

# Precisão e validade de levantamentos epidemiológicos em saúde bucal: cárie dentária na Cidade de São Paulo, 2002

## *Reliability and validity of oral health surveys: dental caries in the city of Sao Paulo, 2002*

**Antonio Carlos Frias**

Departamento de Odontologia Social  
Faculdade de Odontologia  
Universidade de São Paulo  
Av. Lineu Prestes 2227  
05508-900 – São Paulo, SP  
acfrias@usp.br

**José Leopoldo Ferreira Antunes**

Departamento de Odontologia Social  
Faculdade de Odontologia  
Universidade de São Paulo

**Paulo Capel Narvai**

Departamento de Prática de Saúde Pública  
Faculdade de Saúde Pública  
Universidade de São Paulo

## **Resumo**

O presente estudo visa descrever o processo de avaliação da precisão e validade do levantamento de cárie dentária realizado na cidade de São Paulo em 2002, integrado ao Projeto SB2000 do Ministério da Saúde. Para essa finalidade, foi empregada metodologia padronizada internacionalmente pela Organização Mundial da Saúde. Durante o treinamento dos cirurgiões-dentistas que participaram do levantamento, foi realizado exercício de calibração envolvendo nove examinadores e dois grupos de 20 escolares, para a aferição de concordância interobservadores. A validação dos exames bucais usou como critério de "gold standard" a indicação do coordenador do levantamento como examinador padrão. Além disso, durante a coleta de dados, 58 escolares foram reexaminados pelos mesmos profissionais para o cálculo dos indicadores de precisão intra-observador. Foram estimados elevados indicadores sintéticos de precisão para o levantamento: respectivamente, 99,83% e 96,45% de porcentagem geral de concordância intra e interobservadores e 0,996 e 0,943 os correspondentes valores da estatística kappa. A validade do estudo também foi avaliada positivamente, com elevados indicadores globais de sensibilidade (0,972) e especificidade (0,978). A avaliação de precisão e validade para cada observador produziu indicadores igualmente elevados e satisfatórios. A descrição dos procedimentos adotados para avaliação de precisão e validade, bem como a apresentação dos resultados obtidos, podem fornecer indicações metodológicas para estudos subseqüentes dessa natureza.

**Palavras-chave:** Cárie dentária. Coleta de dados. Variações interobservadores. Variações intra-observadores.

**Conflitos de interesse:** nenhum declarado.

**Ética em pesquisa:** O presente estudo utilizou dados originalmente coletados para o estudo "Condições de Saúde Bucal no Estado de São Paulo" (Projeto SB2000), aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Saúde Pública da Universidade de São Paulo.

## Abstract

The present study aims at describing the assessment of the reliability and accuracy of a survey of dental caries performed in the city of São Paulo in 2002, as part of the SB2000 Project of the Ministry of Health. The appraisal of indexes used methodology internationally standardized by the World Health Organization. During the training of examiners participating in the survey, an exercise of calibration was conducted for the gauging of inter-observer agreement, which involved nine dental surgeons and two groups of 20 schoolchildren. The coordinator of the survey was assigned as the gold Standard for the validation of the oral examination records in the calibration process. Furthermore, during the gathering of data, 58 schoolchildren were reexamined by the same professionals, for the assessment of intra-observer agreement. We calculated high synthetic indicators of reliability for the survey: respectively 99.83% and 96.45% for the percentage of overall intra- and inter-observer agreement, and 0.996 and 0.943 for the corresponding values of kappa statistics. The validity of the study was also evaluated positively, with high global indicators of sensitivity (0.972) and specificity (0.978). The evaluation of reliability and accuracy for each observer produced high and satisfactory indicators. The description of the procedures chosen for the assessment of reliability and accuracy, as well as the presentation of results, can provide methodological indications for subsequent studies of dental caries distribution.

**Key Words:** Dental caries. Data collection. Inter-observer variation. Intra-observer variation.

## Introdução

A Organização Mundial da Saúde<sup>1,2</sup> recomenda às autoridades sanitárias a realização de levantamentos epidemiológicos das doenças bucais mais prevalentes com periodicidade entre 5 e 10 anos. Especificamente voltadas para o dimensionamento das condições de saúde bucal e necessidades de tratamento odontológico, essas pesquisas têm focalizado principalmente os seguintes agravos e condições: cárie dentária, fluorose, oclusão dentária, agravos periodontais, uso de prótese e necessidades de tratamento odontológico. Os grupos etários alvo dessas investigações são os seguintes: de 5 a 12 anos, para o estudo de condições afetando a dentição decídua e permanente de crianças; de 15 a 19 anos, como referência para adolescentes; de 35 a 44 anos (adultos) e de 65 a 74 anos (idosos).

Por meio dos levantamentos epidemiológicos é possível estimar a condição de saúde bucal de diferentes grupos populacionais, conhecimento indispensável para a proposição de ações adequadas às suas necessidades e riscos. As informações fornecidas por esses levantamentos possibilitam comparações no tempo e no espaço, e avaliações do impacto diferencial de fatores de risco e proteção, bem como, em certa medida, dos programas de saúde relativos aos agravos e condições considerados.

Seguindo essa recomendação, o serviço público de saúde no Brasil tem realizado diversos levantamentos epidemiológicos em saúde bucal, com amplitude e periodicidade variada, principalmente a partir de meados dos anos 80. Esses levantamentos têm fornecido a base de informações sobre a qual incursões analíticas têm procurado identificar fatores e condições associados aos principais agravos à saúde bucal. Os dados produzidos por tais pesquisas têm sido contemplados por diversos estudos documentando a distribuição de agravos e testando diferentes hipóteses explicativas, principalmente relacionadas à distribuição da cárie dentária em nosso meio<sup>3-7</sup>.

Nesse sentido, é importante que essas bases de dados sejam exploradas de um pon-

to de vista descritivo e analítico, com a finalidade de fornecer parâmetros para futuras análises comparativas e instruir os gestores de saúde na tomada de decisões e no planejamento de serviços. Para essa finalidade, a padronização metodológica desses levantamentos é pré-requisito indispensável para a confiabilidade dos dados obtidos. A OMS<sup>1,2</sup> tem se encarregado dessa atribuição, ao produzir e divulgar métodos padronizados para os levantamentos básicos de saúde bucal.

Dentre os critérios metodológicos cuja padronização é pressuposto para a confiabilidade dos dados produzidos nos levantamentos de saúde bucal, sublinha-se o planejamento da amostra a ser examinada, definições clínicas para os agravos considerados, instrumentos e condições de observação para os exames bucais e – em especial – o dimensionamento da precisão e validade das anotações efetuadas.

Segundo Chaves<sup>8</sup>, a calibração dos examinadores seria o recurso que permite minimizar erros de observação durante a pesquisa epidemiológica em saúde bucal. O método proposto consiste na repetição de exames por diferentes examinadores, para que as diferenças de interpretação sejam discutidas até se firmar um consenso entre os participantes do processo de calibração. Eklund et al.<sup>9</sup> ressaltaram a importância do estudo global de concordâncias dos examinadores num levantamento epidemiológico de saúde bucal, pois a identificação da natureza das discordâncias contribuiu para sua discussão e redução ao mais baixo nível. Com uma compreensão aprimorada dos critérios de observação, os examinadores podem produzir registros mais consistentes, mantendo estáveis os níveis de concordância durante a coleta de dados.

Nesse sentido, o objetivo da calibração seria “assegurar a uniformidade de interpretação, compreensão e aplicação dos critérios para as várias doenças e condições a serem observadas e registradas; assegurar que cada profissional possa examinar de maneira uniformemente padronizada e assim minimizar variações entre os diferentes examinadores”<sup>1</sup>. A OMS<sup>1,2</sup> recomenda que, du-

rante o processo de calibração, diferentes examinadores avaliem os mesmos indivíduos, para se calcular a concordância inter-observadores. A OMS também recomenda o exame em duplicata de cerca de 10% dos indivíduos durante a coleta de dados, para avaliar a concordância intra-observadores<sup>1,2</sup>.

Segundo a OMS<sup>2</sup>, são dois os principais motivos para a variabilidade dos resultados das observações nos estudos de saúde bucal: a dificuldade de classificação das doenças bucais em diferentes níveis, especialmente cárie dentária e doença periodontal, e fatores físicos e psicológicos, como fadiga, flutuações no interesse pela pesquisa e variações na acuidade visual e no senso de tato. Esses fatores, em graus diferenciados, podem afetar o julgamento dos profissionais que participam da coleta de dados para esses estudos.

Nos levantamentos de saúde bucal realizados no Brasil nas últimas décadas, o processo de calibração dos examinadores tornou-se um aspecto importante de planejamento, permitindo o estabelecimento de critérios mais bem definidos para a observação e sua melhor compreensão por parte dos dentistas que atuam como examinadores. Por meio desse procedimento, tem sido possível a apresentação nos relatórios das investigações de resultados sintéticos dos indicadores de concordância intra e interobservadores. Apesar disso, a descrição do processo de calibração dos examinadores nos levantamentos epidemiológicos de saúde bucal realizados recentemente tem sido avaliada como superficial e insuficiente, com a omissão de importantes aspectos metodológicos da avaliação de acurácia e reprodutibilidade<sup>10</sup>.

Tendo em vista essas considerações, o presente estudo visa descrever o processo de avaliação da precisão e validade de um levantamento de cárie dentária realizado na cidade de São Paulo em 2002, com o objetivo mais abrangente de fornecer indicações metodológicas para estudos subsequentes dessa natureza.

## Métodos

Em 2002, a Secretaria de Estado da Saú-

de de São Paulo (por meio de seu Centro Técnico de Saúde Bucal), em convênio com a Faculdade de Saúde Pública da Universidade de São Paulo (por meio do Núcleo de Estudos e Pesquisas de Sistemas de Saúde), promoveu levantamento epidemiológico sobre condições de saúde bucal da população do Estado de São Paulo. Integrada ao Projeto SB2000 do Ministério da Saúde, essa pesquisa<sup>11</sup> forneceu os dados de âmbito estadual para posterior consolidação no quadro mais abrangente do levantamento nacional. O estudo multicêntrico envolveu ainda sete Faculdades de Odontologia: da USP em São Paulo, Bauru e Ribeirão Preto, da UNICAMP em Piracicaba, da UNESP em Araraquara, da PUC em Campinas e a Faculdade de Odontologia de Barretos.

O planejamento da amostra obedeceu aos critérios estipulados pela OMS<sup>2</sup> para garantir a representatividade dos indicadores de prevalência de cárie dentária para o Estado como um todo e para cada um dos 35 municípios participantes. Foi adotada a técnica de amostragem aleatória por conglomerados em três etapas: sorteio dos municípios (unidades amostrais primárias), escolas (unidades amostrais secundárias) e população de referência: crianças, adolescentes, adultos e idosos.

Atendendo às especificações internacionais<sup>2</sup> para esse tipo de estudo, a realização dos exames bucais foi precedida por treinamento e calibração dos dentistas que integraram o estudo na condição de examinadores. O treinamento para os instrutores de calibração foi realizado em fevereiro de 2002, no município de São Paulo, constando de aula teórica ministrada pelo representante do grupo de trabalho constituído pelo Ministério da Saúde e responsável pela elaboração e coordenação do Projeto SB2000 na região Sudeste, sendo seguida de exercícios práticos nos quais eram simuladas situações de exame semelhantes às que seriam encontradas durante a coleta de dados.

Para o processo de calibração dos examinadores foi estipulado um número máximo de cinco dentistas para cada grupo de vinte crianças, visando facilitar o debate so-

bre as discrepâncias de observações e evitar para escolares e profissionais o desgaste devido aos repetidos exames bucais. Para preservar a informação em separado para cada grupo de escolares, os indicadores de precisão do levantamento foram tabulados em dois conjuntos separados.

Todo o processo de calibração foi dimensionado para abranger pelo menos seis turnos de quatro horas de trabalho. A seleção dos escolares que participaram do treinamento de calibração foi efetuada de modo planejado, visando antecipar as condições que os examinadores encontrariam em campo, segundo os critérios recomendados por Pine et al.<sup>12</sup>: inclusão de crianças com cáries, com cáries não tratadas, com dentes restaurados e também crianças livres de cáries.

Na cidade de São Paulo, a calibração dos examinadores foi realizada em março de 2002 na Unidade de Saúde Vila Espanhola QUALIS II – Fundação Zerbini, SES-SP, e na Escola Municipal de Ensino Fundamental Humberto Dantas, na Casa Verde, sob a coordenação de um docente da Faculdade de Odontologia da Universidade de São Paulo e cirurgiões-dentistas das Secretarias municipal e estadual de Saúde. Durante a calibração foram dimensionados os indicadores de concordância interobservadores por meio da repetição de exames e a posterior comparação dos achados. A coleta de dados para o estudo de concordância intra-observadores foi efetuada durante a coleta de dados, através do reexame, pelos mesmos examinadores, em cerca de 20% da amostra pelos mesmos examinadores.

Os dados produzidos em ambos procedimentos foram organizados e transcritos em planilhas eletrônicas fornecidas pelo Ministério da Saúde, e sua avaliação conjunta referenciou o estudo global de precisão do levantamento, com o cálculo da porcentagem geral de concordância – PGC, e da estatística kappa, medida originalmente proposta por Cohen<sup>13</sup> em 1960, que permite avaliar o quanto a concordância obtida excede a que seria obtida apenas pelo acaso na avaliação entre observadores, efetuando registros de uma variável categórica. A informatização

do banco de dados, assim como sua análise estatística, foi efetuada empregando-se programas do pacote Epi Info 6.04d 2001.

A PGC, amplamente empregada em todo o mundo nos levantamentos de cárie dentária até meados dos anos 1980s, refere-se à frequência de registros similares efetuados em campos análogos, correspondentes a determinados espaços dentários de uma mesma criança, preenchidos em cada exame bucal. O cálculo da estatística kappa utilizou uma tabela de dupla entrada, na qual os dados das observações efetuadas em cada conjunto de exames repetidos permitiam discriminar concordâncias e discordâncias nos diferentes campos de registro. Seguindo indicação metodológica fornecida por Eklund et al.<sup>9</sup> e Pine et al.<sup>12</sup>, essa tabela de dupla entrada contemplou as diferentes categorias de variação passíveis de registro para dentes permanentes: hígidos, cariados, extraídos por cárie, restaurados, afetados por trauma, extraídos por outros motivos e não-erupcionados. Esse procedimento possibilitou a comparação dos registros, a quantificação das concordâncias e a identificação da natureza das discordâncias.

Para o estudo de validade dos levantamentos de saúde bucal, a OMS<sup>2</sup> estipula como critério de referência o resultado das observações efetuadas pelo examinador padrão. No âmbito desse estudo, repetindo opção metodológica efetuada em levantamentos realizados anteriormente no país<sup>14,15</sup>, o coordenador assumiu a responsabilidade de “examinador padrão” e estipulou como critério para o registro de suas observações a “técnica do consenso”. Desse modo, cada condição dentária que tenha sido registrada de modo discrepante em qualquer um dos reexames foi matéria de observação e discussão entre o grupo de dentistas, até que eles chegassem a uma conclusão quanto à condição em questão. Essa condição, produto de consenso entre os examinadores e não uma “imposição” do examinador-padrão, era definida como a condição para comparação com as observações dos examinadores no estudo de acurácia das observações. Com base na comparação dos pre-

enchimentos de cada examinador e a ficha estabelecida pelo consenso, foram estimadas a sensibilidade, a especificidade e a estatística J de Youden<sup>16</sup>. Esses indicadores de validade são de uso tradicional em estudos epidemiológicos e se referem à capacidade dos examinadores, ou do levantamento em geral, de identificar corretamente quem tem cáries (sensibilidade) e identificar corretamente quem não tem cáries (especificidade). A estatística J de Youden é calculada através da soma dos indicadores de sensibilidade e especificidade, e subtração de uma unidade, constituindo desse modo uma indicação sintética de validade, na qual as perdas de qualidade devidas a resultados falso-positivos e falso-negativos são contabilizadas em uma única grandeza.

## Resultados

No levantamento epidemiológico de cáries na cidade de São Paulo em 2002, nove dentistas examinadores efetuaram o exame bucal de 272 escolares de 12 e de 15 a 19 anos de idade. Tendo em vista o número de profissionais envolvidos, o procedimento de aferição da precisão interobservadores e da validade do levantamento implicou a constituição de dois grupos, cada um deles abrangendo cinco exames sucessivos de 20 escolares. Além disso, durante a coleta de dados, 58 escolares foram reexaminados pelos mesmos dentistas que efetuaram a primeira observação, para a aferição dos indicadores de precisão intra-observadores.

A Tabela 1 registra os resultados do estudo de precisão para cada um dos dentistas, e para cada dupla de profissionais, segundo o grupo de escolares examinados. Todas as comparações interobservadores resultaram valores de PGC e estatística kappa superiores a, respectivamente, 93% e 0,88. Apenas dois examinadores não obtiveram uma concordância integral em sua checagem intra-observador.

Na seqüência, a avaliação em conjunto das fichas usadas na determinação desses indicadores permitiu o cálculo de indicadores sintéticos da precisão intra e interobser-

**Tabela 1** – Indicadores de precisão inter e intra-observadores para o exame de cárie dentária na cidade de São Paulo, 2002, segundo examinador.

**Table 1** – (Inter- and Intra-observer accuracy indicators for the examination of dental caries in the City of São Paulo, 2002, per examiner)

		Examinador 1	Examinador 4	Examinador 5	Examinador 8
PGC	Examinador 1	100,00%	93,24%	95,99%	96,14%
Kappa		1,000	0,885	0,934	0,934
EP		0,064	0,028	0,029	0,035
N		6	13	13	10
PGC	Examinador 4		100,00%	94,02%	96,78%
Kappa			1,000	0,899	0,940
EP			0,055	0,028	0,035
N			6	14	13
PGC	Examinador 5			99,55%	96,56%
Kappa				0,993	0,947
EP				0,047	0,029
N				7	13
PGC	Examinador 8				100,00%
Kappa					1,000
EP					0,006
N					6

  

		Examinador 2	Examinador 3	Examinador 6	Examinador 7	Examinador 9
PGC	Examinador 2	100,00%	99,26%	98,51%	98,50%	97,38%
Kappa		1,000	0,987	0,977	0,974	0,952
EP		0,042	0,030	0,036	0,038	0,039
N		7	13	9	9	9
PGC	Examinador 3		100,00%	98,46%	95,77%	95,83%
Kappa			1,000	0,977	0,930	0,928
EP			0,048	0,033	0,035	0,034
N			8	11	9	11
PGC	Examinador 6			98,75%	95,70%	94,61%
Kappa				0,970	0,933	0,915
EP				0,058	0,033	0,032
N				5	10	11
PGC	Examinador 7				100,00%	97,81%
Kappa					1,000	0,968
EP					0,054	0,028
N					6	13
PGC	Examinador 9					100,00%
Kappa						1,000
EP						0,053
N						7

PGC: Porcentagem geral de concordância (*General agreement percentage*); Kappa: estatística kappa (*Kappa: kappa statistics*); EP: erro padrão da estatística kappa (*EP: standard error of kappa statistics*); N: número de crianças reexaminadas (*N: number of children re-examined*)

vadores. Os elevados valores de PGC e kappa obtidos através desse procedimento possibilitaram admitir como elevada a reprodutibilidade do estudo (Tabela 2).

Os registros de cada examinador foram cotejados com a ficha-padrão resultante da

construção do consenso entre os examinadores, estabelecida como o “*gold standard*” da pesquisa. A comparação desses registros (Tabela 3) referenciou a avaliação global de validade do levantamento. Todos os examinadores obtiveram índices de sensibilidade

**Tabela 2** – Indicadores globais de precisão inter e intra-observadores para o exame de cárie dentária na cidade de São Paulo, 2002.

**Table 2** – (Inter- and Intra-observer overall accuracy indicators for the examination of dental caries in the City of São Paulo, 2002)

		Intra-observadores	Inter observadores
PGC	Total Parcial 1	99,98%	95,34%
Kappa		0,997	0,922
EP		0,027	0,012
N		25	20
PGC	Total Parcial 2	99,79%	97,25%
Kappa		0,994	0,956
EP		0,021	0,010
N		33	20
PGC	Total Geral	99,83%	96,45%
Kappa		0,996	0,943
EP		0,016	0,008
N		58	40

PGC: Porcentagem geral de concordância (General agreement percentage); Kappa: estatística kappa (Kappa: kappa statistics); EP: erro padrão da estatística kappa (EP: standard error of kappa statistics); N: número de crianças reexaminadas (N: number of children re-examined)

**Tabela 3** – Indicadores de validade para o exame de cárie dentária na cidade de São Paulo, 2002.

**Table 3** – (Validity Indicators for the examination of dental caries in the City of São Paulo, 2002)

Examinador	N	Sensibilidade	Especificidade	J de Youden
<b>1</b>	15	0,961	0,958	0,919
<b>2</b>	13	0,982	0,988	0,970
<b>3</b>	17	0,989	0,985	0,974
<b>4</b>	18	0,955	0,977	0,932
<b>5</b>	18	0,974	0,984	0,958
<b>6</b>	14	0,973	0,970	0,943
<b>7</b>	12	0,960	0,986	0,946
<b>8</b>	15	0,990	0,974	0,964
<b>9</b>	13	0,955	0,975	0,930
<b>Total</b>	40	0,972	0,978	0,951

e especificidade superiores a 0,95. Os elevados valores globais obtidos para esses parâmetros, assim como para cada um dos examinadores, referenciam como elevada a validade do estudo.

## Discussão

É importante estabelecer parâmetros mínimos de precisão dos levantamentos de saúde bucal, para avaliar e discutir os resultados obtidos, reduzindo as discrepâncias a

níveis aceitáveis. Eklund et al.<sup>9</sup> e Pine et al.<sup>12</sup> recomendam que sejam novamente submetidos ao processo de treinamento e calibração, ou então excluídos do levantamento, os examinadores que não atinjam indicadores de precisão dentro do limite indicado pelo intervalo de confiança (95%) das médias apresentadas pelo grupo no treinamento de calibração. Frias<sup>17</sup> estimou em respectivamente 97,5% e 90,0% os limites inferiores aceitáveis para a PGC e intra e interobservadores para o exame de cárie dentária. A

maior amplitude do erro esperado para as observações interobservadores seria justificada pela manifestação mais usual de divergências nesse tipo de determinação. Sob esse ponto de vista, pode-se considerar como satisfatórios os resultados obtidos pelo presente estudo de precisão.

Landis e Koch<sup>18</sup> estabeleceram uma diretriz para a interpretação dos valores obtidos para a estatística kappa, na qual resultados acima de 0,8 são considerados como concordância “quase perfeita”. Outros critérios propostos para esta classificação estipulam valores acima de 0,8 como “muito boa concordância”<sup>19</sup>; acima de 0,75 como “excelente”<sup>20</sup>; e “muito boa” acima de 0,8 e “excelente” acima de 0,92<sup>21</sup>. Seguindo qualquer um desses critérios, também podem ser estimados como satisfatórios os graus de concordância intra e interobservadores obtidos na pesquisa em análise.

Quanto à interpretação do erro padrão da estatística kappa, Fleiss<sup>22</sup> afirmou que o quociente entre o valor calculado para a estatística kappa e seu respectivo erro padrão em comparações envolvendo dois conjuntos de observações cuja concordância fosse apenas devida ao acaso seguiria uma distribuição normal. Nesse sentido, quocientes mais elevados que 1,96 – condição obtida em todas as comparações intra e interexaminadores no presente estudo – indicam concordâncias significativamente mais elevadas que apenas aquela devida ao acaso (nível de significância de 95%).

De modo geral, os resultados de PGC e kappa obtidos na avaliação de concordância intra e interobservadores forneceram indicações favoráveis e análogas. No entanto, existem diferenças de significado e de metodologia de cálculo para ambos os indicadores, e essa similaridade nem sempre é verificada nos valores calculados para esses parâmetros. Enquanto a PGC expressa simplesmente a relação entre o número de registros discordantes e o total de registros considerados, a estatística kappa é influenciada pela distribuição de freqüências das observações nos diferentes campos de registro. Desse modo, as discrepâncias de ob-

servação causam maior ou menor impacto sobre o valor da estatística kappa, segundo a característica diferencial da distribuição de freqüências anotadas nos distintos campos de registro.

Quando a avaliação de concordância resulta em uma distribuição relativamente homogênea nos campos de registro, é maior a probabilidade de ocorrerem divergências nas comparações intra e interobservadores. Em contrapartida, uma distribuição de freqüências mais concentrada em apenas um ou dois campos de registro indica menor probabilidade de discordâncias. Ao contrário da PGC, a estatística kappa permite ponderar esses diferentes padrões, atribuindo um peso mais elevado às divergências observadas nesse último caso. Isto é, quanto menor a probabilidade de ocorrerem discrepâncias de observação, maior será o peso a elas atribuído pela estatística kappa. Por ser influenciada pela distribuição de freqüências registradas nos diferentes campos de observação, a estatística kappa é considerada uma medida mais adequada que a PGC<sup>16</sup> para estudos de precisão.

Peres et al.<sup>10</sup> recomendam a determinação dos valores de estatística kappa para cada dente; desse modo, os resultados divergentes observados não seriam diluídos pelos elementos dentários, como os caninos, para os quais praticamente não há controvérsia nas observações, em função do menor risco de esses dentes serem afetados por cárie. Por meio desse procedimento, seriam estimados indicadores de precisão menos elevados para os dentes com maior probabilidade de registros discrepantes e, conseqüentemente, maior gama de dados poderia ser levada em consideração, permitindo a apreciação de confiabilidade do estudo sob um ângulo diferente. Dentes com maior participação na definição dos escores individuais do índice CPO, como os primeiros molares permanentes, poderiam ser objeto de análise específica.

Para a apreciação de validade do estudo, a OMS recomenda a indicação de um examinador padrão: “um epidemiologista experiente que atuará na validação dos resultados obtidos pela equipe de examinadores”<sup>2</sup>.



A importância de complementar o dimensionamento de reprodutibilidade intra e interobservadores com a aferição de validade do estudo reside no fato de ser possível que observações próximas entre si difiram dos valores presumidos como verdadeiros. No entanto, nem o manual da OMS<sup>2</sup> descrevendo os métodos básicos para levantamentos de saúde bucal, nem os relatórios recentes dos estudos epidemiológicos realizados no Brasil<sup>11,14,15,23,24</sup> forneceram informações suplementares para a comparação entre as observações efetuadas pelo examinador padrão e a equipe de examinadores do estudo.

Visando suprir essa lacuna, Peres et al.<sup>10</sup> propuseram a estimativa da estatística kappa associada à concordância entre o examinador padrão e os membros da equipe de examinadores. No entanto, essa diretriz apenas identifica o grau de concordância geral entre as observações e o padrão estipulado como “*gold Standard*”, sem especificar o sentido dos erros identificados (falso-positivo ou falso-negativo). Ademais, a estatística kappa é influenciada pela prevalência da doença e, deste modo, não é possível utilizá-la em comparações entre estudos, procedimentos ou populações. Nesse sentido, procurou-se no presente estudo dimensionar a sensibilidade e especificidade das observações efetuadas, seguindo indicações metodológicas fornecidas por Fyffe et al.<sup>25</sup> para o cálculo de indicadores de validade.

Para essa finalidade, os campos da tabela de dupla entrada comparando os registros do examinador padrão com os registros de cada um dos examinadores foram consolidados em quatro categorias: (1) negativo, quando as condições associadas à doença cárie foram afastadas por ambos examinadores; (2) positivo, quando essas condições foram identificadas por ambos; (3) falso-negativo, quando o examinador padrão registrou manifestação de cárie não identificada pelo examinador avaliado; e (4) falso-positivo, quando o examinador avaliado identificou manifestação de cárie não identificada pelo examinador padrão. Desse modo, foi possível estimar os indicadores de sensibilidade e especificidade, e também o

índice de Youden (J), que combina essas duas dimensões, e é uma boa alternativa para a estatística kappa na avaliação de acurácia ou validade, uma vez que seus valores independem dos níveis de prevalência da doença em estudo.

Os elevados valores observados para os indicadores de validade indicam como bem sucedido o presente conjunto de observações. Essa constatação deve, no entanto, ser ponderada no âmbito do controle possível de ser efetuado para avaliação de validade dos levantamentos epidemiológicos de cárie dentária, uma vez que a opção metodológica pelo uso do examinador padrão como critério de validade ainda implicaria algum risco de erro nas determinações, pois não existe um “*gold Standard*” verdadeiro para esse tipo de levantamento<sup>25</sup>. Tal restrição não impede, contudo, que a validade de levantamentos epidemiológicos de saúde bucal seja avaliada com o intuito de apreciar diferentes hipóteses de interesse, como a comparação de efetividade entre diferentes métodos de coleta de dados para tais estudos<sup>26</sup>.

Recursos auxiliares de diagnóstico, como a radiografia e equipamentos de laser fluorescente (Diagnodent) poderiam ser considerados na obtenção de dados complementares para a verificação de validade dos levantamentos epidemiológicos de cárie dentária. Outras opções metodológicas poderiam considerar a estimação de máxima verossimilhança ou métodos Bayesianos assumindo a não existência de “*gold standard*” real. No entanto, esses esforços ainda não foram objeto de padronização para levantamentos desse tipo. Nesse sentido, deve-se buscar atender aos requisitos metodológicos estipulados pela Organização Mundial da Saúde, efetuando-se a aferição de validade através do examinador padrão. Independente dos esforços voltados a possíveis avanços nos procedimentos de avaliação da validade, é importante que os estudos de prevalência e severidade de cárie dentária atendam às diretrizes internacionais, de modo a garantir que seus resultados tenham comparabilidade no tempo e no espaço. Cabe ponderar, todavia, que recursos metodológicos

como a técnica do consenso descrita nesse estudo para o preenchimento da ficha padrão, podem configurar contribuição original e útil para aprimorar a qualidade dos levantamentos de cáries e conferir maior credibilidade aos resultados obtidos.

---

## Referências

1. Organização Mundial da Saúde. *Levantamento epidemiológico básico de saúde bucal*. 3ª ed. São Paulo: Santos; 1991.
2. World Health Organization. *Oral health surveys, basic methods*. 4ª ed. Geneva: OMS, 1997.
3. Peres MAA, Narvai PC, Calvo MCM. Prevalência de cárie dentária em crianças aos 12 anos de idade, em localidades do Estado de São Paulo, Brasil, período 1990-1995. *Rev Saúde Pública* 1997; 31: 594-600.
4. Narvai PC, Frazão P, Castellanos RA. Declínio na Experiência de Cárie em Dentes Permanentes de Escolares Brasileiros no Final do Século XX. *Odontologia e Sociedade* 1999; 1: 25-9.
5. Narvai PC, Castellanos RA, Frazão P. Prevalência de cárie em dentes permanentes de escolares do Município de São Paulo, SP, 1970-1996. *Rev Saúde Pública* 2000; 34: 196-200.
6. Antunes JLF, Frazão P, Narvai PC, Bispo C.M, Pegoretti T. Spatial analysis to identify differentials in dental needs by area-based measures. *Community Dent Oral Epidemiol* 2002; 30: 133-42.
7. Antunes JLF, Pegoretti T, Andrade FP, Junqueira SR, Frazão P, Narvai PC. Ethnic disparities in the prevalence of dental caries and restorative dental treatment in Brazilian children. *Int Dent J* 2003; 53: 7-12.
8. Chaves MM. *Odontologia social*. São Paulo: Artes Médicas; 1986.
9. Eklund SA, Moller IJ, Leclercq MH. Calibrating examiners for oral health epidemiological surveys. Geneva: World Health Organization (WHO/ORH/Epid. 93.1); 1996.
10. Peres MA, Traebert J, Marcenes W. Calibração de examinadores para estudos epidemiológicos de cárie dentária. *Cad Saúde Pública* 2001; 17: 153-9.
11. Secretaria de Estado da Saúde de São Paulo. Centro Técnico de Saúde Bucal. Universidade de São Paulo. Faculdade de Saúde Pública. Núcleo de Estudos e Pesquisas de Sistemas de Saúde. *Condições de Saúde Bucal no Estado de São Paulo em 2002: Relatório Final*. São Paulo: SES-SP e FSP-USP; 2002.
12. Pine CM, Pitts NB, Nugent ZJ. British Association for the Study of Community Dentistry (BASCD) guidance on the statistical aspects of examiners for surveys of calibration of examiners for child dental health. A BASCD coordinated dental epidemiology programme quality standard. *Community Dental Health* 1997; 14: 18-29.
13. Cohen JA. Coefficient of agreement for nominal scales. *Education and Psychological Measurement* 1960; 20: 37-46.
14. Secretaria de Estado da Saúde de São Paulo. Centro Técnico de Saúde Bucal. Universidade de São Paulo. Faculdade de Saúde Pública. Núcleo de Estudos e Pesquisas de Sistemas de Saúde. *Levantamento epidemiológico da cárie dentária, oclusopatias e fluorose dentária, em crianças de 5 a 12 anos de idade, em escolas públicas e privadas do município de São Paulo, em 1996*. São Paulo: SES-SP e FSP-USP; 1997.
15. Secretaria de Estado da Saúde de São Paulo. Centro Técnico de Saúde Bucal. Universidade de São Paulo. Faculdade de Saúde Pública. Núcleo de Estudos e Pesquisas de Sistemas de Saúde. *Levantamento Epidemiológico em Saúde Bucal – Estado de São Paulo, 1998*. São Paulo: SES-SP e FSP-USP; 1999.
16. Szklo M, Javier Nieto F. *Epidemiology: beyond the basics*. Gaithersburg: Aspen; 2000.
17. Frias AC. *Estudo de Confiabilidade do Levantamento Epidemiológico de Saúde Bucal – Estado de São Paulo 1998* [dissertação de mestrado]. São Paulo: Faculdade de Saúde Pública da USP; 2000.
18. Landis JR, Kock GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
19. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991.
20. Fleiss JL. *Statistical methods for rates and proportions*. 2<sup>nd</sup> ed. New York: Wiley; 1981.
21. BByrt T. How good is agreement. *Epidemiology* 1996; 7: 561.
22. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971; 5: 378-82.

23. Ministério da Saúde. *Levantamento epidemiológico em saúde bucal: Brasil, zona urbana, 1986*. Brasília: Centro de Documentação do Ministério da Saúde; 1988.
24. Ministério da Saúde. Levantamento Epidemiológico em Saúde Bucal 1996: Cárie dental. Brasília: URL: <http://tabnet.datasus.gov.br/cgi/sbucal/sbdescr.htm>; (06/06/2003).
25. FFyffe HE, Deery C, Nugent ZJ, Nuttall NM, Pitts NB. Effect of diagnostic threshold on the validity and reliability of epidemiological caries diagnosis using the Dundee Selectable Threshold Method for caries diagnosis (DSTM). *Community Dentistry and Oral Epidemiology* 2000; 28(1): 42-51.
26. Beltran ED, Malvitz DM, Eklund SA. Validity of two methods for assessing oral health status of populations. *Journal of Public Health Dentistry* 1997; 57(4): 206-14.

recebido em: 05/01/04  
versão reformulada apresentada em: 22/04/04  
aprovado em: 26/04/04