

Comparação entre medidas de reprodutibilidade para a calibração em levantamentos epidemiológicos da cárie dentária

Comparison of reproducibility measurements for calibration of dental caries epidemiological surveys

Andréa Videira Assaf¹
 Luciane Zanin²
 Marcelo de Castro Meneghim³
 Antonio Carlos Pereira³
 Gláucia Maria Bovi Ambrosano³

Abstract

This study compares three measurements (Kappa, general agreement percentage, or GAP, and dice index) used to determine the reproducibility of caries diagnosis in epidemiological surveys under different clinical diagnostic thresholds. Eleven examiners with previous experience in epidemiological surveys were submitted to a theoretical and clinical calibration process. Data analysis used two caries detection thresholds: World Health Organization (WHO) and WHO with the inclusion of initial enamel lesions (WHO + IL). Twenty-three children 6-7 years of age were examined, with and without caries. Mean values for Kappa index, GAP, and Dice were considered high (≥ 0.90), except for the dice index for the WHO + IL threshold (0.69). Since Kappa is an adjusted agreement index, it can be considered the instrument of choice for calibration of examiners. However, when it is impossible to use, the GAP is recommended together with the dice index in order to orient and improve examiners when examining caries lesions.

Dental Caries; Diagnosis; Reproducibility of Results

Introdução

O papel da saúde coletiva na condução de pesquisas epidemiológicas para a averiguação da saúde bucal nas comunidades tem sido largamente reconhecido^{1,2}. O primeiro passo é o da obtenção de dados para uma futura análise e descrição, com a finalidade de tomar-se alguma decisão de controle ou interventiva. Dessa forma, o controle dos vieses metodológicos gera resultados com maior precisão e fidedignidade, influenciando as respostas do estudo decisivamente³.

Em estudos epidemiológicos, nos quais participam diversos observadores, é de grande importância que as variações na medição sejam também minimizadas mediante uma padronização das medidas de coleta de dados pelos observadores⁴. Procedimentos de estimação têm sido recomendados previamente no período de calibração dos examinadores e durante o levantamento epidemiológico, visando avaliar a variabilidade entre os mesmos. Dentre eles, são frequentemente utilizadas medidas para aferir-se a validade (acurácia) e a reprodutibilidade (precisão).

A validade de um instrumento ou teste é medida pela sua capacidade de classificar corretamente os indivíduos submetidos aos mesmos⁵. Para ser válido, é necessário que o teste meça, de fato, aquilo que se propõe a medir, ou seja, classificar corretamente como positivas as

¹ Superintendência de Recursos Humanos, Universidade Federal Fluminense, Niterói, Brasil.
² Faculdade de Odontologia de Araras, Centro Universitário Hermínio Ometto, Araras, Brasil.
³ Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas, Piracicaba, Brasil.

Correspondência

A. V. Assaf
 Departamento de Assuntos Comunitários, Superintendência de Recursos Humanos, Universidade Federal Fluminense.
 Rua Miguel de Frias 9, Niterói, RJ 24220-000, Brasil.
 avassaf@gmail.com

peças que têm a doença e como negativas as que não têm. A sensibilidade e a especificidade são as medidas que expressam a validade do exame e são definidas através de estudos planejados para esse fim ^{6,7}. A verdade (o verdadeiro estado de doente ou não-doente) é determinada pelo método mais definitivo de diagnóstico, sendo, por isso, denominado de “padrão ouro” (*gold standard*). Por exemplo, o padrão para o diagnóstico de câncer de pulmão pode ser a confirmação histológica. No caso da pesquisa epidemiológica para a cárie dentária, esse tipo de medida não é tão freqüentemente empregado pela inviabilidade de se ter um método padrão para a confirmação da presença ou ausência da doença. Alguns estudos propõem a comparação dos resultados dos exames obtidos pelos examinadores do estudo com o examinador padrão (coordenador) responsável pela calibração dos mesmos. Críticas quanto a essa medida são justificadas, uma vez que o examinador padrão, mesmo com bastante experiência, estaria também sujeito a erros de classificação da doença/saúde nos indivíduos ^{5,8}.

A avaliação da reprodutibilidade (precisão) seria, portanto, a medida de maior emprego nos levantamentos epidemiológicos da cárie dentária durante a calibração dos examinadores. A reprodutibilidade é a capacidade de concordância dos resultados quando o instrumento é aplicado, uniformemente e em repetidas vezes, sobre objetos invariantes ⁵. Medidas de confiabilidade mais empregadas nas pesquisas epidemiológicas da doença cárie são a porcentagem geral de concordância, o índice dice e a estatística kappa ⁹.

Um dos cálculos mais simples para identificar e quantificar a concordância inter e intra-examinadores é a porcentagem geral de concordância, sendo especificada pela porcentagem geral de concordância no diagnóstico entre dois examinadores dentre todas as condições observadas. Apesar de fácil, tal medida não informa como se deu a discordância entre os examinadores, se os resultados positivos e negativos das discordâncias estavam distribuídos uniformemente ou se um examinador teve maior número de resultados positivos do que outro. Além disso, essa medida não leva em consideração a concordância de diagnóstico atribuída ao acaso ^{6,7,9}.

O índice dice, proposto por Dice (1945), pode ser empregado para avaliar a concordância quando apenas uma classe se constitui no objeto de interesse, como o número de dentes cariados. Apresenta vantagens como: facilidade e rapidez de aplicação, fornecendo uma avaliação comparativa razoavelmente adequada, além

de, como na porcentagem geral de concordância, permitir o cálculo manual. Na prática, o índice dice enfatiza a presença da doença, enquanto que a porcentagem geral de concordância e o kappa são medidas mais abrangentes (Tabela 1) ^{7,9,10}.

A estatística kappa, formulada por Cohen (1960), tem sido a medida de concordância mais aceita e utilizada, gerando uma informação de melhor qualidade, permitindo uma melhor avaliação sobre as coincidências e discrepâncias verificadas no processo de calibração. É um índice de concordância ajustado, pois leva em consideração a proporção de concordâncias além da chance, ou seja, atribuídas ao acaso ⁷. Os valores de kappa podem variar de 1 (concordância perfeita) até um valor negativo ^{9,11}. Landis & Koch ¹² propuseram uma escala de seis níveis para interpretar os valores de kappa, nos quais valores de zero são considerados concordância ruim; 0,00 a 0,20, como leve concordância; 0,21 a 0,40, sofrível concordância; 0,41 a 0,60, moderada concordância; 0,61 a 0,80, boa concordância; e valores acima de 0,81, como quase perfeita concordância. A Organização Mundial da Saúde (OMS) atualmente tem preconizado tal estatística para o cálculo da concordância dos examinadores durante o processo de calibração, especialmente para grupos de baixa prevalência de cárie ¹³.

O objetivo deste estudo foi comparar três medidas (kappa, porcentagem geral de concordância e índice dice) utilizadas para determinar a reprodutibilidade do diagnóstico da cárie durante o processo de calibração em levantamentos epidemiológicos, sob diferentes níveis de detecção da doença.

Material e método

Os procedimentos foram iniciados após o projeto ter sido aprovado pelo Comitê de Ética da Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas – FOP/UNICAMP (protocolo n. 068/2002) de acordo com a Resolução 196 de 10 de outubro de 1996 do Conselho Nacional de Saúde do Ministério da Saúde, referente à participação de humanos em pesquisa.

Delineamento do estudo

O estudo consistiu na participação de um grupo de 11 cirurgiões-dentistas, todos com experiência em levantamentos epidemiológicos da cárie dentária, previamente treinados e calibrados de acordo com a codificação e os critérios da OMS ¹³.

Os exames foram realizados em 23 crianças de 6-7 anos de idade de uma Escola Pública Municipal de Piracicaba, São Paulo, Brasil. Essas foram previamente selecionadas por um examinador que não participou diretamente do estudo. Os exames de triagem foram realizados com espelho bucal plano, sonda milimetrada IPC (Índice Periodontal Comunitário – “sonda preconizada pela OMS para levantamentos epidemiológicos”) e escovação dentária prévia. A seleção das crianças foi realizada de modo uniforme, sendo dada preferência àquelas que apresentavam lesões cavitadas bem como lesões iniciais em esmalte (lesões de mancha branca ativa – LI). Algumas crianças livres de cárie também foram selecionadas a fim de que os examinadores não entrassem no exercício com a pré-concepção de que todas as crianças apresentavam lesões cariosas.

Códigos e critérios adotados nos exames

Os códigos e critérios adotados foram os da OMS¹³. Entretanto, incluiu-se o registro das lesões iniciais ativas em esmalte (adaptação dos critérios de Nyvad et al.¹⁴ e Fyfee et al.⁸). Portanto, LI foi definida como uma cárie ativa, a qual, através do exame visual, indica uma superfície intacta, sem perda clínica de tecido dentário, com coloração amarelada/esbranquiçada, de crescente opacidade e rugosa (ao passar-se a sonda de extremo arredondado levemente através da superfície da lesão). Quanto à localização da lesão: superfície lisa – lesão de cárie localizada tipicamente perto da margem gengival; superfície oclusal – morfologia da fôssula e fissura intactas, com a lesão estendendo-se ao longo das paredes da fissura. Manchas brancas ativas, adjacentes ou não a superfícies com selantes, restaurações ou mesmo cáries (cavitações) também foram registradas.

Treinamento e calibração dos examinadores

Todos os examinadores foram auxiliados por anotadores durante a realização dos exames. Um examinador padrão (*gold standard*) conduziu todo o processo de calibração dos examinadores, sendo que as atividades teórico-práticas dos exercícios de treinamento e calibração compreenderam um total de sete períodos de quatro horas (um teórico, quatro treinamentos clínicos e dois exercícios de calibração).

O examinador padrão, dentista com larga experiência em levantamentos epidemiológicos, utilizando critérios da OMS¹³, participou do treinamento e calibração do diagnóstico das LI no estudo piloto, sendo que o mesmo tam-

Tabela 1

Valores médios e intervalo de variação de kappa, dice e porcentagem geral de concordância, segundo critérios de diagnóstico Organização Mundial da Saúde (OMS) e OMS associado a lesões iniciais em esmalte (OMS + LI). Piracicaba, São Paulo, Brasil, 2004.

Estimadores de concordância	Nível de detecção da lesão de cárie	
	OMS*	OMS + LI*
Kappa	0,95 (0,93-0,98)	0,90 (0,86-0,93)
Dice	0,93 (0,89-0,98)	0,69 (0,54-0,77)
Porcentagem geral de concordância	0,96 (0,95-0,99)	0,93 (0,90-0,95)

* Intervalo de variação.

bém empregou critério similar de exame em outros estudos^{15,16}.

As atividades teóricas consistiram de uma aula teórica ministrada pelo examinador padrão sobre condutas de exame e códigos empregados no estudo. Prévia e posteriormente à explanação, exercícios teóricos foram realizados, a partir da exposição de slides, abordando cada situação clínica passível de ser encontrada nos exames. Após essa fase e diante de resultados satisfatórios de acerto no exercício final (90,46%), o grupo iniciou o treinamento clínico.

O treinamento consistiu de 4 períodos de quatro horas (16 horas). Cada participante examinou, por período, de 10 a 12 crianças, com prevalências de cárie diferenciadas, bem como com presença de LI. Discussões entre os examinadores e o examinador padrão, em relação aos achados clínicos, critérios de diagnóstico, codificações e erros de registro, foram realizadas durante o treinamento, com o propósito de alcançar-se um nível aceitável de concordância ($kappa > 0,85$ – OMS¹³).

O exercício final de calibração foi de dois períodos, sendo que cada cirurgião-dentista examinou o grupo de 23 crianças, as quais apresentavam diversos achados clínicos, como já mencionado anteriormente. Nessa fase, os examinadores não discutiram com os outros membros da equipe sobre os achados encontrados.

Condições de exame

Os exames foram realizados sob condições de iluminação natural com espelho e sonda IPC, precedidos por secagem e escovação dentária.

A sonda de extremo arredondado foi utilizada para o diagnóstico, sem a aplicação de pressão sobre a superfície dentária, somente para esclarecer dúvidas advindas do diagnóstico visual, e também como auxiliar na remoção de

biofilme dental ou restos de alimentos que poderiam ter permanecido sobre o dente.

A escovação foi realizada com dentífrico fluoretado, sob a técnica de Bass modificada, por um tempo padronizado de dois minutos. Após a escovação, os indivíduos foram submetidos ao exame.

A secagem para a execução dos exames no pátio da escola foi realizada com o uso do ar comprimido, através de um compressor odontológico (Wetzel: medical line 3,6/30 – 0,5 HP) acoplado a uma peça de metal, a qual era ligada a mangueiras de 0,4cm de diâmetro para a saída de ar, em torno de 3 a 5 segundos por unidade dentária.

Crítérios de diagnóstico utilizados para a avaliação e forma de análise dos resultados

Para o cálculo dos valores de reprodutibilidade interexaminadores, dois níveis diferentes de detecção da doença (ou limiares de diagnóstico) foram empregados: limiar de diagnóstico de cárie a partir da presença da lesão cavitada – referência preconizada pela OMS¹³ e limiar de diagnóstico de cárie com a inclusão das LI (OMS + LI).

Os dados foram utilizados como base para os cálculos de reprodutibilidade interexaminadores durante o período de calibração, por meio da porcentagem geral de concordância, índice dice e estatística kappa, para todos os dentes em conjunto.

Importante relatar que os valores médios dos três instrumentos de medida são o resultado da soma e média final, obtida através de cruzamentos entre examinadores (ex: examinador 1 *versus* examinador 2, 1 x 3, 1 x 4), de acordo com cada nível de detecção da doença, OMS e OMS + LI. A forma de apresentação dos resultados por meio de valores médios justifica-se por ter sido alcançado um grande número de valores de reprodutibilidade obtidos por diferentes duplas de examinadores, o que geraria dificuldades para o entendimento do leitor.

Resultados

Valores dos índices kappa, porcentagem geral de concordância e dice foram considerados altos ($\geq 0,90$), à exceção do índice dice para o limiar de diagnóstico OMS + LI (0,69). De uma maneira geral, os respectivos índices apresentaram valores inferiores para o limiar OMS + LI, quando comparados ao limiar OMS (Tabela 1).

Discussão

Os instrumentos de confiabilidade utilizados para avaliar-se a reprodutibilidade durante o processo de calibração são ferramentas de extrema importância, uma vez que, posterior a esta aplicação, pode-se verificar e melhorar a precisão da pesquisa, levando ao aperfeiçoamento dos examinadores quanto à interpretação dos critérios de exame e situações clínicas vigentes, de forma homogênea e consistente^{3,17,18}.

A porcentagem geral de concordância é um instrumento de medida simples e rápido de ser aplicado no processo de calibração de examinadores, podendo ser calculada manualmente, sem a necessidade da montagem de uma matriz de diagnóstico, ainda em ambiente de campo e após o exercício de calibração dos examinadores⁶. Entretanto, na quarta edição do manual da OMS¹³, recomenda-se o uso do coeficiente kappa para grupos/populações de baixa prevalência de cárie, uma vez que se excluem do cálculo os valores atribuídos ao acaso, apresentando, assim, resultados mais condizentes com a real concordância. Dessa forma, valores de kappa são levemente mais baixos do que os da porcentagem geral de concordância. Diferentemente da porcentagem geral de concordância, o kappa só é viabilizado por meio da construção da matriz de diagnóstico, possibilitando, assim, que sejam analisadas onde estão situadas as discordâncias mais frequentes, gerando, portanto, um melhor direcionamento do processo de calibração e levando, conseqüentemente, a um resultado melhor de calibração dos examinadores.

O índice dice, apesar de fácil cálculo⁷, tem sido preconizado e empregado em poucas pesquisas epidemiológicas, como a de Pine et al.¹⁰. No presente estudo, um valor mais baixo desse índice ($D = 0,69$) foi obtido para o nível de detecção da lesão cariada com a inclusão das LI, podendo isso ser atribuído, primeiramente, ao fato desse instrumento, no caso do presente estudo, enfatizar a presença da doença, ou seja, levar em consideração apenas os resultados de concordâncias e discordâncias entre os examinadores com relação a dentes cariados, e, principalmente, pelo diagnóstico das lesões iniciais, gerar maiores dificuldades e discordâncias entre os examinadores^{8,14}. Entretanto, mesmo tendo sido um valor mais baixo, o mesmo é representativo de uma boa concordância dos examinadores para as lesões de cárie cavitadas e não cavitadas.

Questionamentos podem ser suscitados no sentido de que tanto a porcentagem geral de concordância como o kappa, incluindo-se a

análise de todas as condições clínicas, podem gerar, de uma forma geral, uma boa impressão de concordância entre os examinadores, mesmo mediante o emprego de níveis de detecção de cárie considerados mais complexos, como sob a inclusão de lesões em seus estágios iniciais. Para uma maior compreensão, atualmente, e especialmente em grupos de baixa prevalência de cárie, observa-se a presença de um maior número de dentes hígidos em relação aos dentes cariados. Dessa forma, os dentes hígidos, fáceis de serem diagnosticados, gerariam um efeito de diluição do erro atribuído principalmente aos dentes cariados e dentes com lesões iniciais de cárie, os quais apresentam maiores discordâncias de diagnóstico, levando, assim, a uma visão positiva dos resultados finais da calibração dos examinadores¹⁸.

No entanto, os resultados de reprodutibilidade para os três instrumentos de medida utilizados para o nível de detecção da doença com a inclusão de LI corroboram com diferentes tipos de estudos, como ensaios clínicos^{19,20,21}, estudos específicos de calibração da cárie dentária^{8,14} e pesquisas de cunho epidemiológico^{22,23}. Dessa forma, pode-se afirmar que a calibração de examinadores, a partir de liames de diagnóstico com inclusão de lesões em estágios iniciais, é possível e viável para sua aplicação em diferentes tipos de estudo, mesmo em estudos epidemiológicos. Períodos adicionais de treinamento para reforço do aprendizado teórico e clínico dos critérios de diagnóstico e condições clínicas, bem como a possível utilização de recursos de diagnóstico adicionais, como a luz artificial, poderiam gerar uma melhoria nos resultados de concordância entre os examinadores, especialmente no que se refere ao diagnóstico das lesões iniciais.

Deve-se reforçar o fato de que o emprego dos instrumentos de medida deve estar condizente não só com a confiança inerente nos mesmos, como, por exemplo, com a escolha de um instrumento sofisticado como a estatística kappa, mas também deve estar de acordo com o tipo de pesquisa, com a prevalência da doença, com o grau de precisão pretendido e com os recursos disponíveis para a realização da mesma.

A estatística kappa, bem como os demais instrumentos de medida, é influenciada pela prevalência da doença, havendo mesmo a dificuldade de comparações entre estudos epidemiológicos de populações com diferentes prevalências de cárie¹⁸. Valores de kappa podem ser mais baixos para populações nas quais a prevalência de cárie é mais alta, uma vez que o número de lesões de cárie sob diferentes estágios de evolução estaria mais prevalente do

que em grupos nos quais há maior prevalência de dentes hígidos.

Estudos nos quais há necessidade de maior precisão para a medição da doença, como exemplo, em ensaios clínicos para a avaliação do efeito de medidas preventivas na remineralização de lesões iniciais, devem requerer instrumentos de medida mais eficazes para averiguação da concordância de examinadores do que em outros tipos de estudo, como, por exemplo, em estudos epidemiológicos de abrangência local, devendo, por isso, a estatística kappa ser utilizada, inclusive, com a verificação dos valores de reprodutibilidade de acordo com cada código/condição clínica²⁴.

Entretanto, a consistência dos resultados obtidos pela pesquisa não é garantida apenas pela escolha de um instrumento de medida para concordância de examinadores. Esta busca deve também estar relacionada a uma padronização das condições de exame, como local e recursos de exame, como também ao tipo de população examinada durante o processo de calibração ser similar àquela examinada durante o próprio levantamento epidemiológico.

A utilização da estatística kappa tem sido cada vez maior devido aos inúmeros pacotes estatísticos disponibilizados atualmente para o usuário. A maioria deles apenas exige que os dados sejam registrados, sendo o cálculo gerado automaticamente. Entretanto, na inviabilidade de dispor-se dessa tecnologia, a estatística kappa, apesar de apresentar um padrão mais preciso para a avaliação da reprodutibilidade, poderia ser considerada complexa para o cálculo da reprodutibilidade dos examinadores de forma manual, haja vista a exigência da construção de matrizes de diagnóstico e aplicação de fórmulas extensas, diferentemente da porcentagem geral de concordância e mesmo do índice dice que permitem o cálculo manual. Portanto, diante da dificuldade ou mesmo impossibilidade do emprego da estatística kappa, o que poderia ser indicado seria o emprego da porcentagem geral de concordância associada ao cálculo do índice dice, a fim de objetivar-se a avaliação da calibração com um direcionamento da discordância para a doença, na qual, geralmente, os erros de diagnóstico são mais frequentes.

Como forma de exemplificação, a porcentagem geral de concordância poderia ser empregada, numa primeira etapa do processo de calibração, para analisar o padrão geral de concordância entre os examinadores, e, estando a concordância dentro de expectativas preestabelecidas, por exemplo, concordância acima de 85%, segundo a OMS¹³, o índice dice seria

utilizado, posteriormente, no intuito de analisar se a causa da discordância estaria correlacionada ao diagnóstico da lesão cáriosa. Períodos de calibração adicionais poderiam ser ministrados apenas para averiguar e trabalhar sobre esse tipo de discordância, levando-se, conseqüentemente, ao maior aprimoramento final dos resultados da calibração.

A associação dessas duas ferramentas poderia ser indicada em situações cujas condições e recursos para o levantamento epidemiológico são escassos (ex: falta de computadores e programas para o cálculo do kappa, despreparo do coordenador no emprego dessa tecnologia), como também naqueles grupos/populações cujos índices de cárie (CPO-D e ceo) são especialmente devidos à presença da lesão cáriosa (componentes C e c dos respectivos índices) ^{15,16}.

Dessa forma, o emprego apropriado e bem indicado dos instrumentos de confiabilidade para a calibração de examinadores é de grande

relevância, pois geram respostas e direcionamentos quanto à capacidade e precisão dos examinadores no diagnóstico das condições bucais.

Conclusão

O índice kappa pode ser considerado a medida de concordância de escolha para o cálculo da reprodutibilidade intra/interexaminadores durante a calibração da cárie dentária, por gerar uma informação de melhor qualidade e ser um índice de concordância ajustado, levando em consideração a proporção de concordâncias atribuídas ao acaso. Entretanto, na impossibilidade de seu uso, os autores recomendam a utilização da porcentagem geral de concordância associada, posteriormente, ao índice dice, a fim de direcionar a calibração para o aprimoramento dos examinadores no diagnóstico correto das lesões cárias, nas quais os erros de diagnóstico são geralmente mais freqüentes.

Resumo

Comparar três medidas (kappa, porcentagem geral de concordância e índice dice) utilizadas para se determinar a reprodutibilidade do diagnóstico da cárie em levantamentos epidemiológicos, sob diferentes níveis de detecção da doença. Onze examinadores com experiência prévia em levantamentos epidemiológicos foram submetidos a um processo teórico-clínico de calibração. Dois níveis de detecção da cárie foram adotados: Organização Mundial da Saúde (OMS) e OMS com inclusão de lesões iniciais em esmalte (OMS + LI). Os exames foram realizados em 23 crianças de 6-7 anos de idade com e sem atividade da doença. A análise dos dados foi realizada mediante dois limiares: OMS e OMS + LI. Valores médios dos índices kappa, porcentagem geral de concordância e dice foram considerados altos ($\geq 0,90$), à exceção do índice dice para o limiar OMS + LI (0,69). O kappa, por ser um índice de concordância ajustado, pode ser considerado como instrumento de escolha para a calibração dos examinadores. Entretanto, na impossibilidade de seu uso, recomenda-se a utilização da porcentagem geral de concordância associada ao índice dice, a fim de direcionar e aprimorar os examinadores no exame das lesões cárias.

Cárie Dentária; Diagnóstico; Reprodutibilidade de Resultados

Colaboradores

A. V. Assaf contribuiu na elaboração do desenho do estudo, coleta de dados e redação do manuscrito. L. Zanin colaborou na coleta de dados e seleção da amostra. M. C. Meneghim participou na calibração de examinadores, revisão e aprovação final do texto. A. C. Pereira colaborou na elaboração do desenho do estudo, revisão e aprovação final do texto. G. M. B. Ambrosano contribuiu na análise estatística dos dados.

Agradecimentos

Os autores agradecem aos diretores, professores e estudantes das Escolas Municipais de Piracicaba, aos dentistas pela valiosa participação neste estudo, como também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela bolsa fornecida ao primeiro autor do estudo durante seu curso de doutorado na Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas.

Referências

1. Ministério da Saúde. SB Brasil 2003: Projeto – Condições de saúde oral da população brasileira em 2002-2003. http://portalweb02.saude.gov.br/portal/arquivos/pdf/relatorio_brasil_sorridente.pdf (acessado em 06/Jun/2005).
2. 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.
3. Susin C, Rösing CK. A importância do treinamento, reprodutibilidade e calibragem para a qualidade dos estudos. *Rev Fac Odontol P Alegre* 2000; 41:3-7.
4. Escamila-Cejulo JA, Pena JE, Dominguez RC, Viladomat DC. La variacion inter observador e intra observador, un problema de validez en los estudios epidemiológicos de presion arterial. *Arch Inst Cardiol Mex* 1989; 59:187-93.
5. Klein CH, Costa EA. Os erros de classificação e os resultados de estudos epidemiológicos. *Cad Saúde Pública* 1987; 3:236-49.
6. Pereira AC. Normas operacionais para execução de levantamentos em odontologia. In: Pereira AC, organizador. *Odontologia em saúde coletiva*. São Paulo: Editora Artmed; 2003. p. 83-116.
7. Pinto VG 2000. *Saúde bucal coletiva*. 4ª Ed. São Paulo: Editora Santos; 2000.
8. Fyffe 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 Dent Oral Epidemiol* 2000; 28:42-51.
9. Nuttall NM, Paul JW. The analysis of inter-dentist agreement in caries prevalence studies. *Community Dent Health* 1985; 2:123-8.
10. Pine CM, Pitts NB, Nugent ZJ. British association for the study of community dentistry (BASCD) guidance on the statistical aspects of training and calibration of examiners for surveys of child dental health. A BASCD coordinated epidemiology programme quality standard. *Community Dent Health* 1997; 14 Suppl 1:18-29.
11. Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics* 1975; 31:651-9.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.
13. World Health Organization. *Oral health surveys: basic methods*. 4th Ed. Geneva: World Health Organization; 1997.
14. Nyvad B, Machiulskiene V, Baelum V. Reliability of a new caries diagnostic system differentiating between active and inactive caries lesions. *Caries Res* 1999; 33:252-60.
15. Assaf AV, Meneghim MC, Zanin L, Pardi V, Pereira AC, Ambrosano GMB. Influence of diagnostic adjuncts to the measurement of caries prevalence at different caries thresholds. *Int Dental J* 2003; 53:485-90.
16. Assaf AV, Meneghim MC, Zanin L, Mialhe FL, Pereira AC, Ambrosano GMB. Assessment of different methods for diagnosing dental caries in epidemiological surveys. *Community Dent Oral Epidemiol* 2004; 32:418-25.
17. Peres MA, Traebert J, Marcenes W. Calibração de examinadores para estudos epidemiológicos de cárie dentária. *Cad Saúde Pública* 2001; 17:153-9.
18. Frias AC, Antunes JLF, Narvai PC. Precisão e validade de levantamentos epidemiológicos em saúde bucal: cárie dentária na cidade de São Paulo, 2002. *Rev Bras Epidemiol* 2004; 7:144-54.
19. Deery C, Care R, Chesters R, Huntington E, Stelmachonoka S, Gudkina Y. Prevalence of dental caries in Latvian 11- to 15-year-old children and the enhanced diagnostic yield of temporary tooth separation, FOTI and electronic caries measurement. *Caries Res* 2000; 34:2-7.
20. Forgie AH, Paterson M, Pine CM, Pitts NB, Nugent ZJ. A randomised controlled trial of the caries-preventive efficacy of a chlorhexidine-containing varnish in high-caries-risk adolescents. *Caries Res* 2000; 34:432-9.
21. Chesters RK, Pitts NB, Matuliene G, Kvedariene A, Huntington E, Bendinskaite R, et al. An abbreviated caries clinical trial design validated over 24 months. *J Dent Res* 2002; 81:637-40.
22. Luan WM, Baelum V, Fejerskov O, Chen X. Ten-year incidence of dental caries in adult and elderly Chinese. *Caries Res* 2000; 34:205-13.
23. Warren JJ, Steven ML, Kanellis MJ. Dental caries in the primary dentition: assessing prevalence of cavitated and non-cavitated lesions. *J Public Health Dent* 2002; 62:109-14.
24. Nyvad B, Machiulskiene V, Baelum V. Construct and predictive validity of clinical caries diagnostic criteria assessing lesion activity. *J Dent Res* 2003; 82:117-22.

Recebido em 03/Dez/2004

Versão final reapresentada em 26/Abr/2006

Aprovado em 15/Abr/2006