

Importância da calibração de examinadores para estudos epidemiológicos em odontologia

Fabíola Bof de ANDRADE¹

Roberta MATHIAS²

Maria Helena Monteiro de Barros Miotto³

Josué da Silva COUTINHO⁴

RESUMO

Palavras-chave:

Calibração de examinadores, estatística kappa, estudos epidemiológicos.

A concordância intra e entre examinadores reduz os desvios nos critérios e procedimentos de avaliação do exame epidemiológico, pois erros podem afetar a precisão do estudo. Com o intuito de demonstrar a importância da calibração inter e intra-examinadores para a validade e confiabilidade dos estudos epidemiológicos, oito estudantes de Odontologia foram distribuídos, aleatoriamente, em dois grupos: 1) calibrados e 2) não calibrados. O grupo calibrado submeteu-se a um exercício de calibração e após realizou, junto ao grupo não calibrado, exames clínicos para testar a efetividade da calibração. Verificou-se que houve diferença estatística para os valores de kappa intra ($t=6,27$; $p<0,01$) e interexaminador ($t=5,85$; $p<0,01$) entre os dois grupos e não houve diferença significativa para a consistência intra ($t=2,39$; $p=0,062$) e interexaminador ($t=0,403$; $p=0,714$) entre a primeira e a segunda etapa, para o grupo calibrado. Concluiu-se que a calibração foi efetiva e a reprodutibilidade dos diagnósticos do grupo calibrado manteve-se durante o estudo.

Data de recebimento: 5-1-2004
Data de aceite: 19-4-2004

¹ Cirurgiã-dentista formada pela Universidade Federal do Espírito Santo, pós-graduanda do curso de Especialização em Dentística Restauradora (UFES).

² Cirurgiã-dentista formada pela Universidade Federal do Espírito Santo.

³ Professora adjunta da disciplina Clínica Integrada Infantil - Odontologia - UFES; coordenadora do curso de Especialização em Saúde Coletiva - ABO-ES.

⁴ Professor adjunto da disciplina Escultura Dental - UFES; mestrando em Prótese (UFES).

INTRODUÇÃO

A Epidemiologia Clínica é a ciência que faz predições sobre pacientes individuais contando eventos clínicos em pacientes similares e usando métodos científicos sólidos, para assegurar que as predições sejam corretas (FLETCHER et al., 1996). Os levantamentos básicos de saúde bucal são usados para a coleta de informações sobre o estado de saúde e as necessidades de tratamento de uma população e, posteriormente, para monitorar as mudanças nos níveis e padrões da doença (OMS, 1997). No entanto, muitos artigos na literatura mostram que há divergências nos diagnósticos realizados por dentistas (SILVA et al., 1994; NEWBRUN, 1993) e, além disso, grande parte dos estudos não faz menção a exames de calibração e aos resultados dos testes de concordância intra e interexaminadores.

A concordância intra-examinador e entre examinadores é um processo de redução dos desvios nos critérios e procedimentos de avaliação do exame epidemiológico. O erro do examinador afeta a precisão de um procedimento diagnóstico ou de coleta de dados (OMS, 1997). Assim, para que os levantamentos sejam confiáveis e assumam o papel a que se destinam, faz-se necessário que haja um rigor metodológico para que os dados possam ser reproduzidos com validade e confiabilidade.

Segundo a OMS (1993), uma das maneiras de se reduzir os desvios nos critérios e procedimentos de avaliação do exame epidemiológico é a calibração dos examinadores, que objetiva assegurar uma interpretação, compreensão e aplicação uniformes dos códigos e critérios das várias doenças e condições a serem observadas e registradas por todos os examinadores (OMS, 1997).

A consistência intra e interexaminadores pode ser avaliada pela verificação do percentual de concordância entre escores ou por meio da estatística kappa. As vantagens reconhecidas do teste kappa são a sua capacidade de remover da estatística percentual de concordância aquelas concordâncias que provavelmente foram devidas ao acaso, trazendo uma distribuição mais global das condições dos indivíduos examinados (OMS, 1993).

Fischman et al. (1976), com o intuito de determinar validade e confiabilidade intra e interexaminador no diagnóstico da cárie dentária, utilizaram dois examinadores que foram calibrados e avaliaram anualmente, durante três anos, o Índice CPO-D de

alunos da sexta série. Os resultados mostraram que houve baixa discordância entre os examinadores e concluiu-se que a inconsistência entre os examinadores foi persistente durante os três anos do estudo, porém não significativa.

Heifelts et al. (1984) utilizaram dois investigadores experientes que examinaram 629 crianças de dez a dezessete anos para cálculo do índice CPOD, durante três anos. Ambos os examinadores adotaram os mesmos critérios escrito e visual (*slides*) para diagnóstico, mas não padronizaram ou calibraram seus métodos antes ou durante a pesquisa. Os resultados mostraram que os índices CPOD médio dos dois examinadores foram notavelmente similares, sendo 8,35 e 8,16, respectivamente. Os coeficientes de variação foram idênticos: CV = 87%. O coeficiente de confiabilidade para os dois conjuntos de dados mostrou que somente 4% da variabilidade nos índices de CPOD foram devidos à inconsistência ou a outro erro de mensuração. Os achados indicaram que, sem se submeterem a uma calibração clínica, os dois examinadores atingiram um nível alto de concordância no diagnóstico de cárie somente pela adoção dos mesmos critérios clínicos e visuais claramente definidos.

Scruggs et al. (1989) realizaram um estudo com o objetivo de identificar as diferenças entre examinadores calibrados e não calibrados, examinadores calibrados por experts e por não experts e a confiabilidade entre examinadores estudantes de odontologia e de higiene dental. Cada examinador analisou duas vezes a qualidade do selamento dental em vinte dentes. Oito examinadores, dos dezesseis participaram de uma sessão de calibração entre os exames. Verificou-se que, apesar de não haver significado estatístico, o grupo calibrado demonstrou maior confiabilidade intra e interexaminador que o grupo não calibrado. Além disso, observou-se que, durante o planejamento de uma sessão de calibração, o método utilizado para calibração dos examinadores pode ser mais relevante que a experiência do calibrador.

Cleaton-Jones et al. (1989) investigaram um método de calibração de examinadores para estudos de campo. Para a pesquisa, um total de duzentos dentes cariados foi agrupado em quarenta blocos, então, 920 superfícies foram avaliadas com uma sonda curva, usando o critério da OMS. Os dois exames realizados foram comparados, com intervalo de uma semana, por três examinadores experientes e quatro examinadores novos. O mes-

mo foi feito com os exames executados por dois examinadores experientes, porém com o intervalo de um ano. Os resultados mostraram que a concordância intra-examinador, para os exames realizados dentro do mesmo ano, foi alta para todos os examinadores. Quando foram comparados os exames, com intervalo de um ano, verificou-se que a reprodutibilidade diminuiu, porém permaneceu alta. Além disso, observou-se que a concordância interexaminador foi a que mais variou, sendo maior para os examinadores experientes do que para os não experientes.

Eaton et al. (1997) realizaram um estudo para demonstrar que é possível treinar praticantes da odontologia para realizar e manter altos níveis de consistência interexaminadores no uso de índices periodontais simples por um período de doze meses. Os examinadores foram treinados para usar o índice de placa (PII) e uma versão modificada do índice gengival (mGI) que registra somente o sangramento gengival. Inicialmente, os examinadores atingiram um índice $k = 0,78-0,85$ interexaminadores (média 0,81) para índice de placa e $k = 0,73-0,94$ (média 0,87) para mGI no exame de 168 sítios para cada variável. Durante os doze meses seguintes, os índices k individuais, tomados a cada três meses, em 42 sítios, variaram de 0,51 a 0,90 para índice de placa e de 0,73-1,00 para índice gengival. A média k para índice de placa realizado pelos cinco examinadores caiu durante o estudo de 0,81 para 0,76 enquanto, para o índice gengival, aumentou de 0,87 para 0,92. Conclui-se que é possível treinar examinadores para atingir altos níveis de consistência que se mantenham durante um período de doze meses.

Peres et al. (2001) descreveram um exercício de calibração de examinadores para estudos epidemiológicos de cárie dentária, por meio dos códigos e critérios propostos pela Organização Mundial de Saúde, em 1997, comparando testes estatísticos que medem acurácia e reprodutibilidade. Utilizando oito cirurgiões-dentistas que examinaram independentemente dezessete escolares de doze anos de idade, calcularam o Índice CPO-D médio, as taxas de concordância e o teste Kappa, considerando os dentes em conjunto e dente a dente para cada um dos examinadores, comparando-os com o examinador-padrão. O resultado do estudo mostrou 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 da reprodutibilidade.

O manual de aspectos estatísticos de treinamento e calibração de examinadores para pesquisas em saúde dental da Associação Britânica (BASCD), em 1997 sugere que a estatística deve ser utilizada para suplementar o diagnóstico clínico e permitir um julgamento livre de influências. Na maioria das vezes, os testes irão promover uma confirmação formal da exclusão de um examinador que não consegue adaptar-se aos critérios estabelecidos. Além disso, ressalta-se que a habilidade para encontrar resultados padronizados não tem a ver com a experiência clínica, pois às vezes o "melhor" examinador pode ter um dia ruim.

Considerando as colocações expostas, este estudo foi realizado com o intuito de demonstrar, por meio de um ensaio clínico aleatorizado, a importância da calibração inter e intra-examinadores para a validade e confiabilidade dos estudos epidemiológicos.

MATERIAL E MÉTODOS

Após a aprovação do projeto pelo Comitê de Ética em Pesquisa da Universidade Federal do Espírito Santo, seguiu-se a composição da amostra que foi constituída por 36 pacientes, de seis a doze anos, selecionados de forma aleatória. Dentre eles 21 participaram do exame de calibração e 15 participaram do exame clínico para comparação entre examinadores calibrados e não calibrados. Além disso, para que se pudesse testar a influência da calibração na concordância intra e interexaminador, foram selecionados dezesseis alunos, sem experiência prévia em levantamentos epidemiológicos, que estavam cursando o final do segundo período do Curso de Odontologia da Universidade Federal do Espírito Santo. Inicialmente, esses estudantes foram distribuídos, aleatoriamente, em anotadores (4 para o grupo calibrado e 4 para o grupo não calibrado) e examinadores e depois uma nova distribuição, aleatória, dividiu o grupo de examinadores em 2 grupos: 1) examinadores calibrados; 2) examinadores não calibrados. A amostra obedeceu aos seguintes critérios de inclusão e exclusão:

- critérios de inclusão: pacientes de seis a doze anos, frequentadores da Clínica Integrada Infantil da Universidade Federal do Espírito Santo;
- critérios de exclusão: estudantes que estivessem participando de forma direta ou indireta na elaboração da pesquisa; pacientes que possuíssem

qualquer aparelhagem ortodôntica fixa ou removível; pacientes que apresentassem qualquer enfermidade geral que comprometesse a identificação dos dados necessários ou tivessem desvio de comportamento.

O estudo foi dividido em duas etapas: a primeira destinada à calibração intra e interexaminadores, na qual participaram o grupo de examinadores calibrados e quatro anotadores; e a segunda objetivando comparar os examinadores calibrados e não calibrados, realizada em uma única fase com os dois grupos (calibrados e não calibrados).

Critérios de diagnóstico para cárie dental

Os exames foram realizados no ambulatório da Faculdade de Odontologia com os pacientes deitados nas cadeiras odontológicas, posicionadas de forma que eles ficassem completamente deitados. Os exames foram realizados com a iluminação ambiente, pois não foi permitida a utilização da luz do equipamento. Os examinadores permaneceram posicionados atrás da cabeça do paciente na posição de doze horas e o anotador na posição de nove horas. Utilizou-se, para os exames, espelho clínico, sondas e gaze previamente esterilizados.

Primeira Etapa: calibração intra e interexaminadores

Esta etapa foi subdividida em quatro outras etapas que ocorreram com uma semana de intervalo da outra:

1. Introdução aos procedimentos de exame clínico: etapa teórica com quatro horas de duração, quando foram apresentados os códigos e critérios do Índice CPO-D, segundo a OMS (1997), seguida da apresentação de 28 diapositivos ilustrando as várias condições a serem observadas durante o exame. Cada diapositivo foi projetado por aproximadamente um minuto, tendo sido solicitado aos examinadores que fizessem o diagnóstico, de acordo com os critérios previamente apresentados. Nesta etapa, também foram apresentados os instrumentais a serem usados e as fichas clínicas.

2. Exame clínico pré-teste: dedicado ao exercício clínico para que os examinadores pudessem fixar os critérios apresentados. Cada participante examinou seis pacientes. Durante o exercício, foi permitido e aconselhado que o examinador e o anotador discutissem entre si e com os outros, a fim de alcançar um nível adequado de concordância. Nesta etapa foram observadas as diferenças de concordância entre os examinadores e discutiram-se as divergências até que se julgou que os critérios haviam sido fixados pela equipe.

3. Exame clínico efetivo de concordância: cada examinador avaliou, individualmente, quinze pacientes e os examinadores e anotadores não puderam discutir os seus achados clínicos com os outros. Após cada exame individual a ficha clínica foi entregue ao coordenador do estudo e iniciou-se o reexame dos quinze pacientes.

4. Discussão final: esta etapa teve início depois de verificada a calibração dos examinadores por meio do teste kappa. Enfatizou-se a necessidade de se atingir uma boa concordância e também, foram reforçados os códigos e critérios que mais geraram confusão.

Segunda Etapa: exame clínico para comparação entre examinadores calibrados e não calibrados

Uma semana antes da realização do exame clínico para comparação entre os grupos, os examinadores não calibrados receberam um formulário contendo todo o protocolo para diagnóstico da cárie dental.

Todos os examinadores, contando com anotadores, examinaram duas vezes quinze pacientes e não foi permitido que houvesse qualquer tipo de discussão entre os participantes.

Por razões éticas, todos os pacientes que apresentaram necessidade de tratamento foram encaminhados para atendimento na própria instituição. Além disso, os responsáveis pelos pacientes assinaram um termo de consentimento livre e esclarecido permitindo a participação do menor na pesquisa.

Apuração e análise dos dados

Todos os dados obtidos durante a primeira e segunda etapa foram submetidos a testes estatísticos para análise de sua validade e reprodutibilidade. Utilizou-se o teste kappa para verificar a concordância intra e interexaminador. É importante ressaltar que, antes da realização da estatística, estabeleceu-se o intervalo kappa entre 0,61 – 0,80 como sendo um valor satisfatório em termos de concordância, considerando-se a escala proposta por Landis e Koch (1977) que relatam ser uma concordância substancial.

RESULTADOS

As Tabelas 1 e 2 mostram os valores de kappa para a concordância intra-examinador e interexaminadores durante a primeira etapa destinada à calibração dos examinadores e durante a

segunda etapa destinada à comparação entre o grupo calibrado e o não calibrado.

Nota-se na Tabela 1 que os valores correspondentes à primeira etapa (calibração dos examinadores) estão compreendidos entre 0,88 e 0,92, indicando uma reprodutibilidade quase perfeita, segundo a tabela de Landis e Koch (1977), ocorrendo o mesmo para a segunda etapa (kappa entre 0,87 a 0,93). Na Tabela 2, verifica-se que os valores de concordância interexaminadores diminuíram da primeira etapa (0,79-0,91) para a segunda etapa (0,71-0,83)

Tabela 1 - Índice Kappa intra-examinador obtido durante a primeira etapa

Primeira etapa		Segunda etapa	
Examinador	Kappa	Examinador	Kappa
A x A	0,91	A x A	0,91
B x B	0,94	B x B	0,87
C x C	0,88	C x C	0,93
D x D	0,92	D x D	0,90

Tabela 2 - Índice Kappa interexaminador obtido durante a primeira etapa

Primeira etapa		Segunda etapa	
Examinador	Kappa	Examinador	kappa
A x B	0,91	A x B	0,73
A x C	0,78	A x C	0,82
A x D	0,85	A x D	0,79
B x C	0,80	B x C	0,74
B x D	0,85	B x D	0,71
C x D	0,88	C x D	0,83

As Tabelas 5 e 6 mostram os resultados da comparação entre o grupo calibrado e o não calibrado para os valores de Kappa intra-examinador e interexaminador, respectivamente.

Tabela 5 - Estatística descritiva do índice Kappa intra-examinadores segundo grupos

Grupo	Média	Desvio-padrão	Coeficiente de variação	Intervalo de confiança 95%	
				LI	LS
Não calibrados	0,817	0,017	2,08%	0,790	0,844
Calibrados	0,912	0,0006	0,06%	0,872	0,952

Tabela 6 - Estatística descritiva do índice Kappa interexaminadores segundo grupos

Grupo	Média	Desvio-padrão	Coeficiente de variação	Intervalo de confiança 95%	
				LI	LS
Não calibrados	0,621	0,036	5,79	0,583	0,660
Calibrados	0,770	0,050	6,49	0,717	0,822

Nas Tabelas 3 e 4, observam-se os valores de Kappa intra-examinador e interexaminador, respectivamente, para os examinadores não calibrados obtidos durante a segunda etapa.

Tabela 3 - Índice Kappa intra-examinador para o grupo não calibrado obtido durante a segunda etapa

Segunda etapa	
Examinador	Kappa
E x E	0,80
F x F	0,84
G x G	0,81
H x H	0,82

Tabela 4 - Índice Kappa interexaminador para o grupo não calibrado obtido durante a segunda etapa

Segunda etapa	
Examinador	Kappa
E x F	0,61
E x G	0,58
E x H	0,62
F x G	0,59
F x H	0,66
G x H	0,67

A partir da Tabela 3, verifica-se que os valores 0,80-0,84 para concordância intra-examinador foram altos, em sua maioria, sendo compatíveis com uma concordância quase perfeita. Por outro lado, para concordância interexaminador, os valores indicaram concordância moderada (0,41-0,60 segundo a tabela de Landis e Koch, de 1977) a substancial, estando entre 0,58 e 0,67.

O Teste t de independência mostrou diferença estatística entre não calibrados e calibrados, quanto à concordância intra-examinador ($t=6,27$; $p<0,01$) e interexaminador ($t=5,85$; $p<0,001$).

A comparação dos valores de kappa intra e interexaminador entre a primeira etapa e a segunda etapa está apresentada nas Tabelas 7 e 8. Com relação à concordância intra-examinador, verificou-se um aumento da consistência dos dados da pri-

meira etapa (CV=2,74%) para a segunda (CV=0,06%). Porém, para a concordância interexaminador, observou-se uma maior variação dos valores de kappa em relação à média da primeira para a segunda etapa, todavia mantendo-se a consistência (CV=5,67%-CV=6,49%).

Tabela 7 - Estatística descritiva do índice Kappa intra-examinadores segundo grupos

Grupo calibrado	Média	Desvio-padrão	Coeficiente de variação	Intervalo de confiança 95%	
				LI	LS
Primeira etapa	0,912	0,025	2,74%	0,872	0,952
Segunda etapa	0,912	0,0006	0,06%	0,872	0,952

Tabela 8 - Estatística descritiva do índice Kappa interexaminadores segundo grupos

Grupo calibrado	Média	Desvio-padrão	Coeficiente de variação	Intervalo de confiança 95%	
				LI	LS
Primeira etapa	0,845	0,048	5,67%	0,794	0,895
Segunda etapa	0,770	0,05	6,49%	0,717	0,822

O Teste t pareado ($t=0,403$; $p=0,714$) mostrou não haver diferença estatística para os valores de Kappa dos examinadores calibrados entre a primeira e a segunda etapa, com relação à concordância intra-examinador, ocorrendo o mesmo para a concordância interexaminador ($t=2,39$; $p=0,062$).

DISCUSSÃO

Segundo Peres et al. (2001), para se obter a padronização no uso de critérios diagnósticos para as diversas doenças bucais, é fundamental que os examinadores envolvidos no estudo participem de um treinamento antes do início do trabalho. Portanto, uma etapa necessariamente prévia a todos os estudos epidemiológicos é a calibração dos examinadores. Dessa forma, obtém-se maior acurácia e reprodutibilidade do estudo. A acurácia é o grau pelo qual o instrumento de coleta de dados mede o que se pretende e a reprodutibilidade está relacionada com a capacidade de um instrumento alcançar a mesma medida em tempos diferentes (FLETCHER et al., 1996).

Foi possível verificar, tanto no grupo calibrado quanto no grupo não calibrado, que existiu uma diferença de diagnóstico intra e interexaminador,

como sugeriu Pereira et al. (2003), ao afirmar que, quando um estudo é realizado por uma equipe, a priori, deve-se esperar que os examinadores interpretem e apliquem os critérios para a definição do caso (diagnóstico positivo ou negativo) de forma diferente. Com relação à reprodutibilidade dos dados, verificou-se que todos os examinadores do grupo calibrado alcançaram valores compatíveis com uma boa concordância, na primeira e na segunda etapa, valores estes que podem ser tolerados de acordo com Cleaton-Jones et al. (1989).

Observou-se que houve consistência intra-examinador em relação às médias kappa para os dois grupos, calibrados e não calibrados (CV= 0,06% e CV = 2,08%, respectivamente), no entanto maior variação ocorreu para o grupo não calibrado. Considerando-se a importância dada à etapa da calibração para diminuir os erros de concordância (HOWAT; CANNELL, 1979; OMS, 1997), os achados deste estudo surpreenderam pelos altos valores intra-examinadores encontrados para o grupo não calibrado (Tabela 3).

No entanto, a aplicação do Teste t de independência para a análise dos dados comparativos entre examinadores calibrados e não calibrados mostrou haver diferença estatisticamente significativa tanto para a concordância intra-examinador quanto

interexaminador ($t=6,27$; $p<0,01$ e $t=5,85$; $p<0,01$, respectivamente). A partir desses achados, pode-se afirmar que a calibração teve efeito positivo para a melhoria da concordância.

Além disso, o grupo calibrado revelou, em ambas as situações, uma média do índice kappa mais elevada (Tabelas 5 e 6), concordando com BASCD (1997) e com a OMS (1997) que afirmam que a calibração é um processo de redução dos desvios nos critérios e procedimentos de avaliação do exame epidemiológico. Por outro lado, Heifetz et al. (1985), sabendo da efetividade dos métodos para calibrar examinadores, mostraram-se surpresos ao verificarem, em seu estudo, que os examinadores obtiveram um alto grau de concordância quanto ao diagnóstico da cárie dental sem, ao menos, passarem por um treinamento prévio.

Da mesma forma, Scruggs et al. (1989) observaram ausência de significância estatística quando compararam a concordância entre examinadores calibrados e não calibrados e afirmaram que a calibração promoveu, apenas, um pequeno efeito sobre a reprodutibilidade intra-examinador.

Segundo a escala de Landis e Koch (1977), uma reprodutibilidade excelente é definida por um valor de kappa maior ou igual a 0,81, entretanto Cleaton-Jones et al. (1989) relataram que valores de kappa iguais ou maiores a 0,61 deveriam ser aceitos para estudos epidemiológicos da cárie. Neste estudo, obteve-se uma reprodutibilidade intra-examinador excelente, para todos os examinadores calibrados, nas duas etapas. Porém, para a concordância interexaminador, verificou-se que, apesar da prevalência de valores excelentes na primeira etapa, houve uma queda de valores na segunda etapa, prevalecendo a concordância substancial. Todavia, a aplicação do teste t pareado demonstrou não haver diferença estatisticamente significativa entre os valores de kappa intra ($t=0,403$; $p=0,714$) e interexaminadores ($t=2,39$; $p=0,062$) durante a etapa de calibração e o exercício de comparação entre os grupos, comprovando que a reprodutibilidade foi mantida, coincidindo com os resultados encontrados por Fischman et al. (1972) os quais observaram que a inconsistência entre examinadores foi persistente durante os três anos do estudo, embora apresentando ausência de significância estatística.

Eaton et al. (1997) concluíram ser possível treinar praticantes de odontologia para alcançar e manter altos níveis de consistência interexaminador.

Todavia, a BASCD (1997) ressaltou que variações na concordância intra e interexaminadores podem ocorrer ao longo de um estudo. Assim, relatou que, uma vez incluído em uma pesquisa, o examinador não deve ser excluído do exercício. Uma recalibração anual é necessária para assegurar que o examinador não modifique sua interpretação.

CONCLUSÃO

Com base na análise estatística dos dados, pode-se concluir que

- houve diferença estatisticamente significativa para os valores de kappa intra e interexaminadores entre o grupo calibrado e o não calibrado, obtendo o grupo calibrado uma média de concordância maior;
- não houve diferença estatisticamente significativa para os valores de kappa intra e interexaminadores entre a primeira e segunda etapa do estudo, verificando-se a manutenção da validade e reprodutibilidade dos dados.

ABSTRACT

IMPORTANCE OF THE CALIBRATION OF EXAMINERS FOR EPIDEMIOLOGICAL STUDIES IN DENTISTRY

The intra and interexaminer agreement decrease the deviations in diagnostic criteria and in the evaluations procedures because they can affect the study precision. The purpose of this study was to demonstrate the importance of the intra and interexaminer calibration for the validity and reliability of the epidemiological studies. Eight dental students were randomly allocated into two groups: 1) calibrates and 2) non-calibrated. The calibrated group underwent a calibration period and then, with the non-calibrated group performed a clinical examination with the purpose to test the benefit of a calibration program. The findings indicated that, there was a significant difference in intra ($t=6,27$; $p<0,01$) and interexaminer ($t=5,85$; $p<0,01$) reliability between the two groups and there were no statistically significant differences in intra ($t=2,39$; $p=0,062$) and interexaminer ($t=0,403$; $p=0,714$) reliability, for the calibrated group, between the first and second stage of the study. It was concluded that: 1) the calibration

was effective and 2) the reliability of the calibrated group was kept during the study.

Key words: Calibration of examiners, kappa statistics, epidemiological studies.

REFERÊNCIAS

- 1 British Association for Study of Community Dentistry (BASCD) Guidance on the Statistical Aspects of Training and Calibration of Examiners for Surveys of Child Dental Health. **Community Dental Health**, v. 14, n. 1, p. 18-29, 1997.
- 2 CLEATON-JONES, P. et al. Dental caries diagnosis calibration for clinical field surveys. **Caries Res**, v. 23, n. 3, p.195-199, 1989.
- 3 EATON, K.A. et al. The achievement and maintenance of inter-examiner consistency in the assessment of plaque and gingivitis during a multicentre study based in general dental practices. **J. Clin. Periodontol.**, v. 24, n. 3, p. 183-188, 1997.
- 4 FISCHMAN, S. et al. Examiner standardization for caries Studies. **J. Dent. Res.**, v. 55, n. 6, p. 926-929, 1979.
- 5 FLETCHER, R. H.; FLETCHER, S. W.; WAGNER, E. H. **Epidemiologia clínica: elementos essenciais**. 3. ed. Porto Alegre: Artes Médicas, 1996.
- 6 HEIFELTZ, S. B. et al. Examiner consistency and group balance at baseline of a caries clinical trial. **Community Dent. Oral Epidemiol.**, v. 13, n. 2, p. 82-85, 1985.
- 7 LANDIS, J. R.; KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n.1, p. 159-174, Mar. 1977.
- 8 NEWBRUN, E. Problems in caries diagnosis. **Int. Dent. J.**, v. 43, p. 133-142, 1993.
- 9 PERES, M. A.; TRAEBERT, J.; MARCENES, W. Calibração de examinadores para estudos epidemiológicos de cárie dentária. **Cad. Saúde Pública**, v. 17, n. 1, jan./fev. 2001.
- 10 PEREIRA et al. Normas operacionais para execução de levantamentos em odontologia. In: _____. **Odontologia em saúde coletiva: planejando ações e promovendo saúde**. Porto Alegre: Artmed, 2003. cap. 5, p. 83-116.
- 11 SILVA, B. B.; MALTZ, M.; FRANCO, F. Diagnóstico e tratamento da cárie de superfície oclusal: variação entre examinadores. **Rev. APCD**, v. 48, n. 1, p. 1231-1234, jan. /fev. 1994.
- 12 SCRUGGS, R. R. et al. Effects of specific criteria and calibration on examiner reliability. **J. Dent. Hygien.**, v. 63, n. 3, p. 125-129, Mar./Apr. 1989.
- 13 WORLD HEALTH ORGANIZATION. **Oral health surveys: Basic methods**. 4. ed. Geneva: ORH/EPID, 1997.
- 14 WORLD HEALTH ORGANIZATION. **Calibration of examiners for oral health epidemiological surveys**. Geneva: ORH/EPID, 1993.
- 15 HOWAT, A.P. Value of operating auxiliary personnel for caries diagnosis in dental epidemiology. **Community Dent Epidemiol**, v. 7, n. 3, p. 165-169, 1979.

Correspondência para/ Reprint request to:
Fabiola Bof de Andrade
 R. Waldir Dutra de Freitas, 120 - Mata da Praia
 Vitória/ES - CEP: 29066-280
 Tel.: (27) 3324-4941 e-mail: ffob@terra.com.br