

Manual de Calibração de Examinadores

Brasília-DF
2009



Ministério da Saúde
Secretaria de Vigilância à Saúde
Secretaria de Atenção à Saúde
Departamento de Atenção Básica
Coordenação Nacional de Saúde Bucal



Manual de Calibração de Examinadores

Brasília, DF
2009

Projeto SBBrasil 2010

José Gomes Temporão
Ministro da Saúde

Alberto Beltrame
Secretário de Atenção à Saúde

Claunara Schilling Mendonça
Diretora do Departamento de Atenção Básica

Gilberto Alfredo Pucca Jr.
Coordenador de Saúde Bucal

Comitê Técnico Assessor em Vigilância em Saúde Bucal

Angelo Giuseppe Roncalli da Costa Oliveira
Universidade Federal do Rio Grande do Norte (UFRN)

Lenildo de Moura
Coord. Geral Doenças e Agravos Não-Transmissíveis/SVS

Marco Aurélio Peres
Universidade Federal de Santa Catarina

Maria do Carmo Matias Freire
Universidade Federal de Goiás

Maria Ilma de Souza Gruppioni Côrtes
Pontifícia Universidade Católica de Minas Gerais – PUC-MG

Mario Vianna Vettore
Escola Nacional de Saúde Pública – ENSP-Fiocruz

Paulo Capel Narvai
Faculdade de Saúde Pública da USP – FSP/USP

Paulo Sávio Angeiras de Góes
Universidade de Pernambuco – FOP-UPE

Samuel Jorge Moysés
Pontifícia Universidade Católica do Paraná – PUC-PR

Coordenação Executiva do SBBrasil nos Centros Colaboradores e Ministério da Saúde

Cláudia Helena Soares de Moraes Freitas
Centro Colaborador Univ. Fed. do Rio G. do Norte – UFRN

Nilcema Figueiredo
Centro Colaborador Univ. de Pernambuco – UPE

Sandra Cristina Guimarães Bahia Reis
Centro Colaborador da Universidade Federal de Goiás – UFG

Rafaela da Silveira Pinto
Centro Colaborador Pontifícia Univ. Católica de Minas Gerais

Elisete Casotti
Centro Colaborador Escola Nacional de Saúde Pública – ENSP

Regina Auxiliadora de Amorim Marques
Centro Colaborador Faculdade de Saúde Pública da USP

Antonio Carlos Nascimento
Centro Colaborador Pontifícia Universidade Católica do Paraná

Karen Glazer de Anselmo Peres
Centro Colaborador Universidade Federal de Santa Catarina

Moacir Paludetto Jr.
Coordenação de Saúde Bucal/DAB/SAS/MS

Lenildo de Moura
Coord. Geral Doenças e Agravos Não-Transmissíveis/SVS

Coordenação Geral

Angelo Giuseppe Roncalli da Costa Oliveira
Universidade Federal do Rio Grande do Norte – UFRN

Assessoria em Amostragem

Nilza Nunes da Silva
Faculdade de Saúde Pública da Universidade de São Paulo (FSP-USP)



A série de manuais de apoio à realização do Projeto SBBrasil 2010 – Pesquisa Nacional de Saúde Bucal está disponível para download no endereço www.saude.gov.br/bucal e pode ser utilizada livremente, desde que citada a fonte.

Sumário

1. Introdução	5
2. Alguns conceitos	5
3. O Processo de Calibração de Examinadores	7
3.1. Preparação do processo	7
3.2. Discussão teórica das variáveis utilizadas, códigos e critérios de exame (4 horas).....	8
3.3. Discussão Prática (8 horas)	8
3.4. Calibração propriamente dita (14 horas)	9
3.5. Discussão Final da Calibração (2 horas)	10
4. Calculando a concordância interexaminador	11
4.1. A Concordância Simples ou Concordância Percentual	11
4.2. O Coeficiente Kappa	13
4.3. A Concordância por Escore	14
5. Utilizando a planilha eletrônica	14
5.1. Funcionamento das pastas de trabalho	15
6. Verificação da reprodutibilidade em campo.....	19
7. Valores recomendados para a Concordância	19
8. Relatório	19
9. Bibliografia	19

1. Introdução

Os três grandes levantamentos nacionais realizados em 1986, 1996 e 2003 foram de grande relevância para a construção de uma consistente base de dados relativa ao perfil epidemiológico de saúde bucal da população brasileira. Contudo, é fundamental que a realização destes estudos faça parte de uma estratégia inserida no componente de vigilância à saúde da Política de Saúde, na perspectiva da construção de uma série histórica de dados de saúde bucal com o objetivo de verificar tendências, planejar e avaliar serviços.

A Política Nacional de Saúde Bucal – Brasil Sorridente – se constitui num marco na história das Políticas Públicas no Brasil na medida em que incorpora uma agenda em discussão desde o Movimento pela Reforma Sanitária Brasileira e traduz, em seus pressupostos operacionais, os princípios do Sistema Único de Saúde. Ao trabalhar os eixos da atenção à saúde bucal a partir do incremento da atenção básica por meio da Estratégia Saúde da Família, da implementação dos Centros de Especialidades Odontológicas como elemento estruturante da atenção secundária, além das ações de caráter coletivo, o Brasil Sorridente se insere no conjunto de programas estratégicos na atual Política de Saúde.

Dentre os pressupostos dessa política, que visam à reorientação do modelo de atenção à saúde bucal, destacam-se: (a) “utilizar a Epidemiologia e as informações sobre o território subsidiando o planejamento” e (b) “centrar a atuação na Vigilância à Saúde, incorporando práticas contínuas de avaliação e acompanhamento dos danos, riscos e determinantes do processo saúde doença”. Tais pressupostos devem, portanto, ser postos em prática a partir de diversas estratégias, dentre elas a realização de pesquisas epidemiológicas de base nacional.

Desse modo, a Pesquisa Nacional de Saúde Bucal – Projeto SBBrasil 2010 faz parte de um processo histórico que se ampliou e se aprofundou com o Projeto SBBrasil 2003, o qual proporcionou um dos mais completos diagnósticos da saúde bucal dos brasileiros. Pretende-se, com este projeto, dar continuidade a esse processo, realizando uma pesquisa em moldes semelhantes, de maneira a construir uma série histórica, contribuindo para as estratégias de avaliação e planejamento dos serviços, ao mesmo tempo em que consolida um modelo metodológico e demarca o campo de atuação do componente de vigilância à saúde da Política Nacional de Saúde Bucal.

Do ponto de vista operacional, o SBBrasil 2010 possui uma lógica de organização em que os diferentes níveis de gestão do SUS (Municipal, Estadual e Federal) trabalham articuladamente, com a coordenação técnica dos Centros Colaboradores em Vigilância à Saúde e com o apoio de entidades de classe, universidades e centros de pesquisa.

Este *Manual de Calibração de Examinadores* visa subsidiar os instrutores de calibração com relação às técnicas para a realização do treinamento das equipes de campo, especificamente no que diz respeito à calibração. A sua base teórica é a proposta da Organização Mundial da Saúde, de 1993, descrita na publicação *Calibração de Examinadores para Levantamentos Epidemiológicos em Saúde Bucal*, de autoria de Stephen Eklund, Ingolf Möller e Marie-Hélène Leclercq (OMS, 1993). Alguns trechos desse documento, bem como alguns exemplos, foram transcritos integralmente. Lembramos, também, que as situações eventualmente não contempladas aqui deverão ser discutidas e resolvidas com as coordenações estaduais e com a coordenação técnica do Centro Colaborador.

2. Alguns conceitos

Os levantamentos epidemiológicos em saúde bucal envolvem, na maioria das vezes, um grande número de examinadores. É praticamente impossível realizar estudos abrangendo grandes amostras e vários índices em diferentes grupos etários, com a participação de apenas um examinador.

Dadas as características dos índices utilizados em saúde bucal e da subjetividade inerente ao exame dos tecidos bucais, a manutenção de uma boa reprodutibilidade das observações é uma condição fundamental para a confiabilidade dos dados. Neste estudo, a extensa variabilidade dos eventos que serão observados exige o treinamento de calibração dos examinadores a fim de se conhecer o grau de confiabilidade alcançado durante a coleta dos dados nos diferentes municípios participantes do estudo.

Calibração

A despeito das críticas que podem ser feitas ao termo *calibração* (é uma expressão usada na indústria, ou seja, calibram-se máquinas, não pessoas), pela consagração de seu uso em epidemiologia bucal, ele será utilizado neste manual 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.**

Objetivos da Calibração

Em linhas gerais, pode-se dizer que os principais objetivos da calibração de examinadores em levantamentos epidemiológicos são (WHO, 1993):

- 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.
- Minimizar variações entre os diferentes examinadores

Ou seja, trata-se de proporcionar consistência aos exames epidemiológicos realizados por um ou mais examinadores. Para isso, é importante que eles adotem na maioria das observações o mesmo critério, atuando como se fossem um único examinador.

Concordância Interexaminador

A equipe deve funcionar de modo a examinar sempre a partir de um mesmo critério, concordando, o máximo possível, em seus diagnósticos. A equipe deve examinar um mesmo grupo de indivíduos e os resultados devem ser comparados entre si. É sempre bom lembrar que o objetivo da calibração interexaminador não é estabelecer quem está certo e quem está errado, mas verificar quais examinadores estão diferindo de forma mais significativa, reduzindo a variabilidade entre eles. Uma vez detectadas grandes discrepâncias, todo o exercício deve ser repetido, permanecendo na equipe apenas aqueles examinadores que apresentarem um grau de consistência aceitável com a equipe.

Concordância Intra-examinador

É bastante comum entre os dentistas a argumentação de que seus exames são consistentes, ou seja, o diagnóstico feito em um paciente num dado momento é sempre o mesmo em outra ocasião. Os estudos sobre a consistência interna dos exames em saúde bucal, contudo, não mostram isso. Em função do reconhecido grau de subjetividade do qual é revestido o diagnóstico das doenças bucais, é comum que um por cento dos exames repetidos por um dentista não sejam iguais. Na situação específica de um estudo epidemiológico, há ainda o agravante de serem feitas várias observações em um mesmo dia, deste modo a fadiga pode levar a diagnósticos diferentes em distintos momentos. A verificação da concordância intra-examinador busca aferir o quanto o examinador concorda com ele mesmo em diferentes momentos. Ela deverá ser feita durante a coleta dos dados, no sentido de aferir se a consistência obtida no treinamento está sendo mantida em campo. O processo de verificação da concordância intra-examinador durante o processo de trabalho de campo é função do coordenador municipal e/ou supervisor das equipes.

Durante o levantamento, a concordância deve ser feita utilizando os *exames em duplicata*, ou seja, cada examinador deve reexaminar 5% da amostra (repete um indivíduo a cada grupo de 20). Na medida do possível, o examinador não deve identificar o paciente que está sendo reexaminado. O fato de o examinador saber que

o indivíduo foi examinado previamente pode influenciar o grau de atenção e, conseqüentemente, as observações. O anotador pode ser convidado para a tarefa de anotar os indivíduos (1 a cada 20) para o reexame durante o levantamento. Neste estudo é recomendada a realização de exames em duplicata durante todo o período da pesquisa, organizada pelo supervisor de campo.

3. O Processo de Calibração de Examinadores

A calibração deve ser realizada para alguns agravos e seus respectivos grupos etários. Uma recomendação de caráter geral é que se procure reproduzir, ao máximo as condições que serão encontradas em campo. Com relação às idades e grupos etários, por razões operacionais, será utilizada a idade de 5 anos e os grupos etários de 12 a 19 anos e de 30 anos ou mais. O quadro a seguir sumariza os agravos a serem calibrados e as respectivas idades.

Idade / Grupo Etário	Agravo	Técnica
5 anos	<ul style="list-style-type: none"> • Cárie de coroa • Condição de Oclusão 	<i>In vivo</i>
12 a 19 anos	<ul style="list-style-type: none"> • Cárie de Coroa • DAI • CPI Cálculo 	<i>In vivo</i>
	<ul style="list-style-type: none"> • Traumatismo Dentário • Fluorose 	<i>In lux</i> ^(*)
30 anos e mais	<ul style="list-style-type: none"> • Cárie de coroa • CPI Cálculo • CPI Bolsa 	<i>In vivo</i>

(*) Por analogia, adotaremos o termo "in lux" para designar a calibração que não é realizada pelo exame em indivíduos, mas com a utilização de imagens.

O processo de calibração deve ser planejado considerando-se um número máximo de 10 (dez) examinadores. Números maiores que esse normalmente tornam a prática de difícil execução, entretanto alternativas diferentes devem ser discutidas com a coordenação técnica do estudo. Todo o processo de calibração da equipe de examinadores foi dimensionado para durar, pelo menos 32 horas de trabalho (8 turnos de 4), assim divididas:

3.1. Preparação do processo

Antes de se proceder ao treinamento teórico e prático, a coordenação do levantamento SBBrasil no município e da calibração devem organizar toda a infra-estrutura necessária. Isto inclui um contato prévio com as instituições (escolas, quartéis, centros de saúde etc.) nas quais se procederá aos exames para efeito de treinamento e calibração. Nessas instituições, deve ser verificada a existência de pessoas dentro das faixas etárias exigidas para o treinamento, devendo-se solicitar sua colaboração, explicando-lhes como se dará o processo, obtendo-se assim o consentimento. Além disso, deve também ser observada a disponibilidade de ambientes bem iluminados e adequados para a realização dos exames.

O instrutor deve também procurar selecionar os voluntários de modo que estes apresentem uma maior variedade de situações relativas às condições bucais pesquisadas. No caso da cárie, por exemplo, é importante que existam, em número razoável, casos de dentes cariados, restaurados, restaurados com cárie, dentre outros.

Nestes casos, indivíduos com baixa prevalência de cárie não são adequados para atuarem como voluntários no treinamento de calibração.

Com relação à discussão teórica, é prudente atentar para o espaço onde esta acontecerá, de modo que apresente condições adequadas, tais como possibilidade de escurecimento da sala (para a projeção de slides), cadeiras que possibilitem fazer anotações, entre outros aspectos, que facilitem o melhor processo de aprendizagem.

3.2. Discussão teórica das variáveis utilizadas, códigos e critérios de exame (4 horas)

Nesta fase deve ser feita exposição teórica e discussão junto à equipe de examinadores, por parte da equipe de instrutores, sobre todos os índices, códigos e critérios a serem utilizados. Recomenda-se utilizar slides e fazer a leitura prévia do *Manual da Equipe de Campo*. O importante é que, neste momento, seja esclarecido o maior número possível de dúvidas relativas aos critérios e que sejam exercitadas situações em que uma regra de decisão seja exigida. Convém sempre lembrar que, apesar dos critérios de exame epidemiológico serem os mais rígidos possíveis, existem situações-limite, que devem ser amplamente discutidas no grupo para que, durante a coleta dos dados, prevaleça o consenso da equipe sobre o critério individual de cada examinador. É importante ressaltar que, ao contrário da clínica, onde o diagnóstico precede a realização de determinada conduta ou procedimento, em epidemiologia é mais importante o consenso diagnóstico e que o mesmo seja seguido por todos os examinadores, mesmo que contrário às suas convicções pessoais.

3.3. Discussão Prática (8 horas)

Cada participante da equipe deve examinar em torno de 10 pessoas nos seguintes grupos etários: 5 anos, 12 a 19 anos e 35 a 44 anos.

O processo todo se dá a partir da seguinte sequência (ver Figura 1):

- Formam-se blocos de 5 equipes no máximo (examinador/anotador). Caso tenhamos 10 equipes serão dois blocos de 5 equipes, 8 equipes conformarão dois blocos de 4 equipes e assim por diante.
- Cada cadeira deve ser identificada por um número (1, 2, 3 etc.)
- São chamados cinco (se for o caso) voluntários para ocupar as cadeiras.
- Cada examinador deve passar por cada um dos voluntários e realizar o exame.
- Ao final da rodada, os voluntários não devem ser dispensados ainda. Os examinadores, orientados pelo instrutor, deverão examinar suas fichas e observar as discordâncias em cada um dos exames que cada um realizou.
- Ao serem observadas discordâncias, todas são discutidas exaustivamente de modo a chegar a um consenso sobre o diagnóstico ao qual todos concordam.



Por exemplo...

Durante o exame do garoto Marcos, de 12 anos, 3 examinadores divergem a respeito do diagnóstico do dente 25. Um dos examinadores diagnosticou o dente como cariado, outro como hígido e outro como com presença de selante. Ao discutir o caso, os três examinadores podem chegar a um consenso de que o dente, na verdade, tinha um selante. Pode ocorrer, entretanto, que, apenas observando as fichas, não se chegue a um consenso e daí, nestes casos, o instrutor deve orientar para que os códigos e critérios sejam relidos com atenção e a criança deve ser observada novamente.

- Ainda mantendo o voluntário em sua cadeira, o instrutor de calibração realiza a aferição das discordâncias entre os exames.

- h) Todas as discordâncias são discutidas com as equipes e, se necessário, deve ser realizado novo exame no voluntário a partir da releitura dos códigos e critérios do índice, expostos no manual.
- i) O instrutor preenche uma "Ficha de Consenso" do voluntário, resultado da discussão com as equipes. Calcula-se a concordância simples (que será explicada adiante, no item 4 deste manual), entretanto ainda sem preocupação margens de erro aceitáveis, servindo apenas para orientar as equipes a respeito de sua evolução em todo o processo.
- j) Repete-se este procedimento até que cada equipe tenha examinado 10 pessoas de cada grupo etário.

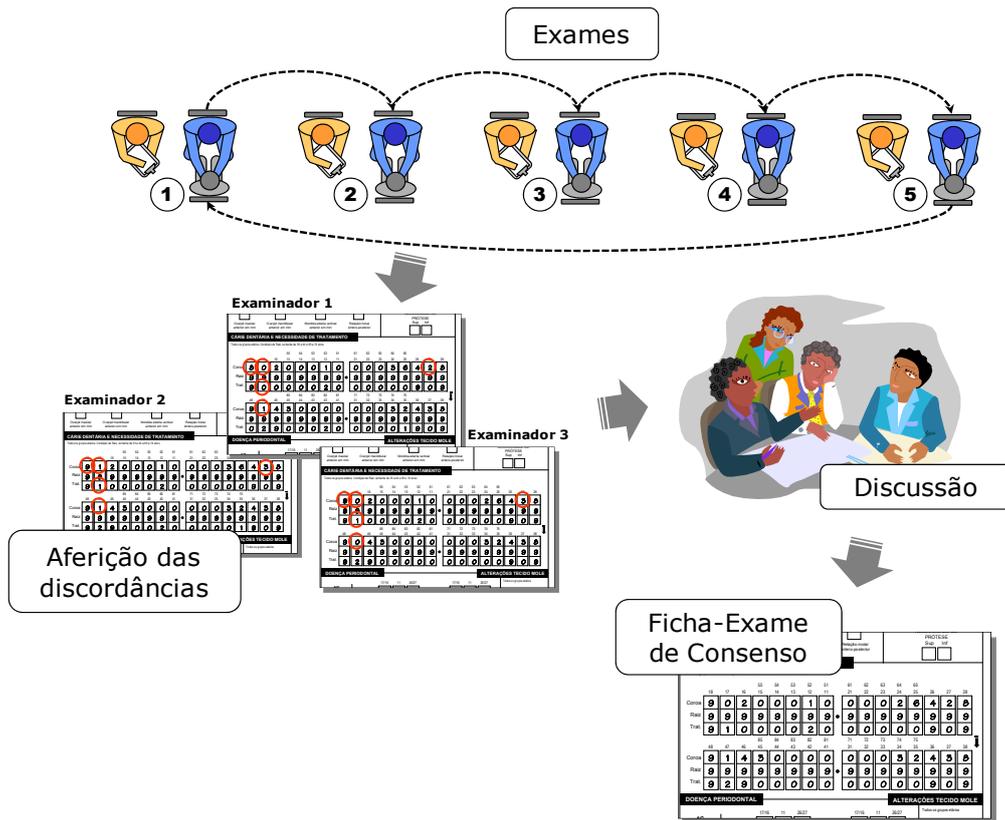


Figura 1. Esquema do processo de discussão prática. Exemplo para um bloco de 5 examinadores.

Importante ressaltar que esta é a fase da calibração em que se busca, ao máximo, a assimilação uniforme de todos os critérios e regras de decisão. Portanto, só deve ser concluída após os examinadores terem assimilado todos os critérios, pois somente desta forma se pode passar à próxima fase, onde a concordância entre os examinadores é medida estatisticamente.

3.4. Calibração propriamente dita (14 horas)

A calibração deverá ser feita da mesma maneira que o exercício anterior, exceto pelo número de pessoas examinadas, que deve ser maior (em torno de 20 a 25 de cada grupo etário).

O princípio básico é que, nesta fase, todas as eventuais discrepâncias deverão ocorrer não mais por interpretações equivocadas dos critérios, pois estas, em sua maioria, já devem ter sido resolvidas na fase anterior, mas única e exclusivamente pela visão peculiar de cada examinador. Os dados encontrados nesta fase é que servirão de base para os cálculos de concordância interexaminador descritos adiante.

O procedimento é semelhante ao item anterior (observe na Figura 1):

- a) Formam-se blocos de, no máximo, cinco equipes (examinador e anotador).
- b) São chamados cinco voluntários para ocupar as cinco cadeiras (se for o caso).
- c) Cada examinador deve passar por cada um dos cinco voluntários e realizar o exame.
- d) Ao final da rodada, os voluntários não devem ser dispensados ainda. Os examinadores, orientados pelo instrutor, deverão examinar suas fichas e observar as discordâncias em cada um dos cinco exames que cada um realizou.
- e) Ao serem observadas discordâncias, todas são discutidas exaustivamente de modo a chegar a um consenso sobre o diagnóstico ao qual todos concordam.
- f) Após chegar ao consenso com relação a todas as discordâncias de um determinado voluntário, o instrutor de calibração preenche uma ficha, chamada de *Ficha Padrão*, a qual servirá como referência para os cálculos de concordância. Assim, para cada grupo de cinco fichas de um determinado voluntário (referente ao exame por parte dos cinco examinadores), uma sexta ficha, resultado do consenso, deve ser preenchida.
- g) Ao final de uma rodada, supondo uma equipe de cinco examinadores, deverá haver seis fichas para cada criança (sendo uma a Ficha Padrão), totalizando 30 fichas.
- h) Novas rodadas são conduzidas, recomeçando o processo desde o item "a", de modo a termos, no final, pelo menos 20 voluntários examinados para cada bloco de cinco examinadores.
- i) Ao final, o instrutor deve preencher as planilhas para o cálculo das concordâncias (ver adiante instruções sobre uso das planilhas).

3.5. Discussão Final da Calibração (2 horas)

A última parte do exercício de calibração deve ser usada para se certificar de que a equipe de examinadores está completamente familiarizada com todos os procedimentos de exame e de registro, critérios de diagnóstico, formulários de registro e o manejo de instrumentos e materiais. Também são discutidos os resultados dos cálculos de concordância. No caso de serem verificados valores altos de discrepância (veja item 7 adiante), decide-se pela repetição do exercício e/ou substituição de algum examinador da equipe.

O último turno (4 horas) deve ser dedicado à discussão com a equipe das estratégias a serem desenvolvidas em campo, tais como o processo de percurso, o uso do PDA, o processo de supervisão, dentre outras.

Um bom resumo de todas estas atividades encontra-se no quadro a seguir.

Quadro 1. Resumo das atividades do processo de treinamento e calibração de examinadores.

	Atividade	Nº Horas	Local	Índices	Idades	Nº Voluntários
1º Turno	Discussão Teórica	4	Sala de Aula	Todos	-	
2º Turno	Discussão Prática	2	Escolas de Ensino Fundamental	Cárie, e Condição de oclusão	5 anos	10
		2	Escolas de Ensino Fundamental e Médio	Cárie, Fluorose, CPI e Condição de Oclusão	12 a 19	10
3º Turno	Discussão Prática	2	Unidade de Saúde, Escola ou Fábrica	Cárie	30 anos e mais	10
		2	Unidade de Saúde, Escola ou Fábrica	CPI Cálculo, CPI Bolsa	30 anos e mais	10
4º Turno	Calibração	4	Escolas de Ensino Fundamental e Médio	Todos	12 a 19	De 20 a 25 em cada idade / grupo etário
5º Turno	Calibração	4	Escola de Ensino Fundamental, Unidade de Saúde, Escola ou Fábrica	Todos	5 anos e 30 anos e mais	
6º Turno	Calibração	4	Escolas de Ensino Fundamental e Médio	Todos	12 a 19	
7º Turno	Calibração	2	Unidade de Saúde, Escola ou Fábrica	Todos	30 anos e mais	
	Discussão Final	2				
8º Turno	Discussão das estratégias do levantamento em campo					

4. Calculando a concordância interexaminador

4.1. A Concordância Simples ou Concordância Percentual

O processo para o cálculo da concordância interexaminador é relativamente simples. Parte-se do princípio de que, de um conjunto de diagnósticos dados aos dentes (ou gengivas, ou esmalte, a depender do índice), por dois examinadores, teremos inevitavelmente alguns destes diagnósticos que não são coincidentes. O grau de concordância nada mais é do que a razão entre o número de diagnósticos coincidentes e o total de diagnósticos.

De acordo com a técnica preconizada pelo SBBrasil 2010, será calculada a concordância de cada examinador com o examinador padrão. Este "examinador" na verdade não existe, ele é o produto do consenso, obtido nos exames realizados em cada voluntário, conforme descrito nos itens 3.1 a 3.5. Para cada grupo de (no máximo) cinco examinadores existirá, portanto, um "examinador-padrão".

Para compreender esse processo, antes de iniciarmos as instruções de uso da planilha eletrônica, vejamos como seria o cálculo de concordância realizado de forma manual. Tudo começa com uma tabela 2 x 2, em que os códigos do índice que está sendo avaliado são colocados nas linhas e colunas. Se estabelece que o examinador padrão terá seus dados nas colunas e um determinado examinador (o examinador 1, por exemplo) terá seus dados expressos nas linhas. Veja na Figura 2, como isso funciona para o Índice de Traumatismo.

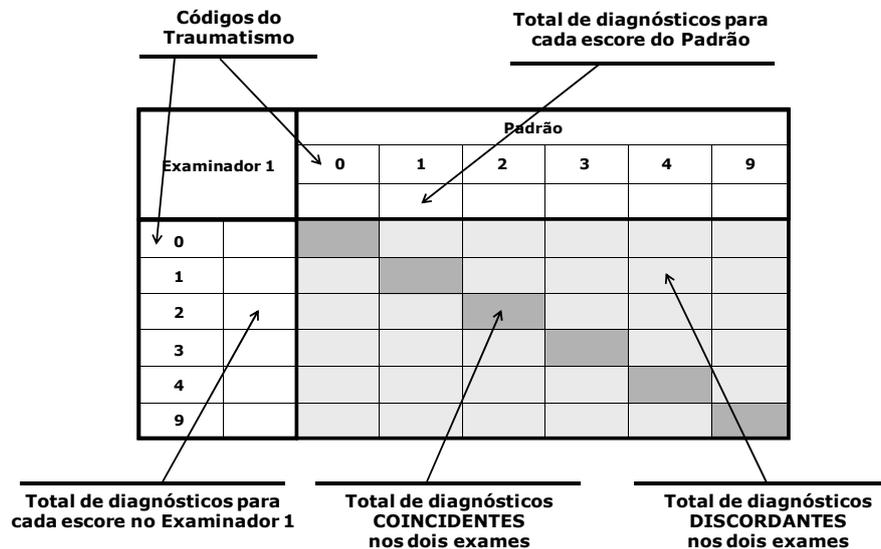


Figura 2. Esquema de uma Tabela 2 x 2 para cálculo de concordância.

Numa situação hipotética em que 30 diagnósticos fossem dados por um examinador e os mesmos fossem comparados com o padrão, poderíamos ter a situação ilustrada na Tabela 1.

Tabela 1. Exemplo de uma tabela 2 x 2 comparando os diagnósticos de um examinador em relação ao padrão.

		Padrão						
		0	1	2	3	4	9	
Examinador 1 N = 30	0	3	13	5	9	0	0	
	0	5	2	3	-	-	-	-
	1	12	1	10	1	-	-	-
	2	5	-	-	3	2	-	-
	3	8	-	-	1	7	-	-
	4	0	-	-	-	-	-	-
	9	0	-	-	-	-	-	-

Obs: A concordância do examinador é indicada pelas células na diagonal (em negrito)

Pode-se observar que, dos 30 diagnósticos, 22 foram coincidentes (a soma dos valores da diagonal, 2 + 10 + 3 + 7). A taxa de concordância é dada por $22/30 = 0,73$. Em outras palavras, o examinador 1 concordou em 73% dos diagnósticos com o padrão.

4.2. O Coeficiente Kappa

Tem-se admitido que o dado percentual pode mostrar um valor demasiado otimista, isto porque, se o examinador segue examinando aleatoriamente, de acordo com as probabilidades das categorias do índice, algum nível de concordância pode se dar devido ao acaso. Com o propósito de ajustar o percentual de concordância descontando esta concordância do acaso, utilizam-se as frequências para calcular a faixa de concordância que seria, em média, uma ocorrência do acaso, "corrigindo" assim a estatística de concordância percentual. É o que se chama de *Coeficiente Kappa** o qual é definido como:

$$\text{Coeficiente Kappa} = \frac{\text{Concordância observada} - \text{Concordância Esperada}}{N - \text{Concordância Esperada}}$$

A vantagem usualmente reconhecida no Coeficiente Kappa é sua capacidade de remover da concordância percentual aquelas concordâncias que provavelmente foram devidas ao acaso, trazendo uma distribuição mais global das condições dentro dos indivíduos examinados.

No caso do exemplo ilustrado na Tabela 1, a concordância observada é igual a 0,73 e a concordância esperada é dada por: $[(3 \times 5) + (12 \times 13) + (5 \times 5) + (8 \times 9) + (0 \times 0) + (0 \times 0)] / (30)^2$ que é igual a 0,2977 ou 29,78%. O Kappa encontrado, portanto, é igual a **0,62**. Trata-se de um valor menor que o da concordância simples, entretanto mais confiável.

O Coeficiente Kappa pode ser *Simple* ou *Ponderado*. No primeiro caso, todas as discordâncias (entre os códigos 0 e 1 e entre 0 e 2, por exemplo) têm um mesmo peso e contribuem com o mesmo grau de importância para o cálculo do coeficiente.

Esta modalidade do Kappa é mais utilizada quando temos variáveis nominais, ou seja, quando não existe uma ordenação entre as categorias. No caso dos índices utilizados em pesquisas epidemiológicas de saúde bucal, a maioria é de natureza ordinal, ou seja, temos uma hierarquia natural (relativa à gravidade do evento) entre as categorias do índice.

Continuando com o exemplo do Traumatismo, uma discordância entre o código 1 (fratura de esmalte) e o código 2 (fratura de esmalte e dentina) pode ser considerada menos comprometedor do que uma discordância entre o mesmo código 1 e o código 3 (fratura de esmalte e dentina com exposição pulpar). Neste caso, o peso da primeira discordância deve ser menor do que o da segunda e, com base neste raciocínio, calculamos o *Kappa Ponderado*. O Coeficiente *Kappa ponderado* é calculado pela mesma expressão do Coeficiente *Kappa simples*, apresentado acima. Entretanto, o valor de cada célula é multiplicado por um peso estatístico arbitrário que varia conforme o número de categorias da variável em análise. A Concordância observada então é calculada: a) Multiplicando-se o número de cada célula pelo correspondente "peso", b) somando-se os produtos obtidos e c) dividindo-se pelo total (n= 30 no exemplo do Traumatismo). A concordância esperada é calculada: a) multiplicando-se o total correspondente das células com o mesmo "peso", b) Somando-se esses produtos e multiplicando-se a soma com o "peso" correspondente e, finalmente c) somando-se tudo e dividindo pelo grande total ao quadrado (n= 30). No exemplo do Traumatismo o valor do Coeficiente *Kappa ponderado* é igual 0,765 ou 76,5%.

(*) Maiores discussões de caráter teórico sobre o Kappa bem como o detalhamento do seu processo de cálculo podem ser encontrados na versão anterior do Manual de Calibração do Projeto SB2000 (disponível em www.saude.gov.br/bucal) e também na publicação da OMS *Calibration of examiners for oral health epidemiological surveys*, descrita nas referências.

4.3. A Concordância por Escore

Uma extensão recomendada para o percentual de concordância global é o cálculo da concordância de cada escore em cada índice, como um complemento para a concordância global do índice. Continuando com o exemplo da Tabela 1, a concordância global é de 0,73, entretanto, se tomarmos somente o escore "0" por exemplo, veremos que o valor é de 0,33. Ou seja, em termos globais, pode-se considerar como razoável o valor encontrado, porém, especificamente com relação ao escore "0" o valor da concordância foi bem baixo. Este tipo de análise é particularmente útil para o instrutor de calibração orientar as equipes em treinamento no que diz respeito a *onde* estão ocorrendo as discordâncias.

5. Utilizando a planilha eletrônica

O cálculo manual da concordância é trabalhoso, pois exige o preenchimento de tantas tabelas quantas forem necessárias para avaliar o grau de concordância. O cálculo manual do Coeficiente Kappa é igualmente complicado de se fazer, exigindo uma série de operações matemáticas.

Para auxiliar o instrutor de calibração nesta tarefa foram elaboradas planilhas eletrônicas no Programa Excel® da Microsoft Corporation, versão 2007. Cada arquivo contém planilhas específicas para cada situação de saúde bucal a ser avaliada e são nomeados de modo característico (o arquivo para avaliação de condição de coroa é, por exemplo, "Calibracao_Carie_Permanente.xlsx"). As planilhas estão disponíveis no sítio do Projeto SBBrazil e serão fornecidas aos instrutores por ocasião das oficinas de treinamento.

Uma vez que o exercício de calibração foi concluído, todas as fichas devem ser agrupadas, separando-as por examinador. Este agrupamento é importante, pois facilita a digitação dos dados, visto que as planilhas estão com as entradas de dados organizadas desta forma (veja esquema na Figura 3).

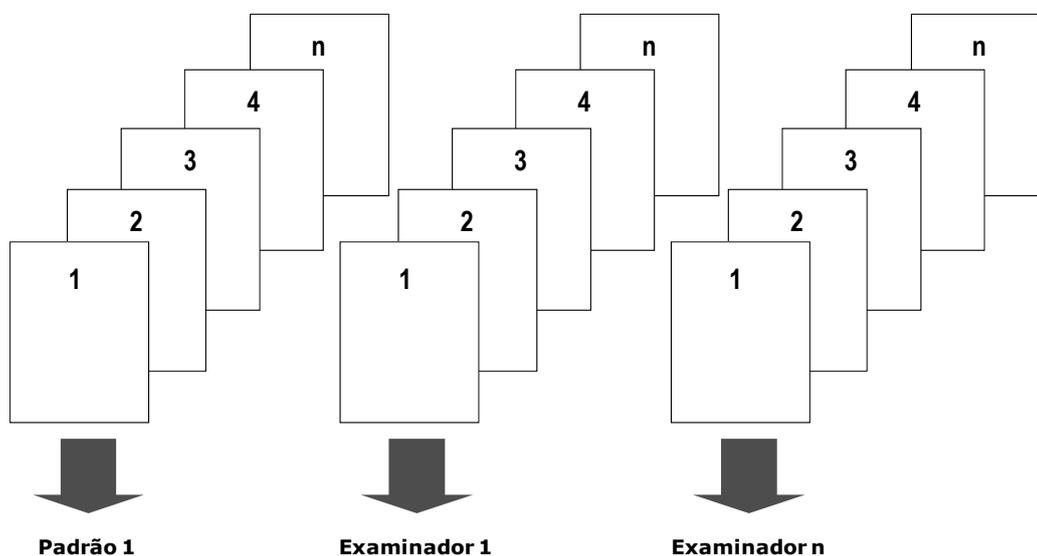


Figura 3. Esquema ilustrando o modo de agrupamento das fichas de calibração.

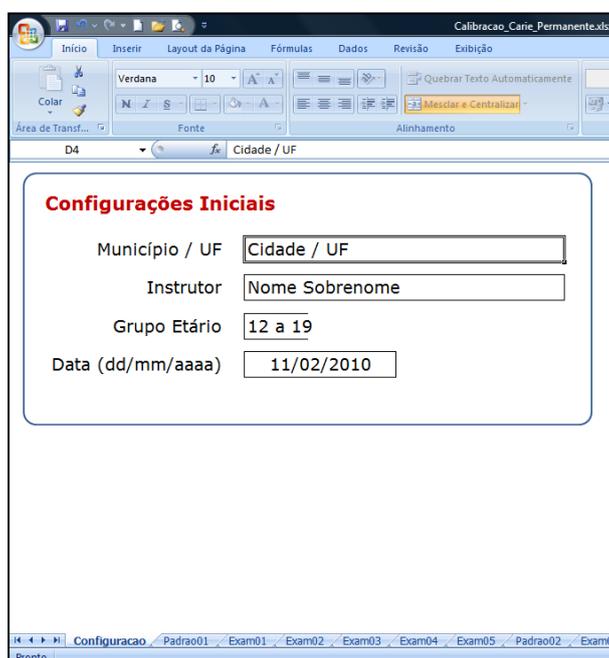
Agrupadas as fichas, procede-se à entrada de dados nas planilhas. A primeira coisa a fazer é salvar cada planilha com um novo nome, acrescentando, por exemplo, o nome do município, para que o modelo de arquivo original seja preservado. Abre-se o arquivo desejado (vamos exemplificar com o arquivo Calibracao_Carie_Permanente.xlsx) e, em seguida, clica-se em "Arquivo" e escolhe "Salvar como...". Em seguida identifica-se o

arquivo com o nome do município (Calibracao_Carie_Permanente_Brasilia.xlsx, por exemplo). Dessa forma, os arquivos ficarão identificados com o índice que está sendo avaliado e a cidade na qual a calibração foi feita.

5.1. Funcionamento das pastas de trabalho

Cada arquivo do Excel apresenta diversas planilhas. Exemplificaremos o funcionamento delas a partir do arquivo utilizado para a calibração de cárie dentária em dentes permanentes (*Calibracao_Carie_Permanente.xlsx*).

A primeira planilha, nomeada como "Configuração", deve ser preenchida com os dados gerais da calibração (Cidade/UF, nome do instrutor, o grupo etário que está sendo avaliado e a data). O preenchimento desta primeira planilha é importante, pois estes dados ficarão presentes em todas as outras planilhas, facilitando a identificação e a confecção do relatório final.



Configurações Iniciais	
Município / UF	Cidade / UF
Instrutor	Nome Sobrenome
Grupo Etário	12 a 19
Data (dd/mm/aaaa)	11/02/2010

Figura 4. Exemplo do preenchimento da planilha inicial de configuração.

Em seguida, todas as outras planilhas, com os dados presentes nas fichas devem ser preenchidas, começando pela planilha "Padrão01" a qual é resultante de todas as fichas de consenso, conforme detalhado anteriormente. Depois, segue-se preenchendo as outras planilhas relativas a cada um dos examinadores (Exam01, Exam02 e assim por diante). O preenchimento da casela relativa ao número da ficha se dá de modo automático, mantendo uma sequência correta, ou seja, todas as fichas "1" (do Padrão e de todos os examinadores) pertencem a um mesmo indivíduo e o mesmo vale para todas as fichas "2" e assim por diante.

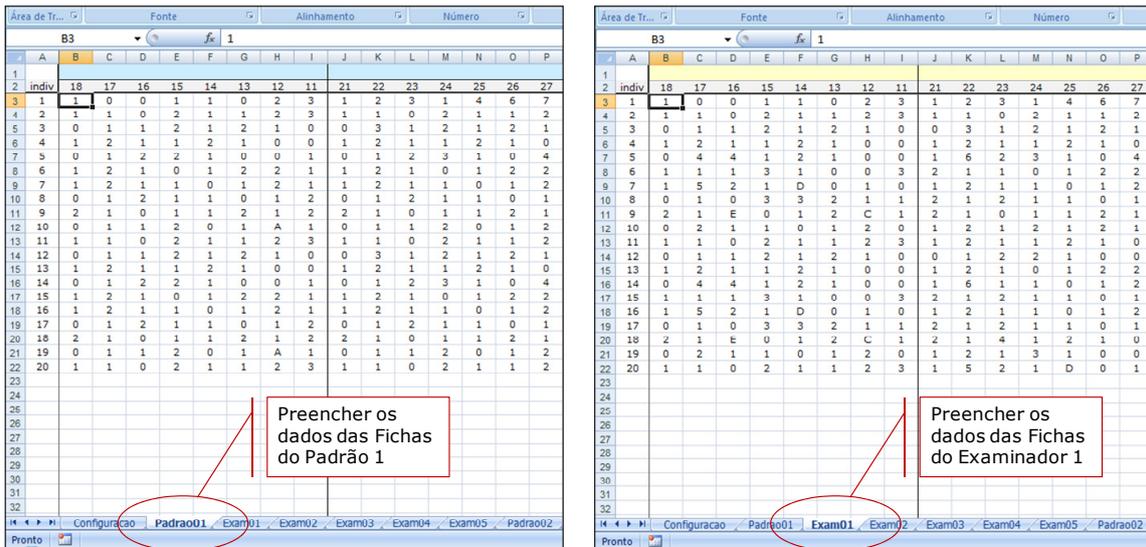


Figura 5. Exemplo do preenchimento dos dados para as fichas padrão e para o examinador 1.

Importante lembrar que, para cada bloco de cinco examinadores, temos um padrão. Assim, se a equipe a ser treinada é formada por, até, 5 examinadores, teremos apenas uma planilha padrão preenchida. Caso tenhamos, por exemplo, 8 examinadores, os examinadores 1 a 4 terão como padrão o “Padrão01”. Deixa-se a planilha “Exam05” em branco e preenche-se a planilha “Padrao02” que servirá como referência para os examinadores 6 a 9 (Exam06, Exam07, Exam08 e Exam09). Ou seja, os blocos se organizam como no Quadro 2 a seguir.

Quadro 2. Organização das planilhas no arquivo.

Bloco 1	Bloco 2
Padrao01	Padrao02
Exam01	Exam06
Exam02	Exam07
Exam03	Exam08
Exam04	Exam09
Exam05	Exam10

Depois de digitadas todas as fichas de todos os examinadores e do padrão, basta observar as planilhas de resultados (os cálculos são feitos automaticamente), nomeadas com as combinações possíveis entre o padrão e os examinadores. Assim temos as planilhas “Pad1 x Exam1”, “Pad1 x Exam2”, “Pad1 x Exam3”, “Pad1 x Exam4” e “Pad1 x Exam5” para o primeiro bloco e as planilhas “Pad2 x Exam6”, “Pad2 x Exam7”, “Pad2 x Exam8”, “Pad2 x Exam9”, “Pad2 x Exam10” para o segundo bloco. Veja na Figura o exemplo para a combinação “Pad1 x Exam1”.

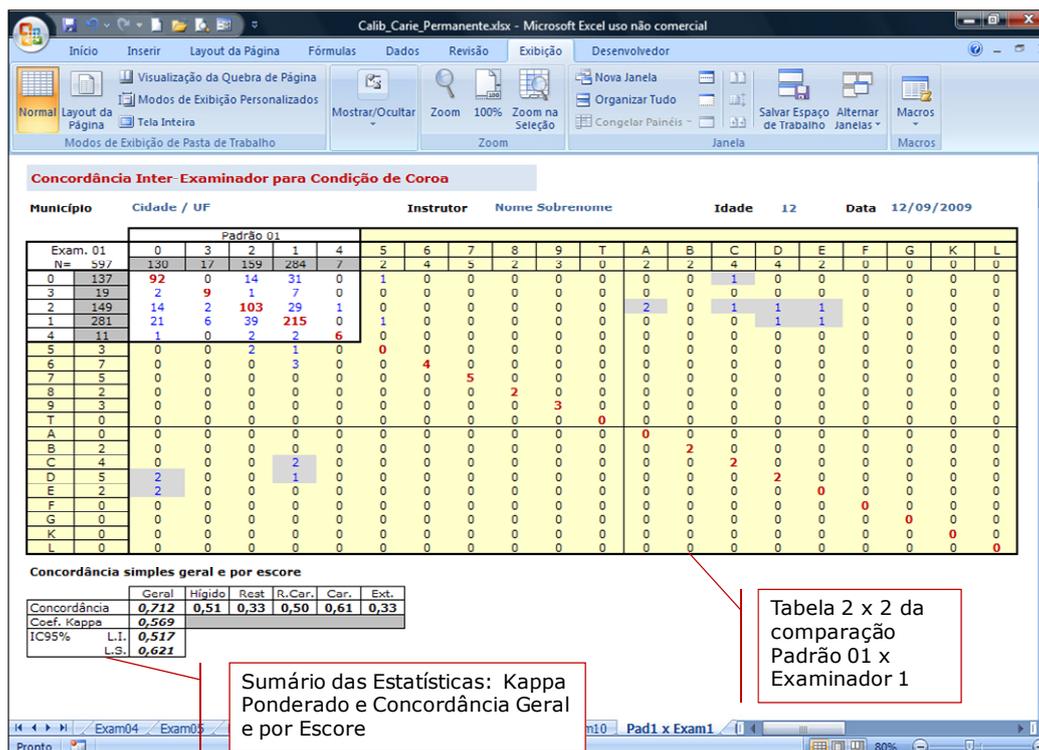


Figura 6. Planilha de resultados. Exemplo para a concordância entre o padrão 1 e o examinador 1.

Pode-se observar que, na planilha temos todas as análises possíveis da concordância entre o examinador e o padrão. O exemplo contém dados fictícios de 20 pessoas e algumas discordâncias foram artificialmente colocadas. Para o caso específico de cárie dentária, a tabela 2 x 2 apresenta a relação entre todos os códigos e critérios do CPO para que o instrutor possa avaliar onde residem as discordâncias.

O destaque dado pelas células em cinza chama a atenção para um tipo de discordância inaceitável, que é a confusão entre os códigos para dentes decíduos e permanentes. Em uma situação real, deve-se estar atento para que situações como esta sejam resolvidas ainda na fase de treinamento.

As células em amarelo indicam as caselas que não são levadas em conta para o cálculo das estatísticas de concordância (estas estão indicadas em branco). Considera-se, portanto, para efeito de cálculo das estatísticas de concordância, somente os escores "0" (Hígido), "1" (Cariado), "2" (Restaurado e Cariado), "3" (Restaurado) e "4" (Extraído). A ordem colocada na planilha (0, 3, 2, 1 e 4) indica a escala nominal dada pela gravidade crescente dos códigos do CPO para que seja calculado o Coeficiente Kappa ponderado.

A tabela 2 x 2, portanto, serve como um guia para o instrutor discutir com a equipe e fazer as devidas correções. Abaixo da tabela podem ser observados os resultados dos cálculos para a concordância simples (geral e por escore) e do Coeficiente Kappa ponderado e seu respectivo intervalo de confiança de 95%.

Tomando como exemplo os resultados da Figura 6, pode-se dizer que o examinador concorda em 71,2% dos diagnósticos com o padrão. Corrigindo pelo Kappa ponderado, esta concordância é de somente 56,9% (com um intervalo de confiança de 51,7% a 62,1%). A pior concordância é entre os escores "Restaurado" e "Extraído", com 33% e a melhor concordância está o escore "Cariado" (61%).

Finalmente, depois da entrada de todos os dados e da avaliação de todas as concordâncias pelas planilhas de resultados, o instrutor deverá observar a última planilha, denominada "Sumário", onde todos os resultados, de todas as concordâncias, estão sintetizados. Esta última planilha deverá ser o documento final que será impresso e anexado ao relatório de treinamento e calibração (veja exemplo na Figura 7).

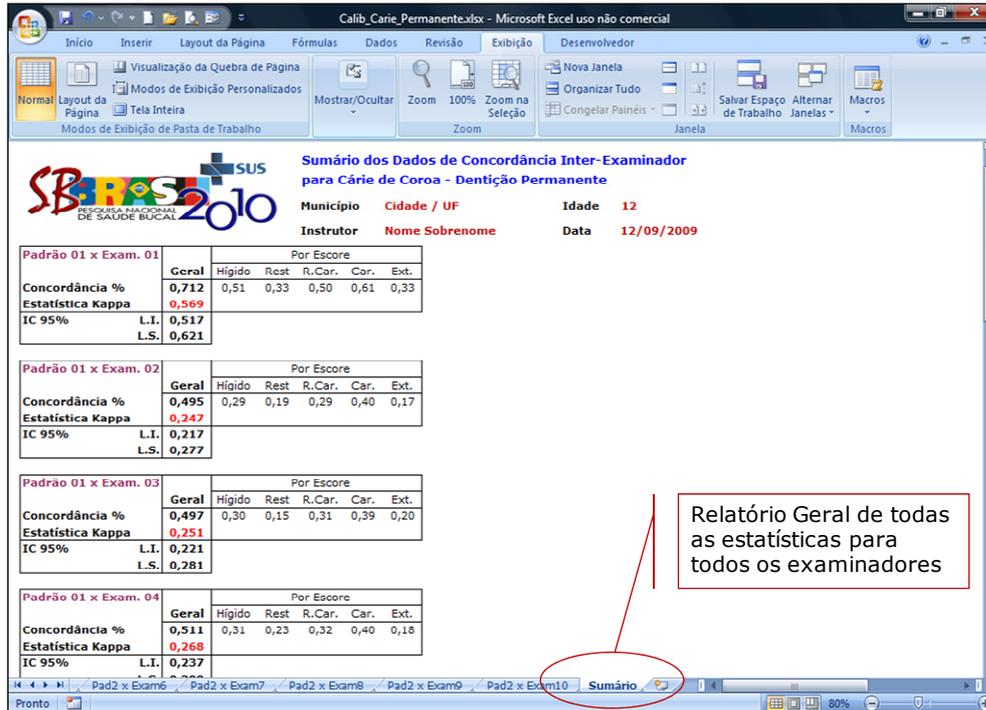


Figura 7. Exemplo da planilha do sumário de resultados.

Todos os outros arquivos em Excel para o cálculo das concordâncias possuem a mesma lógica. Sempre deveremos entrar com os dados dos examinadores e do padrão e depois avaliar os resultados. Apenas para exemplificar, observe como se estrutura a planilha para cálculos do Índice de Estética Dental (DAI).

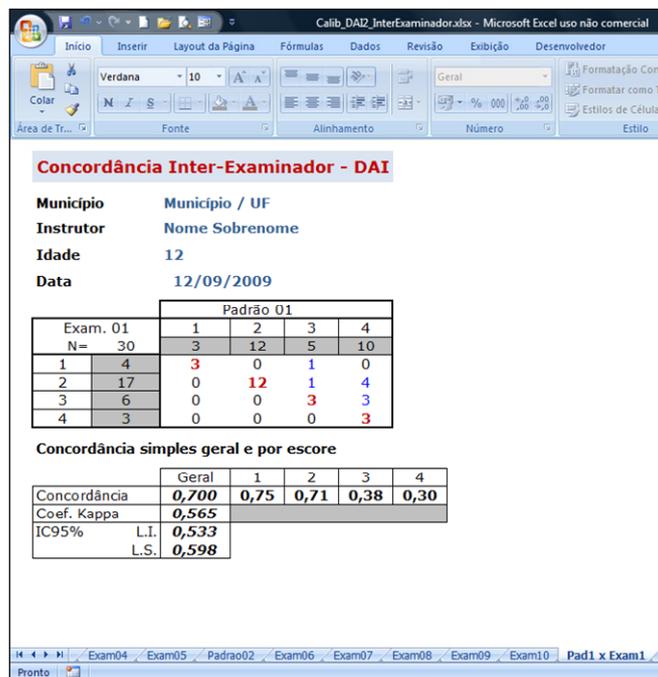


Figura 8. Exemplo da planilha de resultados para o DAI.

6. Verificação da reprodutibilidade em campo

A verificação da reprodutibilidade intra-examinador deverá ser realizada em campo. O examinador deverá reexaminar aproximadamente **5% de toda a amostra**, tendo o cuidado de marcar na ficha que o exame é duplicata. Uma vez que a coleta será realizada eletronicamente, o *software* do PDA deverá calcular automaticamente a concordância para os mesmos índices utilizados na calibração inter-examinador. O procedimento é o seguinte:

- a) Após examinar cerca de 20 indivíduos, ainda no mesmo setor, o examinador pede para o anotador escolher, na Folha de Arrolamento, um domicílio e um indivíduo a ser reexaminado.
- b) O anotador, ao fazer a escolha, de preferência entre os primeiros do grupo dos 20, circula a casela do mesmo na Folha de Arrolamento.
- c) A equipe retorna ao domicílio e solicita que seja realizado novo exame no indivíduo escolhido. O novo exame é realizado, tendo o cuidado de, no PDA, **colocar o mesmo número de domicílio que se encontra na folha de arrolamento** e marcar exame em "duplicata" em vez de "original".
- d) Este procedimento deve ser repetido a cada bloco de aproximadamente 20 indivíduos examinados, tendo-se o cuidado de alternar entre as várias idades

Após todos estes procedimentos, quando da sistematização de todos os bancos de dados dos PDAs, a verificação da concordância intra-examinador será aferida pelo coordenador municipal ou pelo Centro Colaborador. Caso esta consistência apresente valores globais abaixo de 85% os dados do examinador serão descartados e uma nova equipe será composta para realizar os exames.

7. Valores recomendados para a Concordância

Embora não haja consenso na literatura sobre qual o valor mínimo de concordância considerado adequado, a grande maioria dos autores reconhece o valor de kappa de 0,60 como o valor mínimo aceitável. Está sendo recomendado para este estudo um limite de concordância mínimo de **0,65** para o Kappa ponderado. Obviamente, este é um valor limite a ser considerado para a não aceitação de um examinador como membro da equipe, contudo, deve-se buscar o maior valor de concordância possível para todos os membros da equipe.

Vale lembrar que o Coeficiente Kappa vai de "0" a "1", onde "0" significa absoluta discordância entre os examinadores e o "1" a absoluta concordância entre os mesmos, esta última uma situação obviamente idealizada, mas na prática impossível quando tratamos com padrões diagnósticos onde a subjetividade é um fator não completamente controlado.

8. Relatório

Ao final de todo o exercício de calibração dos examinadores, o instrutor de calibração deverá elaborar o Relatório de Calibração. Esse documento técnico é fundamental para o conhecimento do grau de confiabilidade dos dados coletados. Nesse relatório devem constar todas as dificuldades surgidas durante o processo, bem como as tabelas com os resultados de concordância inter e intra-examinadores. No anexo encontra-se um modelo de relatório de calibração.

9. Bibliografia

1. ALTMAN, D. Practical Statistics for Medical Research. London:Chapman and Hall, 1991. p. 406-7.

2. BERQUÓ, E.S., SOUZA, J.M.P., GOTLIEB, S.L.D. Bioestatística. 2ed. São Paulo: EPU, 1981.
3. BULMAN, J.S., OSBORN, J.F. Measuring diagnostic consistency. Br Dent J. v. 166. p. 377-81. 1989.
4. CONS, N. C. et al. Utility of the dental aesthetic index in industrialized and developing countries. J Pub Health Dent. v. 49, n. 3, 1989, p. 163-6.
5. FRIAS, A. C., ANTUNES, J.L.F., NARVAI, P.C. Precisão e validade de levantamentos epidemiológicos em saúde bucal: cárie dentária na cidade de São Paulo, 2002. Rev. Bras. Epidemiol. v. 7, n. 2, 2004. P.144-154.
6. LANDIS, J.R., KOCH, G.G. The measurement of observer agreement for categorical data. Biometrics. v. 33, p. 159-174. 1977.
7. LUIZ, R.R. Erros de mensuração. In: LUIZ, R.R.; COSTA, A.J.L.; NADANOVSKY, P. Epidemiologia e bioestatística na pesquisa odontológica. São Paulo: Atheneu, 2005. Cap.5, p.91-124.
8. PERES, M. A.; TRAEBERT, J.; MARCENES, W. Calibração de examinadores para estudos epidemiológicos de cárie dentária. Cad. Saúde Pública v.17, n. 1, 2001, p. 153-159.
9. PINTO, V.G. Saúde bucal coletiva. São Paulo: Santos, 2000. 541p.
10. PITTS, N. Diagnostic tools and measuments - impact on appropriate care. Comm Den Oral Epidemiol. v. 25, p. 24-35. 1997.
11. SZKLO, M., JAVIER NIETO F. Epidemiology. Beyond the basics. Sudbury, Massachussets: Jones and Bartlett Publishers, 2004. p363-377.
12. UNIVERSIDADE DE SÃO PAULO, Faculdade de Saúde Pública, Secretaria de Estado da Saúde de São Paulo. Levantamento das Condições de Saúde Bucal - Estado de São Paulo, 1998. Caderno de Instruções. São Paulo, 1998. [mimeo]
13. VIEIRA, S. Introdução à bioestatística. Rio de Janeiro: Campus, 1981. 293p.
14. WORLD HEALTH ORGANIZATION. Calibration of examiners for oral health epidemiological surveys. Geneva: ORH/EPID, 1993.
15. WORLD HEALTH ORGANIZATION. Oral health surveys: basic methods. 4 ed. Geneva: ORH/EPID, 1997.

Anexo 1

Modelo de relatório de calibração

Centro Colaborador: _____

Município / UF: _____

Local do treinamento: _____

Participantes

Nome	CRO	R.G.	Residência	Telefones	Função(*)

(*) Instrutor, Examinador

Atividades realizadas

Indicar local, data, horário, tipo de atividade, além do número de exames e idade/faixa etária utilizada.

Resultados

Transcrever as tabelas com os valores dos cálculos de concordância intra e interexaminador segundo as condições examinadas. Anexar as planilhas de sumário do arquivo Excel, caso tenha sido utilizada.

Observações adicionais

Indicar as facilidades/dificuldades encontradas, e outros aspectos importantes para a compreensão das circunstâncias do treinamento.

Data e Assinatura do instrutor

O instrutor é o responsável técnico pelas informações contidas no relatório