

## Calibração de examinadores para estudos epidemiológicos de cárie dentária

Calibration of examiners for dental caries epidemiology studies

Marco Aurélio Peres <sup>1</sup>  
Jefferson Traebert <sup>2</sup>  
Wagner Marcenes <sup>3</sup>

---

<sup>1</sup> Departamento de Saúde Pública, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina. Campus Universitário de Trindade, Florianópolis, SC 88049-970, Brasil.  
peresp@repensul.ufsc.br

<sup>2</sup> Programa de Pós-Graduação em Odontologia, Universidade Federal de Santa Catarina. Rua Cel. Teixeira de Oliveira 288, Biguaçu, SC 88160-000, Brasil.  
traebert@brasilnet.com.br

<sup>3</sup> Department of Epidemiology & Public Health, Faculty of Clinical Sciences, Royal Free and University College Medical School, University College of London. Gower Street Campus, 1-19 Torrington Place, London WC1E 6BT, United Kingdom.

**Abstract** *The aim of this study was describe a training and calibration exercise for dental caries and to discuss the use of validity and reliability statistical tests. Eight dentists examined seven-teen 12 year-old schoolchildren using World Health Organization (1997) criteria. Data analysis included calculation of the DMF-T index, the percent agreement, and the kappa scores on a tooth-by-tooth basis. Inter-examiner agreement was tested against a gold standard examiner. Scores for measures of agreement calculated on a tooth-by-tooth basis were lower than those calculated for the whole mouth. The training and calibration exercise is a crucial step in cross-sectional epidemiological surveys, and the kappa test calculated on a tooth-by-tooth basis is the appropriate measurement to test agreement between examiners for dental caries.*

**Key words** DMF-T Index; Dental Caries; Evaluation

**Resumo** *O objetivo do estudo foi descrever um exercício de calibração de examinadores para estudos epidemiológicos de cárie dentária, comparando os testes estatísticos que medem acurácia e reprodutibilidade. Oito cirurgiões-dentistas examinaram independentemente 17 escolares de 12 anos de idade, utilizando os códigos e critérios propostos pela Organização Mundial da Saúde em 1997. Calculou-se o índice CPO-D médio, as taxas de concordância e o teste kappa considerando os dentes no seu conjunto e dente a dente para cada um dos examinadores comparando-os com o examinador padrão. As taxas de concordância e os valores de kappa quando considerou-se todos os dentes em conjunto, foram superiores aos valores do teste kappa quando calculado dente a dente. A padronização de critérios diagnósticos, o rigoroso treinamento e exercícios de calibração de examinadores e a mensuração e apresentação adequadas dos erros de diagnóstico são fundamentais para o julgamento da validade dos resultados de estudos epidemiológicos. A revisão de literatura assim como os resultados deste estudo mostram que a utilização do teste kappa e seu cálculo com base em cada um dos dentes separadamente é a forma mais adequada de mensuração da acurácia e reprodutibilidade.*

**Palavras-chave** Índice CPO-D; Cárie Dentária; Avaliação

## Introdução

A padronização de critérios diagnósticos para estudos epidemiológicos de saúde bucal foi claramente definida pela Organização Mundial da Saúde (OMS, 1991; WHO, 1997), entretanto formas de mensuração da acurácia e reprodutibilidade diagnóstica não têm sido abordadas de maneira clara (Cleaton-Jones et al., 1989). A acurácia é o grau pelo qual o instrumento de coleta de dados mede o que se pretende, sendo reprodutibilidade a extensão em que medidas de um fenômeno alcançam resultados semelhantes quando coletadas sucessivas vezes (Fletcher et al., 1996).

Um dos requisitos para se assegurar a fidedignidade dos achados é a minimização de variação e erros de diagnóstico, utilizando-se para tanto critérios padronizados, como os estabelecidos pela OMS (WHO, 1997).

Entretanto, algum grau de variação ou erro de diagnóstico pode ocorrer quando examinamos muitas pessoas, sendo portanto necessária a mensuração do mesmo. As divergências de diagnóstico podem ocorrer entre os diferentes examinadores (erros entre examinadores), ou entre diferentes exames de um mesmo grupo de pessoas, por um examinador ao longo do tempo (erros intra examinador). Em publicações científicas ou relatórios de estudos oficiais, a apresentação da quantificação das divergências de diagnóstico ocorridas durante o estudo é recomendável, pois facilita o julgamento do leitor quanto à confiabilidade dos resultados.

Para se obter uma padronização no uso de critérios de diagnóstico para as diversas doenças bucais, é fundamental que os examinadores envolvidos no estudo participem de um treinamento antes do início do mesmo. Portanto, uma etapa necessariamente prévia a todos os estudos epidemiológicos é a calibração dos examinadores.

Conceitua-se calibração como sendo a repetição de exames nas mesmas pessoas pelos mesmos examinadores, ou pelo mesmo examinador em tempos diferentes, a fim de diminuir as discrepâncias de interpretação nos diagnósticos (WHO, 1993). Os objetivos da calibração são: a) assegurar uniformidade de interpretação, entendimento e aplicação dos critérios das várias doenças e condições a serem observadas e registradas; b) assegurar que cada um dos examinadores possa trabalhar consistentemente com o padrão adotado; c) minimizar variações entre diferentes examinadores (WHO, 1993).

O presente artigo relata um exercício de calibração de examinadores, anterior a estudos epidemiológicos de cárie dentária, discutindo

as principais utilizações e propriedade de aplicação de testes estatísticos que buscam medir a consistência diagnóstica intra e entre examinadores.

## Métodos

Um grupo de oito cirurgiões-dentistas, sem prévia experiência em levantamentos epidemiológicos, participou de um exercício de calibração com o objetivo de assegurar uniformidade diagnóstica de cárie dentária. Um total de 17 escolares de 12 anos de idade foram examinados por todos os cirurgiões-dentistas envolvidos neste estudo, utilizando os códigos e critérios propostos pela OMS (WHO, 1997). A OMS (WHO, 1997) preconiza a utilização de sondas periodontais, especialmente desenhadas para estudos epidemiológicos (sonda *ball point*) de cárie dentária, com o objetivo de eliminar detritos situados sobre os dentes e assim melhorar a visualização. No presente estudo foram utilizadas gazes esterilizadas para esta finalidade. O trabalho foi realizado em uma escola da rede estadual de ensino do município de Florianópolis, Santa Catarina, sob a orientação de dois dos autores do presente artigo (MAP e JT). Os exames foram realizados em ambiente amplo, com boa iluminação, com os examinados deitados em bancadas simulando macas. Os examinadores posicionaram-se em 12 horas em relação à cabeça do examinado e os anotadores permaneceram na posição de 9 horas. Foram utilizados nos exames clínicos, espelhos bucais planos previamente esterilizados, espátulas de madeira, e gaze.

A atividade foi dividida em quatro etapas:

a) Primeira etapa: uma etapa teórica com quatro horas de duração, que consistiu da apresentação dos critérios do índice CPO-D (dentes cariados, perdidos e obturados na dentição permanente) (WHO, 1997), seguida da apresentação de vinte diapositivos que exibiam as várias condições a serem observadas nos exames (dentes cariados, dentes restaurados sem cárie, restaurados com cárie, dentes perdidos, com fraturas de coroa e dentes não erupcionados). Cada diapositivo foi projetado por aproximadamente um minuto, tendo sido solicitado aos examinadores que após cada projeção fosse realizado o respectivo diagnóstico, coerentemente com os critérios previamente apresentados. Na seqüência, apresentamos a ficha que seria utilizada e explicamos a rotina do exame clínico.

b) Segunda etapa: foi dedicada ao exercício clínico, cujo objetivo era propiciar aos examinadores a fixação dos critérios apresentados.

Foram realizados inicialmente exames individuais em seis crianças. Nas que foram observadas diferenças de diagnóstico entre um examinador e um dos orientadores do exercício (MAP), realizamos novo exame. Com base no menor número de divergências observadas entre os resultados obtidos pelos examinadores e o orientador do exercício, escolhemos um examinador padrão dentre os oito cirurgiões-dentistas da equipe. Repetimos os exames clínicos até quando julgamos que os critérios tinham sido fixados pela equipe. Depois, os responsáveis pela condução da calibração selecionaram 17 crianças de 12 anos de idade a serem examinadas individualmente por cada um dos examinadores. A variedade de manifestações clínicas que compõem os critérios do índice utilizado foi adotada para a seleção das crianças. A autorização para a realização dos exames foi obtida através de carta enviada aos pais ou res-

ponsáveis pelas mesmas, com explicações sobre os objetivos e a rotina do estudo.

A seguir, cada um dos oito examinadores e um dos orientadores (MAP) realizaram os exames contando com anotadores que transcreviam os códigos dos diagnósticos para fichas padronizadas. Após a conclusão de todos os exames, verificou-se através das fichas, as divergências ocorridas. Montaram-se matrizes (Tabelas 1 e 2) para comparar as divergências e concordâncias, e permitir a realização do teste *kappa* considerando-se todos os dentes em conjunto, e também para cada um dos dentes separadamente. É importante ressaltar que a comparação e discussão dos resultados através do uso das matrizes (Tabelas 1 e 2) é o elemento mais importante desta fase de treinamento, sendo secundário o cálculo dos testes estatísticos. A comparação das matrizes permite identificar os critérios dos índices, nos quais os exa-

Tabela 1

Exemplo do cálculo do *kappa* considerando-se todos os dentes em conjunto em exames realizados com 17 escolares de 12 anos de idade. Florianópolis, Santa Catarina, 1999.

	0	A	1	B	2	C	3	D	4	E	5	6	F	7	G	8	T	9	Total
0	365		4	1								5				1			376
A		10																	10
1	2		14																16
B	1			14															15
2																			
C																			
3					1		12												13
D																			
4																			
E																1			1
5																			
6												4							4
F																			
7																			
G																			
8	1															36			37
T	1																3		4
9																			
<b>Total</b>	<b>370</b>	<b>10</b>	<b>18</b>	<b>15</b>	<b>1</b>		<b>12</b>					<b>9</b>				<b>38</b>	<b>3</b>		<b>476</b>

Códigos: 0: dente permanente hígido; A: dente decíduo hígido; 1: dente permanente cariado;

B: dente decíduo cariado; 2: dente permanente restaurado sem cárie; C: dente decíduo restaurado sem cárie;

3: dente permanente restaurado com cárie; D: dente decíduo restaurado com cárie;

4: dente permanente perdido por cárie; E: dente decíduo perdido por cárie;

5: dente permanente perdido por outras razões; 6: dente permanente com selante;

F: dente decíduo com selante; 7: dente permanente apoio de ponte ou coroa;

G: dente decíduo apoio de ponte ou coroa; 8: dente permanente não erupcionado;

T: dente permanente com fratura coronária; 9: dente decíduo com fratura coronária.

Cálculo:  $kappa = po - pe / 1 - pe$ , onde: po: proporção de concordâncias observadas; pe: proporção de concordâncias esperadas; po = somatória das diagonais / total; pe = somatória de cada linha \* cada coluna / total<sup>2</sup>.

No exemplo tem-se que:  $po = 365 + 10 + 14 + 14 + 12 + 4 + 36 + 3 / 476 = 0,98$ ;  $pe = 376 * 370 + 10 * 10 + 16 * 18 + 15 * 1 + 13 * 12 + 4 * 9 + 37 * 38 / 476^2 = 0,62$ ;  $kappa = 0,98 - 0,62 / 1 - 0,62 = 0,94$ .

Tabela 2

Exemplo do cálculo do *kappa* considerando-se dente a dente examinando-se 17 escolares de 12 anos de idade. Dente 16. Florianópolis, Santa Catarina, 1999.

	0	A	1	B	2	C	3	D	4	E	5	6	F	7	G	8	T	9	Total
0	5		1									1							7
A																			
1			3																3
B																			
2																			
C																			
3							6												6
D																			
4																			
E																			
5																			
6																			1
F																			
7																			
G																			
8																			
T																			
9																			
<b>Total</b>	5		4	1			6					1							17

Cálculo:  $kappa = p_o - p_e / 1 - p_e$ ;  $p_o = 5 + 3 + 6 / 17 = 0,82$ ;  $p_e = 7 * 5 + 3 * 4 + 6 * 6 + 1 * 1 / 17^2 = 0,29$ ;  $kappa = 0,82 - 0,29 / 1 - 0,29 = 0,75$ .

minadores tiveram maior dificuldade de fixação, enquanto que o resultado do teste *kappa* não fornece esta informação. Esta etapa teve duração de três dias.

c) Terceira etapa: foram realizados novos exames, desta vez incluindo todas as 17 crianças de 12 anos de idade, sem que houvesse qualquer comunicação entre os examinadores, tendo por finalidade avaliar o resultado da calibração. Os sete cirurgiões-dentistas e o examinador padrão examinaram duas vezes cada criança. Esta etapa durou aproximadamente quatro horas. Novamente, montaram matrizes (Tabelas 1 e 2) para verificar as concordâncias e permitir a realização do teste *kappa*, considerando-se todos os dentes em conjunto e também para cada um dos dentes separadamente. Os índices CPO-D médios de cada um dos examinadores, do grupo como um todo e do examinador padrão foram calculados.

d) Quarta etapa: A última etapa do exercício foi destinada à discussão final dos resultados, enfatizando-se a necessidade de se atingir uma boa concordância antes do início do trabalho de campo.

Por razões éticas, asseguramos o atendimento odontológico às crianças que apresenta-

vam necessidade de tratamento na Unidade de Saúde próxima à escola, observando-se a ordem crescente de necessidades de tratamento.

## Resultados

As Tabelas 3 e 4 mostram os valores de *kappa* para todos os examinadores e para o examinador padrão, obtidos nos exames clínicos das 17 crianças incluídas nesta pesquisa, considerando-se todos os dentes das arcadas superiores e inferiores respectivamente.

Nota-se que na arcada superior obtivemos concordâncias acima de 0,6, considerada boa (Landis & Koch, 1977), nos dentes 17, 16, 13, 12, 21, 22, 23, e 27 para todos os examinadores, sendo o examinador de número seis o que apresentou os menores valores de *kappa* (-0,12 no dente 14; 0,00 no 12; 0,29 no 11; 0,11 no 24; 0,18 no dente 25 e 0,31 no 26) (Tabela 3). Resultados similares foram encontrados para as arcadas inferiores (Tabela 4).

A Tabela 5 mostra os menores valores de *kappa* com os respectivos dentes onde os mesmos ocorreram e os valores de *kappa* quando consideramos todos os dentes em conjunto. O

Tabela 3

Valores de *kappa*, dente a dente, das arcadas superiores de 17 crianças de 12 anos de idade segundo cada examinador, comparados com o examinador padrão. Florianópolis, Santa Catarina, 1999.

Examinador	17	16	15	14	13	12	11	21	22	23	24	25	26	27
1	0,88	1,00	1,00	0,65	1,00	1,00	0,77	1,00	1,00	1,00	0,65	0,86	0,82	1,00
2	0,88	1,00	1,00	0,29	1,00	1,00	1,00	1,00	1,00	1,00	0,65	1,00	0,88	1,00
3	0,88	0,91	0,69	0,56	1,00	1,00	0,77	0,83	1,00	1,00	0,77	0,74	0,69	1,00
4	0,88	0,81	0,82	0,18	1,00	1,00	1,00	1,00	1,00	1,00	0,65	0,86	0,88	1,00
5	0,88	1,00	1,00	0,65	0,86	1,00	0,78	1,00	1,00	1,00	0,29	1,00	0,74	1,00
6	0,76	0,60	0,37	-0,12	0,61	0,00	0,29	0,85	1,00	1,00	0,11	0,18	0,31	0,87
7	0,88	0,90	1,00	0,65	0,83	1,00	1,00	1,00	1,00	1,00	0,65	1,00	0,74	1,00

Tabela 4

Valores de *kappa*, dente a dente, das arcadas inferiores de 17 crianças de 12 anos de idade segundo cada examinador, comparados com o examinador padrão. Florianópolis, Santa Catarina, 1999.

Examinador	47	46	45	44	43	42	41	31	32	33	34	35	36	37
1	1,00	0,65	0,45	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,83	0,75	0,77
2	1,00	0,82	0,65	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,83	0,83	0,69
3	0,75	0,74	0,29	0,46	1,00	1,00	1,00	1,00	1,00	1,00	0,45	0,82	0,75	0,89
4	0,90	0,84	0,65	0,65	0,65	1,00	1,00	1,00	1,00	1,00	0,65	0,69	0,83	0,67
5	1,00	0,63	0,45	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,83	0,84	0,68
6	0,88	0,31	0,26	0,65	0,33	0,00	1,00	1,00	1,00	1,00	0,45	0,22	0,40	0,89
7	1,00	0,73	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,80	0,82	0,67

menor valor de *kappa* observado, quando o cálculo foi feito separadamente para cada dente, foi igual a -0,12. Já quando considerados todos os dentes em conjunto o menor valor de *kappa* obtido foi 0,53. O mesmo comportamento em relação à taxa de concordância geral foi observado (taxa de concordância geral = concordância observada/total de diagnósticos realizados). Observamos ainda, como esperado, que a taxa de concordância foi maior que os valores do teste *kappa*.

A Tabela 6 apresenta os valores do CPO-D médio para cada um dos examinadores e do examinador padrão; o desvio dos valores do CPO-D médio de cada um dos examinadores em relação ao CPO-D médio do examinador padrão. Verificamos que houve pouca variação em relação ao CPO-D médio.

## Discussão

No primeiro estudo nacional oficial de 1986, apesar de extenso relatório elaborado com os resultados, pouca atenção foi destinada à metodologia utilizada para o treinamento e exercício de calibração dos examinadores, desconhecendo-se até o número total de examinadores envolvidos (MS, 1988).

No segundo levantamento epidemiológico nacional realizado em 1996, após conceituar e elencar os objetivos da calibração, o relatório final mostra superficialmente o exercício realizado de calibração de examinadores. Omite como foram calculadas a acurácia e a reprodutibilidade e como estas se comportaram durante o trabalho de campo. A omissão destes aspectos metodológicos aumenta de importância na medida que o referido estudo contou com a participação de 108 equipes de examinadores (Oliveira, 1998).

A escolha de um examinador padrão é uns dos principais objetivos da calibração (WHO,

Tabela 5

Menores valores de *kappa* obtidos segundo o(s) dente(s) (entre parêntesis) e valores de *kappa* considerando-se todos os dentes em conjunto e taxas de concordâncias, em 17 crianças de 12 anos de idade. Florianópolis, Santa Catarina, 1999.

Examinador	Menores valores de <i>kappa</i> (dente(s))	Valores de <i>kappa</i> considerando-se todos os dentes	Valores de concordância considerando-se todos os dentes
1	0,45 (15)	0,89	0,97
2	0,29 (14)	0,91	0,97
3	0,29 (15)	0,85	0,94
4	0,18 (14)	0,95	0,95
5	0,29 (24)	0,90	0,96
6	- 0,12 (14)	0,53	0,82
7	0,65 (14 e 24)	0,89	0,97

Tabela 6

Índice CPO-D médio, desvio do CPO-D médio de cada examinador em relação ao CPO-D médio do examinador padrão em 17 crianças de 12 anos de idade. Florianópolis, Santa Catarina, 1999.

Examinador	CPO-D médio	Desvio do CPO-D em relação ao padrão
Padrão	1,82	-
1	1,65	- 0,17
2	1,88	+ 0,06
3	1,88	+ 0,06
4	1,82	0,00
5	1,71	- 0,11
6	2,00	+ 0,18
7	2,00	+ 0,18
Grupo*	1,85	+ 0,03

\* Exceto o examinador padrão.

1993), a fim de mensurar a que distância cada um dos examinadores encontra-se em relação ao padrão que assume como sendo o verdadeiro valor. Quando não se fixa o examinador padrão, é possível que todos os valores observados pelos diferentes examinadores estejam próximos entre si (altos valores de *kappa*), porém distantes do presumido valor verdadeiro (o do examinador padrão). Em artigos ou relatórios de estudos que não adotam examinadores padrão e não medem reprodutibilidade entre os diversos examinadores, não é possível aos leitores conhecerem quais foram a concordância entre os examinadores e a distância entre os

valores aferidos por estes, e o valor presumidamente verdadeiro.

Em 1991, a OMS preconizou a utilização da taxa de concordância como parâmetro para aferir a acurácia e reprodutibilidade (OMS, 1991). Já em 1997, a OMS recomendou a utilização da estatística *kappa* e destinou maior atenção para a mensuração da acurácia e reprodutibilidade entre examinadores (WHO, 1997). Esta recomendação expressa em linhas muito gerais, omitiu aspectos importantes, como a necessária utilização de um examinador padrão nos casos em que vários examinadores são envolvidos no levantamento epidemiológico e a recomendação de se calcular o *kappa* tomando-se como unidade o dente, já que esta é a unidade de medida para os índices mais utilizados em estudos epidemiológicos de cárie, como os índices ceo-d e CPO-D. Em outras palavras, da forma como está preconizado, o teste *kappa* medirá o nível de concordância entre os examinadores ou entre o mesmo examinador em momentos distintos.

Segundo Bulman & Osborn (1989) a maior parte das publicações abordando estudos epidemiológicos de cárie dentária menciona vagamente que consistência entre examinadores foi testada e encontrou-se dentro dos limites aceitáveis omitindo portanto a forma de mensuração e os testes que foram aplicados. Segundo os mesmos autores, a OMS e a Federação Dentária Internacional reconheceram o problema e têm sugerido diferentes soluções, ainda que em termos muito gerais e insuficientes.

Os estudos epidemiológicos nacionais realizados no Brasil em 1986 (MS, 1988) e 1996 (Oliveira, 1998) corroboram com as afirmações de Bulman & Osborn (1989) ao não explicarem com detalhes, como foi realizado o processo de treinamento e exercício de calibração dos examinadores e a aferição da acurácia e da reprodutibilidade obtidas ao longo do estudo.

Há diversas maneiras de verificar a concordância de resultados entre medidas de um mesmo evento, e assim, estimar o erro de sua aferição (Pereira, 1995). As variáveis expressas em categorias, como casos de diagnóstico e da presença ou ausência de sintomas, são comuns em estudos epidemiológicos de saúde bucal. Este estudo levou em consideração apenas o cálculo da acurácia em relação ao índice CPO-D, o mais comumente utilizado em levantamentos epidemiológicos de cárie dentária.

A forma mais simples de apresentar os resultados é através da taxa de concordância observada entre os examinadores. Esta maneira de expressar concordâncias tem sido criticada por não levar em consideração a concordância

devida ao acaso e, conseqüentemente, outras formas surgiram para substituí-la, como o teste *kappa*. Este teste é um indicador de concordância ajustado, pois leva em consideração, descontando no resultado final, a concordância devida ao acaso (Pereira, 1995). Conseqüentemente, o nível de concordância obtido com o teste *kappa* tende a ser menor que o obtido com a taxa de concordância. Kingman (1997) em um estudo anterior a este relatou que enquanto a taxa de concordância observada foi de 0,64, o valor de *kappa* foi de 0,04. Os resultados do nosso estudo confirmam estes achados.

O valor de *kappa* informa a proporção de concordâncias além da esperada pelo acaso e varia de “menos 1” a “mais 1” sendo que, “menos 1” significa completa discordância e “mais 1”, concordância perfeita. Valores de *kappa* considerados como indicativos de boa concordância situam-se entre 0,61 e 0,80, sendo valores superiores a 0,80 indicativos de ótima concordância (Landis & Koch, 1977).

No nosso estudo pode-se verificar que as taxas gerais de concordância e o *kappa* medido pelo conjunto de dentes pode dar uma falsa impressão de boa concordância. Tome-se por exemplo o examinador número seis que apresentou baixos níveis de acurácia diagnóstica (Tabelas 3 e 4) quando analisou dente a dente. No entanto, quando o mesmo examinador teve sua acurácia medida através do *kappa* geral (todos os dentes em conjunto) apresentou valores bastante razoáveis. O mesmo ocorreu em relação à taxa de concordância.

A redução da prevalência e severidade da cárie dentária, conseqüentemente afetando

poucos dentes em um indivíduo, explica as diferenças obtidas entre o *kappa* e a taxa de concordância geral. Quando medimos a concordância geral ou mesmo o *kappa* geral (todos os dentes em conjunto), os dentes sem nenhuma lesão (hígidos) são em número muito superior aos dentes atacados pela cárie o que acaba diminuindo o impacto dos erros de diagnóstico naqueles poucos dentes que apresentaram a doença, oferecendo assim uma visão exageradamente otimista da concordância obtida. Por outro lado, comparando-se dente a dente contorna-se este problema, pois um valor baixo de *kappa* em um dente já é indicativo da necessidade de continuidade do treinamento, independente dos demais valores obtidos.

Os níveis de acurácia e reprodutibilidade dependem não só da prevalência da doença, mas também do tipo de evento de estudo e de inúmeros fatores, tais como os relacionados ao examinador, ao procedimento em questão e ao ambiente onde as observações foram feitas. Para elevar a reprodutibilidade são essenciais a clara definição do evento, regras de medida e de classificação diagnóstica apropriada, como ocorre com os critérios adotados pela OMS (OMS, 1991; WHO, 1997). Diminuindo-se o número de categorias para as respostas aumentase a reprodutibilidade (Pereira, 1995). Infelizmente, os manuais elaborados pela OMS acerca de métodos de estudos epidemiológicos em saúde bucal, que são referência para muitos levantamentos epidemiológicos em vários países, apresentam estas recomendações de forma insuficiente.

## Referências

- BULMAN, J. S. & OSBORN, J. E., 1989. Measuring diagnostic consistency. *British Dental Journal*, 166: 377-381.
- CLEATON-JONES, P.; FATTI, L. P.; CHANDLER, H. D. & GROSSMAN, E. S., 1989. Dental caries diagnosis calibration for clinical fields surveys. *Caries Research*, 23:195-199.
- FLETCHER, R. H.; FLETCHER, S. W. & WAGNER, E. H., 1996. Introdução. In: *Epidemiologia Clínica: Elementos Essenciais* (R. H. Fletcher, S. W. Fletcher & E. H. Wagner, org.), pp. 32-33, 3ª Ed., Porto Alegre: Artes Médicas.
- KINGMAN, A., 1997. Statistics in community oral health. In: *Community Oral Health* (C. N. Pine, ed.), pp. 147-162, Oxford: Wright.
- LANDIS, J. R. & KOCH, G. G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174.
- MS (Ministério da Saúde), 1988. *Levantamento Epidemiológico em Saúde Bucal: Brasil, Zona Urbana*, 1986. Brasília: Centro de Documentação, Ministério da Saúde.
- OLIVEIRA, A. R. C., 1998. *Dados Básicos do Levantamento Epidemiológico em Saúde Bucal (Primeira Etapa – Cárie Dental) Realizado pela Coordenação de Saúde Bucal do Ministério da Saúde em 1998*. Araçatuba: Faculdade de Odontologia de Araçatuba, Universidade Estadual de São Paulo. (mimeo.)
- OMS (Organização Mundial da Saúde), 1991. *Levantamento Epidemiológico Básico de Saúde Bucal. Manual de Instruções*. 3ª Ed., São Paulo: Santos.
- PEREIRA, M. G., 1995. *Epidemiologia. Teoria e Prática*. Rio de Janeiro: Guanabara Koogan.
- WHO (World Health Organization), 1993. *Calibration of Examiners for Oral Health Epidemiology Surveys*. Technical Report. Geneva: WHO.
- WHO (World Health Organization), 1997. *Oral Health Surveys. Basic Methods*. Geneva: WHO.