN failed to reflect the real periodontal status of the sample.

Background

Instruments for periodontitis screening are commonly applied to populations in epidemiological studies of oral health. Several tests or indexes have been developed for this purpose, mainly due to a growing impression of inadequacy in the existing instruments. Another reason for the development of new instruments has been the search for efficient and simple ways to screen populations for periodontitis 1,2,3,4. The criteria used for periodontitis diagnosis vary greatly, but it is agreed that monitoring the attachment level is the most reasonable way to access activity of periodontal destruction 5,6,7. Since it is known that the feasibility of repeated measurements in healthy populations to assess periodontal disease progression is difficult, mainly due to economic, logistic, and statistical reasons, researchers are expected to conducted screening of periodontal disease with the available instruments. Few instruments available for periodontal screening have been validated against a standard exam, and the performance of these instruments for screening purposes is unpredictable when the disease status of the population is unknown.

Validating a test is necessary for screening, and it has been determined that some of the periodontal screening tests tend to underestimate disease prevalence 1,3,6,8,9,10,11,12,13. It has also been reported that in some situations pe-
periodontal screening instruments may overestimate disease.\cite{6,8,14,15,16,17,18,19}

Operational characteristics of screening instruments can be assessed by validation studies. It is taken for granted that periodontal disease varies considerably among populations\cite{16,17,20,21,22} and that severe limitations in the current disease descriptors exist\cite{23,24,25}. These points should be considered when evaluating data from studies describing the periodontal health status of populations through these instruments.

The use of screening instruments for periodontitis is not questioned, although the adequacy of the instruments should be considered more carefully. Several authors have proposed short tests for periodontal screening\cite{2,4,26,27}. The Community Periodontal Index of Treatment Needs (CPITN) proposition in 1982 and the rapid acceptance of this instrument provided the scientific community with an enormous amount of epidemiological data collected through it. Based on the limitations identified by several authors during years of research, the World Health Organization (WHO) proposed a few changes to the CPITN in 1987 and again in 1997. Nevertheless, the instrument's diagnostic criteria were basically unaffected by the modifications.

There is a lack of studies comparing CPITN to standard exams/diagnostic criteria. The purpose of the present study is to assess CPITN's operational characteristics. Understanding the performance of the instrument and specifically the assessment of its sensitivity and specificity will improve interpretations of the disease prevalence rates estimated with the CPITN.

Methods

A cross-sectional study was performed in a small rural population in southern Brazil including individuals over 18 years of age for whom disease status was unknown. Full mouth examinations were performed from January to March 2000. To calculate a sample size that would allow all the necessary calculations of operational characteristics of the CPITN, sensitivity and specificity of the index were estimated in a pilot study including 50 randomly selected individuals from the same location. Sample size calculation was assessed and included an alpha error of 5%, 80% power, estimated sensitivity of 40%, and specificity of 80% (data from pilot study) and a 10-point margin of error. Two sample calculations were performed (one using sensitivity and another with specificity) and the largest sample size was chosen. This number plus 20% for losses and refusals resulted in a sample size of 400.

Individuals were selected from a primarily rural population consisting of 4,600 individuals. Individuals were selected according to availability in local industries, public health centers, public services, rural unions, high schools, the university, and commercial establishments. Inclusion criteria were: age $\geq$ 18 years; presence of at least six natural teeth excluding third molars; and formal consent for participation. Four-hundred and fifty-seven individuals were examined. Fifty-seven individuals were excluded for not meeting the minimum number of natural teeth. Table 1 shows the sample's demographic characteristics.

A full periodontal examination was performed in a dental office using a standard manual periodontal probe (Newmar – Brazil). Data on probing depth and clinical attachment level was collected from six sites on every tooth present and registered on a chart by a trained team. The information was used to assess the individuals' periodontal health, and diagnosis based on these data resulted in the gold standard. For quality control, 10% of the exams were repeated randomly, and examiner agreement was consistent during the study. Periodontitis was determined as the presence of at least three sites, in different teeth, presenting three or more millimeters of clinical attachment level. Presence of gingival recession was considered and excluded for diagnostic purposes when present in buccal or lingual surfaces. Periodon-

### Table 1

Demographic characteristics and periodontal health conditions of the sample.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>169</td>
<td>42.3</td>
</tr>
<tr>
<td>Female</td>
<td>231</td>
<td>57.8</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-35</td>
<td>246</td>
<td>61.5</td>
</tr>
<tr>
<td>36-59</td>
<td>124</td>
<td>31.0</td>
</tr>
<tr>
<td>&gt; 59</td>
<td>30</td>
<td>7.5</td>
</tr>
<tr>
<td>Periodontitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>124</td>
<td>31.0</td>
</tr>
<tr>
<td>Mild</td>
<td>187</td>
<td>46.9</td>
</tr>
<tr>
<td>Moderate</td>
<td>72</td>
<td>17.8</td>
</tr>
<tr>
<td>Severe</td>
<td>17</td>
<td>4.3</td>
</tr>
</tbody>
</table>
tis was categorized as mild (3 sites/3mm-4.9mm), moderate (3 sites/5-6.9mm), and severe (3 sites/7mm or more). Intra-examiner agreement was assessed through the differences by means plot 28.29. This analysis revealed that more than 95% of the observations were between ± 2 standard deviations consistently throughout the study set.

The CPITN 4 diagnosis and diagnosis from its reduced version – partial CPITN 26 – were derived from the full exam using criteria proposed by Ainamo & Ainamo 30 and Diamanti-Kipioti et al. 26, respectively. Partial CPITN included the examination of 10 index teeth (17, 16, 11, 26, 27, 36, 37, 31, 46, 47), when present, for the evaluation of the sextant’s score. In the absence of the index teeth, all teeth in the sextant were examined to generate the score, as in the original CPITN. Contingency tables were constructed to assess sensitivity, specificity, predictive values, likelihood ratios, and test-estimated prevalence (prevalence estimated by the instrument). Stratification by age groups, disease severity, and gender was conducted. Consistency and coherence of the dataset was systematically evaluated prior to data analysis until a previously determined pattern was achieved.

Data were entered using Epi Info 6.0 (Centers for Disease Control and Prevention, Atlanta, United States) and analyzed using Intercooled Stata version 7.0 (Stata Corporation, College Station, United States).

Results

Periodontal disease prevalence in the sample was 69% (gold standard). The screening tests operated similarly. The following measurements are percent values with the respective 95% confidence intervals. The original CPITN’s sensitivity was 58% (51.9%-63.8%), and specificity was 80.6% (72.4%-87%). Positive predictive value reached 87% (81%-91.3%). Periodontitis prevalence according to the original CPITN was 46%. The difference between the actual prevalence of disease in the sample and that estimated through the test was 23 percentage points. The odds of a positive test in a diseased individual was 3.8, having a healthy subject as the reference (likelihood ratio = 3.8). Table 2 presents these data in details.

Figure 1 shows the ROC curves for both versions evaluated herein. The full CPITN appears to have performed slightly better than the partial CPITN. The areas under the curves are statistically equal according to the test of equality for areas under ROC curves (p = 0.605).

When the operational characteristics of the CPITN were evaluated according to different cutoff points of disease severity, the best balance between sensitivity and specificity appeared in the moderate category (Figures 2 and 3). This assumption is based on the largest area under the ROC curves. The best sensitivity values were observed for the severe cutoff point both for the full and partial CPITN. On the other hand the highest specificity values were observed for the mild cutoff for both versions of the instrument.

Discussion

Periodontitis severity and prevalence in this sample were similar to the values observed in similar populations, although the present study used a convenience sample 17,31,32 which could have influenced the predictive values but not the sensitivity and specificity if the prevalence was artificially modified by selection bias. The use of probing depth as the clinical parameter for diagnosis in some studies may account for some of the variability. This aspect has been previously discussed by several authors 2,6,24,25,26,33,34. Most periodontitis screening studies

<table>
<thead>
<tr>
<th>Measure</th>
<th>Full CPITN (%)</th>
<th>Partial CPITN (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>58.0 (51.9-63.8)</td>
<td>50.0 (44.0-56.0)</td>
</tr>
<tr>
<td>Specificity</td>
<td>80.6 (72.4-87.0)</td>
<td>87.1 (79.6-92.2)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>87.0 (81.0-91.3)</td>
<td>89.6 (83.4-93.8)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>46.3 (39.5-53.2)</td>
<td>43.9 (37.6-50.4)</td>
</tr>
<tr>
<td>Estimated prevalence</td>
<td>46.0 (41.0-51.0)</td>
<td>38.9 (33.7-43.5)</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>3.0</td>
<td>3.8</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

The reference value for periodontitis prevalence is 69% (64.2%-72.5%). Sensitivity, specificity, and positive and negative predictive values are expressed in percentage points.

Table 2

CPITN (Community Periodontal Index of Treatment Needs) operating characteristics for the full sample, and 95% confidence intervals.
using the probing depth as the main clinical parameter have shown overestimated prevalence measures 2,6,24,33,34. The main clinical parameter accounted for the periodontitis diagnosis through the CPITN is probing depth. This may be the most reasonable explanation for the operational characteristics observed. The low values for sensitivity (58%) result in a high rate of false negative values and consequently are not suitable for exclusion diagnosis. Although the specificity values reached higher scores (80.6%), increasing the odds of identifying a non-diseased individual under a negative test, for screening purposes higher sensitivity values are desirable, due to the goal of identifying disease rather than non-disease. The balance between sensitivity and specificity was attained when the cutoff point was set at moderate. The highest values of sensitivity were obtained when the cutoff point was set at the severe disease status (82.4%). This apparently reasonable performance leaves room for an
even higher rate of false-negative results reflected by the low specificity of the test (55.6%). The CPITN, to some extent, could be considered a partial exam since the main clinical parameter accounting for the diagnosis is the probing depth, and the attachment level is not considered in the diagnosis, although the WHO recommends its use concomitant to the CPITN application.

Underestimation of periodontitis prevalence was observed in the evaluation of other partial instruments published previously. Hunt reports an underestimation of periodontitis prevalence when performing half mouth exams. Even the full CPITN version used in the present study, although examining all teeth, could be considered a partial instrument since the multiplicity of sites is not considered for the diagnosis when a single score is applied for each sextant.

Diamanti-Kipioti et al. compared the CPITN to a full exam and reported underestimation of sextants presenting scores higher than four and also of the frequency of individuals presenting scores higher than four and number of deep pockets.

The trend for lower values observed for the partial version of the CPITN in the present study may reflect the fact that only ten teeth contribute to the diagnosis. The analysis of the operational characteristics of the instruments under different severities shows the best balance when the moderate level is chosen as the cutoff. This provides no advantage, since early stages of disease are not diagnosed and will account for an unidentified repressed demand in the years that follow the screening process. For screening purposes, the CPITN has been reported previously as having better suitability for severe disease diagnosis. The diagnosis provided by the instruments, when evaluated under the kappa statistics, shows the disparity between the two CPITN versions and the gold standard. It might not be accurate to compare periodontitis prevalence directly when measured by the instrument. Population data collected through the CPITN should be carefully interpreted, especially when drawing conclusions in the field of descriptive periodontal epidemiology targeting public health strategies, service cost studies and service demand estimation.
Resumo

Visando validar duas versões do CPITN para o diagnóstico de periodontite em estudos populacionais, 400 indivíduos foram submetidos a exames de perda de inserção clínica, profundidade de sondagem e presença de cálcio subgengival. Uma versão parcial e completa do CPITN foi derivada deste exame e comparada com o mesmo (padrão-ouro). Tabelas de contingência foram elaboradas e obtiveram-se as características operacionais do CPITN. Os resultados mostram sensibilidade de 58% e especificidade de 80,6% para a versão completa. Os valores preditivos positivo e negativo foram 67% e 45,3% respectivamente. A prevalência de periodontite estimada por esta versão foi 46% enquanto que o valor de referência segundo o padrão-ouro foi 69%. A versão parcial demonstrou uma sensibilidade de 50% e uma especificidade de 87,1%. Os valores preditivos positivo e negativo foram, respectivamente, 89,6% e 43,9%. A prevalência estimada de periodontite foi 30,3%. A taxa global ajustada de concordância (kappa) para a versão parcial e completa foi, respectivamente, 0,32 e 0,29. Ambas as versões discordaram dos resultados do padrão-ouro (qui-quadrado p < 0,001). Ambas as versões do CPITN falharam em representar o status periodontal da amostra.

Estudos de Validação; Índice Periodontal; Periodontite

Contributors

D. G. Bassani reviewed the literature, elaborated the protocol, coordinated field work, collected data, analyzed data, and drafted the manuscript. C. M. Silva assisted the elaboration of the protocol, conducted field work, collected data, and assisted in drafting the manuscript. R. V. Oppermann supervised the literature review, assisted in the elaboration of the protocol, supervised the data analysis, and supervised and contributed to drafting of the manuscript.

Referências

16. Benigeri M, Brodeur JM, Payette M, Charbonneau A, Ismail AI. Community periodontal index of