

Desafios de um exercício de calibração para estudo epidemiológico envolvendo variáveis quantitativas e categóricas ordinais: um exemplo

Challenges of a calibration process for an epidemiological study involving quantitative and ordinal categorical variables: one example

Andréa Maria Eleutério de Barros Lima Martins¹, Marise Fagundes Silveira², Carolina Vieira de Freitas³, Núbia Barbosa Eleutério⁴, Pablo Henrique Ataíde Oliveira⁵, Raquel Conceição Ferreira¹

RESUMO

Objetivo: Relatar os desafios de um exercício de calibração de examinadores para estudo epidemiológico sobre maloclusão envolvendo variáveis quantitativas e categóricas ordinais, apresentando os resultados da consistência diagnóstica intra e interexaminadores. **Materiais e Métodos:** Voluntários foram selecionados antecipando condições que os examinadores encontrariam em campo. Maloclusão foi avaliada pelo DAI, cujos critérios foram expostos no treinamento teórico e reforçados no prático. Os examinadores foram divididos em grupos e os dados coletados por exame. As concordâncias inter e intraexaminadores foram estimadas pelo CCI (Coeficiente de Correlação Intraclasse) para o DAI geral e o CCI ou Kappa ponderado (K_w) para os componentes do DAI, utilizando-se os programas PASW e *Excel*. Os examinadores com concordância inferior a 0,60 foram considerados inaptos e deveriam ser submetidos a mais um exercício de calibração. **Resultados:** Em três exercícios de calibração, demorados e dispendiosos, 33 examinadores em sete grupos, examinaram e re-examinaram 315 escolares. As concordâncias interexaminadores para o DAI geral, entre os pares de examinadores, variaram de -0,9 a 0,99 (pobre à excelente). A concordância de todos os examinadores em conjunto, por grupo, variou de 0,51 a 0,96, (razoável a excelente). Na análise por componentes, concordâncias entre os pares de examinadores $>0,60$ nos sete grupos foram encontradas para dentição, apinhamento e diastema. Discordâncias entre pares de examinadores para alguns dos componentes foram observadas mesmo em grupos cuja concordância entre o mesmo par de examinadores foi satisfatória para DAI geral. Pela concordância intra-examinadores para o DAI geral, de 0,37 a 1 (pobre a excelente), dois examinadores foram inaptos; na avaliação por componentes do DAI, 11 apresentaram CCI e $K_w < 0,60$. **Conclusão:** A calibração de examinadores configura-se como um desafio para os pesquisadores demandando tempo e esforço. A concordância obtida por componentes do DAI identifica maior número de examinadores inaptos, garantindo assim maior acurácia aos resultados do levantamento epidemiológico.

Descritores: Avaliação. Epidemiologia. Reprodutibilidade dos testes.

INTRODUÇÃO

A padronização de critérios diagnósticos para estudos epidemiológicos em saúde bucal foi proposta pela Organização Mundial da Saúde (OMS) em 1991¹ e constitui um pré-requisito para avaliar a confiabilidade dos dados coletados². Entretanto, algum grau de variação ou de erros de diagnóstico pode ocorrer quando vários avaliadores examinam muitos indivíduos, situação comum

nesses levantamentos. Dessa forma, antes da coleta dos dados, os examinadores devem participar de um treinamento e de exercício de calibração para padronização dos critérios de diagnósticos das doenças^{3,4}, minimizando os erros de observação e permitindo a mensuração dos mesmos^{1,5}.

A calibração consiste na repetição de exames nas mesmas pessoas por diferentes examinadores, ou pelo mesmo examinador nas mesmas pessoas

¹Departamento de Odontologia, Universidade Estadual de Montes Claros (Unimontes), Montes Claros, MG, Brasil

²Departamento de Matemática, Universidade Estadual de Montes Claros (Unimontes), Montes Claros, MG, Brasil

³Cirurgião-dentista

⁴Residente Multiprofissional em Saúde da Família, Universidade Estadual de Montes Claros (Unimontes), Montes Claros, MG, Brasil

⁵Matemático

Contatos: martins.andreambl@gmail.com, ciaestatistica@yahoo.com.br, carolinavfreitas@yahoo.com.br, nubiaeuterio@yahoo.com.br, pablomoc@globo.com, ferreira_rc@hotmail.com

em diferentes momentos, para que as diferenças de interpretação nos diagnósticos sejam discutidas até se firmar um consenso^{3,6,7}. É definida como o “processo que visa estabelecer padrões uniformes para o exame epidemiológico em saúde bucal e determina parâmetros aceitáveis de consistência interna e externa aos examinadores”, tendo como objetivos: assegurar uma interpretação, entendimento e aplicação uniformes dos critérios para as doenças e condições a serem observadas e registradas; assegurar que cada examinador possa examinar dentro de um padrão consistente e minimizar variações entre os diferentes examinadores³. Ao final da calibração espera-se obter concordâncias intra e interexaminadores satisfatórias. A concordância interexaminador é necessária para que os resultados do mesmo grupo de indivíduos sejam comparados entre si quando examinados por equipes diferentes. O objetivo não é estabelecer quem está certo e quem está errado, mas verificar quais examinadores estão diferindo, reduzindo a variabilidade entre eles. Os examinadores que não atingirem níveis de concordância satisfatórios devem ser novamente treinados e calibrados ou então excluídos do levantamento⁶. Dentre as estatísticas utilizadas para avaliar a consistência intra e interexaminadores destacam-se os coeficientes Kappa (simples e ponderado) e de correlação intra-classe (CCI), com métodos para cálculo específicos.


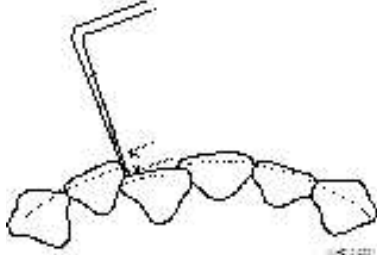
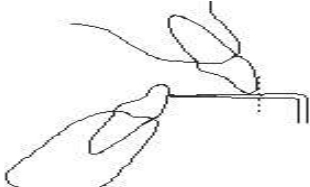
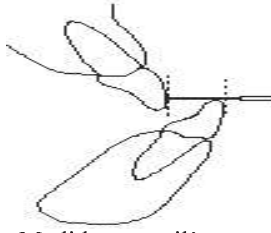
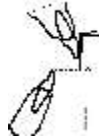
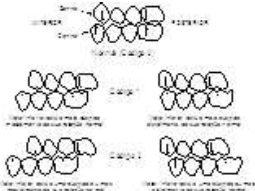
Dentre as diversas condições identificadas em levantamentos epidemiológicos, destaca-se a maloclusão, considerada um problema de saúde pública, pois apresenta alta prevalência e pode interferir negativamente na qualidade de vida dos indivíduos acometidos^{8,9}. Essa alteração, do ponto de vista de diagnóstico, não é classificada como uma doença, sendo de difícil definição¹⁰. A Organização Mundial da Saúde preconiza o levantamento das maloclusões pelo Índice de Estética Dental (Dental Aesthetic Index - DAI) em adolescentes de 12 anos e de 15 a 19 anos^{5,11}. O princípio básico do DAI é de uma combinação de medidas as quais, em seu conjunto, expressam o estado oclusal do indivíduo e, consequentemente, sua necessidade de tratamento, devido à composição do índice que considera comprometimento estético além da oclusão¹²⁻¹⁴. São consideradas três dimensões a partir de dez características oclusais de acordo com critérios definidos¹⁵, incluindo variáveis categóricas ordinais e quantitativas¹⁶. A estimativa é feita a partir de uma equação de regressão logística que relaciona, matematicamente, a gravidade da maloclusão a partir de medidas físicas objetivas de características

occlusais associadas à maloclusão^{13,17}. Para calcular seu valor, essas características são medidas e os valores a elas atribuídos são multiplicados pelos seus respectivos coeficientes (pesos) e os resultados são totalizados com a adição de uma constante, conforme equação de regressão para o cálculo do valor do DAI (Quadro 1). O resultado, um valor numérico entre treze a, aproximadamente, oitenta pode ser considerado uma escala de intervalo, que pode ser categorizado através de pontos de corte: ausência de anormalidade ou maloclusão leve, cujo tratamento ortodôntico é desnecessário (DAI < 25), maloclusão definida, cujo tratamento é eletivo (DAI = 26 a 30), maloclusão grave, cujo tratamento é altamente desejável (DAI = 31 a 35) e maloclusão muito grave, cujo tratamento ortodôntico é fundamental (DAI > 36)^{11,15,17}.

O DAI tem sido utilizado em diferentes populações^{7,9,11,14,18-22}. Embora a importância do treinamento e da calibração de examinadores seja reconhecida, em alguns deles não há relato dessa etapa da pesquisa¹⁹. Nos estudos sobre maloclusão empregando o DAI, verificou-se uma variabilidade nos procedimentos estatísticos para a estimativa da concordância entre os examinadores, sendo observados o uso da estatística Kappa^{9,14}, do coeficiente de correlação de Pearson²⁰, da concordância geral²¹ e do CCI^{18,22}. No entanto, nenhum desses estudos refere-se especificamente à calibração, limitando-se a apresentarem os valores de concordância, sem descrever com detalhes os procedimentos para atingi-los.

A calibração dos examinadores para aferição da maloclusão é uma etapa complexa, apesar de indispensável. Pode-se avaliar a confiabilidade dos dados referentes à maloclusão a partir do valor total do DAI, chamado nesse artigo de DAI geral, e/ou de seus componentes. Em publicações científicas ou relatórios de estudos oficiais, a apresentação da quantificação das divergências de diagnóstico ocorridas durante o estudo é recomendável, pois facilita o julgamento do leitor quanto à confiabilidade dos resultados⁶. Nesse contexto, buscou-se relatar os desafios enfrentados em um exercício de calibração de examinadores, anterior a um estudo epidemiológico sobre a maloclusão a partir do DAI, variável quantitativa, e de seus componentes, variáveis quantitativas ou categóricas ordinais, apresentando as principais utilizações e propriedades de aplicação de testes estatísticos que buscam medir a consistência diagnóstica intra e interexaminadores, assim como seus resultados.

Quadro 1: Dimensões, características e critérios de diagnóstico do Dental Aesthetic Índice – DAI.

Dimensão	Características	Código/critério	Nota / Exemplo
Dentição	Condição da Dentição	Número de incisivos, caninos e pré-molares perdidos que causam problemas estéticos no arco superior e no arco inferior.	Não considerar perdidos quando o respectivo espaço estiver fechado, o decíduo ainda estiver em posição, ou apresentas prótese(s).
Espaço	Apinhamento no segmento incisal (Segmento de canino a canino)	00 - Sem apinhamento 01 - Apinhamento em um segmento 02 - Apinhamento em dois segmentos	Há dentes com giroversão ou mal posicionados no arco. Não se considera quando incisivos estão alinhados e um ou ambos os caninos estão deslocados.
	Espaçamento no segmento incisal	00 - Sem espaçamento 01 - Espaçamento em um segmento 02 - Espaçamento em dois segmentos	
	Diastema Incisal	Espaço, em milímetros, entre os dois incisivos centrais superiores permanentes, quando estes perdem o ponto de contato	
	Desalinhamento maxilar anterior Desalinhamento mandibular anterior	Giroversões ou deslocamentos em relação ao alinhamento normal entre incisivos, registra-se a maior irregularidade. A medida é feita, em mm, com a sonda CPI, cuja ponta é posicionada sobre a superfície vestibular do dente posicionado mais para lingual, num plano paralelo ao plano oclusal e formando um ângulo reto com a linha do arco. Pode ocorrer com ou sem apinhamento	
Oclusão	Overjet maxilar Anterior	A relação horizontal entre os incisivos é medida com os dentes em oclusão cêntrica, utilizando-se a sonda CPI, posicionada em plano paralelo ao plano oclusal. O overjet é a distância, em mm, entre as superfícies vestibulares do incisivo superior mais proeminente e do incisivo inferior correspondente.	 Medida, em milímetros
	Overjet mandibular anterior	Caracterizado quando algum incisivo inferior se posiciona anteriormente ou por vestibular em relação ao seu correspondente superior. A protrusão mandibular, ou mordida cruzada, é medida com a sonda CPI e registrada em milímetros. Desconsideradas as situações em que há giroversão de incisivo inferior, com apenas parte do bordo incisal em cruzamento	 Medida, em milímetros
	Mordida aberta vertical anterior	Se há falta de ultrapassagem vertical entre incisivos opostos caracteriza-se uma situação de mordida aberta. O tamanho da distância entre os bordos incisais é medido com a sonda CPI e o valor, em mm, registrado no campo correspondente	 Medida, em milímetros
	Relação molar antero-Posterior	0 – Normal 1 – Meia Cúspide. O 1º molar inferior está deslocado meia cúspide para mesial ou distal, em relação à posição normal. 2 – Cúspide Inteira. O 1º molar inferior está deslocado uma cúspide para mesial ou distal, em relação à posição normal	

Fonte: Projeto SB Brasil 2002/2003

MATERIAIS E MÉTODOS

A calibração dos examinadores foi conduzida em cinco etapas: seleção de voluntários, treinamento teórico, treinamento prático, coleta de dados e cálculo da concordância. Contou com a participação de quatro instrutores (cirurgiões-dentistas e pesquisadores com experiência em levantamentos epidemiológicos).

Inicialmente, o número de examinadores necessários para coleta de dados foi definido a partir do tamanho da amostra a ser examinada no Levantamento Epidemiológico das condições de Saúde Bucal do município de Montes Claros, MG – Projeto SBMOC: 4852 indivíduos. Nesse sentido, estimou-se a necessidade de treinamento e calibração de quinze equipes de campo (um cirurgião-dentista da rede municipal de saúde que era o examinador, um acadêmico e/ou auxiliar de saúde bucal da rede municipal de saúde que eram anotadores). Para isso, foram convidados vinte profissionais da rede municipal de saúde, que atuariam como examinadores. Para estimativa da concordância, os examinadores foram divididos em grupos com o máximo de cinco indivíduos em função da previsão de utilização do programa de computador desenvolvido pelo Ministério da Saúde Brasileiro para condução do SB Brasil 2002/2003⁷. Contou ainda com a participação de uma estatística e de dez acadêmicos da graduação em matemática para construção do banco de dados e cálculos das concordâncias.

Seleção de voluntários: Adolescentes voluntários foram selecionados por cirurgiões-dentistas da rede municipal de saúde, apresentando diferentes maloclusões, visando antecipar as

condições que os examinadores encontrariam em campo. A estimativa do número de voluntários a serem examinados considerou que cada voluntário fosse examinado no máximo cinco vezes, com o intuito de se evitar o cansaço e desconforto dos mesmos. Estimou-se, portanto, a necessidade de seleção de 15 voluntários para cada grupo de cinco examinadores, totalizando sessenta para o treinamento prático, outros sessenta para a coleta de dados e caso não fossem alcançados níveis de concordância superiores a 0,60^{23,24} seriam necessários mais sessenta. Enfim, foi estimada a necessidade de 180 voluntários.

Treinamento teórico e prático: O treinamento teórico consistiu na exposição dos critérios de diagnóstico do DAI⁵. No treinamento prático, os critérios de diagnóstico adotados foram reforçados para fixação dos mesmos e dúvidas a respeito desses critérios foram esclarecidas. Na fase de exposição teórica e treinamento prático dos examinadores houve ampla discussão dos critérios adotados para que, durante a coleta dos dados, prevalecesse o consenso da equipe sobre o critério individual de cada examinador, mesmo que contrário às suas convicções pessoais⁷. Após os exames, foram estimadas as concordâncias interexaminadores para identificar os componentes do DAI que poderiam comprometer o processo de calibração, incrementando as discussões e esclarecimentos.

Coleta de dados: A coleta de dados para o cálculo da concordância foi idealizada mediante exame de outros 60 adolescentes voluntários que seriam examinados e reexaminados num intervalo de oito dias, para estimativa das concordâncias interexaminadores e intra-examinadores.

kappa ponderado (K_w)

$$K_w = \frac{p_e(w) - p_e(w)}{1 - p_e(w)}$$

$$p_e(w) = \frac{1}{n} \sum_{i,j=1}^c n_{ij} w_{ij}$$

$$p_e(w) = \frac{1}{N} \sum_{i=1}^c S_i w_i$$

n_{ij} = frequência na i -ésima linha e j -ésima coluna;

S_i = somatório das frequências na i -ésima linha;

s_j = somatório das frequências na j -ésima coluna;

c = nº de categorias da variável;

w_{ij} = peso da i -ésima linha e j -ésima coluna proposto por Cicchetti.

Coefficiente de Correlação Intra-Classe (CCI)

$$CCI = \frac{\sigma_x^2}{\sigma_x^2 + \chi_{c-2}} = \frac{MSB - MSW}{MSB + (K-1)MSW}$$

$$CCI = \frac{\frac{MSB}{MSW} - F_{N-1, K-N, \alpha}}{\frac{MSB}{MSW} + (K-1)F_{N-1, K-N, \alpha}}$$

σ_x^2 = variância das médias

χ_{c-2} = média das variâncias.

MSB = quadrado médio entre os objetos medidos

MSW = quadrado médio intra-objetos medidos

K_0 = número de medidas realizadas em cada objeto

$N-1$ e $K-N$ = graus de liberdade da distribuição F correspondente à proporção α na cauda superior.

Cálculo da concordância: Os cálculos foram conduzidos empregando-se os *softwares* PASW® (*Predictive Analytics Software*) versão 17.0 *for Windows e Excel*®.

Foram conduzidos cálculos estatísticos considerando as propostas enumeradas a seguir:

Os coeficientes de correlação intraclasse (CCI) foram calculados para a estimativa da concordância interexaminadores para o diagnóstico do DAI geral para os pares de examinadores por grupo: AxB, AxC, AxD, AxE, BxC, BxD, BxE, CxD, CxE e DxE. A concordância interexaminador total, considerando todos os examinadores em conjunto, foi também estimada para cada um dos grupos. Além disso, a concordância entre os pares de examinadores foi estimada em cada grupo, para cada um dos componentes do DAI. A concordância intra-examinadores foi calculada para todos os examinadores, considerando o DAI geral e cada um dos seus componentes separadamente. A concordância para os componentes categóricos ordinários do DAI (apinhamento, espaçamento, relação molar ântero posterior) foi medida pelo kappa ponderado (k_w) e para os quantitativos (dentição, diastema, desalinhamento maxilar, desalinhamento mandibular, overjet maxilar, overjet mandibular, mordida aberta vertical anterior) pelo CCI^{25,26}. Os valores de kappa podem variar de -1 a +1, no caso de concordância completa, $k = +1$; quando a concordância observada é maior ou igual à esperada pela chance (ao acaso), $k \geq 0$ e se a concordância observada for menor ou igual a esperada pela chance, $k \leq 0$. Os valores de CCI podem variar de zero a +1, valor igual a zero significa que as medidas são pouco confiáveis isto é, que as diferenças entre as medidas são creditadas exclusivamente aos erros aleatórios^{22,27}. Após os cálculos da concordância foram considerados aptos os examinadores que apresentassem valores de concordância superiores a 0,60^{23,24}. Caso alguns examinadores fossem considerados inaptos, todos deveriam ser novamente treinados e submetidos a mais um exercício de calibração com outros voluntários, distribuídos da mesma maneira entre os grupos.

Os exames do treinamento prático e para a coleta de dados foram realizados em ambiente amplo, quadras cobertas da Unimontes e de escolas públicas do município, sob boas condições de iluminação, luz natural, com espelho e sonda específica preconizada pela OMS⁵, previamente esterilizados. Foram utilizados formulários específicos para o registro dos dados. Após os exames, os voluntários participantes foram orientados quanto à higiene bucal e receberam escovas e creme dental. Além disso, o atendimento dos voluntários com necessidade de tratamento

odontológico foi garantido no serviço odontológico de referência na rede municipal. Todos os exames foram precedidos por consentimentos livres e esclarecidos obtidos dos pais ou responsáveis e anuência dos menores de 18 anos ou do próprio participante com mais de 18 anos.

O “Levantamento epidemiológico das condições de saúde bucal da população de Montes Claros - Minas Gerais, 2008/2009 - Projeto SBMOC” foi aprovado pelo Comitê de Ética em Pesquisa da Unimontes (parecer nº 318/06), foi fruto de uma parceria entre a Unimontes e a Prefeitura Municipal de Montes Claros, contou com financiamento da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG e adotou metodologia baseada no Projeto SB Brasil¹¹.

RESULTADOS

Inicialmente, vinte cirurgiões-dentistas da rede municipal de saúde foram convidados a participar da pesquisa como examinadores. Apenas dezenove compareceram e desses, somente oito permaneceram no projeto. Foi realizada uma segunda calibração com mais cinco, dos quais apenas três permaneceram no projeto. Realizou-se uma terceira e última calibração com mais nove profissionais sendo que somente três continuaram no projeto. Portanto, houve a participação total de trinta e três cirurgiões-dentistas. Os examinadores eram funcionários públicos municipais, alguns deles foram demitidos; outros se afastaram por motivos pessoais. Portanto, somente quatorze examinadores treinados permaneceram no projeto do início até o final da coleta de dados. Os examinadores foram divididos em sete grupos. Os grupos 1, 2, 3, 5 e 6 foram compostos de cinco indivíduos e os grupos 4 e 7 de quatro indivíduos.

O exercício de calibração para avaliação da maloclusão demandou dez meses, incluindo a seleção dos voluntários, os treinamentos teóricos e práticos, a coleta de dados e o cálculo da concordância. Para os primeiros quatro grupos foram necessários seis meses, para o grupo cinco dois meses e para os grupos seis e sete dois meses. Foi necessária a participação de 315 adolescentes voluntários (cento e cinco no treinamento prático, cento e cinco na primeira calibração e cento e cinco na última), selecionados a partir do exame de, aproximadamente, mil adolescentes.

Os treinamentos teóricos demandaram sessenta horas (20 horas por treinamento). No treinamento prático, foram examinados 15 voluntários por grupo, totalizando 105 voluntários (60 no primeiro treinamento, 15 no segundo e 30 no terceiro). Essa etapa demandou trinta horas (10 horas

por treinamento).

As coletas de dados para o cálculo da concordância do DAI demandaram sessenta horas (20 horas por treinamento) e foram realizadas entre outros 105 voluntários. O último exercício de calibração, realizado somente quanto aos domínios do DAI que apresentaram discordância no primeiro exercício, contou com a participação de outros 105 voluntários. Os resultados desse último exercício não serão apresentados porque os valores de concordância foram superiores a 0,60.

Os níveis das concordâncias interexaminadores para avaliação do DAI geral entre

os pares de examinadores para cada um dos grupos variaram de -0,09 (AxB grupo 7) a 0,99 (BxC grupo 4) (concordância pobre à excelente) e por grupos, considerando todos examinadores juntos, variou de 0,51 (grupo 7) a 0,96 (grupo 3), (concordância razoável a excelente) (Tabela 1). Observou-se que a concordância intra-examinadores para o DAI geral, avaliada por meio do CCI, variou de 0,37 (examinador E, grupo 3) a 1,00 (examinador A, grupo 1), de concordância pobre a excelente (Tabela 2). Por meio dessa avaliação, os examinadores A do grupo 7, E do grupo 3 e C do grupo 4 foram considerados inaptos.

Tabela 1: Concordância – CCI - interexaminador entre os 10 pares de examinadores e para todos os examinadores juntos, por grupo de examinadores, para o DAI geral

	Grupo 1 (5 exam.)	Grupo 2 (5 exam.)	Grupo 3 (5 exam.)	Grupo 4 (4 exam.)	Grupo 5 (5 exam.)	Grupo 6 (5 exam.)	Grupo 7 (4 exam.)
A x B	0,97	0,98	0,98	0,94	0,90	0,95	-0,09
A x C	0,93	0,97	0,96	0,94	0,90	0,94	0,12
A x D	0,97	0,91	0,96	-	0,86	0,90	-0,12
A x E	0,97	0,82	0,98	0,94	0,86	0,88	-
B x C	0,92	0,97	0,96	0,99	0,97	0,95	0,75
B x D	0,98	0,93	0,94	-	0,94	0,97	0,72
B x E	0,98	0,87	0,97	0,95	0,93	0,90	-
C x D	0,88	0,93	0,94	-	0,97	0,94	0,92
C x E	0,94	0,85	0,95	0,94	0,95	0,91	-
D x E	0,96	0,84	0,97	-	0,91	0,86	-
TODOS	0,95	0,91	0,96	0,95	0,92	0,92	0,51

Tabela 2: Concordância – CCI - intra-examinadores para cada um dos examinadores para o DAI geral

Concordância intraexaminadores	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
A	1,00	0,88	0,95	0,97	0,92	0,87	0,85
B	0,98	0,88	0,97	0,95	0,91	0,83	0,85
C	0,92	0,93	0,93	0,41	0,86	0,92	0,94
D	0,97	0,85	0,91	-	0,88	0,94	0,82
E	0,93	0,81	0,37	0,89	0,92	0,90	-

Na avaliação da concordância interexaminadores dos pares de examinadores para os 10 componentes do DAI, por grupos, concordâncias superiores a 0,60 nos sete grupos foram encontradas somente para três componentes do DAI (dentição, apinhamento e diastema) (Tabela 3). As concordâncias interexaminadores para os pares de examinadores foram também obtidas separadamente para cada um dos componentes do DAI, em cada um dos grupos, permitindo identificar

qual (is) examinador (es) apresentavam discordância e em qual componente do DAI a discordância acontecia. Por essa análise, discordâncias entre pares de examinadores para alguns dos componentes do DAI foram observadas mesmo naqueles grupos cuja concordância entre o mesmo par de examinadores foi satisfatória considerando o DAI geral. Em função da grande extensão das tabelas, optou-se pela não apresentação desses dados.

Tabela 3: Valor mínimo e máximo do CCI e do Coeficiente Kappa ponderado obtidos pelos pares de examinadores para os 10 componentes do DAI nos sete grupos de examinadores

Componentes do DAI	Grupo 1 Min-Max	Grupo 2 Min-Max	Grupo 3 Min-Max	Grupo 4 Min-Max	Grupo 5 Min-Max	Grupo 6 Min-Max	Grupo 7 Min-Max
Dentição*	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00
Apinhamento**	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00
Espaçamento**	0,30-1,00	0,16-1,00	0,42-0,71	0,12-0,62	0,72-1,00	0,53-0,86	0,63-1,00
Diastema*	0,86-1,00	0,62-0,87	0,48-0,91	0,72-0,92	1,00-1,00	0,64-0,96	0,67-1,00
Desalinhamento maxilar*	0,93-1,00	0,50-0,90	0,60-0,92	0,81-0,90	0,04-0,93	0,55-0,79	0,52-0,89
Desalinhamento mandibular*	0,75-0,97	0,03-0,64	0,31-0,84	0,36-0,77	1,00-1,00	0,57-0,92	0,71-0,97
Overjet maxilar*	0,63-0,91	0,33-0,71	0,41-0,87	0,44-0,82	0,32-0,82	0,56-0,95	0,37-0,97
Overjet mandibular*	0,96-0,97	0,61-0,99	0,40-0,94	0,22-0,92	0,79-1,00	0,32-0,92	0,32-0,98
Mordida aberta vertical anterior*	1,00-1,00	0,04-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00	1,00-1,00
Relação molar antero posterior **	0,14-1,00	0,59-1,00	0,86-1,00	0,80-1,00	0,40-0,93	1,00-1,00	1,00-1,00

*Coeficiente de Correlação Intra classe; ** Coeficiente Kappa ponderado

Na avaliação da concordância intra-examinadores para os 10 componentes do DAI foi constatado que os examinadores A, C e E do grupo 1; A, B, C, D e E do grupo 2, E do grupo 3 e B e C do grupo 4, ou seja, 11 examinadores foram considerados inaptos para a coleta de dados. Vale ressaltar que somente dois dos examinadores considerados inaptos por esse método de avaliação (C do grupo 3 e E do grupo 4), haviam sido considerados inaptos pelo cálculo da concordância intra a partir dos valores gerais do DAI (Tabela 4 e 5).

DISCUSSÃO

Um dos grandes e primeiros desafios enfrentados nesse exercício de calibração foi a seleção de voluntários dentro das faixas etárias e características exigidas para o treinamento. Nesse estudo, somente 31,5% (n = 315) dos voluntários examinados apresentaram as características exigidas, o que demandou tempo e esforço. Embora trabalhosa, essa etapa é importante já que a consistência dos resultados obtidos pela pesquisa está, entre outros fatores, relacionada ao tipo de população examinada durante o processo de calibração que deve ser similar àquela examinada durante o próprio levantamento epidemiológico^{4,7}.

Outra dificuldade enfrentada foi a perda de examinadores durante o exercício de calibração. Alguns examinadores do presente estudo eram funcionários públicos municipais e foram demitidos, afastando-se da pesquisa; outros passaram em concursos públicos em outros municípios; outros afastaram por considerarem o processo cansativo. Os estudos epidemiológicos são comumente

direcionados para avaliar o nível de saúde de uma comunidade. Como no caso do SBMOC, tais estudos são ligados, em sua maioria, ao sistema público de saúde, tornando-os vulneráveis às mudanças políticas, correndo-se o risco da política (“politics”) ser mais forte do que as políticas (“policies”) que originaram os projetos de pesquisa²⁸. A não liberação dos funcionários municipais, de forma integral, por um determinado período para atuar na coleta de dados foi um dificultador, pois levou ao afastamento de examinadores já treinados. Além da coleta, os examinadores tinham que desenvolver suas atividades rotineiras por, aproximadamente, quatro turnos dentre os dez turnos semanais. O treinamento de recursos humanos de serviços públicos, envolvidos em suas atividades rotineiras, pode não ser a melhor opção em função do acúmulo de atividades e da falta de garantia de continuidade, pela instabilidade no emprego, comum neste cenário.

No presente estudo, o tempo despendido para cada processo de calibração de examinadores (\pm 30 horas) foi considerado satisfatório, pois de acordo com o Manual de Calibração SB Brasil 2010⁷ todo o processo de calibração da equipe de examinadores foi dimensionado para durar, pelo menos 32 horas de trabalho considerando-se um número máximo de dez examinadores.

Nesse estudo, as concordâncias inter e intra examinadores foram calculadas empregando o K_w ou o CCI. O K_w foi indicado para os componentes ordinais do DAI, pois são atribuídos pesos (w) aos diferentes níveis de concordância dentro da escala ordinal^{23,24}. O k_w tem a vantagem de “penalizar” as maiores discordâncias mais severamente que as

Tabela 4: Concordância intra-examinadores (CCI e Coeficiente Kappa ponderado) para os componentes do DAI (dentição, alinhamento, espaçamento, diastema e desalinhamento maxilar) nos sete grupos de examinadores (Ex)

Grupo	Ex	Dentição*	Alinhamento**	Espaçamento**	Diastema*	Desalinhamento Maxilar*
1	A	1,00	1,00	0,90	0,77	0,80
	B	1,00	1,00	0,94	1,00	0,97
	C	1,00	1,00	0,95	1,00	0,90
	D	1,00	1,00	0,73	1,00	0,94
	E	1,00	1,00	0,88	0,76	0,96
2	A	1,00	1,00	0,41	0,66	0,88
	B	1,00	1,00	0,37	0,77	0,63
	C	1,00	1,00	0,64	0,83	0,91
	D	1,00	1,00	0,66	0,49	0,58
	E	1,00	1,00	0,63	0,85	0,94
3	A	1,00	1,00	0,73	1,00	0,92
	B	1,00	1,00	0,96	0,91	0,81
	C	1,00	1,00	0,90	0,87	0,98
	D	1,00	1,00	0,88	0,92	0,98
	E	1,00	1,00	0,86	0,95	0,97
4	A	1,00	1,00	0,63	0,79	0,95
	B	1,00	1,00	0,32	0,91	0,95
	C	1,00	1,00	0,86	0,36	0,93
	D	-	-	-	-	-
	E	1,00	1,00	0,61	0,83	0,94
5	A	1,00	1,00	0,90	0,93	0,94
	B	1,00	1,00	0,89	0,86	0,93
	C	1,00	1,00	0,86	0,85	0,90
	D	1,00	1,00	0,90	0,87	0,91
	E	1,00	1,00	0,91	0,92	0,90
6	A	1,00	1,00	0,67	0,68	0,87
	B	1,00	1,00	0,65	0,85	0,92
	C	1,00	1,00	0,87	0,84	0,91
	D	1,00	1,00	0,90	0,91	0,87
	E	1,00	1,00	0,89	0,90	0,85
7	A	1,00	1,00	0,85	0,84	0,91
	B	1,00	1,00	0,86	0,86	0,90
	C	1,00	1,00	0,92	0,90	0,94
	D	1,00	1,00	0,84	0,83	0,92
	E	-	-	-	-	-

*Coeficiente de Correlação Intra classe; ** Coeficiente Kappa ponderado

discordâncias menos críticas²⁶. A estatística kappa considera a concordância atribuível, ou seja, a concordância observada menos aquela obtida pela chance ($po - pe$) e corresponde a proporção da concordância atribuível em relação ao seu valor máximo possível ($1 - pe$). O Kappa simples pode ser alternativamente empregado quando o DAI é categorizado com o estabelecimento de pontos de corte, indicando a severidade da maloclusão e a

necessidade de tratamento ortodôntico. Nesse estudo, o CCI foi utilizado para avaliar as concordância intra e interexaminadores no diagnóstico do DAI e de seus componentes numéricos, uma vez que os valores desse índice são obtidos através da soma de componentes ponderados, com resultados variando dentro de uma escala de intervalo que pode ser tratada como variável numérica²⁹. O Coeficiente de Correlação Intra-Classe é recomendado para avaliar

Tabela 5: Concordância intra-examinadores (CCI e Coeficiente Kappa ponderado) para os componentes do DAI (desalinhamento mandibular, overjet maxilar, overjet mandibular, mordida aberta vertical anterior e relação molar ântero-posterior) nos sete grupos de examinadores.

Grupo	Exam.	Desalinhamento Mandibular	Overjet Maxilar*	Overjet Mandibular*	Mordida Aberta vertical anterior*	Relação Molar ântero-posterior**
1	A	0,80	0,56	0,90	0,81	0,93
	B	0,86	0,97	0,90	10	10
	C	0,80	0,56	0,97	10	10
	D	0,83	0,92	0,96	10	10
	E	0,27	0,68	0,90	10	0,98
2	A	0,87	0,61	0,85	10	10
	B	01	0,42	0,85	10	0,82
	C	0,72	0,40	0,92	10	0,81
	D	0,27	06	0,91	10	10
	E	0,52	0,55	0,83	10	0,98
3	A	0,84	0,73	0,97	10	10
	B	0,90	0,90	0,96	10	10
	C	0,93	0,98	0,95	10	10
	D	0,88	0,92	0,94	10	10
	E	0,55	0,91	0,57	10	10
4	A	0,73	0,94	0,96	10	0,95
	B	0,88	0,89	0,94	10	0,96
	C	0,73	0,76	0,21	0,97	0,98
	D	-	-	-	-	-
	E	0,71	0,97	0,85	10	0,96
5	A	0,93	0,90	0,94	0,92	0,92
	B	0,92	0,90	0,92	0,90	0,90
	C	0,91	0,89	0,90	0,89	0,89
	D	0,92	0,90	0,90	0,89	0,88
	E	0,93	0,92	0,93	0,90	0,93
6	A	0,88	0,68	0,89	0,98	0,98
	B	0,62	0,68	0,83	10	0,98
	C	0,89	0,87	0,91	0,89	0,90
	D	0,88	0,86	0,90	0,89	0,92
	E	0,83	0,85	0,92	0,88	0,91
7	A	0,92	0,88	0,92	0,90	0,90
	B	0,89	0,89	0,88	0,89	0,90
	C	0,90	0,93	0,94	0,95	0,96
	D	0,93	0,87	0,90	0,88	0,86
	E	-	-	-	-	-

a concordância entre examinadores quando a variável investigada for quantitativa ou numérica²⁵ e avalia o grau de concordância desconsiderando aquela obtida além das chances, ou seja, a concordância atribuída ao acaso.

Conforme critérios de interpretação dos resultados do CCI e do K_w adotados, a maioria dos níveis de concordância intra e interexaminadores

obtidos foram considerados satisfatórios tendo como referência o cálculo do CCI para os valores do DAI geral ou seus componentes separadamente. Entretanto, ao se considerar os componentes do DAI separadamente, algumas discrepâncias não observadas de outra maneira foram detectadas, indicando a necessidade de novos treinamentos, culminando com melhor consistência diagnóstica

e maior confiabilidade dos resultados. Da mesma maneira, para a concordância intra-examinadores, nove examinadores a mais foram avaliados como inaptos analisando os componentes do DAI separadamente, determinando a necessidade de participação em novos exercícios de calibração. Com base nessa experiência, os pesquisadores recomendam o uso do CCI e do K_w para estimar a concordância intra e interexaminadores, considerando as características oclusais do DAI separadamente. Esse método de avaliação atribuiu maior acurácia no exercício de concordância, bem como maior confiabilidade nos resultados da pesquisa.

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 bucais⁴.

Uma vez detectadas discrepâncias, todo o exercício deve ser repetido, permanecendo no grupo apenas aqueles examinadores que apresentarem um grau de consistência aceitável com a equipe⁷. Assim, períodos adicionais de treinamento para reforço do aprendizado teórico e prático dos critérios de avaliações estabelecidos, podem gerar uma melhoria nos resultados de concordância entre os examinadores⁴. Por isso, nesse estudo, para alguns examinadores foi necessária a repetição do exercício de calibração por três vezes. Além disso, os três treinamentos foram necessários para obtenção de um número satisfatório de examinadores, devido ao afastamento de alguns. Vale ressaltar que períodos de calibração adicionais podem ser ministrados apenas para averiguar e trabalhar sobre a discordância encontrada, levando-se, conseqüentemente, ao maior aprimoramento final dos resultados da calibração⁴.

Nesse estudo, foram necessários dez meses até que os 33 examinadores fossem treinados; o que atrasou a coleta de dados do levantamento epidemiológico. Identifica-se a realização de outras investigações para o desenvolvimento de metodologias e ou de programas de computadores para agilizar e facilitar esta etapa das investigações epidemiológicas sobre as condições bucais.

CONCLUSÃO

A calibração de examinadores, recomendada e necessária em pesquisas epidemiológicas, configura-se como um desafio para os pesquisadores já que demanda tempo e esforço para se alcançar os resultados almejados. A concordância obtida por componentes do DAI identifica maior número de examinadores inaptos. Apesar de todas as dificuldades do processo de calibração, a estimativa

da concordância pelo CCI e pelo Kappa ponderado para o DAI e seus componentes garantiu uma maior acurácia aos resultados do levantamento epidemiológico.

AGRADECIMENTOS

Ao apoio logístico da Unimontes e da Prefeitura Municipal de Montes Claros-MG, ao fomento da Fundação de Amparo a Pesquisa do Estado de Minas Gerais-FAPEMIG e a colaboração dos participantes. Os pesquisadores Andréa Maria Eleutério de Barros Lima Martins, Marise Fagundes da Silveira e Raquel Conceição Ferreira receberam bolsa da FAPEMIG.

ABSTRACT

Aim: To report the challenges faced in a calibration process of examiners for an epidemiological study on malocclusion involving quantitative and ordinal categorical variables, showing the results of inter and intra-examiner diagnostic consistency. **Materials and Methods:** Volunteers were selected to anticipate the conditions the examiners would find in field. Malocclusion was assessed by the DAI, whose criteria were exhibited in the theoretical training and strengthened in practical training. The examiners were divided into groups and the data collected by examination. Inter and intra-examiner agreement was estimated by means of the Intra-class Correlation Coefficient (ICC) for the general DAI and and the ICC or weighted kappa (K_w) for the DAI components, using the PASW and Excel programs. The examiners with an agreement of less than 0.60 were considered inept and should be recommended for another calibration process. **Results:** In three calibration exercises, which proved to be lengthy and costly, 33 examiners divided into 7 groups, examined and re-examined 315 students. The inter-examiner agreement for the general DAI, among pairs of examiners, ranged from -0.09 to 0.99 (poor to excellent). The agreement of all examiners together, by group, ranged from 0.51 to 0.96 (reasonable to excellent). In the analyses of components from the seven groups, an agreement of > 0.60 among the pairs of examiners was found for dentitions, crowding and spacing. Disagreements among examiner pairs regarding some components were observed even in groups whose agreement among the same pair of examiners was satisfactory for the general DAI. In the intra-examiner agreement for the general DAI, from 0.37 to 1.00 (poor to excellent), two examiners were considered inept; in the evaluation for DAI components, 11 presented ICC and $K_w < 0.60$. **Conclusion:** The calibration process of examiners configured itself as a challenge

to researchers, demanding time and effort. The agreement for the components of DAI identified a higher number of inept examiners, ensuring greater accuracy regarding the results of the epidemiological survey.

Uniterms: Evaluation. Epidemiology. Reproducibility of results.

REFERÊNCIAS

1. Organização Mundial da Saúde. Levantamento epidemiológico básico de saúde bucal. 3ª ed. São Paulo: Santos; 1991.
2. Frias AC, Antunes JLF, Narval 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.
3. Peres MA, Traebert J, Marcenes W. Calibração de examinadores para estudos epidemiológicos de cárie dentária. *Cad Saude Publica.* 2001; 17:153-9.
4. Assaf AV, Zanin L, Meneghim MC, Pereira AC, Ambrosano GMB. Comparação entre medidas de reprodutividade para calibração em levantamentos epidemiológicos da cárie dentária. *Cad Saude Publica.* 2006; 22:1901-07.
5. World Health Organization. Oral Health surveys: basic methods. 4th ed. Geneva: ORH EPID; 1997.
6. World Health Organization. Calibration of examiners for oral health epidemiology surveys. technical report. Geneva: ORH/EPID; 1993.
7. Brasil. Ministério da Saúde. SB Brasil 2010. Manual de Calibração de Examinadores. Brasília: Ministério da Saúde; 2010.
8. Oliveira CM, Sheiham A. Orthodontic treatment and its impact on oral health-related quality of life in Brazilian adolescents. *J Orthod.* 2004; 31:20-7.
9. Marques LS, Barbosa CC, Ramos-Jorge ML, Pordeus IA, Paiva SM. Prevalência da maloclusão e necessidade de tratamento ortodôntico em escolares de 10 a 14 anos de idade em Belo Horizonte, Minas Gerais, Brasil: enfoque psicossocial. *Cad Saúde Pública.* 2005; 21:1099-106.
10. Richmond S, O'Brien KD, Roberts CT, Andrews M. Dentists variation in the determination of orthodontic treatment need. *Br J Orthod.* 1994; 21:65-8.
11. Brasil, Ministério da Saúde. Projeto SBBrasil 2003: condições de saúde bucal da população brasileira 2002-2003: resultados principais. Brasília: Ministério da Saúde; 2004.
12. Munoz SR, Bangdiwala SI. Interpretation of Kappa and B Statistics Measures of Agreement. *J Appl Stat.* 1997; 24:105-11.
13. Cons NC, Jenny J, Kohout FJ. DAI: the dental aesthetic index. Iowa City: Iowa College of Dentistry; 1986.
14. Peres KG, Traebert ESA, Marcenes W. Diferenças entre autopercepção e critérios normativos na identificação das oclusopatias. *Rev Saúde Pública.* 2002; 36:230-6.
15. Beglin FM, Firestone AR, Vig KWL, Beck FM, Kuthy RA, Wade D. A comparison of the reliability and validity of 3 occlusal indexes of orthodontic treatment need. *Am J Orthod Dentofacial Orthop.* 2001; 120:240-6.
16. Jenny J, Cons NC. Guidelines for using the DAI: a supplement to DAI—the dental aesthetic index. Iowa: College of Dentistry; 1988.
17. Jenny J, Cons NC. Establishing malocclusion severity levels on the dental aesthetic index (DAI) scale. *Aust Dent J.* 1996; 41:43-6.
18. Hamamci N, Basaran G, Uysal E. Dental Aesthetic Index scores and perception of personal dental appearance among Turkish university students. *Eur J Orthod.* 2009; 31:168-73.
19. Gábris K, Márton S, Madléna M. Prevalence of malocclusions in Hungarian adolescents. *Eur J Orthod.* 2006; 28:467-70.
20. Alves JAO, Forte FDS, Sampaio FC. Condição socioeconômica e prevalência de más oclusões em crianças de 5 e 12 anos na USF Castelo Branco III – João Pessoa/Paraíba. *Rev Dent Press Ortodon Ortopedi Facial.* 2009; 14:52-9.
21. Lopes LS, Cangussu MCT. Prevalência e severidade das alterações oclusais em escolares de 12 a 15 anos de Salvador – BA, 2004. *J Med Biol Sci.* 2005; 4:105-12.
22. Costa RN, de Abreu MH, de Magalhães CS, Moreira AN. Validity of two occlusal indices for determining orthodontic treatment needs of patients treated in a public university in Belo Horizonte, Minas Gerais State, Brazil. *Cad Saúde Pública.* 2011; 27:581-90.

23. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons, 1981.
24. Cicchetti DV, Volkmar F, Sparrow SS, Cohen D, Fermanian J, Rourke BP. Assessing the reliability of clinical scales when the data have both nominal and ordinal features: proposed guidelines for neuropsychological assessments. *J Clin Exp Neuropsychol*. 1992; 14:673-86.
25. Bartko JJ, Carpenter WT. On the methods and theory of reliability. *J Nerv Ment Dis*. 1976; 163:307-17.
26. Bulmann JS, Osborn JF. Measuring diagnostic consistency. *Br Dent J*. 1989; 166:377-81.
27. Fleiss JL. The Design and analysis of clinical experiments. New York: John Wiley & Sons, 1986.
28. Franco LJ. A Pesquisa em epidemiologia: dificuldades e perspectivas. *Saúde Soc*. 1995; 4:31-4.
29. Roberts CT, Richmond S. The design and analysis of reliability studies for the use of epidemiological and audit indices in orthodontics. *Br J Orthod*. 1997; 24:139-47.

Recebido em 26/06/2011 – Aceito em 12/09/2011

Autor correspondente:

Andréa Maria Eleutério de Barros Lima Martins
Universidade Estadual de Montes Claros
Campus Universitário Professor Darcy Ribeiro - Vila Mauricéia
CEP: 39401-089 - Montes Claros - Minas Gerais – Brasil
E-mail martins.andreamebl@gmail.com