

UNIVERSIDADE VEIGA DE ALMEIDA

Liliane Mouta

**REPRODUTIBILIDADE DO EXAME HISTOLÓGICO COMO
CRITÉRIO DE VALIDAÇÃO PARA MÉTODOS DE
DIAGNÓSTICO DE CÁRIE**

Rio de Janeiro
2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Liliane Mouta

**REPRODUTIBILIDADE DO EXAME HISTOLÓGICO COMO
CRITÉRIO DE VALIDAÇÃO PARA MÉTODOS DE
DIAGNÓSTICO DE CÁRIE**

Dissertação apresentada ao curso de pós-graduação em Odontologia da Universidade Veiga de Almeida, como parte dos requisitos para obtenção do Grau de Mestre. Área de concentração: Reabilitação Oral.

Orientadora: Prof^a. Dr^a. Denise Fonseca Côrtes

Rio de Janeiro
2006

P 436r Mouta, Liliane

Reprodutibilidade do exame histológico como critério de validação para métodos de diagnóstico de cárie. / Liliane Mouta, 2006.

132p.; 30cm

Tese (Mestrado) – Universidade Veiga de Almeida,
Mestrado Profissionalizante em Odontologia, Reabilitação Oral,
Rio de Janeiro, 2006.

Orientação: Prof^a Denise Fonseca Côrtes.

1. Cáries dentárias. 2. Diagnóstico. I. Côrtes, Denise Fonseca (orientadora). II. Universidade Veiga de Almeida, Mestrado Profissionalizante em Odontologia, Reabilitação Oral. III. Título.

CDD – 617.67

Ficha catalográfica elaborada pela Biblioteca Central/UVA
Biblioteca Maria Anunciação Almeida de Carvalho

Liliane Mouta

REPRODUTIBILIDADE DO EXAME HISTOLÓGICO COMO CRITÉRIO DE VALIDAÇÃO PARA MÉTODOS DE DIAGNÓSTICO DE CÁRIE

Dissertação apresentada ao curso de pós-graduação em Odontologia da Universidade Veiga de Almeida, como parte dos requisitos para obtenção do Grau de Mestre. Área de concentração: Reabilitação Oral.

Aprovada em 28 de abril de 2006.

Banca Examinadora:

Prof. João Galan Jr. – UVA

Prof^ª. Fátima Maria Namen D'Ávila – UVA

Prof^ª. Ivete Pomarico Ribeiro de Souza – UFRJ

A Deus, Senhor da vida e fonte
de toda a sabedoria humana.

Aos meus pais que me proporcionaram toda minha formação e me ensinaram os principais valores da vida. A eles o meu eterno agradecimento.

Ao meu marido Marcelo Sobral e ao meu filho Paulo Guilherme, com a esperança de poder devolver-lhes os momentos de convivência perdidos.

À minha orientadora, Prof^a. Denise Fonseca Côrtes, pessoa tão especial, que mais do que tudo se mostrou amiga. A ela meu reconhecimento e admiração pelo amor com que tem se dedicado à profissão que abraçou.

À Angela Maria do Couto Martins, minha dupla constante e amiga de todas as horas, que muito me ensinou, principalmente a lutar e a vencer barreiras.

À Roselene Trevisan Nespoli Jacintho, a terceira componente do “trio inseparável”, que com seu jeitinho meigo estará para sempre em meu coração.

Ao Coordenador do Mestrado Profissionalizante em Odontologia, Prof. Cláudio Pinheiro Fernandes, por sua garra, espírito empreendedor e inteligência incomparável. Pessoa que soube fazer de um sonho, realidade. A ele todo o meu respeito.

Ao Prof. João Galan Jr., símbolo de competência, dedicação e simplicidade. A ele o meu carinho especial.

À Prof^a. Fátima Maria Namen D’Ávila, em reconhecimento ao maravilhoso trabalho desenvolvido na elaboração do Manual que nos guiou na formatação das dissertações.

Ao Prof. Achilles Dutra por ter aberto as portas do Laboratório de Eletroquímica Aplicada da COPPE (UFRJ), sob sua responsabilidade, tendo possibilitado a realização da parte experimental deste trabalho.

Ao Carlos Falcão Júnior, aluno de doutorado da UFRJ, sempre tão solícito às minhas necessidades para a utilização do laboratório na COPPE. A ele o meu sincero agradecimento.

A Prof^a. Adriana Terezinha Novellino, por ter sacrificado o seu pouco tempo livre para participar com tanta boa vontade de nossa parte experimental.

A Juliana Guerra e Renata Araújo que tive o prazer de conhecer durante o desenvolvimento do trabalho. Agradeço pela disponibilidade e participação na metodologia. É sempre bom fazer novas amizades!

A Luciene Carvalho e ao Dr. Pedro Carvalho Rodrigues que me auxiliaram, com extrema boa vontade, na estatística deste trabalho. Agradeço a vocês, de todo o coração!

A todos os professores que passaram por este curso tão marcante e que de alguma forma, certamente, mudaram a minha vida. O meu muito obrigada!

Ao meu querido amigo Agostinho Artur de Queiróz Pereira Filho, que muito me ensinou e a quem devo o meu despertar para o magistério.

A todos os meus colegas do Mestrado, componentes de uma turma inesquecível, que certamente terão lugar especial em meu coração. Foi muito bom conhecer cada um de vocês!

“Se metade do orçamento dos gastos militares no mundo fosse investido na educação, os generais se tornariam jardineiros; os policiais, poetas; os psiquiatras, músicos. A violência, a fome, o medo, o terrorismo e os problemas emocionais estariam nas páginas dos dicionários e não nas páginas da vida ...”

Augusto Cury

RESUMO

Este estudo teve como objetivo avaliar a reprodutibilidade do exame histológico tradicionalmente utilizado como método de validação para métodos de diagnóstico de cárie e o impacto da experiência prévia e da calibração dos examinadores na mesma. Secções histológicas previamente preparadas (350 μm) em 123 sítios oclusais de 89 molares permanentes extraídos foram examinadas independentemente por três examinadores através de estereomicroscópio (aumento de 4X). Um examinador era cirurgião-dentista e histologista (examinador A) e os outros dois eram cirurgiões-dentistas que não tinham experiência prévia com o exame histológico (examinadores B e C). Exames das secções histológicas foram realizados antes e após calibração dos examinadores utilizando sete escores de profundidade de lesão de cárie. Em ambos os exames pré e pós-calibração, uma segunda avaliação de toda amostra foi executada uma semana após a primeira para avaliação da reprodutibilidade intra-examinador. Na etapa pré-calibração os examinadores receberam somente instruções escritas sobre os critérios de diagnóstico. Um examinador treinado externo conduziu a sessão de calibração com os três examinadores utilizando 40 secções histológicas. Os resultados da reprodutibilidade intra-examinador pelo índice de Kappa de Cohen no ponto de corte esmalte para os examinadores A, B e C foram, respectivamente, 0,73, 0,71, 0,58 antes da calibração e 0,63, 0,61, 0,56 após calibração. No ponto de corte em dentina os resultados foram 0,79, 0,78, 0,53 antes da calibração e 0,97, 0,86, 0,87 após calibração. Os resultados da reprodutibilidade intra-examinador pelo índice de Kappa Ponderado e índice de Kappa não Ponderado para os examinadores A, B, e C foram, respectivamente, 0,85, 0,85, 0,71 e 0,56, 0,54, 0,45 antes da calibração e 0,74, 0,85, 0,56 e 0,60, 0,53, 0,46 após calibração. Os resultados da reprodutibilidade inter-examinador pelo índice de Kappa de Cohen no ponto de corte esmalte entre os examinadores A-B, A-C e B-C foram respectivamente 0,64, 0,43, 0,55 antes da calibração e 0,47, 0,38, 0,56 após calibração. No ponto de corte em dentina os resultados foram 0,75, 0,33, 0,38 antes da calibração e 0,69, 0,79, 0,73 após calibração. Os resultados da reprodutibilidade inter-examinador pelo índice de Kappa Ponderado e índice de Kappa não Ponderado entre os examinadores A-B, A-C e B-C foram, respectivamente, 0,63, 0,30, 0,50 e 0,29, 0,16, 0,27 antes da calibração e 0,68, 0,52, 0,55 e 0,38, 0,29, 0,40 após calibração. Isto representa uma variação de moderada a quase perfeita para a reprodutibilidade intra-examinador e de substancial a fraca para a reprodutibilidade entre examinadores. Contudo, somente poucas avaliações foram registradas como reprodutibilidade quase perfeita (índice de Kappa > 0,80). Em conclusão, foi verificada uma reprodutibilidade inaceitável e com grande variabilidade na validação histológica tanto intra-examinador quanto entre os examinadores. Tanto a experiência prévia do

examinador quanto o exercício de calibração não melhoraram a reprodutibilidade dos examinadores. O conceito tradicionalmente estabelecido que o exame histológico seria o “padrão-ouro” para validação dos métodos de diagnóstico de cáries deve ser considerado com cautela. É essencial que pesquisas sobre métodos de diagnóstico de cárie relatem a reprodutibilidade dos examinadores dos métodos de validação.

Palavras-chave: Padrão-ouro; diagnóstico de cárie; validação; reprodutibilidade; estereomicroscópio.

ABSTRACT

This study aimed to evaluate the reproducibility of the histological examination traditionally used as validation method for caries diagnostic methods. Previously prepared histological sections (350 μm) of 123 occlusal sites of 89 extracted permanent molars were examined independently by 3 persons using stereomicroscope (4X). An examiner was dentist and histologist (A) and the other two were dentists who had no previous experience in histological examination (B and C). Pre- and post-calibration examination of the histological sections were performed. For both, the second examination of all sample was carried out one week after the first one for evaluation of the reproducibility. The pre-calibration assessment included only written instructions of the diagnostic criteria. Another trained examiner conducted a calibration session with the three examiners using 40 histological sections. The results of the intra-examiner reproducibility using Cohen's Kappa statistics in the enamel threshold for the examiners A, B and C had been respectively 0.73, 0.71, 0.58 before calibration and 0.63, 0.61, 0.56 after calibration. At dentin threshold the results were 0.79, 0.78, 0.53 before calibration and 0.97, 0.86, 0.87 after calibration. Results of intra-examiner reproducibility using Weighted and Unweighted Kappa tests for examiners A, B, and C were respectively 0.85, 0.85, 0.71 and 0.56, 0.54, 0.45 before calibration and 0.74, 0.85, 0.56 and 0.60, 0.53, 0.46 after calibration. Results for inter-examiner reproducibility by Cohen's Kappa statistics at enamel threshold for examiners A-B, A-C and B-C were respectively 0.64, 0.43, 0.55 before calibration and 0.47, 0.38, 0.56 after calibration. At dentin threshold the results were 0.75, 0.33, 0.38 before calibration and 0.69, 0.79, 0.73 after calibration. Results of inter-examiner reproducibility using Weighted and Unweighted Kappa tests for examiners A-B, A-C and B-C were 0.63, 0.30, 0.50 and 0.29, 0.16, 0.27 before calibration and 0.68, 0.52, 0.55 and 0.38, 0.29, 0.40 after calibration. This represents a ranging from moderate to almost perfect for intra-examiner reproducibility and from substantial to weak for inter-examiner reproducibility. However, only few assessments were recorded as almost perfect reproducibility (Kappa tests > 0.80). In conclusion, a large unacceptable intra- and inter-examiner reproducibility for histological validation was observed. Examiner experience and the calibration exercise did not improve the examiner reproducibility. The traditional assumption that histological examination is the "gold standard" for evaluation of caries diagnostic methods should be considered with caution. Future studies for caries diagnostics methods should report the reproducibility of the validation methods.

Key words: Gold standard; caries diagnosis; validation; reproducibility; stereomicroscopy.

LISTA DE TABELAS

- Tabela 1 – Modelo para avaliação da validade de um teste diagnóstico, p. 39
- Tabela 2 – Cálculo da sensibilidade, especificidade e valores preditivos, p. 40
- Tabela 3 – Matriz de decisão para o cálculo do índice de Kappa de Cohen, p. 77
- Tabela 4 – Interpretação do Índice de Kappa de Cohen (Landis; Koch, 1977), p. 78
- Tabela 5 – Valores observados por um imaginário examinador A para o cálculo do índice de Kappa não Ponderado, p. 79
- Tabela 6 – Valores esperados de um imaginário examinador A para o cálculo do índice de Kappa não Ponderado, p. 80
- Tabela 7 – Pesos atribuídos para cada par de concordância em uma escala ordinal contínua de sete pontos para o cálculo do índice de Kappa Ponderado, p. 82
- Tabela 8 – Critérios histológicos adotados no exame ao estereomicroscópio, de acordo com a profundidade da lesão cariosa, p. 87
- Tabela 9 – Reprodutibilidade intra-examinador nos exames pré e pós-calibração, nos pontos de corte esmalte e dentina, obtida pelo Índice de Kappa de Cohen e classificada por Landis e Koch (1977), p. 90
- Tabela 10 – Reprodutibilidade intra-examinador nos exames pré e pós-calibração obtida através do Índice de Kappa não Ponderado e classificada por Landis e Koch (1977), p. 91
- Tabela 11 – Reprodutibilidade intra-examinador, nos exames pré e pós-calibração, através do Índice de Kappa Ponderado e classificada por Landis e Koch (1977), p. 92
- Tabela 12 – Reprodutibilidade inter-examinador nos exames pré e pós-calibração, nos pontos de corte esmalte e dentina, obtida através do Índice de Kappa de Cohen, e classificada por Landis e Koch (1977), p. 93
- Tabela 13 – Reprodutibilidade inter-examinador nos exames pré e pós-calibração obtida através do Índice de Kappa não Ponderado e classificada por Landis e Koch (1977), p.94
- Tabela 14 – Reprodutibilidade inter-examinador nos exames pré e pós-calibração obtida através do Índice de Kappa Ponderado e classificada por Landis e Koch (1977), p. 95

LISTA DE FIGURAS

Figura 1 – Cárie de esmalte (CE) na superfície oclusal de molar inferior., p. 33

Figura 2 – Cárie de esmalte e dentina na superfície oclusal de molar inferior, p. 33

LISTA DE SÍMBOLOS

X = indica aumento em número de vezes

μm = micrometro

> = maior do que

% = porcentagem

LISTA DE ABREVIATURAS E SIGLAS

ROC	– Curva ROC (Receiver Operating Characteristic)
WHO	– World Health Oral = Organização Mundial de Saúde
CPO-D	– Índice de Dentes Cariados, Perdidos e Obturados
ECM	– Electrical Caries Monitor = Monitor Elétrico de Cárie
FOTI	– Fiber Optic Transillumination = Transiluminação por Fibra Ótica
QLF	– Quantitative Light-induced Fluorescence = Fluorescência Quantitativa Induzida por Luz
EUA	– Estados Unidos da América
ACTA	– Academic Centre for Dentistry Amsterdam = Centro Acadêmico para Odontologia de Amsterdam
D1	– Ponto de corte em esmalte
D3	– Ponto de corte em dentina
NIH	– National Institutes of Health = Institutos Nacionais de Saúde dos Estados Unidos
ICW-CCT	– International Consensus Workshop on Caries Clinical Trials = Workshop de Consenso Internacional de Estudos Clínicos sobre Cárie
LF	– Fluorescência por Laser Diodo
K	– Índice de Kappa
P _o	– Proporção de concordância observada
P _e	– Proporção de concordância esperada
C-O	– Escala ordinal contínua
k	– Pontos de critérios diagnósticos da escala ordinal
N	– Tamanho mínimo da amostra

SUMÁRIO

RESUMO

ABSTRACT

LISTA DE TABELAS

LISTA DE FIGURAS

LISTA DE SÍMBOLOS

LISTA DE ABREVIATURAS E SIGLAS

1 INTRODUÇÃO, p. 16

2 REVISÃO DA LITERATURA, p. 22

2.1 DIAGNÓSTICO DA CÁRIE OCLUSAL, p. 22

2.2 HISTOLOGIA DA CÁRIE, p. 27

2.2.1 Cárie de esmalte, p. 27

2.2.2 Cárie de dentina, p. 30

2.3 AVALIAÇÃO DOS MÉTODOS DE DIAGNÓSTICO DE CÁRIE, p. 35

2.4 MÉTODOS DE VALIDAÇÃO DOS MÉTODOS DE DIAGNÓSTICO DE CÁRIE, p. 46

2.5 ANÁLISE ESTATÍSTICA, p. 75

2.5.1 Índice de Kappa de Cohen, p. 76

2.5.2 Índice de Kappa não Ponderado, p. 78

2.5.3 Índice de Kappa Ponderado, p. 80

3 OBJETIVOS, p. 83

4 HIPÓTESES, p. 84

5 MATERIAIS E MÉTODOS, p. 85

5.1 AMOSTRA, p. 85

5.2 EXAMINADORES, p. 85

5.3 EXAME HISTOLÓGICO, p. 86

5.4 ANÁLISE ESTATÍSTICA, p. 88

6 RESULTADOS, p. 90

7 DISCUSSÃO, p. 96

8 CONCLUSÕES, p. 115

9 REFERÊNCIAS BIBLIOGRÁFICAS, p. 116

10 ANEXOS, p. 124

1 INTRODUÇÃO

Somente através de um correto diagnóstico das lesões de cárie pode-se indicar um tratamento adequado para os pacientes. Um erro no diagnóstico de uma lesão restrita ao esmalte, por exemplo, pode resultar em um desnecessário tratamento restaurador e no início de um ciclo restaurador repetitivo (ELDERTON,1985; DOWNER, 1989; FERREIRA et al. 1999). Da mesma forma, lesões cariosas não diagnosticadas teriam impacto pela continuidade do processo carioso.

Há um consenso sobre as dificuldades em se diagnosticar a presença de cáries em cicatrículas e fissuras (WEERHEIJM; van AMERONGEN; EGGINK, 1989; KIDD; RICKETTS; PITTS, 1993; NEWBRUN, 1993; EKSTRAND; RICKETTS; KIDD,1997), especialmente devido à morfologia da região e às limitações dos métodos de diagnóstico rotineiramente utilizados (MOTA et al., 2002). O diagnóstico das cáries oclusais mais avançadas geralmente não representa um problema (ELDERTON,1985; NEWBRUN, 1993) pois, quanto maior a severidade de uma doença, mais óbvio o diagnóstico, pois a mesma vem acompanhada de uma série de sinais e sintomas através dos quais faz-se o diagnóstico diferencial (FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, 1996; FERREIRA et al., 1999). Entretanto essa dificuldade é bem evidente em lesões incipientes confinadas ao esmalte (ELDERTON, 1985; NEWBRUN, 1993). Para estes casos, há a necessidade de

métodos diagnósticos mais sensíveis que sejam capazes de determinar pequenas perdas minerais.

Nas regiões com água fluoretada o diagnóstico da superfície oclusal tem se tornado ainda mais complexo (SAWLE; ANDLAW, 1988), já que os fluoretos promovem remineralização do esmalte, por vezes mascarando lesões em dentina (KIDD; RICKETTS; PITTS, 1993). O exame clínico pode mostrar esmalte aparentemente intacto, mesmo em presença de cárie dentinária subjacente (DOOLAND; SMALES, 1982; KIDD; RICKETTS; PITTS, 1993). Apesar da redução da prevalência da cárie dental, a oclusal tem sido a superfície mais suscetível à doença, possivelmente por ser a área das cicatrículas e fissuras menos protegida pela ação do flúor do que as superfícies lisas (LUSSI, 1991).

Diferentes métodos de diagnóstico para a cárie dental têm sido propostos (MOTA et al., 2002; CÔRTEZ; ELLWOOD; EKSTRAND, 2003), variando desde o exame clínico convencional até métodos mais sofisticados, como a fluorescência induzida pelo laser. Para que se possa medir a acurácia de tais métodos, seus resultados precisam ser validados contra um método confiável, denominado na literatura de “padrão-ouro” (*gold standard*) (PEREIRA, 1995; FERREIRA et al., 1999; ten BOSCH; ANGMAR-MANSSON, 2000).

Padrão-ouro ou *gold standard* seria um método de diagnóstico perfeito para se predizer presença ou ausência de doença (WENZEL; HINTZE, 1999b; HINTZE; WENZEL, 2003). A fim de que o padrão-ouro seja realmente representativo, é necessário o preenchimento de três critérios: (1) deve ser estabelecido através de um método reproduzível; (2) deve refletir a aparência anátomo-patológica da doença a ser detectada; e (3) deve ser estabelecido independentemente do método diagnóstico sob avaliação (WULFF, 1981 *apud* WENZEL; HINTZE, 1999; HINTZE;

WENZEL, 2003). A literatura, no entanto, tem evidenciado que os padrões-ouro já estabelecidos também estão sujeitos a falhas (WENZEL et al., 1994; WENZEL; VERDONSCHOT, 1994; ten CATE et al., 1999), tendo Wenzel e Verdonschot (1994) recomendado o abandono da terminologia “padrão-ouro” por passar indevidamente a idéia de “um absoluto”, dando-se preferência aos termos “método ou critério de validação”.

Diversos testes têm sido usados como critérios de validação para os métodos de diagnóstico da cárie dental. Entre eles encontram-se a inspeção visual, radiografia convencional e digital, preparo cavitário, microrradiografia transversa, microscopia de luz polarizada e estereomicroscopia (DOWNER, 1975; DOWNER, 1989; LUSSI, 1991; ten BOSCH, 1991; WENZEL et al., 1990; WENZEL; FEJERSKOV, 1992; WENZEL; VERDONSCHOT, 1994; WENZEL et al., 1994; van RIJKOM; VERDONSCHOT, 1995; HINTZE; WENZEL; LARSEN, 1995; EKSTRAND et al., 1995; HINTZE; WENZEL, 1996; EKSTRAND; RICKETTS; KIDD, 1997; ASHLEY; BLINKHORN; DAVIES, 1998; TOVO et al., 1998; ten CATE et al., 1999; WENZEL; HINTZE, 1999; DEERY et al., 2000; PITTS et al., 2001; HINTZE; WENZEL, 2002; HINTZE; WENZEL, 2003; CÔRTEZ; ELLWOOD; EKSTRAND, 2003; CASTRO, 2004; ELLWOOD; CÔRTEZ, 2004; HUYSMANS; LONGBOTTOM, 2004; ROCHA et al., 2005; MESTRINER; VINHA; MESTRINER JÚNIOR, 2005; BENGTON et al., 2005).

O exame ao estereomicroscópio de cortes dentais poderia preencher os critérios estabelecidos para um padrão-ouro ideal. A estereomicroscopia é citada na literatura desde os primórdios das pesquisas sobre métodos de diagnóstico de cárie (DOWNER, 1975) e tem sido utilizada em diversos estudos com esta finalidade (DOWNER, 1975; WENZEL et al., 1994; HINTZE; WENZEL; LARSEN, 1995;

EKSTRAND et al., 1995; HINTZE; WENZEL, 1996; EKSTRAND; RICKETTS; KIDD, 1997; ten CATE et al., 1999; DEERY et al., 2000; PITTS et al., 2001; HINTZE; WENZEL, 2002; HINTZE; WENZEL, 2003; MESTRINER; VINHA; MESTRINER JÚNIOR, 2005; BENGTON et al., 2005). Uma das vantagens deste método de validação é a simplicidade da metodologia. O estereomicroscópio é um microscópio binocular que possui conjuntos óticos independentes para cada olho. Desta maneira permite uma visão tridimensional e a dimensão vertical do objeto em estudo fica bastante realçada (OPTOVAC, 2006). Este aspecto é particularmente interessante para avaliação da profundidade de lesões de cárie em secções dentárias. Estes aparelhos possuem uma luz halógena que incide sobre o objeto a ser examinado, cuja intensidade pode ser regulada pelo operador. Eles permitem a utilização de diferentes aumentos, que variam de 2,5 a 100X.

Embora a literatura demonstre uma clara predileção dos autores pelo exame de cortes dentais ao estereomicroscópio como critério de validação dos métodos de diagnóstico de cárie, os trabalhos apresentam diferenças metodológicas que dificultam a comparação entre os estudos, como relatado por Huysmans e Longbottom (2004). Os autores verificaram que a estereomicroscopia foi o método de validação de escolha em 14 dentre 18 trabalhos *in vitro* sobre diagnóstico de cárie publicados no período de quatro anos, tendo cada um deles adotado metodologia própria.

Para que o método de validação seja confiável, faz-se essencial que ele seja reprodutível. ten Bosch e Angmar-Mansson (2000) definiram reprodutibilidade como a capacidade de um método em produzir os mesmos resultados sob diferentes circunstâncias. Quando se comparam resultados obtidos pelo mesmo examinador em diferentes ocasiões encontra-se a reprodutibilidade intra-examinador. Ao serem

confrontados resultados entre diferentes examinadores obtém-se a reprodutibilidade inter-examinador. Embora a reprodutibilidade do padrão-ouro seja considerada fundamental para que ele possa validar adequadamente novos métodos de diagnóstico de cárie, a literatura reporta poucos trabalhos sobre a mesma (ten BOSCH; ANGMAR-MANSSON, 2000; PITTS et al., 2001; HINTZE; WENZEL, 2003; HUYSMANS; LONGBOTTOM, 2004). PITTS et al. (2001) utilizaram três examinadores para avaliar o exame de cortes dentais ao estereomicroscópio no diagnóstico da cárie oclusal e a reprodutibilidade dos examinadores da validação. Os resultados foram bastante insatisfatórios, com intensa variabilidade e baixa reprodutibilidade dos validadores.

A calibração dos examinadores tem sido considerada de grande importância ao se avaliar um determinado método de validação (HINTZE; WENZEL; LARSEN, 1995; PERES; TRAEBERT; MARCENES, 2001; PITTS et al., 2001; HUYSMANS; LONGBOTTOM, 2004). De acordo com Hintze; Wenzel e Larsen (1995), a calibração acarretaria uma melhoria na reprodutibilidade dos examinadores. Entretanto, a calibração dos validadores não foi reportada ou avaliada na metodologia dos estudos anteriormente citados.

A experiência prévia apresentada pelos examinadores pode ter impacto nos resultados dos estudos (BENGTSON et al., 2005). Seria importante avaliar se este fator também influenciaria os resultados dos estudos de validação dos métodos de diagnóstico de cárie.

Em face a estas considerações, o presente trabalho se propõe a avaliar a reprodutibilidade dos examinadores do exame histológico de secções dentárias, utilizado como critério de validação para métodos de diagnóstico de cárie, assim

como o efeito da calibração e experiência prévia dos examinadores na referida reprodutibilidade.

2 REVISÃO DA LITERATURA

2.1 DIAGNÓSTICO DA CÁRIE OCLUSAL

Clínicos e pesquisadores já atingiram consenso a respeito da dificuldade em se realizar o diagnóstico da cárie oclusal (WEERHEIJM; van AMERONGEN; EGGINK, 1989; KIDD; RICKETTS; PITTS, 1993; NEWBRUN, 1993; EKSTRAND; RICKETTS; KIDD, 1997). Clinicamente tais lesões podem progredir na dentina sem o colapso da superfície do esmalte, havendo necessidade da associação de métodos diagnósticos para se aumentar a taxa de detecção da doença (SAWLE; ANDLAW, 1988; WEERHEIJM; van AMERONGEN; EGGINK, 1989; WENZEL; FEJERSKOV, 1992).

A complexidade da anatomia oclusal, assim como a forma e profundidade dos sulcos e fissuras poderia dificultar o diagnóstico e explicar a progressão da lesão nesta região. König (1963) observou 140 secções com descoloração escura de variadas profundidades em fissuras, com um aumento que variou de 4 a 6,4X. Esta série de secções permitiu o estudo da extensão das lesões iniciais em relação às características dos sulcos e fissuras. Com este trabalho, pôde ser observado que sulcos formados por paredes que se uniam em ângulo de 90-70°, mostraram baixa suscetibilidade à cárie. Já naqueles formados por ângulos menores do que 70°, as descalcificações do esmalte foram mais comuns e usualmente iniciavam no ponto mais profundo do sulco ou próximo a este ponto. De acordo com as observações do

autor, o ataque cariioso severo, com envolvimento de dentina, esteve localizado nos elementos onde o sulco terminava em uma fissura profunda e estreita. As áreas severas de descalcificação, nestes casos, foram encontradas em ambos os lados da fissura em sua porção mais estreita, fazendo com que a junção amelodentinária fosse alcançada em diferentes pontos a partir do ataque inicial.

Juhl (1983) desenvolveu um trabalho com a finalidade de descrever a localização da lesão cariiosa em área de fissuras através do seccionamento seriado de premolares. A lesão inicial de cárie, por ocorrer na parede da fissura, pode não ser vista facilmente. Em muitos casos ela poderia se desenvolver independentemente em diferentes sítios ao longo das fissuras e depois coalescer, unificando a lesão. A fim de registrar a posição da lesão, o esmalte ao redor da fissura foi dividido em quatro zonas: (1) acima da entrada da fissura; (2) parte de cima da parede; (3) parte de baixo da parede e (4) abaixo da base da fissura. Foi demonstrado que nas secções selecionadas, 61% das lesões iniciais de cárie estavam localizadas na subsuperfície ao redor da parte mais baixa da fissura. Em 36% foram encontrados múltiplos focos tanto nas porções superiores quanto profundas das fissuras. Em somente 13% dos espécimes os focos cariosos estavam situados acima da entrada da fissura, estando a lesão diretamente acessível ao exame tátil e visual.

Sawle e Andlaw (1988) realizaram um estudo com o propósito de investigar se as cáries oclusais em molares tinham se tornado mais difíceis de serem diagnosticadas através do exame com sonda e espelho. Na primeira parte do estudo, lesões oclusais diagnosticadas radiograficamente foram comparadas com o diagnóstico clínico feito em 1974 e 1982. O número de lesões não detectadas clinicamente foi significativamente maior em 1982. Como os critérios utilizados nas

duas datas foram idênticos e um dos examinadores foi comum aos estudos, foi reforçada a idéia que as cáries oclusais tornaram-se mais difíceis de serem diagnosticadas clinicamente. Este resultado poderia ser explicado pelo mais longo período de exposição dos dentes das crianças que participaram da segunda pesquisa clínica a dentifrícios fluoretados. Foi sugerido que o flúor ajudou a manter a integridade do esmalte que cobria as lesões dentinárias, o que tornou o diagnóstico mais difícil na etapa seguinte. Na segunda parte do estudo, as radiografias dos dentes que tinham sido diagnosticados como clinicamente saudáveis em 1974 e 1982 foram examinadas, sendo a proporção dos dentes com evidência radiológica de cárie oclusal similar. Sugeriu-se que as áreas radiolúcidas detectadas nestes elementos poderiam estar associadas com o padrão anatômico da fissura, que dificultaria o diagnóstico de cárie.

Weerheijm, van Amerongen e Eggink (1989) desenvolveram um estudo piloto para entenderem melhor a situação do cirurgião-dentista face ao diagnóstico de cáries abaixo de uma aparentemente intacta superfície de esmalte. O estudo tentou avaliar os procedimentos clínicos usados para o diagnóstico da cárie oclusal e também considerou a disponibilidade de outros métodos diagnósticos mais confiáveis ou que complementassem o exame com espelho, sonda e luz. Foram selecionados 26 molares e premolares que apresentassem descolorações de diferentes níveis em suas fissuras. O diagnóstico foi baseado no exame clínico, análises de radiografias interproximais e observação de *slides*, tendo sido confirmado após a abertura da superfície oclusal. Na discussão do artigo, os autores afirmaram que o diagnóstico da cárie oclusal permanecia incerto, atestando a necessidade de novos desenvolvimentos neste campo. Foi sugerido que até que surgisse um método mais efetivo para a determinação da extensão das cáries

oclusais, seria útil incluir a informação da radiografia interproximal para o fechamento do diagnóstico. A ausência de radiolucidez discernível na radiografia não justificaria, segundo os autores, a conclusão de que a cárie oclusal em dentina está ausente. Já uma radiolucidez oclusal neste tipo de radiografia, mereceria atenção. Neste estudo piloto, os dentes tiveram sua superfície oclusal aberta quando se teve dúvida da extensão da lesão, e nestes casos, todas as cáries estavam em dentina.

Wenzel, Larsen e Fejerskov (1991) realizaram um estudo com 166 molares e premolares extraídos, sem cavitação macroscópica, com o objetivo de comparar inspeção visual, radiografia convencional, xerorradiografia e radiografia digital para diagnóstico da cárie oclusal. Secções histológicas dos elementos serviram como padrão-ouro para a presença e profundidade das lesões, tendo sido 82 dentes considerados cariados e 84 saudáveis. Pelo exame visual, cerca de 20% das lesões dentinárias foram detectadas, enquanto aproximadamente 5% dos dentes sem lesão em dentina tenham sido considerados cariados. As técnicas radiográficas aumentaram a detecção dos verdadeiro-positivos, embora tenham persistido os resultados falso-positivos, confirmando a dificuldade de diagnóstico nesta região.

Weerheijm, Gruijthuysen e van Amerongen (1992) realizaram um estudo com o objetivo de investigar com que frequência lesões dentinárias oclusais não eram detectadas nos primeiros e segundos molares sob ótimas condições de exame clínico. Os resultados do estudo mostraram que apesar de um completo exame clínico, 15% dos dentes classificados como saudáveis apresentaram áreas radiolúcidas nas radiografias, sugerindo lesões de cárie. Segundo os autores, o subestimado número de lesões oclusais em dentina, detectadas quando a avaliação se deu apenas clinicamente, sugeriu que o diagnóstico das cáries oclusais tem se

tornado mais complexo. O termo “cárie oculta” tem sido usado para descrever lesões encontradas em radiografias interproximais na ausência de óbvios sinais visuais de cárie. A prevalência de cárie oculta no presente trabalho fez com que fosse ressaltada a importância das radiografias interproximais no diagnóstico de cárie e que a política de selamento de todas as superfícies oclusais logo após a erupção dental, adotada em alguns países, ganhasse maior respaldo.

Kidd, Ricketts e Pitts (1993) relataram que lesões cariosas visíveis na dentina em radiografias interproximais eram extensas quando tratadas operatoricamente, confirmando que este tipo de exame parece ser de grande valor no diagnóstico das áreas oclusais.

Newbrun (1993) reviu os problemas encontrados no diagnóstico de lesões cariosas, principalmente no que diz respeito àquelas restritas ao esmalte ou às não cavitadas. O autor enfatizou que embora se esperasse que as lesões oclusais, por sua localização, fossem mais fáceis de serem diagnosticadas, existem clinicamente muitas dificuldades para fazê-lo. Foi comentado que numa fissura típica, a maioria dos estágios iniciais da lesão é difícil de ser observada a olho nu. Em uma fissura limpa e seca seria possível visualizar, em suas paredes, lesões de mancha branca nos estágios iniciais da doença, mas em seguida elas poderiam dar lugar a uma linha escura na base da fissura, o que dificultaria a diferenciação para uma pigmentação exógena. A importância que assume o diagnóstico das lesões oclusais, se deve ao fato, segundo o autor, delas serem responsáveis por cerca de 80% de todas as novas lesões de cárie.

Ekstrand, Ricketts e Kidd (1997) afirmaram que o que é definido como “cárie oculta” ocorreria devido a uma dificuldade dos clínicos em identificar largas lesões a partir do exame visual. Isto poderia refletir um erro na seleção de critérios visuais

apropriados ou devido a falha ao se examinar um dente que não estivesse devidamente limpo e seco.

Mota et al. (2002) revisaram a literatura com relação ao diagnóstico da cárie oclusal incipiente. Relataram que as cáries de cicatrículas e fissuras são vistas como de difícil diagnóstico e de elevado grau de subjetividade entre os profissionais, tanto no que se refere à presença da lesão quanto ao tratamento de eleição a ser adotado. Ressaltam que este diagnóstico tem se tornado mais complexo, em virtude da elevada administração de fluoretos, que promoveriam intensa remineralização do esmalte, mascarando as lesões extensas em dentina.

2.2 HISTOLOGIA DA CÁRIE

2.2.1 Cárie de esmalte

Holmen, Thylstrup e Artun (1987) comentaram que a transição de uma lesão ativa para inativa ou paralisada seria acompanhada por mudanças nas características superficiais da lesão. De acordo com a literatura, uma típica lesão de cárie ativa em esmalte exibe um aspecto opaco de giz.

Thylstrup e Fejerskov (1988) explicam as características próprias da cárie de esmalte, já que em virtude da anatomia peculiar da área das fissuras, a lesão se alargaria ao se aproximar da dentina subjacente. As lesões de cárie não começariam usualmente no fundo das fissuras. Mais freqüentemente ocorreriam de maneira bilateral ao longo das paredes. A localização poderia, entretanto, variar grandemente. Graças ao padrão peculiar de espalhamento, seriam observadas mais freqüentemente duas lesões de cárie independentes que se expandiriam

divergentemente em direção à dentina. Com a expansão lateral na junção amelodentinária, a área de dentina envolvida seria muito maior do que a de uma única lesão em uma superfície lisa. Dentro desse contexto, parte do esmalte circunjacente poderia ser solapada, e isso seria visto clinicamente como uma extensão de uma área opaca ou pigmentada ao longo da fissura oclusal.

A cárie de esmalte, no sentido da superfície em direção a junção amelodentinária, apresentaria as seguintes características, antes da formação de cavidade (HOFFMAN, 1984; THYLSTRUP e FEJERSKOV, 1988; CONSOLARO, 1996):

- 1) Camada superficial aprismática ou camada de Darling;
- 2) Corpo da lesão, com 10% a 25% de seu volume constituído por microporos;
- 3) Zona escura, com 2% a 4% de seu volume constituído por microporos;
- 4) Zona translúcida, com 1% de seu volume ocupado por microporos.

Segundo Consolaro (1996) a progressão da cárie de esmalte se faria na mesma direção dos prismas e das estrias incrementais de Retzius. Assim, o direcionamento em um corte microscópico de cárie de esmalte quando observado ao microscópio ótico, encontrar-se-ia com seu contorno determinado pela distribuição destas duas estruturas. Quando ocorresse o comprometimento da junção amelodentinária, independentemente da existência ou não de cavidade, muitos ácidos alcançariam a dentina por difusão, podendo ser encontradas alterações dentinárias e/ou pulpares. Algumas características morfológicas locais e propriedades inerentes à saliva poderiam determinar um equilíbrio momentâneo ou duradouro entre o processo de desmineralização e remineralização, podendo levar à

regressão da própria cárie de esmalte ou a sua paralisação no estágio em que se encontrasse. Este tipo de cárie seria denominado de inativa, crônica, marrom ou mais comumente cárie paralisada.

Ainda de acordo com Consolaro (1996), na patogenia da cárie de esmalte, poderiam ser estabelecidos seis estágios, determinados de forma mais precisa do ponto de vista microscópico e assim distribuídos seqüencialmente:

- 1) Estabelecimento da zona translúcida: anteriormente a ela, somente a microscopia eletrônica de transmissão permitiria a detecção de alterações no esmalte. A zona translúcida seria a primeira alteração morfológica específica do esmalte na cárie, visualizável ao microscópio ótico;
- 2) Surgimento da zona escura;
- 3) Visualização da formação do corpo da lesão;
- 4) Aparecimento da lesão branca: a parte orgânica estaria modificada e a dentina poderia apresentar-se alterada;
- 5) Estabelecimento de uma superfície gredosa no esmalte: nesta fase, e menos comumente na fase anterior, seria possível obter o sinal radiográfico inicial da cárie;
- 6) Formação de cavidade.

Na superfície oclusal, a progressão do processo seguiria a orientação dos prismas lateralmente, assumindo uma forma triangular com base voltada para a junção amelodentinária, pela qual ocorreria o espalhamento da cárie, bem como resultaria em comprometimento da dentina, em maior ou menor grau (CONSOLARO, 1996).

2.2.2 Cárie de dentina

Fusayama, Okuse e Hosoda (1966) realizaram um trabalho relacionando dureza, descoloração e invasão microbiana da dentina. Os autores mostraram que a dentina cariada poderia apresentar diferentes colorações (preto-amarronzada, marrom, marrom-amarelada, amarelo e amarelo claro). Foi citado que o amolecimento e a descoloração da dentina seriam muitas vezes critérios que nos fariam optar por sua remoção. Para realçar que estes parâmetros seriam enganosos, chamaram a atenção para os casos de cárie crônica, onde devido a deposição de dentina secundária haveria uma forte descoloração da dentina. Já em cáries agudas, caracterizadas pela rapidez da progressão, não ocorreria deposição deste tipo de dentina, havendo somente uma leve descoloração, o que não estaria em nenhum dos casos relacionado à infecção microbiana

Hoffman (1984) realçou que no processo cariogênico, muito antes da ocorrência de cavitação do esmalte, já existiria evidência de envolvimento dentinário.

De acordo com Thylstrup e Fejerskov (1988) quando a vanguarda de uma lesão cariosa se aproximasse da junção amelodentinária, ácidos, enzimas e outros estímulos atingiriam a dentina periférica como resultado do aumento da permeabilidade do esmalte. Nas imediações do vértice da lesão do esmalte ocorreria uma desmineralização da dentina, que se tornaria acastanhada. A desmineralização estender-se-ia ao longo da junção amelodentinária. Essa área seria chamada de zona de desmineralização. Nos túbulos dentinários correspondentes à área desmineralizada, assim como na periferia imediata, uma esclerose tubular seria observada. Essa reação daria origem à chamada zona translúcida ou transparente. Os autores ressaltaram que os odontoblastos, que seriam possivelmente os responsáveis por essa reação iniciariam essa resposta em torno da zona de

desmineralização numa tentativa de emparedar a lesão, enquanto no centro dela o processo destrutivo ocorreria de modo tão intenso que os processos citoplasmáticos aparentemente teriam que se abrigar próximo à polpa, antes de promover uma resposta intensa dentro do tecido. A zona de translucidez poderia se estender até quase a polpa, embora fosse possível encontrar dentina não afetada entre essa zona e a polpa.

Thylstrup e Fejerskov (1988) chamaram atenção para os grupos de túbulos dentinários localizados no centro da dentina desmineralizada que se apresentariam vazios em consequência da esclerose ter ocorrido em um estágio anterior mais distante em relação à polpa. Esse grupo de túbulos poderia formar o chamado trato morto, o qual poderia ser invadido por microorganismos. Entre a zona de invasão bacteriana e a esclerótica existiria sempre uma desmineralização em andamento como resultado da produção de ácidos pela massa microbiana dentro da cavidade cariada. Nessas áreas a desmineralização e a proteólise apareceriam tão rapidamente que muitos grupos de túbulos coalesceriam para dar origem ao chamado foco de liquificação. Adicionalmente, a desmineralização e a destruição poderiam seguir as linhas incrementais de maneira que apareceriam fendas transversas no tecido.

Consolaro (1996) ressaltou que quando a cárie dentária compromettesse a junção amelodentinária, o processo sofreria espalhamento em grande parte da superfície dentinária, como se estabelecesse um plano de clivagem entre o esmalte e a dentina, aumentando significativamente a área de contato e em consequência a disseminação dos produtos bacterianos e das próprias bactérias pelo tecido dentinário, minando e solapando gradativamente o esmalte subjacente. De acordo

com o autor, o complexo dentinopulpar reagiria às agressões, especialmente à cárie dentária, por meio de três mecanismos principais:

- 1) Submetendo seus canalículos à esclerose;
- 2) Estabelecendo a formação de dentina reacional na superfície interna;
- 3) Induzindo um processo inflamatório no tecido pulpar.

Consolaro (1996) citou que a esclerose dentinária seria o primeiro aspecto que se estabeleceria na cárie de dentina (figuras 1 e 2). A região da dentina esclerosada também poderia ser denominada de zona translúcida ou transparente da cárie de dentina. A seguir ocorreria, freqüentemente, em áreas de esclerose dentinária, uma desmineralização inicial, promovida por produtos bacterianos, tais como ácidos e enzimas, a qual precederia a invasão bacteriana propriamente dita dos canalículos dentinários, para promover a desmineralização avançada e posterior destruição e desorganização total da dentina. Embora estes aspectos não pudessem ser identificados individualmente, em um mesmo corte microscópico de dentina, em decorrência das várias técnicas de preparo dos espécimes, a cárie poderia ser dividida morfológicamente, no sentido da polpa para a superfície em cinco regiões:

1. Zona de esclerose dentinária;
2. Zona de desmineralização inicial ou profunda;
3. Zona de invasão bacteriana;
4. Zona de desmineralização avançada ou superficial;
5. Zona de destruição e desorganização total.

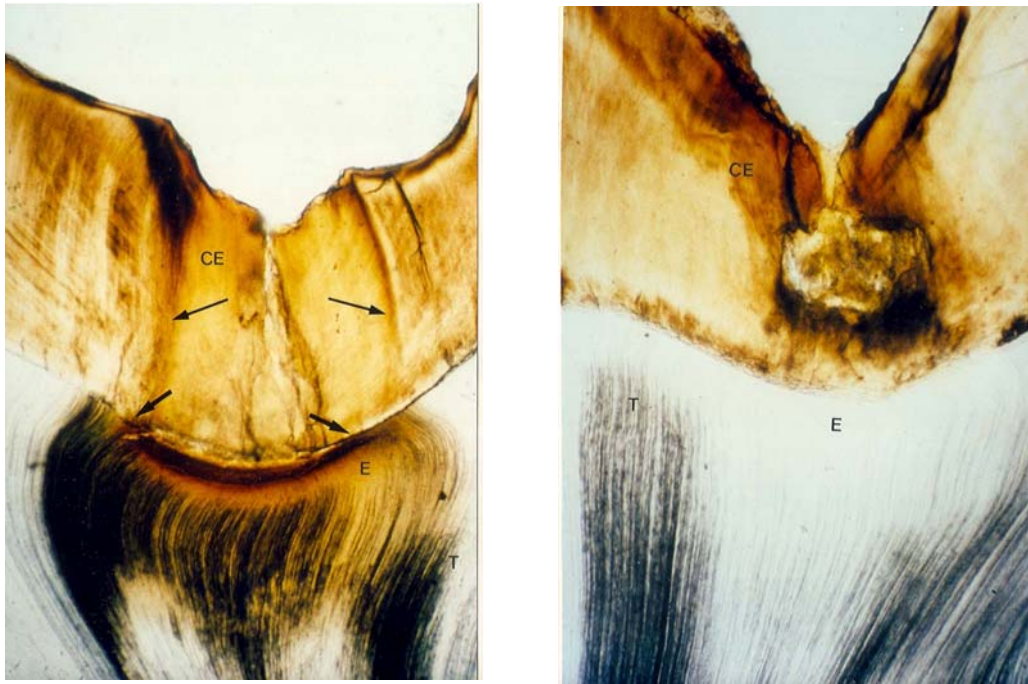


Figura 1: Cárie de esmalte (CE) na superfície oclusal de molar inferior. A progressão do processo segue a orientação dos prismas lateralmente (setas menores), assumindo uma forma triangular com base voltada para a junção amelodentinária, pela qual ocorre o espreado da cárie, bem como resulta em comprometimento superficial maior da dentina (setas maiores). Observam-se ainda esclerose dentinária (E) e tratos mortos da dentina (T) (corte por desgaste; aumento = 40X). (Fonte: CONSOLARO, A.: Cárie dentária: histopatologia e correlações clínico-radiográficas, 1996)

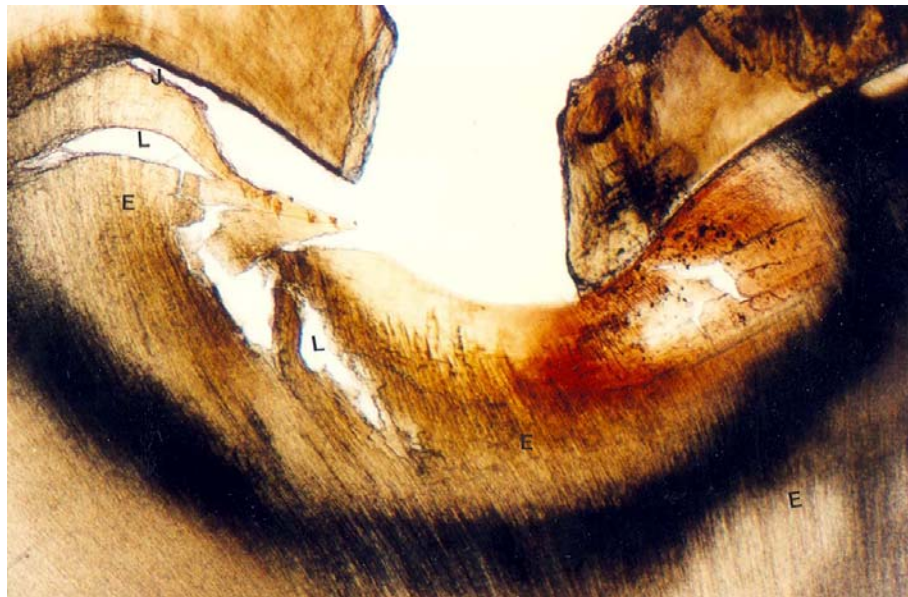


Figura 2: Cárie de esmalte e dentina na superfície oclusal de molar inferior, destacando-se o socavamento do esmalte na junção amelodentinária (J) e a esclerose dentinária (E) caracterizada pelo tamponamento dos canalículos dentinários. Notam-se também focos de liquefação da dentina (L), longitudinais e transversais, e os tratos mortos (corte por desgaste; aumento = 40X). (Fonte: CONSOLARO, A.: Cárie dentária: histopatologia e correlações clínico-radiográficas, 1996)

O autor continuou esclarecendo que a impossibilidade de distinção entre as várias zonas de evolução da cárie de dentina, microscópica e/ou clinicamente, levou à necessidade de outros critérios de distinção entre uma dentina inicialmente cariada e uma dentina totalmente destruída. Assim, na cárie, poderiam ser distinguidos dois tipos de dentina comprometida, no sentido da superfície para a polpa:

- Dentina infectada: consistência mole, de cor amarelada, assemelhando-se a um queijo. Nesta dentina poderia ser distinguida uma zona de dentina necrótica, desorganizada e amolecida. Microscopicamente, seria observado um material granuloso, eosinófilo e em quase sua totalidade sem qualquer preservação estrutural, ou seja, praticamente amorfo. Uma outra região, mais profundamente localizada corresponderia à zona de dentina desmineralizada superficial, caracterizada pela deformação da morfologia canalicular e manutenção da matriz orgânica. A dentina infectada, desta forma, não poderia ser remineralizada (HOFFMAN, 1984; CONSOLARO, 1996).
- Dentina contaminada: apareceria com uma estrutura distorcida, mas com textura parcialmente mantida, assemelhando-se à do couro. Na dentina contaminada poderia ser distinguida, da superfície para a polpa três regiões: uma zona de desmineralização profunda, localizada abaixo da zona de desmineralização mais superficial; uma zona de esclerose dentinária e freqüentemente a dentina reacional, que juntas constituiriam a zona de dentina hipermineralizada. Na zona de dentina desmineralizada profunda, a matriz canalicular seria encontrada preservada e intacta, podendo ser considerada remineralizável (HOFFMAN, 1984; CONSOLARO, 1996).

A esclerose dentinária seria uma tentativa de bloqueio da evolução da cárie. Os mecanismos envolvidos na formação da dentina esclerosada teriam como fenômeno principal a deposição acelerada e contínua de matriz dentinária, indistinguível da dentina peritubular, e sua posterior mineralização. A vitalidade do odontoblasto seria essencial para a ocorrência de esclerose dentinária, bem como a velocidade do processo. Em cáries agudas, principalmente, inexistem dentina esclerosada, devido à rapidez com que ocorre a destruição dos tecidos dentários duros.

2.3 AVALIAÇÃO DOS MÉTODOS DE DIAGNÓSTICO DE CÁRIE

Ferreira et al. (1999) apresentaram uma revisão onde foram discutidos os conceitos epidemiológicos com relação aos testes de diagnóstico, suas propriedades e interrelações, considerando como exemplo prático o diagnóstico de cárie. Inicialmente os autores comentaram que o sucesso de um profissional de saúde seria diretamente proporcional às condutas acertadas que tomasse em sua vida profissional, o que dependeria, sobretudo, de um correto diagnóstico. Segundo eles, um diagnóstico correto seria composto, na verdade, de resultados mais prováveis do que de certezas, na medida em que se deveria identificar presença ou ausência de doença, estimando-se o seu estágio evolutivo. Com relação à cárie dental, o grau de evolução determinaria diferentes tipos de tratamento e um diagnóstico impreciso poderia levar a erros irremediáveis.

Ao se coletar dados seria possível a ocorrência de erros que poderiam estar relacionados ao paciente ao pesquisador, ao instrumento (testes de diagnóstico)

e/ou circunstâncias nas quais fossem feitas as medidas (PEREIRA 1995; FERREIRA et al., 1999). Dois tipos de erros do examinador poderiam depreciar a validade de um procedimento diagnóstico. O erro ao acaso ocorreria devido à imperfeita repetibilidade na aplicação do teste, podendo ser estimado pela estatística. Ele afetaria a confiabilidade do diagnóstico e conseqüentemente sua validade. O erro sistemático (*bias*) resultaria de falhas do método usado ou de flutuações sistemáticas nos padrões diagnósticos do examinador. Isto depreciaria diretamente a validade do teste (DOWNER, 1989; FERREIRA et al., 1999). Quando um único examinador experiente e bem treinado fosse usado, o efeito do erro ao acaso seria pequeno em relação ao erro sistemático. Contudo ele se tornaria grandemente obstrutivo onde múltiplos examinadores não padronizados participassem. A importância do treinamento dos examinadores no campo dos estudos epidemiológicos para minimizar tanto o erro ao acaso quanto o erro sistemático tem sido universalmente estabelecida (DOWNER, 1989; DOOLAND; SMALES, 1982; FERREIRA et al., 1999)

De acordo com Douglass (1993) os dois padrões básicos que deveriam ser apresentados por um teste seriam validade e confiabilidade, os quais têm sido descritos na literatura médica como acurácia e precisão.

Douglass (1993) esclareceu que confiabilidade e precisão traduziriam se os resultados dos testes poderiam ser repetidos pelo mesmo ou por diferentes examinadores.

Segundo Pereira (1995) validade ou acurácia fariam referência ao grau em que o exame estaria apropriado para medir o verdadeiro valor daquilo que fosse medido, observado ou interpretado. A validade informaria se os resultados representariam a “verdade” ou o quanto se afastariam dela. Em um teste

diagnóstico, a questão a ser investigada passaria a ser a sua capacidade de discriminar corretamente doentes e sadios. Um instrumento seria válido quando se pudesse obter através dele a maior exatidão possível em um diagnóstico. A validade poderia ser definida como a medida da verdade, a exatidão.

Douglass (1993) também definiu validade como a relação entre os resultados dos testes e os resultados do padrão-ouro.

Na avaliação de um teste, muitas expressões são definidas de modo semelhante. É o que acontece com os termos: precisão, credibilidade, confiabilidade e reprodutibilidade.

Wenzel e Verdonschot (1994) afirmaram que o teste diagnóstico seria preciso quando ele medisse consistentemente. Precisão seria medida pela comparação de resultados a partir de repetidas medições do mesmo objeto.

Segundo Fletcher R.H, Fletcher S.W. e Wagner (1996) e Ferreira et al. (1999), a precisão poderia ser avaliada pelos indicadores de credibilidade e confiabilidade. A credibilidade estaria na dependência do paciente, na medida em que ele poderia ou não dizer a verdade, devendo o profissional se valer de métodos para obter o máximo de exatidão possível. Já a confiabilidade estaria mais relacionada com o examinador, que deveria estar preocupado com a padronização dos instrumentos ou testes, além de realizar um bom treinamento antes de ser iniciado o teste (FERREIRA et al., 1999).

Reprodutibilidade poderia ser definida, segundo ten Bosch e Angmar-Mansson (2000) como a capacidade de um método de produzir os mesmos resultados sob diferentes circunstâncias. Quando em tal variação somente um examinador fosse envolvido, a reprodutibilidade intra-examinador seria encontrada. Quando fossem utilizados mais examinadores, o processo de variação levaria a

reprodutibilidade inter-examinador, que usualmente seria maior do que a primeira. Sempre que possível, a medição da variação intra e inter-examinador deveria ser conduzida sem o conhecimento dos participantes (VERDONSCHOT; WENZEL; BRONKHORST, 1993), como isto seria difícil de ser alcançado, poderíamos admitir que o grau de variação descrito em estudos de pesquisa diagnóstica seria menor do que o obtido na prática clínica (WENZEL; VERDONSCHOT, 1994).

O método seria considerado melhor quando a variação de seus resultados fosse pequena. A padronização das circunstâncias de medição contribuiria para uma pequena variação (ten BOSCH; ANGMAR-MANSSON, 2000).

Pereira (1995) chamou atenção para o fato de que nem sempre existiria uma relação entre reprodutibilidade e validade. Um teste de baixa reprodutibilidade forçosamente acarretaria em baixa validade, o que significaria pouca utilidade. Já um teste de alta reprodutibilidade, no entanto, não asseguraria alta validade. Ele poderia dar resultados idênticos ou próximos quando o processo diagnóstico fosse repetido, mas todos os resultados poderiam estar errados.

Para se validar um teste, uma comparação do diagnóstico dele obtido poderia ser feita com o diagnóstico estabelecido por um teste de exatidão máxima, denominado de padrão-ouro. Se o teste em questão indicasse que haveria doença e o padrão-ouro confirmasse, teria ocorrido um acerto (verdadeiro-positivo). Se o teste indicasse a doença, mas o padrão-ouro negasse, estaríamos diante de um erro (falso-positivo). Numa série de aplicações de um teste, quanto mais resultados verdadeiro-positivos, mais sensível seria o teste. Em contrapartida, quanto mais verdadeiro-negativos fossem obtidos, mais específico seria o teste (PEREIRA, 1995; FERREIRA et al., 1999; ten BOSCH; ANGMAR-MANSSON, 2000).

Para o diagnóstico de lesão cariiosa, o padrão-ouro mais utilizado, por sua precisão tem sido o exame histológico baseado principalmente no exame ao estereomicroscópio (LUSSI et al., 1991; ESPELID; TVEIT; TJELTVEIT, 1994; FERREIRA et al., 1999), embora permita somente o exame *in vitro* do elemento ou *in vivo* quando há indicação de extração.

Através de uma tabela de matriz de decisão de diagnóstico (tabela 1) poderíamos observar melhor os resultados de um teste e compará-los com aqueles fornecidos pelo padrão-ouro, medindo-se as proporções entre acertos e erros. Na verdade seria uma tabela onde as colunas representariam a presença ou ausência de doença e as linhas, os testes com seus resultados positivos e negativos (FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, 1996; FERREIRA et al., 1999). Ao se analisar a tabela, observamos que existiriam duas situações, onde seria alcançada a exatidão: o verdadeiro-positivo, onde a lesão foi diagnosticada pelo teste e confirmada pelo padrão-ouro, e o verdadeiro-negativo, tendo o teste excluído a possibilidade de doença, onde ela realmente não existia. Em relação aos possíveis erros, teríamos o falso-positivo, representados pelos sadios considerados doentes, e os falso-negativos, correspondendo aos doentes que não foram detectados pelo teste (FERREIRA et al., 1999).

Tabela 1 – Modelo para avaliação da validade de um teste diagnóstico

TESTE	DOENTES	SADIOS	TOTAL
Positivo	Verdadeiro-positivo (a)	Falso-positivo (b)	a + b
Negativo	Falso-negativo (c)	Verdadeiro-negativo (d)	c + d
Total	a + c	b + d	a + b + c + d

Fonte: PEREIRA, M.G. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara Kooogan, 1995.

Cap. 17: Aferição dos Eventos, p. 358-76

Conhecendo-se a proporção de acertos (verdadeiro-positivos e verdadeiro-negativos) e de erros (falso-positivos e falso-negativos), seriam estabelecidos os diversos parâmetros pelos quais a validade seria expressa: sensibilidade, especificidade e valores preditivos positivos e negativos (PEREIRA, 1995), como visto na tabela 2.

Tabela 2 – Cálculo da sensibilidade, especificidade e valores preditivos.

Sensibilidade (%)	$= \frac{\text{verdadeiro-positivos}}{\text{verdadeiro-positivos} + \text{falso-negativos}} \times 100$
Especificidade (%)	$= \frac{\text{verdadeiro-negativos}}{\text{falso-positivos} + \text{verdadeiro-negativos}} \times 100$
Valor preditivo positivo (%)	$= \frac{\text{verdadeiro-positivos}}{\text{verdadeiro-positivos} + \text{falso-positivos}} \times 100$
Valor preditivo negativo (%)	$= \frac{\text{verdadeiro-negativos}}{\text{falso-negativos} + \text{verdadeiro-negativos}} \times 100$

Fonte: PEREIRA, M.G. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara Kooogan, 1995.

Cap. 17: Aferição dos Eventos, p. 358-76

Sensibilidade poderia ser definida como a capacidade que o teste apresentaria de detectar os indivíduos verdadeiramente positivos, ou seja, diagnosticar corretamente os doentes (DOWNER, 1975; PEREIRA, 1995; FERREIRA et al., 1999). De acordo com Fletcher, R.H., Fletcher, S.W. e Wagner (1996) um teste sensível deveria ser o teste de escolha quando a penalidade por deixar de diagnosticar a doença fosse grande. Testes sensíveis também seriam úteis nas fases iniciais de um processo diagnóstico, quando um grande número de possibilidades estariam sendo consideradas, e se quisesse reduzi-las.

Especificidade seria a capacidade que do teste em detectar os verdadeiro-negativos, isto é, de diagnosticar corretamente os indivíduos sadios (DOWNER,

1975; PEREIRA, 1995; FERREIRA et al., 1999). Fletcher, R.H., Fletcher, S.W. e Wagner (1996) comentaram que testes altamente específicos seriam particularmente necessários quando os resultados falso-positivos poderiam lesar o paciente física, emocional ou financeiramente.

Downer (1989) apresentou algumas considerações sobre a validade diagnóstica em um momento de redução da prevalência de cáries. Foi relatado que um método diagnóstico de cárie que enfatizasse especificidade às expensas de alguma perda de sensibilidade seria preferível. Ressaltou que as conseqüências de uma decisão falso-positiva em termos clínicos é uma restauração desnecessária de um dente saudável e o início de um ciclo restaurador repetitivo. Considerou também como ideal que a amostra investigada no estudo devesse refletir a verdadeira distribuição da doença ocorrida *in vivo*. Continuou sugerindo que se devesse selecionar para a investigação uma amostra de sítios com lesões limítrofes onde o número de verdadeiro-positivos e verdadeiro-negativos estivesse contrabalançado. O autor concluiu levantando a necessidade de futuras pesquisas para elucidar a relação matemática entre os valores de sensibilidade e especificidade.

O ideal seria utilizar um teste em que ambas as propriedades, sensibilidade e especificidade, fossem de 100%. Na prática, isto raramente seria possível, pois elas estariam relacionadas de maneira inversa. A tentativa de melhorar a sensibilidade freqüentemente resultaria em piora da especificidade (PEREIRA, 1995; FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, 1996).

Valor preditivo positivo seria a proporção de doentes entre os considerados positivos ao teste. Valor preditivo negativo tem sido definido como a proporção de sadios entre os negativos ao teste (PEREIRA, 1995).

Wenzel e Verdonschot (1994) lembraram que os problemas da interpretação da sensibilidade e especificidade de um teste diagnóstico têm sido discursados por muitos autores. Geralmente os testes têm avaliado sensibilidade independentemente de especificidade, embora sensibilidade seja altamente dependente de especificidade.

De acordo com Berkey et al. (1990) e van Rijkon e Chot (1995) dois métodos estariam atualmente disponíveis para avaliar sensibilidade e especificidade simultaneamente: regressão logística e análise ROC. A análise de regressão logística, contudo, seria somente confiável quando largas bases de dados estivessem disponíveis. A análise da curva ROC (*receiver operating characteristic*) seria um método muito apropriado quando bases de dados menos extensas estivessem presentes (VERDONSCHOT; WENZEL; BRONKHORST, 1993b; van RIJKON; CHOT, 1995; FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, 1996), sendo uma curva que plotaria sensibilidade contra 1-especificidade (VERDONSCHOT; WENZEL; BRONKHORST, 1993a, b).

A partir da curva ROC seria possível decidir qual o melhor ponto de corte de uma distribuição, de modo a definir os níveis de sensibilidade e especificidade mais adequados. Quanto mais a curva se afastasse do ângulo superior esquerdo do gráfico, mais inadequado seria o ponto de corte da distribuição. A visualização concomitante dos dois parâmetros de diagnóstico na curva permitiria escolher aquele que fosse mais útil, isto é, o que tivesse melhor poder diagnóstico, em função de seus níveis de sensibilidade e especificidade. O melhor teste seria o que mais se aproximasse do limite superior esquerdo da figura (PEREIRA, 1995; FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, 1996).

Verdonschot, Wenzel e Bronkhorst (1993a) realizaram um estudo com o objetivo de investigar se a aplicação da análise ROC poderia se constituir como um parâmetro válido para acurácia diagnóstica na avaliação de cáries proximais, em uma escala de graus de profundidade de cárie assim como em uma escala de graus de probabilidade. Neste artigo os autores relataram que em pesquisas sobre o desempenho de sistemas diagnósticos de cárie, uma escala de graus tem sido freqüentemente utilizada, permitindo o registro da profundidade da lesão. A análise destes dados tem produzido valores de sensibilidade e especificidade. No entanto, a grande desvantagem desta medição seria a necessidade de se definir critérios para a presença ou ausência de cárie. Conseqüentemente, os valores de sensibilidade e especificidade dependeriam dos limites definidos e da interpretação destes pelos examinadores, criando-se a impossibilidade de comparação da acurácia de sistemas diagnósticos.

O desempenho do examinador em diagnóstico de cárie tem tradicionalmente sido avaliado usando-se uma tabela de contingência 2 X 2 contendo as distribuições dos verdadeiro-positivos, verdadeiro-negativos, falso-positivos e falso-negativos. As respostas dos examinadores têm sido obtidas a partir de uma escala ordinal da profundidade da lesão cariada partindo-se da superfície externa em direção às camadas internas do dente (BERKEY et al., 1990; VERDONSCHOT; WENZEL; BRONKHORST, 1993). Baseados em tais tabelas de contingência, o desempenho diagnóstico poderia ser expresso em termos de sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo (DOUGLASS; McNEIL, 1983; VERDONSCHOT; WENZEL; BRONKHORST, 1993). A maior falha desses indicadores de acurácia diagnóstica seria que seus valores dependeriam da distribuição das superfícies cariadas e não cariadas na amostra investigada, dos

tamanhos das lesões incluídas e das características dos examinadores (BERKEY et al. 1990; VERDONSCHOT; WENZEL; BRONKHORST, 1993). Para interpretar significativamente sensibilidade e especificidade do desempenho do examinador a partir de um sistema diagnóstico, a distribuição das superfícies cariadas e não cariadas assim como a representação dos vários estágios de profundidade de cárie deveria refletir a distribuição entre a população na qual se pretendesse utilizar o sistema diagnóstico. A inclusão de muitas lesões largas, por exemplo, iria resultar em sobrestimação da sensibilidade, enquanto a inclusão de um grande número de superfícies não cariadas na amostra causaria sobreestimação da especificidade (BERKEY et al., 1990; VERDONSCHOT; WENZEL; BRONKHORST, 1993). Uma segunda desvantagem do uso de parâmetros de especificidade e sensibilidade para avaliação da acurácia diagnóstica seria a aplicação obrigatória de limites diagnósticos indicando quais estágios da lesão cariada seriam denominados “doentes” e “não doentes”. A análise ROC não teria nenhuma das desvantagens mencionadas anteriormente (VERDONSCHOT; WENZEL; BRONKHORST, 1993).

Peres, Traebert e Marcenes (2001) realizaram um estudo cujo objetivo foi descrever um exercício de calibração de examinadores para estudos epidemiológicos de cárie. De acordo com os autores, para se obter uma padronização no uso de critérios de diagnóstico para diversas doenças bucais seria fundamental para que os examinadores envolvidos no estudo participassem de um treinamento antes do início do mesmo. Portanto, uma etapa necessariamente prévia a todos estes estudos seria a calibração dos examinadores. Foi citada a definição dada pela Organização Mundial de Saúde (WHO, 1993) para calibração, que seria a repetição de exames nas mesmas pessoas pelos mesmos examinadores, ou pelo mesmo examinador em tempos diferentes, a fim de diminuir as discrepâncias de

interpretação nos diagnósticos. Ainda segundo a WHO (1993), os objetivos da calibração seriam: (1) assegurar uniformidade de interpretação, entendimento e aplicação dos critérios das várias doenças e condições a serem observadas e registradas; (2) assegurar que cada um dos examinadores pudesse trabalhar consistentemente com o padrão adotado e (3) minimizar as variações entre diferentes examinadores. O grupo de examinadores do presente trabalho contou com oito cirurgiões-dentistas, sem experiência prévia em levantamentos epidemiológicos, tendo sido a amostra composta por 17 escolares com 12 anos de idade. A atividade foi dividida em quatro etapas. A primeira delas consistiu de uma apresentação dos critérios a serem adotados (Índice CPO-D), seguida da apresentação de 20 *slides* para os quais os examinadores deveriam realizar o diagnóstico. Nesta fase foi também apresentada a ficha que seria utilizada e a seqüência do exame clínico. Na segunda etapa foi realizado um exercício clínico para fixação dos critérios apresentados, tendo sido feitos exames individuais em seis crianças. Quando eram observadas diferenças entre o examinador e um dos orientadores do exercício, era realizado um novo exame. Foi escolhido como examinador-padrão aquele que, dentre os oito, teve menos divergências com o orientador. Os exames clínicos foram repetidos até que se julgasse que os critérios haviam sido fixados pela equipe. Dando continuidade, 17 crianças de 12 anos de idade foram selecionadas para serem examinadas individualmente por cada um dos oito examinadores e um dos orientadores. Foram montadas matrizes que permitiam a comparação das divergências e concordâncias, tendo sido realizado o teste Kappa não Ponderado considerando-se todos os dentes e cada um separadamente. Na terceira etapa foram realizados dois novos exames pelos sete cirurgiões-dentistas e pelo examinador-padrão, com todas as 17 crianças, sem que houvesse

comunicação entre os examinadores, para que se pudesse avaliar o resultado da calibração. Novamente foram montadas matrizes para verificar concordâncias e permitir a realização do teste Kappa. Na quarta etapa realizou-se uma discussão dos resultados. De acordo com os autores, há diversas formas de verificar a concordância de resultados, sendo a mais simples dela através da taxa de concordância observada entre os examinadores. Esta maneira, no entanto, teria sido criticada por não levar em consideração a concordância devida ao acaso e, outras formas surgiram para substituí-la, como o teste Kappa. Este seria um teste de concordância ajustado, pois levaria em consideração, descontando no resultado final, a concordância devida ao acaso, o que acarretaria em menores níveis de concordância do que os obtidos com a taxa de concordância. Após a análise dos resultados os autores concluíram que a utilização do teste Kappa e seu cálculo com base em cada um dos dentes separadamente seria a forma mais adequada de mensuração da acurácia e reprodutibilidade.

2.4 MÉTODOS DE VALIDAÇÃO DOS MÉTODOS DE DIAGNÓSTICO DE CÁRIE

Downer (1975) investigou a habilidade de um sistema de diagnóstico epidemiológico em distinguir sítios dentais com cavidades cariosas daqueles que não as possuíam. Os dados diagnósticos foram obtidos durante um exame clínico visual anual complementado com radiografias interproximais de 551 adolescentes. A amostra consistiu de 109 premolares e molares permanentes, extraídos durante o tratamento subsequente dos pacientes. Secções seriadas não descalcificadas, com

espessura variando entre 150-200 μm , foram cortadas no sentido mesio-distal, estando cada sítio de predileção incluído entre a terceira e a quinta secção. As secções foram examinadas em água ao estereomicroscópio com um aumento de 2,5X, para comparar os achados diagnósticos originais com a extensão histológica das cáries. A detecção de cavidades atingiu 84%, enquanto 22% dos sítios sem cavitação foram erroneamente diagnosticados como tendo cavidades. Embora 97% das superfícies sem cavitação tenham sido corretamente identificadas, somente 73% das cavidades proximais foram detectadas radiograficamente. De acordo com o autor, o exame histológico de dentes extraídos seria o melhor método de validação disponível para se validar técnicas de diagnóstico de cárie. A técnica de inspeção histológica foi, no entanto, criticada pelo autor já que durante o seccionamento dos dentes com discos de diamante cerca de 300 μm de espessura do material seria perdido entre cada secção. Haveria também uma falha técnica, por ser inaceitável tomar ao acaso uma amostra de todos os dentes examinados na boca para investigação. No estudo realizado pelo autor, os dentes foram extraídos devido a cáries, e muitos possuíam duas ou mais superfícies com cavitação. Além disso, somente oito das 230 fissuras e 18 das 185 superfícies proximais estavam livres de qualquer evidência histológica de cárie. Assim poucos sítios saudáveis estiveram disponíveis para uma segura avaliação de sensibilidade e especificidade. O autor recomendou ainda cuidado em qualquer extrapolação desses achados, pois eles deveriam ser aplicáveis somente a amostra de dentes com uma alta prevalência de cárie.

Wulff (1981 *apud* WENZEL; HINTZE, 1999) propôs que para que o padrão-ouro fosse realmente representativo, seria necessário o preenchimento de três critérios: (1) deveria ser estabelecido através de um método reproduzível; (2) deveria

refletir a aparência anátomo-patológica da doença a ser detectada; e (3) deveria ser estabelecido independentemente do método diagnóstico em avaliação.

De acordo com Wefel, Clarkson e Heilman (1985) a microscopia de luz polarizada e a microrradiografia transversa, utilizadas como padrões-ouro para os métodos de diagnóstico de cárie seriam usadas principalmente com o propósito de detectar zonas de diferentes conteúdos minerais dentro da lesão, sendo a microrradiografia a técnica mais informativa. Para este propósito, a microscopia de luz polarizada requereria embebição da secção em diferentes meios, que limitassem os dados a cerca do conteúdo mineral, não melhorando no entanto, a detecção da profundidade da lesão.

Mitropoulos e Downer (1987) afirmaram que o diagnóstico da cárie em estágio inicial de desenvolvimento aumentaria problemas de reprodutibilidade e validade. Segundo eles, os dentistas poderiam concordar sobre a presença da lesão cariosa quando ela está bem estabelecida e óbvia a olho nu, mas algumas vezes iriam discordar deles mesmos quando reexaminassem uma superfície com lesão inicial, discordando, certamente também um dos outros. De acordo com o que foi relatado neste artigo, 22 oficiais dentistas participaram de um treinamento para atualizar seus conhecimentos quanto ao diagnóstico da cárie dental, assim como para padronizar suas aplicações de métodos e técnicas. A primeira sessão começou com uma descrição de critérios para diagnóstico de cárie suportada por fotografias de secções histológicas de dentes extraídos. Em um primeiro momento, entre 20 e 25 adolescentes e adultos jovens foram examinados e as diferenças de diagnóstico foram resolvidas com o examinador-tutor. Em outra etapa, de 25 a 27 pacientes (14 a 15 anos de idade) foram examinados individualmente e discussões relativas ao diagnóstico foram realizadas em seminário posterior. Ao final do trabalho, excelente

concordância pôde ser alcançada pelos oficiais dentistas quando comparados com o examinador-tutor, em todos os aspectos de diagnóstico. Os resultados deste exercício demonstraram que a maioria dos oficiais foi bem calibrada para o diagnóstico da cárie dental e estavam capacitados para diagnosticar lesões incipientes.

ten Bosch e Angmar-Mansson (1991) relataram que estudos atuais sobre cárie requererem a medição de pequenas mudanças no conteúdo mineral do dente, sendo desejáveis medições quantitativas. Alguns desses métodos, como a microscopia de luz polarizada e microrradiografia transversa, viriam sendo utilizados como padrões-ouro. Ao compararem microrradiografia transversa com microscopia de luz polarizada para a análise de secções dentais, consideraram o primeiro método preferível, por ser mais direcionado ao conteúdo mineral.

Lussi (1991) realizou um estudo *in vitro* com o propósito de testar a acurácia, a reprodutibilidade do diagnóstico e decisão de tratamento de cáries de fissuras com e sem o uso do explorador. Foram selecionados 61 dentes de modo que nenhuma superfície lisa apresentasse lesão cariosa. Os dentes foram montados em blocos, tendo sido cuidadosamente limpos e mantidos hidratados. As superfícies oclusais foram fotografadas com um aumento de 5X, para que se pudesse fazer um desenho de tal superfície e marcar o ponto exato a ser diagnosticado. O questionamento quanto ao diagnóstico dos dentes foi feito a 34 dentistas, que decidiram sobre o possível tratamento. Deste total, oito profissionais utilizaram o explorador, tendo os demais empregado somente a técnica visual. Dos 34 dentistas, 12 foram convidados a repetir o exame após o período mínimo de uma semana. Seguindo-se à segunda inspeção, os dentes foram histologicamente preparados e a área a ser diagnosticada foi seccionada perpendicularmente a superfície de modo a produzir três cortes

histológicos incluindo a área diagnosticada. *Slides* foram realizados e a extensão das cáries foi avaliada. A avaliação da área seccionada foi o critério de validação para a avaliação da porcentagem de dentes corretamente diagnosticados, da sensibilidade e especificidade. A porcentagem de dentes diagnosticados corretamente foi de aproximadamente 42%. A porcentagem de decisões de tratamento clinicamente corretas, contudo, foi de 73%. O teste de reprodutibilidade (Kappa de Cohen) apresentou valores de 0,47 para o diagnóstico e 0,44 para as decisões de tratamento.

Verdonschot et al. (1992) utilizaram a leitura de dois radiologistas dentais como padrão-ouro, tendo os examinadores graduado individualmente as superfícies a partir de radiografias. Posteriormente, em uma sessão conjunta, foi estabelecido um consenso entre os diagnósticos atribuídos pelos dois radiologistas. Os autores justificaram a metodologia adotada alertando-nos para o fato de que os radiologistas dentais estariam mais atentos do que os clínicos para o fato de que as radiografias subestimam a verdadeira profundidade da lesão e que este conhecimento teria sido incluído na leitura das radiografias.

Douglass (1993) definiu padrão-ouro como sendo um diagnóstico definitivo da presença ou ausência da doença, obtido preferencialmente a partir de resultados histopatológicos. A questão primária seria: poderia o resultado positivo e negativo do teste diagnóstico ser comparável com os dados obtidos a partir do padrão-ouro? No entanto, segundo o autor, a partir dessa comparação seriam definidos os parâmetros-chave de um teste de diagnóstico (sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo). Foi relatado que a parcela da população portadora de doenças orais tem sofrido redução. A necessidade de se aprender a identificar esses pacientes de alto risco nos estágios iniciais da doença requeria o

desenvolvimento de testes mais sensíveis e que pudessem prever o aparecimento de futuras patologias orais. Segundo o autor, mesmo tendo surgido inúmeros testes, muitos têm falhado em sua sensibilidade, especificidade ou em sua habilidade em prever o percentual dos casos em que a doença evoluirá. Douglass relatou que devido ao fato das novas tecnologias poderem identificar estágios muito precoces da doença, onde não existiria nenhum padrão-ouro, seria mais importante mostrar relação direta entre os resultados do teste diagnóstico e os resultados de interesse do paciente. Comentou finalmente que se nova e mais sensível tecnologia fosse desenvolvida e que pudesse ser de fato acurada e confiável preditora dos estágios futuros da doença, ela poderia se tornar útil como um novo padrão-ouro. Na conclusão deste artigo foram levantadas cinco questões, que segundo o autor, deveriam direcionar as pesquisas para novos testes: (1) a teoria científica do teste se adapta com os conhecimentos atuais? (2) os parâmetros de eficácia do teste foram confiavelmente determinados? (3) como poderia o teste afetar a decisão de tratamento? (4) o uso do teste melhorará a saúde do paciente? (5) o custo adicionado ao teste foi justificado pelo aumento da efetividade ou por se evitar outras despesas com saúde?

Wenzel e Verdonshot (1994) relataram que o método de validação por si só seria sujeito a incertezas desde que dependesse de uma avaliação subjetiva do examinador ou de alguma medição quantitativa a qual seria também sempre sujeita a erros. Sugeriram que a expressão “padrão-ouro” (*gold standard*) fosse evitada, já que ela criaria uma imagem de “um absoluto”. Alguns métodos de validação poderiam ser mais válidos do que outros para propósitos específicos. Assim na avaliação da acurácia dos testes diagnósticos, a escolha do método de validação requereria muito cuidado, do mesmo modo que não se poderia esquecer de que o(s)

validador(es), poderiam influenciar no resultado do teste. Continuando, afirmaram que alguma forma de validação seria essencial quando for comparado o desempenho de métodos diagnósticos em termos de acurácia e que o padrão-ouro deveria preencher tanto quanto possível aos requisitos propostos por Wulff (1981 *apud* WENZEL; HINTZE, 1999). Concluíram lembrando que deveria ser um requerimento ético que um sistema diagnóstico fosse avaliado *in vitro* antes de ser testado em pacientes. Para tal, os passos *in vitro* deveriam assemelhar-se o mais possível das condições *in vivo*, seguindo as regras para avaliação de um teste diagnóstico.

Wenzel et al. (1994) realizaram um estudo com o objetivo de comparar o impacto do validador e do sistema de validação (estereomicroscopia ou radiografia de secções dentais) com o resultado de testes diagnósticos para cáries oclusais. O material consistiu de 60 terceiros molares que foram seriadamente seccionados e avaliados por quatro examinadores que examinaram as amostras independentemente através dos dois métodos de validação em questão. Os valores para concordância entre os examinadores foram muito maiores dentro de um mesmo método de validação do que na comparação entre os dois métodos. Para radiografia das secções, os valores variaram de 0,44 a 0,76 (média 0,55) e para estereomicroscopia de 0,47 a 0,60 (média de 0,53). Os valores para concordância intra-examinador entre os dois métodos foram baixos, variando de 0,31 a 0,49. Também foram baixos os valores da concordância inter-examinador ao se comparar os dois métodos, variando de 0,16 a 0,53. O trabalho mostrou que as diferenças entre examinadores dentro de um método de validação foi maior com microscopia do que com radiografia. Além disso, concluíram que os resultados dos testes de diagnóstico de cárie poderiam ser influenciados tanto pelo validador quanto pelo

método de validação. Ao final de seu artigo ressaltaram que os métodos de validação são avaliações subjetivas realizadas por observadores humanos e como tais, sujeitos a variabilidade. Foi sugerido que a palavra padrão-ouro não fosse usada em estudos que avaliassem o resultado de testes diagnósticos, já que ela envolveria um absoluto e inequívoco padrão. Chamaram a atenção para o fato de que em estudos de diagnóstico de cárie uma completa descrição do método de validação em uso e do(s) validador(es) deveria ser dada, já que estes fatores poderiam influenciar o resultado dos testes diagnósticos em estudo. Outro ponto para o qual os autores chamaram atenção foi para o fato de terem optado pelo seccionamento seriado dos dentes. Foi comentado que quando a opção é feita pela hemissecção, existe a possibilidade de se perder lesões nas partes periféricas das fissuras, levando a amostra sob estudo a uma mais baixa prevalência de cáries.

Hintze, Wenzel e Larsen (1995) compararam a acurácia de quatro métodos utilizados para validação de lesões de cáries oclusais: estereomicroscopia, filme radiográfico, microrradiografia e inspeção a olho nu. Neste estudo os cortes foram avaliados por três examinadores tendo sido a reprodutibilidade inter-examinador também checada. O material consistiu de 18 terceiros molares não irrompidos, sabendo-se serem saudáveis com respeito às cáries, já que eles estavam cobertos por osso antes de sua extração, e 20 terceiros molares com um desconhecido “estado real de doença”. Os dentes foram cortados seriadamente, codificados e examinados cegamente com relação às cáries oclusais pelos três examinadores independentes, através dos quatro métodos de validação em estudo. A estereomicroscopia demonstrou ser o único método pelo qual todos os observadores identificaram corretamente todos os 18 dentes não irrompidos como saudáveis, resultando numa especificidade de 1,00. Pelos outros métodos, um ou mais dentes

não irrompidos foram falsamente classificados como cariados. Nos dentes irrompidos, 73%, 58%, 52% e 32%, respectivamente, dos dentes foram determinados como cariados pela estereomicroscopia, microrradiografia, inspeção a olho nu e filme radiográfico. A concordância inter-examinador foi avaliada pelo Índice de Kappa de Cohen, tendo alcançado o valor médio de 0,73, variando de 0,43 (filme radiográfico) a 0,89 (microrradiografia e inspeção a olho nu). Como foi possível através da estereomicroscopia detectar cáries em aproximadamente três quartos dos dentes irrompidos e ao mesmo tempo identificar todos os dentes não irrompidos como saudáveis, este pareceu, segundo os autores, ser o mais confiável dos métodos de validação em estudo. Outra observação dos autores foi que radiografias normais de secções mais grossas (com cerca de 700 μm) proveram um inadequado padrão-ouro.

Ekstrand et al. (1995) relacionaram características histológicas e externas dos estágios progressivos da cárie na fossa oclusal. O material compreendeu 140 terceiros molares extraídos. A fossa central foi examinada com um estereomicroscópio (aumento de 16X) e macroscopicamente sob condições padronizadas após limpeza e secagem com ar. Os sinais de cárie foram classificados usando um detalhado sistema envolvendo 12 critérios microscópicos, oito macroscópicos e seis escores radiográficos. Secções de 250 μm de espessura foram cortadas no sentido vestibulo-lingual a partir da fossa central e a secção da fossa com alterações mais extensas foi selecionada para o exame histológico. Os exames macroscópico e ao estereomicroscópio foram realizados independentemente por dois examinadores, tendo também sido repetidos por ambos. Os critérios usados para o exame ao estereomicroscópio foram: (0) esmalte com translucidez e configuração anatômica mantida após cinco segundos de

secagem com ar; (1) esmalte com translucidez mantida após cinco segundos de secagem com ar com destruição localizada da superfície; (2) opacidade do esmalte se torna visível após cinco segundos de secagem com ar, nenhuma alteração anatômica; (3) opacidade do esmalte se torna visível após cinco segundos de secagem com ar, uma ou mais áreas da superfície com aparência erodida; (4) opacidade do esmalte visível sem secagem com ar, áreas localizadas de superfície erodida tornam-se visíveis após secagem com ar; (5) descoloração amarronzada do esmalte com opacidade visível sem secagem por ar, nenhuma alteração anatômica de superfície; (6) opacidade do esmalte visível sem secagem por ar, uma ou mais destruições localizadas de superfície; (7) opacidade do esmalte visível sem secagem por ar, uma ou mais microcavidades; (8) opacidades descoloridas no esmalte, destruições localizadas de superfície e/ou pequenas microcavidades; (9) cavidades demarcadas no esmalte opaco; (9) cavidades demarcadas no esmalte descolorido; (10) cavidades demarcadas na dentina com uma entrada relativamente pequena. O Índice de Kappa foi calculado para avaliar a reprodutibilidade inter e intra-examinador. A reprodutibilidade inter e intra-examinador com relação a inspeção macroscópica foi 0,91, 0,90 e 0,90, respectivamente. Para a estereomicroscopia os valores foram 0,92, 0,92 e 0,91. Devido à alta reprodutibilidade, as dúvidas em relação aos escores foram reexaminadas por ambos examinadores e o escore final foi estabelecido por consenso. A classificação histológica, em conjunção com as observações macroscópicas, tornou possível demonstrar uma clara relação entre o grau externo de progressão da cárie e as reações internas no esmalte e dentina. Os dados não suportaram o uso rotineiro do exame radiográfico para diagnóstico da cárie oclusal.

Segundo Jones e Hunter (1995) os métodos de consenso seriam uma maneira de se negociar com evidências científicas conflitantes. O objetivo dos métodos de consenso seria determinar a extensão na qual *experts* ou pessoas leigas concordam sobre um determinado tema. O foco dos métodos de consenso repousaria onde a unanimidade de opinião não existe devido à falta de evidência científica ou onde haveria evidência contraditória sobre um assunto. Os métodos tentariam avaliar a extensão de concordância (medição do consenso) e resolver as discordâncias (desenvolvimento de consenso).

Fletcher, R.H, Fletcher, S.W. e Wagner (1996) relataram que a avaliação da acurácia de um teste estaria baseada na sua relação com algum meio de saber se a doença estaria ou não realmente presente – um indicador mais fiel da verdade, geralmente definido como “padrão-ouro” (“*gold standard*”, “teste padrão”). Os autores comentaram, no entanto que o “padrão-ouro” freqüentemente seria difícil de ser encontrado. Uma das dificuldades seria a ausência de critérios diagnósticos exatos para uma determinada doença. Foi relatado que não seria sempre possível, na prática, saber quão bem os testes em uso seriam comparados a um padrão inteiramente fidedigno. Seria preciso escolher como padrão de validade um teste que, sabidamente, não seria perfeito, mas que seria o melhor disponível. Isso, segundo os autores, poderia levar a comparação de um teste fraco com outro, sendo um deles considerado como padrão de validade por ter maior uso, ou por ter sido julgado superior por um consenso de peritos. Poderia então ser originado um paradoxo. Se um teste novo fosse comparado com um teste padrão antigo, o novo teste poderia parecer pior, mesmo sendo realmente melhor.

Hintze e Wenzel (1996) comparando a acurácia diagnóstica de radiografias obtidas tanto *in vivo* quanto *in vitro* dos mesmos dentes para detecção de cáries

oclusais e proximais também optaram pela validação histológica após seccionamento seriado. A estereomicroscopia foi escolhida por ter, em estudo anterior (HINTZE; WENZEL; LARSEN, 1995), tido perfeita especificidade quando comparada com microrradiografia, filme radiográfico e inspeção a olho nu.

Ekstrand, Ricketts e Kidd (1997) realizaram um estudo laboratorial em 100 superfícies oclusais (80 molares e 20 premolares) onde foi investigada a reprodutibilidade e acurácia de um sistema de registro visual de cárie, um sistema de registro eletrônico de cárie (ECM) usando uma escala de corrente contínua e um sistema de registro radiográfico de cárie. O exame histológico ao estereomicroscópio (aumento de 5X) dos dentes serviu como padrão-ouro para validar a habilidade de cada sistema em avaliar a profundidade da lesão e predizer amolecimento e dentina desmineralizada. Os autores citaram as desvantagens de um estudo laboratorial, onde a amostra poderia não ser representativa por conter predominantemente terceiros molares, além de muitos fatores clínicos (idade do paciente, higiene oral, estado dos demais dentes) serem desconhecidos. A escolha da validação histológica foi feita pelos autores baseando-se nos resultados de estudos prévios (WENZEL et al., 1994; HINTZE; WENZEL; LARSEN, 1995), onde foi demonstrada a maior acurácia da estereomicroscopia para validação de lesões oclusais.

Côrtes (1998) se propôs a validar *in vivo* o diagnóstico de cárie oclusal dentinária e de lesões cavitadas e pré-cavitadas em proximais de dentes posteriores através de exame clínico visual, transiluminação por fibra ótica (FOTI) e radiografias interproximais. As superfícies oclusais e proximais em pré-molares e molares permanentes de 36 pacientes (idade entre 11 e 21 anos) foram examinadas independentemente por dois examinadores calibrados por método. As superfícies julgadas cariadas por pelo menos um dos examinadores dentre os métodos de

diagnóstico foram agendadas para validação. A validação da presença de cárie foi obtida pelo diagnóstico de consenso entre dois examinadores após a invasão oclusal por meio da inspeção visual-tátil e de corante fucsina básica em propilenoglicol a 0,5% e após separação dentária pelo exame visual direto e moldagem em silicone de adição das proximais afastadas com separador de Elliot. Para as superfícies oclusais de molares, foram encontrados na validação visual-tátil: sensibilidade e especificidade de 0,36 e 1,00 para exame clínico visual, 0,44 e 0,99 para o exame com o FOTI e 0,42 e 0,87 para o exame radiográfico; áreas sob as curvas ROC de 0,77, 0,86 e 0,63, respectivamente. Os valores de sensibilidade indicaram que os métodos não detectaram um número considerável de superfícies com cáries dentinárias oclusais e cavidades proximais. Para as superfícies oclusais, o exame clínico visual e FOTI apresentaram um desempenho moderado e similar para o diagnóstico de cárie dentinária; o exame radiográfico interproximal detectou quase tantas lesões de cárie dentinárias oclusais quanto o FOTI, porém a expensas de uma proporção considerável de resultados falso-positivos, o que na prática induziria a sobretratamento restaurador; o exame clínico visual e FOTI apresentaram-se superiores no diagnóstico de oclusais saudáveis.

Janhom et al. (1999) propuseram-se a avaliar a acurácia diagnóstica para detecção de cáries em 66 radiografias *bitewings* digitais. A leitura inicial foi feita por sete examinadores. Um painel de *experts*, composto por três radiologistas orais e um cariologista, verificou a presença de cárie em cada superfície pela observação das radiografias originais com o auxílio, se necessário, de uma lente de aumento. Quando os *experts* discordavam sobre a presença ou profundidade da lesão, estas superfícies eram avaliadas em uma segunda e em alguns casos, em uma terceira rodada. O consenso final era alcançado através de discussão. Nenhum dado

histológico estava disponível devido à natureza das imagens utilizadas. Segundo os autores, o uso de um painel de *experts* para fazer um diagnóstico clínico seria considerado como um padrão-prata nos estudos *in vivo*, sendo um método aceito para se alcançar um julgamento clínico.

Wenzel e Hintze (1999a) afirmaram que nenhum método diagnóstico para cárie seria acurado o suficiente para atuar como padrão-ouro na avaliação *in vivo* de novos testes. A utilização de termos como “padrão-prata” utilizado por Janhom et al. em 1999, ou outros metais menos preciosos para justificar o uso de examinadores para validar cálculos de acurácia de diagnóstico, deveriam ser abandonados na medida em que eles somente confundem os achados.

Alguns trabalhos têm relatado o grave erro metodológico que existe quando um teste em avaliação é ao mesmo tempo o padrão-ouro, não se conseguindo, assim o preenchimento de nenhum dos critérios essenciais (WULFF, 1981 *apud* WENZEL; HINTZE, 1999) para um padrão-ouro ideal (VERDONSCHOT et al., 1992; WENZEL e HINTZE, 1999a).

Wenzel e Hintze (1999b) discutiram a escolha do padrão-ouro para se avaliar um teste para diagnóstico de cárie. Dissertaram sobre o paradigma existente no diagnóstico radiográfico da cárie dental, onde áreas de radiolucidez poderiam representar apenas uma cicatriz de uma lesão ativa, graças ao bem estabelecido efeito dos fluoretos em manter a superfície do esmalte intacta. Mostraram que a radiografia estaria longe de ser um método acurado de diagnóstico, a não ser em casos de lesões mais profundas, justificando-se a necessidade de desenvolvimento de novos métodos. Confirmaram que através do exame de cortes dentais ao estereomicroscópio, as cáries poderiam ser acuradamente avaliadas e que a microscopia deveria ser o padrão-ouro em estudos sobre a acurácia dos testes de

diagnóstico de cárie. Foi feita uma crítica aos estudos que utilizam lesões artificiais, criadas com o uso de ácidos ou brocas, com o objetivo de avaliar o diagnóstico de cáries. Isto, segundo as autoras, poderia a primeira vista ser um aceitável padrão-ouro. Contudo estes estudos falhariam no que diz respeito ao segundo critério para um padrão-ouro ideal, que seria refletir a aparência anátomo-patológica da doença (WULFF, 1981 *apud* WENZEL; HINTZE, 1999). Ressaltaram que até aquele momento não existiam métodos para diagnóstico de cárie que fossem acurados o suficiente para atuar como padrão-ouro para avaliação *in vivo* de novos testes. Relataram, citando um de seus estudos (HINTZE; WENZEL, 1996), que resultados obtidos a partir de bons estudos laboratoriais com uma verdadeira validação poderiam ser transferidos para uma situação clínica.

ten Cate et al. (1999) realizaram um experimento clínico de um ano, desenhado para avaliar e comparar a habilidade de vários métodos de diagnóstico em monitorar estágios precoces do processo de cárie e o grau de concordância entre quatro Institutos de Pesquisa em diferentes países. Foram utilizados 100 dentes decíduos exfoliados, selecionados aleatoriamente, muitos dos quais apresentavam áreas de lesões reversíveis de cárie. O estudo se propôs a validar a capacidade da fluorescência de luz quantitativa (QLF) em detectar acuradamente a presença e severidade de tais lesões na estrutura dental, sendo utilizados como padrão-ouro as técnicas de histologia (estereomicroscopia com aumento de 40X e 100X), microrradiografia transversa e microscopia de luz polarizada. Os dentes foram identificados e classificados pelo *Oral Health Research Institute of the Indiana University School of Dentistry* (Indianápolis – EUA), onde foram realizadas imagens de QLF nas superfícies vestibular, lingual e oclusal, tendo sido cada uma delas dividida em nove sítios para análise. Após serem obtidas tais imagens, os dentes

foram encaminhados para o Departamento de Cariologia do *Academic Centre for Dentistry Amsterdam* (ACTA – Holanda), onde novas análises de QLF foram feitas, sendo cada superfície cortada em três secções histológicas no sentido vestibulo-lingual, totalizando 300 secções. As secções foram identificadas e individualmente analisadas (estereomicroscopia, microrradiografia e microscopia de luz polarizada). O ACTA (Holanda) usou técnicas histológicas e microrradiografia para a análise, assim como o *Glasgow Dental Hospital & School* (Escócia). Após a completa análise em Glasgow, os espécimes foram encaminhados para o *Dows Institute for Dental Research*, da Universidade de Iowa (EUA), onde os espécimes foram analisados através de luz polarizada. Como último passo, as secções retornaram para Indiana onde foram realizadas análises por histologia e microrradiografia. Analisando-se a concordância entre as três técnicas (histologia, microrradiografia transversa e microscopia de luz polarizada) utilizadas como padrão-ouro, obteve-se uma pobre correlação entre os testes. Quando os dados foram analisados através do Índice de Kappa de Cohen valores entre 0,15 e 0,33 foram encontrados, indicando uma fraca a sofrível correlação entre os testes. A concordância entre o QLF e os três métodos individualmente resultou nos seguintes valores: 0,26, 0,33 e 0,32 para QLF quando comparado com histologia, microrradiografia transversa e microscopia de luz polarizada. Segundo os autores, o achado mais surpreendente do estudo foi a inconsistência entre as várias técnicas de avaliações laboratoriais na determinação do estado da cárie nos dentes estudados. Algumas lesões foram encontradas pelas três técnicas, enquanto algumas por apenas duas. Houve ainda lesões que só foram identificadas por uma das técnicas. O número de consenso pleno entre as três técnicas realizadas nos quatro Institutos foi da ordem de 18%. A concordância entre o QLF e os outros três métodos também foi pequena.

ten Bosch e Angmar-Mansson (2000) levantaram uma série de considerações a respeito dos métodos de diagnóstico, chamando a atenção para os critérios de validação adotados. Relataram que o progresso na ciência diagnóstica traria novos métodos, melhores do que os existentes, incluindo os estabelecidos padrões-ouro, o que conflitaria com a noção que qualquer novo método deveria ser validado contra padrões-ouro já conhecidos. Na opinião dos autores, os seguintes argumentos deveriam ser considerados quando a hipótese de substituição do padrão-ouro for levantada: reprodutibilidade, quantidade e qualidade de informação provida e validade no sentido de relação ao objetivo diagnóstico.

Deery et al. (2000) realizaram um estudo com o objetivo de investigar o efeito de diferentes critérios de diagnóstico de cárie ou do sistema de validação sobre a interpretação de resultados diagnósticos de cárie. As superfícies oclusais de 52 dentes permanentes humanos extraídos (26 molares e 26 premolares) foram examinadas independentemente por dois examinadores calibrados e treinados usando dois critérios de diagnóstico. O sítio da superfície a ser analisado foi predeterminado e marcado através de fotografia. Os dentes foram subsequentemente seccionados nos sítios identificados e examinados independentemente por dois examinadores sob um microscópio binocular (4X de aumento). Esta validação histológica foi conduzida em duas ocasiões usando o apropriado sistema de classificação relacionado ao critério de cárie. A estratégia de tratamento foi sugerida (nenhum cuidado adicional, prevenção ou restauração) associada com cada um dos códigos que foram usados com o propósito de comparação. Em termos de tratamento apropriado proposto, os dois sistemas diagnósticos levaram a diferentes resultados em 23 das 52 ocasiões (Índice de Kappa = 0,27) e o sistema de validação histológica em 5 ocasiões (Índice de Kappa

= 0,84). Os autores concluíram que diferentes critérios diagnósticos ou de validação têm uma maior influência sobre os resultados das investigações de prevalência de cárie e necessidade de tratamento. Enfatizaram a importância em se alcançar um consenso quanto aos mais válidos e implementáveis métodos para facilitar a comparação e avaliação das ferramentas diagnósticas.

Pitts et al. (2001) relataram que o aumentado enfoque sobre a avaliação do desempenho dos examinadores e das novas tecnologias no diagnóstico da cárie seria acompanhado da necessidade de se sintetizar resultados em revisões sistemáticas, fazendo comparações válidas com referência aos padrões-ouro histológicos. Reforçaram também que pouco tem sido relatado sobre a reprodutibilidade do examinador. O objetivo deste estudo foi explorar a reprodutibilidade inter e intra-examinador de três examinadores que avaliaram individualmente 480 secções seriadas para validar os diagnósticos clínicos obtidos em estudo prévio. Após observarem por duas vezes cada secção, a concordância foi calculada usando-se o Índice de Kappa de Cohen. A reprodutibilidade intra-examinador foi: examinador I = 0,58 no ponto de corte D1 e 0,88 no ponto de corte D3; examinador II = 0,74 no D1 e 0,83 no D3; examinador III = 0,82 no D1 e 0,37 no D3. A reprodutibilidade inter-examinador no ponto de corte D1 foi: examinador I X II = 0,58; examinador I X III = 0,63; examinador II X III = 0,46. No ponto de corte D3: examinador I X II = 0,60; examinador I X III = 0,44; examinador II X III = 0,64. Houve uma considerável variação na reprodutibilidade tanto intra quanto inter-examinadores. A direção e a magnitude da variação não foi sempre consistente com o tipo de superfície examinada ou com o ponto de corte empregado. Os autores concluíram que a variação relacionada ao examinador poderia impactar a validade das comparações entre métodos e entre estudos. Foi sugerido também a realização

de trabalhos que explorassem o efeito do treinamento e calibração, conferindo a discordância e a possibilidade de desenvolvimento de padrões metodológicos e de relatos para auxiliar a comparabilidade entre estudos.

Hintze e Wenzel (2002) em trabalho onde avaliaram a influência do método de validação sobre a acurácia diagnóstica, compararam oito sistemas radiográficos para detecção de cárie, tecendo algumas considerações a respeito do método de validação empregado. Expuseram que haveria um consenso geral de que a histologia de cortes dentais seriados seria um método confiável para o diagnóstico de cárie. Concordaram que a validação histológica preencheria os critérios para um robusto padrão-ouro. Foi relatado que em muitos estudos *in vitro* leitura de radiografias convencionais por *experts* tem sido usada como alternativa (VERDONSCHOT et al., 1992; VERDONSCHOT; WENZEL; BRONKHORST, 1993b; JANHOM et al., 1999). Contudo, segundo as autoras, tal método falharia no preenchimento de qualquer dos critérios universais para um forte padrão-ouro na medida em que: diagnósticos radiográficos seriam pobremente reproduzíveis, falhariam em demonstrar desmineralização (restaurações, hipomineralização e artefatos poderiam imitar cáries) e não seriam independentes dos métodos radiográficos sob avaliação. As autoras citam, no entanto, trabalhos onde a leitura de radiografias por *experts* obteve maior acurácia na detecção de cáries do que a microrradiografia ou exame histológico ao estereomicroscópio com um aumento de 20-25X quando estes foram utilizados como padrão-ouro. Neste estudo conduzido por Hintze e Wenzel (2002) para validação da presença ou ausência de cárie, os dentes foram serialmente seccionados e avaliados, em conjunto, por dois examinadores com experiência em microscopia utilizando-se para tal finalidade um estereomicroscópio com um aumento de 12,5-16X. Para o estabelecimento do

padrão-ouro radiográfico, dois radiologistas examinaram, em conjunto, as radiografias. A conclusão alcançada a partir da análise das áreas sob as curvas ROC foi que o método radiográfico demonstrou maior sensibilidade do que a validação histológica ao estereomicroscópio. Foi ressaltado ainda que desde que a validação histológica se apresente como o procedimento mais correto, os resultados obtidos desse modo devem ser tomados como mais confiáveis. Recomendaram que a validação radiográfica deva ser tratada com cautela na medida em que ela possa conter erros.

Pitts e Stamm (2002) relataram as conclusões do *International Collaborative Workshop on Caries Clinical Trials*: o processo de cárie ocorreria como uma interação entre biofilme, superfície dental e subsuperfície. A lesão de cárie seria a manifestação de um estágio do processo num determinado intervalo de tempo; dada a deficiência dos critérios de validação usados até então e o desafio de apropriada validação, novos padrões de referência e protocolos de validação deveriam ser desenvolvidos. Na determinação da acurácia e precisão de novos métodos diagnósticos seria necessário que a distribuição das lesões de cárie em avaliações *in vitro* / *in situ* incluísse todos os estágios da doença onde o método diagnóstico fosse usado *in vivo*; deveria ser verificado que estudos de validação usando dentes imaturos ou somente premolares e terceiros molares poderiam não ser amplamente representativos; para futuros experimentos clínicos o registro de lesões cavitadas somente, como um resultado de medição estaria se tornando em desuso; a avaliação de lesões não cavitadas seria essencial para futuros estudos clínicos sobre cárie. Novos métodos de avaliação de cárie teriam o propósito de ajudar nesta tarefa; todas as novas metodologias deveriam ter uma avaliação *in vitro* sobre lesões de cárie naturais como parte do processo de validação; algumas novas técnicas

diagnósticas poderiam exceder a acurácia, sensibilidade e resolução dos padrões de referência de validação existentes.

Hintze e Wenzel (2003) realizaram um estudo *in vitro*, com o objetivo de comparar os resultados de métodos de diagnóstico, freqüentemente usados como validação para cárie dental (exame clínico, análise radiográfica e exame histológico ao estereomicroscópio). Na primeira parte do estudo um total de 373 superfícies proximais e 158 oclusais não restauradas foi disponibilizado. Para o exame histológico, os dentes foram seccionados em secções com 700 μm de espessura e então examinados ao estereomicroscópio com um aumento de 12,5-16X. O mesmo examinador realizou os exames clínicos, radiográficos e microscópicos. Em uma segunda etapa, houve comparação dos resultados da validação histológica de outras 113 superfícies proximais e 53 oclusais após hemissecção e cortes seriados. O mesmo examinador que participou da primeira fase do estudo realizou também estas análises valendo-se de um estereomicroscópio com um aumento de 12,5X. Para a avaliação da reprodutibilidade intra-examinador 23-60% das superfícies foram reavaliadas pelos diferentes métodos. Dois pontos de corte para cada método foram avaliados (exame clínico: (1) saudável X todos os registros de cárie, e (2) não cavitado X lesões cavitadas; radiografia e estereomicroscopia: (1) saudável X todos os registros de cárie, e (2) sem lesões em dentina X lesões em dentina). Em geral, grandes diferenças foram observadas com os vários métodos. No ponto de corte 1, o exame clínico resultou em mais lesões do que a radiografia, cortes seriados das superfícies proximais e radiografias das superfícies oclusais. No ponto de corte 2, não houve diferenças significantes entre exame clínico, radiografia e cortes seriados nas superfícies proximais, mas nas superfícies oclusais, mais lesões foram diagnosticadas com radiografia e cortes seriados do que com o exame clínico.

Significativamente mais lesões oclusais foram encontradas através dos cortes seriados do que pela radiografia em ambos os limites. Nas superfícies proximais, mais lesões foram diagnosticadas com os cortes seriados do que com a hemissecção nos pontos de corte 1 e 2. Nas superfícies oclusais o mesmo foi encontrado somente para o ponto de corte 1. Os valores obtidos através do Índice de Kappa de Cohen para a reprodutibilidade intra-examinador de 95 superfícies oclusais reexaminadas foram 0,67, 0,83 e 0,90 para exame clínico, exame radiográfico e exame histológico ao estereomicroscópio através de cortes seriados, respectivamente. Com o exame histológico de 12 superfícies oclusais hemisseccionadas, a reprodutibilidade intra-examinador foi 1,00. A reprodutibilidade intra-examinador foi mais alta usando cortes seriados do que radiografia ou exame clínico. As autoras concluíram que os métodos testados usados individualmente como padrão-ouro, alcançaram resultados diagnósticos significativamente diferentes, tanto para superfícies oclusais quanto para proximais. Houve um alto número de diagnósticos contraditórios, dependendo do método e de qual ponto de corte diagnóstico foi usado para um diagnóstico positivo. Segundo o estudo, a histologia através de cortes seriados resultou em uma maior identificação de lesões do que ao se utilizar apenas o hemisseccionamento. A reprodutibilidade intra-observador do exame histológico ao estereomicroscópio foi considerada “quase perfeita”, sendo seguida pela radiografia e exame clínico. Ainda foi mencionada uma falha do estudo, pelo fato deste ter sido conduzido com um único examinador, o que segundo as próprias autoras tem sido comum neste tipo de trabalho. Hintze e Wenzel comentaram que geralmente quando se opta por mais de um examinador, o resultado diagnóstico seria obtido através de um consenso entre eles, o que tornaria

os dados quanto a reprodutibilidade inter e/ou intra-examinadores bastante raros na literatura.

Hintze e Wenzel (2003) ainda afirmaram que existiria grande dificuldade de um determinado método preencher todos os requisitos estabelecidos para um padrão-ouro ideal. Foi comentado que a inspeção visual direta não satisfaria o primeiro critério, já que escores clínicos não são reproduzíveis. Além disso, haveria dificuldade em se estabelecer a profundidade da lesão, e se a presença de cavidade marcasse o limite para doença, muitas lesões, diagnosticadas por outros métodos passariam não detectadas. Em relação à radiografia, mesmo quando realizada por profissionais experientes, nenhum dos critérios universais para um impactante padrão-ouro seria totalmente preenchido. Isto porque diagnósticos radiográficos são pobremente reproduzíveis, poderia haver dúvida em relação a áreas radiolúcidas e o método não seria independente de outros. Já a validação histológica, baseada em exame ao microscópio ou microrradiográfico de dentes hemisseccionados ou cortados serialmente, poderia ser capaz de preencher completamente os requisitos para um padrão-ouro ideal, pois se acreditaria que reunisse todos os três critérios. As autoras, no entanto, ressaltaram as desvantagens do método, que além de mais dispendioso, só poderia ser realizado em dentes extraídos.

Nyvad, Machiulskiene e Baelum (2003) relataram que embora a maioria dos critérios de diagnóstico de cárie enfocasse a profundidade de penetração da lesão e fosse validado pelo uso da aparência histológica das lesões como padrão-ouro, alguns aspectos da cárie não teriam padrão-ouro histológico. Assim, outras formas de validação deveriam ser consideradas, como os critérios clínicos de diagnóstico, incluindo uma avaliação da atividade da lesão. Os autores ressaltaram a importância de tal atitude porque a avaliação da atividade da lesão determinaria a necessidade

de tratamento, partindo-se do princípio que lesões requereriam intervenção, fosse de caráter preventivo ou restaurador.

Huysmans e Longbottom (2004) realizaram uma revisão sobre as mudanças dos métodos diagnósticos e seleção de apropriados padrões-ouro. Relataram que os padrões-ouro disponíveis para lesões de cárie seriam discutíveis e que sua validade estaria longe de ser conhecida ou de poder ser inferida. A partir da definição de cárie como doença de perda mineral, uma acurada medição da perda de pequenas quantidades de mineral a partir de uma superfície dental deveria também produzir um método diagnóstico válido. A verdadeira validação de tais métodos de diagnóstico, de acordo com os autores, poderia somente ser alcançada na clínica, onde a doença real ocorre. Foram relacionadas dezoito publicações de estudos *in vitro* de sobre diagnóstico de cáries em um período de quatro anos (1996-1999) mostrando que a estereomicroscopia de secções dentais foi o método de validação de escolha de quatorze desses trabalhos. Apesar deste fato, foram feitas críticas em relação a metodologia adotada na validação histológica, a qual segundo os autores mostraria grandes variações: (1) uso de diferentes tipos de microscópio para o exame (estereomicroscópio, microscópio de luz refletida ou transmitida); (2) fator de aumento (2,5 a 100 vezes); (3) modo de inspeção (seca ou úmida); (4) uso ou não de corantes; (5) número de critérios fixados (3 a 6); e (6) espessura das secções (100 μm a dentes hemisseccionados). Diversos desses fatores poderiam exercer influência sobre a detecção das lesões, sendo clara, para os autores a necessidade de padronização. A microscopia de luz polarizada e a microrradiografia transversa têm sido usadas como padrão-ouro principalmente com o propósito de detectar zonas de diferentes conteúdos minerais dentro da lesão. Para esta finalidade, a microscopia de luz polarizada requereria a embebição da secção em diferentes

meios, produzindo dados limitados sobre o conteúdo mineral e não melhorando a detecção da profundidade da lesão. A microrradiografia transversa seria menos adequada para um completo processo de validação, desde que o procedimento usaria diferentes espessuras de secção e parâmetros distintos de exposição para esmalte e dentina. Foi comentado ainda que poucos estudos teriam incluído relatos sobre a reprodutibilidade dos padrões-ouro utilizados. Os métodos diagnósticos poderiam depender de fatores clínicos em diferentes graus e estudos clínicos deveriam ser realizados para explorar se os métodos mediriam a extensão da lesão tão acuradamente *in vivo* como eles o fazem *in vitro*. Alguns pontos foram realçados pelos autores: a inspeção microscópica (usualmente estereomicroscopia) das secções seria o mais usado padrão-ouro para validação da profundidade da lesão, embora fosse necessária uma padronização da técnica; a determinação visual da extensão da lesão em dentina não tem sido bem definida ou calibrada; o seccionamento da amostra, embora necessário para quase todos os padrões-ouro, permaneceria como fator de variação, na medida em que a secção analisada poderia não corresponder exatamente ao sítio medido pelo método diagnóstico ou representar somente parte dele, além de consideráveis partes do dente serem perdidas durante o processo de seccionamento (120-250µm por corte). Os autores sugeriram que estudos sobre validação de cáries fossem realizados, incluindo: (1) calibração e padronização da inspeção microscópica e radiográfica; (2) padronização da constituição da amostra; (3) validação de critérios de validação clínicos; (4) desenvolvimento de critérios de validação não destrutivos; e (5) desenvolvimento de padrões para outros parâmetros da cárie (lesão), por exemplo, atividade ou infecção.

Horowitz (2004) relatou que o *National Institutes of Health* (NIH) dos Estados Unidos tem organizado conferências a fim de estabelecer um consenso para resolver controvérsias sobre os mais diversos temas na área de saúde. A conferência “Diagnóstico e Manejo da Cárie Dental através da Vida”, realizada em março de 2001, foi a primeira conferência para desenvolvimento de consenso, direcionada a cárie dental. Um dos pontos abordados foi a questão do padrão-ouro. Foi colocado que a maior limitação de qualquer revisão sobre diagnóstico de cárie seria a falta de um padrão-ouro *in vivo* aceitável. As seguintes questões foram levantadas: (1) deveria um padrão-ouro ser validado? (2) se deveria, como validar um padrão-ouro? (3) deveria um padrão-ouro que não fosse usado no cenário clínico ser utilizado? Segundo o autor, a validação histológica limitaria os estudos *in vivo* a terceiros molares e premolares, e questionou quem iria aderir a um padrão-ouro que utilizasse exame histológico de terceiros molares ou premolares extraídos, os quais não seriam tipicamente os dentes mais suscetíveis a cárie?

Nyvad (2004) relatou que a validade do método de diagnóstico poderia ser avaliada de diferentes maneiras, embora o critério de validade para cárie tenha sido mais comumente aplicado. A avaliação do critério de validade requereria a existência de um critério externo para o fenômeno – denominado “padrão-ouro”. A maioria das avaliações de critérios diagnósticos de cárie tem sido focada na profundidade de penetração da lesão através de avaliações histológicas, clínicas ou radiográficas. Não haveria uma regra rigorosa em relação a qual deveria ser o padrão-ouro, sendo uma questão de escolha. Segundo o autor, o pesquisador deveria escolher o padrão-ouro que representasse o mais alto nível de precisão para o fenômeno sob estudo. Foi mencionado que a profundidade de penetração da lesão poderia não necessariamente ser o mais apropriado padrão-ouro para cáries. Além disso, por

razões práticas ou éticas poderia não ser sempre possível usar o mais exato padrão-ouro. Foi feito um alerta para que pesquisadores e clínicos ficassem atentos quanto a isto quando avaliassem o desempenho de testes diagnósticos de cárie. Novos métodos diagnósticos que se igualassem favoravelmente com um aceito padrão-ouro poderiam não ser melhores do que o padrão-ouro que estivesse sendo usado. Do mesmo modo, se o novo método falhasse na comparação com o padrão-ouro, o teste poderia não necessariamente ser menos útil. Ele poderia, de fato, ser superior ao padrão-ouro utilizado. Nyvad (2004) mencionou que alguns estudos têm mostrado pobre validade com baixa sensibilidade e moderada especificidade para os mais utilizados métodos de diagnóstico clínico de cárie (por exemplo, exame clínico convencional e radiografia). Isto implicaria, segundo o autor, que o diagnóstico da cárie, como normalmente tem sido realizado na prática clínica, seria um procedimento inexato que resultaria em sobre ou sub-diagnóstico. Na realidade, o desempenho diagnóstico poderia ser muito mais pobre do que o admitido porque a maioria dos estudos de validação tem sido realizada sob condições laboratoriais, excluindo as dificuldades introduzidas pela presença de placa e saliva.

Pitts e Stamm (2004) relataram as declarações finais do *International Consensus Workshop on Caries Clinical Trials (ICW-CCT)*. Vale ressaltar alguns pontos, como por exemplo em relação à avaliação e validação dos métodos de detecção de cáries, tendo ficado estabelecido que: (1) Estudos *in vitro* / *in situ* seriam ainda necessários, ambos para desenvolver e avaliar novas técnicas diagnósticas; (2) Para extrapolação dos resultados de estudos laboratoriais sobre técnicas diagnósticas de cárie para o uso clínico, o meio ambiente clínico – tecidos moles, saliva, etc. – deveria ser simulado nos passos laboratoriais; (3) Haveria um interesse sobre os chamados “padrões-ouro” atualmente usados em pesquisa para

detecção e diagnóstico de cárie, particularmente para avaliação das superfícies oclusais e proximais; (4) Devido a deficiência dos atuais “padrões-ouro” e os desafios de se alcançar uma apropriada validação, novos padrões de referência e protocolos de validação deveriam ser desenvolvidos. Quando a acurácia e precisão dos novos métodos de diagnóstico de cáries fossem determinadas, seria necessário que a distribuição das lesões de cárie nas avaliações *in vitro* / *in situ* incluíssem todos os estágios da doença nos quais os métodos de diagnóstico seriam usados *in vivo*; (5) Deveria ser notado que estudos de validação utilizando dentes imaturos ou somente premolares e terceiros molares poderiam não ser suficientemente representativos para prover dados precisos e acurados para propósitos de validação. No que se refere aos novos métodos de detecção e avaliação de cáries chamaram a atenção para o fato de terem considerado importante que todas as novas metodologias devessem ter uma avaliação das lesões naturais de cárie como parte do processo de validação, além de realçarem que algumas novas técnicas diagnósticas poderiam exceder a acurácia e sensibilidade dos padrões de validação existentes. Ao comentarem sobre os métodos estatísticos adotados em testes clínicos de cárie, lembraram que a dicotomização dos resultados seria com freqüência ineficiente e deveria ser evitada a menos que especificamente justificada através dos objetivos da pesquisa. Foi ainda estabelecido que métodos capazes de registrar a continuidade do processo de cárie (incluindo lesões não-cavidades) deveriam ser avaliados e seus resultados comparados com aqueles métodos de avaliação convencionais em um período de estudo de dois a três anos.

ROCHA et al. (2005) realizaram uma pesquisa com o objetivo de comparar dois sistemas digitais de armazenamento de fósforo e filme convencional, assim como verificar a concordância inter-examinadores em diagnóstico radiográfico de

cáries oclusais e proximais. Dois sistemas radiográficos digitais e filme convencional foram usados para avaliar 144 superfícies dentais com e sem cavitação; as radiografias foram analisadas e classificadas em uma escala de quatro pontos. Participaram como examinadores 13 alunos de graduação e um radiologista odontológico. Para validar esta pesquisa foram realizadas análises em microscopia ótica (32X). Os resultados mostraram não haver diferença significativa entre os sistemas radiográficos, entretanto, todos tiveram diferenças significantes quando comparados a microscopia ótica, com exceção do sistema Digora nas superfícies proximais, quando o radiologista foi o examinador. Na avaliação inter-examinadores, foi obtida concordância moderada e entre alunos e o radiologista obteve-se nível sofrível a moderado (Índice de Kappa). Com relação a validade (sensibilidade, especificidade, valores preditivos positivos e negativos), os valores foram semelhantes entre os alunos e o radiologista, com exceção dos valores de sensibilidade nas superfícies oclusais. Os estudantes obtiveram um maior número de resultados falso-positivos e o radiologista odontológico obteve um maior número de resultados falso-negativos.

Bengtson et al. (2005) avaliaram o efeito da experiência clínica do examinador na detecção de lesões de cárie oclusal em dentes decíduos usando inspeção visual e a Fluorescência por Laser Diodo (LF). Foram avaliados 87 sítios suspeitos em 50 dentes decíduos exfoliados ou extraídos através de inspeção visual e LF (DIAGNOdent). A avaliação foi realizada por três estudantes de Odontologia, três dentistas recém-formados e três especialistas em Odontopediatria. Após o exame, os dentes foram serialmente seccionados e avaliados em estereomicroscópio. No ponto de corte esmalte não houve diferença na acurácia entre os diferentes grupos de examinadores usando ambos os métodos. No ponto de corte dentina, contudo, os

estudantes alcançaram os piores resultados utilizando a inspeção visual. Ainda com relação a inspeção visual, os estudantes apresentaram maior sensibilidade e mais baixa especificidade do que os outros grupos em ambos os pontos de corte. Quando foi utilizado o método de Fluorescência por Laser Diodo (LF), não houve diferenças significantes entre os parâmetros, e a concordância inter-examinador (Índice de Kappa de Cohen) foi maior.

Mestriner, Vinha e Mestriner Júnior (2005) realizaram um estudo cujo objetivo foi comparar, *in vitro*, a efetividade de métodos de diagnóstico de lesões de cáries oclusais em dentina. Foram avaliados 38 sítios em terceiros molares sem cavidades macroscópicas, de indivíduos adultos. Os exames inspeção visual, endoscópio (AccuCam), inspeção visual e uso de sonda exploradora de ponta romba (táctil), radiografia interproximal, radiografia digital direta e radiografia digital direta com manipulação do contraste e brilho foram realizados por cinco examinadores não calibrados. Para validação dos dados os dentes foram serialmente seccionados e avaliados histologicamente em estereomicroscópio. Foi concluído que, comparativamente, a efetividade entre os métodos clínicos foram similares, tanto quanto os radiográficos. Os métodos clínicos apresentaram um menor número de diagnósticos falso-positivos do que os radiográficos.

2.5 ANÁLISE ESTATÍSTICA

Bulman e Osborn (1989) escreveram um artigo sobre consistência diagnóstica em Odontologia onde comentaram sobre a falta de confiabilidade relacionada à análise de dados que envolvessem diagnósticos clínicos subjetivos. Neste texto vale ressaltar as seguintes afirmativas: (1) em pesquisa odontológica clínica ou

epidemiológica, variações na interpretação do critério de diagnóstico teriam um marcante efeito sobre a confiabilidade da análise subsequente; (2) a pesquisa dental dependeria do modo como o pesquisador interpretasse o critério diagnóstico e de como sua interpretação seria comparada com a interpretação de outro; e (3) a adoção de índices diagnósticos na epidemiologia odontológica seria uma tentativa para se padronizar os critérios de diagnóstico a fim de que todos os examinadores, usando aquele índice, pudessem relatar a doença consistentemente. Contudo, segundo os autores, mesmo com o mais detalhado critério diagnóstico não se teria garantia de consistência do examinador, e mesmo um único examinador poderia ser inconsistente de uma ocasião para a outra. Os autores afirmaram ainda que até aquele momento era aceito que a comparação de dados envolvendo diagnóstico clínico subjetivo por examinadores independentes não seria confiável.

Ainda segundo Bulman e Osborn (1989) alguns métodos estatísticos atribuiriam valores numéricos para a reprodutibilidade intra ou inter-examinador, onde somente um examinador está envolvido em ocasiões distintas ou quando mais de um examinador faria parte da pesquisa. O diagnóstico poderia ser dicotômico, ou seja, baseado no “sim/não”, “presente/ausente” ou “positivo/negativo”. Ele poderia também estar relacionado à severidade da doença, profundidade da lesão ou extensão do dano.

2.5.1 Índice de Kappa de Cohen

Esta estatística (COHEN, 1960) relaciona a medida de concordância obtida, excluindo o grau de concordância alcançado tendo o diagnóstico sido feito ao acaso. Segundo Bulman e Osborn (1989) este seria, provavelmente, o mais confiável meio

de se avaliar a concordância total do examinador. É um método dicotômico e no caso do diagnóstico de cárie, é baseado em “saudável/cariado”.

Os dados obtidos são registrados em uma tabela 2 X 2, composta por quatro células, que contêm os seguintes dados (tabela 3):

- **Célula a:** proporção de cortes em que ambos os examinadores concordaram como sendo saudáveis;
- **Célula b:** proporção de cortes onde um dos examinadores considerou como sendo saudável e o outro considerou como cariado;
- **Célula c:** proporção de cortes onde um examinador avaliou como cariado e o outro como saudável;
- **Célula d:** proporção de cortes que ambos os examinadores avaliaram como cariados.

Tabela 3 – Matriz de decisão para o cálculo do Índice de Kappa de Cohen

		Examinador I ou 1º Exame		
		SAUDÁVEL	CARIADO	TOTAL
Examinador II ou 2º Exame	SAUDÁVEL	a	c	a + c
	CARIADO	b	d	b + d
	TOTAL	a + b	c + d	a + b + c + d

A fórmula geral para o cálculo do Índice de Kappa de Cohen é:

$$K = \frac{P_o - P_e}{1 - P_e}, \text{ onde:}$$

$$P_o = a + d$$

$$P_e (\text{saudável}) = (a + c) \times (a + b)$$

$$P_e (\text{cariado}) = (b + d) \times (c + d)$$

Logo:

$$Pe = [(a + c) (a + b)] + [(b + d) (c + d)]$$

Po é a proporção de concordância observada e **Pe** é a proporção de concordância esperada.

Landis e Koch (1977) propuseram uma nomenclatura para descrever a concordância associada com a estatística de Kappa, que compreende faixas de valores (tabela 4).

Tabela 4 – Interpretação do Índice de Kappa de Cohen (Landis; Koch, 1977)

Índice de Kappa	Concordância
< 0,00	Pobre
0,00 – 0,20	Fraca
0,21 – 0,40	Sofrível
0,41 – 0,60	Moderada
0,61 – 0,80	Substancial
0,81 – 1,00	Quase perfeita

2.5.2 Índice de Kappa não Ponderado

O Índice de Kappa de Cohen pode ser usado na maioria dos estudos de cárie dental onde os registros estão baseados em doença presente ou ausente, ou seja, em dados dicotômicos, sendo o seu cálculo bastante simples.

Kingman (1986) propôs uma modificação do Índice de Kappa de Cohen para ser aplicada em estudos que envolvessem diferentes graus de severidade da doença. Dessa forma poderia ser obtida uma maior quantidade de informações sobre a reprodutibilidade dos examinadores. O Índice de Kappa não Ponderado considera somente áreas de total concordância, sendo seu cálculo mais complexo.

Imaginemos que queiramos calcular a reprodutibilidade intra-examinador após exames de 127 cortes histológicos, que foram classificados de acordo com sete critérios de diferentes profundidades de lesão cariiosa, como na presente pesquisa. A primeira tabela construída conterá os valores observados no primeiro exame e na repetição do mesmo (2º exame) pelo examinador A (tabela 5).

Tabela 5 – Valores observados por um imaginário examinador A para o cálculo do Índice de Kappa não Ponderado

Examinador A - Valores Observados no 1º Exame

	0	1	2	3	4	5	6	TOTAL
0	11 (0,09)	02 (0,02)		02 (0,02)				15 (0,12)
1	01 (0,01)	08 (0,06)	06 (0,05)	02 (0,02)				17 (0,13)
2	01 (0,01)	05 (0,04)	14 (0,11)	02 (0,02)				22 (0,17)
3	02 (0,02)	03 (0,02)	06 (0,05)	30 (0,24)	03 (0,02)			44 (0,35)
4			02 (0,02)	05 (0,04)	08 (0,06)	01 (0,01)		16 (0,13)
5					01 (0,01)	04 (0,03)		05 (0,04)
6						04 (0,03)	04 (0,03)	08 (0,06)
TOTAL	15 (0,12)	18 (0,14)	28 (0,22)	41 (0,32)	12 (0,09)	09 (0,07)	04 (0,03)	127 (1,00)

Examinador A
Valores
Observados
no 2º Exame

Interpretando a tabela acima, 11 cortes foram classificados pelo examinador A como “0” tanto no 1º quanto no 2º exame. Dois cortes foram avaliados como “1” no 1º exame e como “0” no 2º, e assim sucessivamente. Estes dados foram convertidos em proporções (entre parênteses na tabela), através da divisão pelo número total da amostra, no caso 127.

O cálculo da probabilidade de concordância observada (**P_o**), será feito a partir do somatório das proporções observadas nas células “0-0”, “1-1”, “2-2”, “3-3”, “4-4”, “5-5” e “6-6”.

$$P_o = 0,09 + 0,06 + 0,11 + 0,24 + 0,06 + 0,03 + 0,03 = \mathbf{0,62}$$

Uma segunda tabela será construída com os valores esperados nos exames (tabela 6).

Tabela 6 – Valores esperados de um imaginário examinador A para o cálculo do Índice de Kappa não Ponderado
Examinador A - Valores Esperados no 1º Exame

	0	1	2	3	4	5	6	TOTAL
0	1,77 (0,01)	2,13 (0,02)	3,31 (0,03)	4,84 (0,04)	1,42 (0,01)	1,06 (0,01)	0,47 (0,00)	15 (0,12)
1	2,01 (0,02)	2,41 (0,02)	3,75 (0,03)	5,49 (0,04)	1,61 (0,01)	1,20 (0,01)	0,54 (0,00)	17 (0,13)
2	2,60 (0,02)	3,12 (0,02)	4,85 (0,04)	7,10 (0,06)	2,08 (0,02)	1,56 (0,01)	0,69 (0,01)	22 (0,17)
3	5,20 (0,04)	6,24 (0,05)	9,70 (0,08)	14,20 (0,11)	4,16 (0,03)	3,12 (0,02)	1,39 (0,01)	44 (0,35)
4	1,89 (0,02)	2,27 (0,02)	3,53 (0,03)	5,17 (0,04)	1,51 (0,01)	1,13 (0,01)	0,50 (0,00)	16 (0,13)
5	0,59 (0,01)	0,71 (0,01)	1,10 (0,01)	1,61 (0,01)	0,47 (0,00)	0,35 (0,00)	0,16 (0,00)	05 (0,04)
6	0,94 (0,01)	1,13 (0,01)	1,76 (0,01)	2,58 (0,02)	0,76 (0,01)	0,57 (0,00)	0,25 (0,00)	08 (0,06)
TOTAL	15 (0,12)	18 (0,14)	28 (0,22)	41 (0,32)	12 (0,09)	09 (0,07)	04 (0,03)	127 (1,00)

O valor esperado será calculado a partir da tabela referente aos valores observados, multiplicando-se o total da linha pelo total da coluna e dividindo-se o resultado pelo total da amostra (127). O número obtido será também convertido em proporção através de sua divisão pelo valor total (127).

A probabilidade de concordância esperada (**Pe**) é o somatório das proporções esperadas nas células “0-0”, “1-1”, “2-2”, “3-3”, “4-4”, “5-5” e “6-6”.

$$Pe = 0,01 + 0,02 + 0,04 + 0,11 + 0,01 + 0,00 + 0,00 = \mathbf{0,19}$$

$$K = \frac{Po - Pe}{1 - Pe} = \frac{0,62 - 0,19}{1 - 0,19} = \frac{0,43}{0,81} = \mathbf{0,53} \text{ (concordância moderada)}$$

2.5.3 Índice de Kappa Ponderado

O Índice de Kappa não Ponderado não leva em consideração a gravidade do erro. Mas segundo Bulman e Osborn (1989), um erro entre escores diagnósticos próximos, embora não represente uma concordância total, seria melhor do que uma

total discordância. Por esta razão alguma espécie de sistema de peso deveria ser aplicada a tais escores, de modo que eles pudessem ter uma apropriada e realística contribuição para o Índice de Kappa estatístico.

O Índice de Kappa Ponderado foi introduzido por Cohen (1968), mas segundo Cicchetti (1976), ainda existiriam dois problemas a serem considerados: (1) Quais pesos deveriam ser usados para se determinar o nível de concordância? e (2) Qual seria o tamanho mínimo da amostra para que fosse válida a aplicação do Índice de Kappa Ponderado?

De acordo com Cicchetti (1976), no caso de uma escala ordinal contínua (C-O), como a que adotamos em nosso estudo, o peso seria determinado simplesmente pelo número de pontos ordinais (k) na escala. Cada peso corresponderia a uma categoria de concordância. Seguindo o exemplo do nosso estudo, que adota uma escala de sete pontos, existiriam sete pesos lineares: o primeiro para os pares de completa concordância (0-0; 1-1; 2-2; 3-3; 4-4; 5-5; 6-6); o segundo para os pares de concordância mais aproximada (0-1; 1-0; 1-2; 2-1; 2-3; 3-2; 3-4; 4-3; 4-5; 5-4; 5-6; 6-5), seguindo até o sétimo peso para os pares de completa discordância (0-6; 6-0). Os pesos, para qualquer tamanho k de escala ordinal, seguem a fórmula proposta por Cicchetti (1976), onde k se refere ao número de pontos de uma dada escala ordinal contínua (C-O):

$$\text{Pesos lineares (C-O)} = \frac{k-1}{k-1}; \frac{k-2}{k-1}; \frac{k-3}{k-1}; \dots \frac{k-k}{k-1}$$

A partir desta fórmula podemos atribuir pesos para cada par de concordância. Na tabela abaixo estão relacionados os pesos adotados em nosso estudo.

Tabela 7: Pesos atribuídos para cada par de concordância em uma escala ordinal contínua de sete pontos para o cálculo do Índice de Kappa Ponderado.

	0	1	2	3	4	5	6
0	1,00	0,83	0,67	0,50	0,33	0,17	0,00
1	0,83	1,00	0,83	0,67	0,50	0,33	0,17
2	0,67	0,83	1,00	0,83	0,67	0,50	0,33
3	0,50	0,67	0,83	1,00	0,83	0,67	0,50
4	0,33	0,50	0,67	0,83	1,00	0,83	0,67
5	0,17	0,33	0,50	0,67	0,83	1,00	0,83
6	0,00	0,17	0,33	0,50	0,67	0,83	1,00

A vantagem desse sistema de pesos imaginado por Cicchetti (1976) é a penalização dos erros de acordo com sua gravidade.

Os pesos estabelecidos para cada grau de concordância são multiplicados pela proporção encontrada em cada célula das tabelas referentes aos valores observados e aos valores esperados. Em nosso caso, a proporção na célula “0-0” deverá ser multiplicada por 1,00. A da célula “0-1” será multiplicada por 0,83 e assim sucessivamente. Dessa forma, a maior contribuição para o escore total será dada pelas células onde a concordância foi mais alta (“0.0”; “1.1”; “2.2”...) e terão uma mínima participação as células onde a concordância foi baixa (“0.6”; “1.6”; “5.0”, “6.0”, “6.1”...).

Para o cálculo do Índice de Kappa Ponderado temos:

Po = soma de todos os pesos referentes aos valores observados.

Pe = soma de todos os pesos referentes aos valores esperados.

Segundo Bulman e Osborn (1989), a grande vantagem desta análise de Índice de Kappa é que se torna possível, através do uso de pesos discriminatórios sobre os dados originais, analisar onde repousa a área mais frágil do diagnóstico.

3 OBJETIVOS

O presente trabalho tem como objetivo a avaliação da reprodutibilidade de examinadores durante a validação histológica de métodos de diagnóstico da cárie dental. Tem ainda como objetivos específicos:

- Avaliar a reprodutibilidade intra e inter-examinadores durante o exame histológico utilizado como critério de validação para os métodos de diagnóstico de cárie;
- Analisar o impacto que a calibração dos examinadores pode exercer sobre a reprodutibilidade do método de validação;
- Verificar se a experiência prévia do examinador pode melhorar os resultados da reprodutibilidade do critério de validação em estudo.

4 HIPÓTESES

- A reprodutibilidade intra e inter-examinador do exame histológico como critério de validação para os diferentes métodos de diagnóstico da cárie dental apresenta variabilidade.
- O exercício de calibração melhora a reprodutibilidade dos examinadores do exame histológico utilizado como critério de validação para métodos de diagnóstico da cárie dental.
- A experiência prévia do examinador tem impacto na reprodutibilidade dos examinadores durante a validação histológica de métodos de diagnóstico da cárie dental.

5 MATERIAIS E MÉTODOS

5.1 AMOSTRA:

A avaliação foi realizada em 123 sítios obtidos a partir de superfícies oclusais de 89 molares permanentes extraídos. As secções histológicas analisadas foram doadas pela orientadora deste projeto, tendo sido utilizadas pela mesma em pesquisas anteriores (CÔRTEZ; ELLWOOD; EKSTRAND, 2003; ELLWOOD; CÔRTEZ, 2004). A seleção dos dentes, na pesquisa inicial, teve a intenção de prover uma gama de elementos saudáveis e com lesões tanto em esmalte como em dentina. Extensas lesões oclusais em dentina, assim como nas superfícies lisas e proximais foram excluídas da amostra. Os sítios de interesse foram selecionados e os elementos foram então seccionados em secções com cerca de 350 μm de espessura.

5.2 EXAMINADORES:

O estudo foi desenvolvido com a participação de três examinadores, com diferentes graus de experiência em leitura de cortes histológicos. Um dos examinadores era experiente, com formação em Patologia Oral e atuava como professor de histologia (examinador A). Os dois outros examinadores eram inexperientes em histologia, cirurgiões-dentistas e alunos do curso de Mestrado Profissionalizante da Universidade Veiga de Almeida (examinadores B e C).

5.3 EXAME HISTOLÓGICO:

Para o exame histológico, realizado no Laboratório de Eletroquímica Aplicada da COPPE (UFRJ), foi utilizado um estereomicroscópio da marca OLYMPUS (modelo SZ-CTV - Japão). Inicialmente foi feita uma calibração do aparelho para que todos os examinadores pudessem obedecer aos mesmos parâmetros. A intensidade de luz foi regulada no nível 2; a haste de onde provinha a iluminação ficou localizada a uma altura de 3,5 cm do local onde seriam colocados os cortes; o aumento foi regulado em 4 X, tendo sido todos os cortes examinados úmidos. O lado do corte a ser examinado já havia sido pré-determinado, assim como o sítio da oclusal a ser analisado.

Os cortes histológicos foram fotografados por uma câmera da marca SONY (Color Video Camera-CCD-IRIS/RGB – modelo: DXC-151A – Japão) acoplada ao estereomicroscópio e as imagens foram transferidas para o computador através do programa ATI VIDEO PLAYER, podendo ser observadas neste trabalho (anexo I). As fotos não tiveram influência no exame, na medida em que nenhum dos examinadores teve acesso a elas.

Numa primeira fase, denominada de Pré-Calibração, os três examinadores receberam por escrito, os critérios para diagnóstico da profundidade das lesões de cárie nos cortes histológicos (tabela 8), segundo Côrtes, Ellwood e Ekstrand (2003), assim como as regras a serem observadas durante a avaliação da profundidade das lesões. Os critérios de diagnóstico adotados nos exames podem ser observados na tabela 8.

Tabela 8: Critérios histológicos adotados no exame ao estereomicroscópio, de acordo com a profundidade da lesão cariada.

CRITÉRIOS HISTOLÓGICOS PARA EXAME AO ESTEREOMICROSCÓPIO	
ESCORE	DESCRIÇÃO
0	Sem desmineralização
1	Desmineralização estendendo-se até a metade externa do esmalte
2	Desmineralização estendendo-se até a metade interna do esmalte, sem incluir a junção amelodentinária
3	Desmineralização estendendo-se até a junção amelodentinária
4	Desmineralização estendendo-se até o terço externo da dentina
5	Desmineralização estendendo-se até o terço médio da dentina
6	Desmineralização estendendo-se até o terço interno da dentina

Fonte: CÔRTEZ; ELLWOOD; EKSTRAND. An in vitro comparison of a combined FOTI/visual examination of occlusal caries with other caries diagnostic methods and the effect of stain on their diagnostic performance. *Caries Res*, v. 37, p. 8-16, 2003.

As regras gerais de exame foram:

- A área a ser lida deveria ser aquela que apresentasse lesão mais profunda;
- Em caso de dúvida, optar pelo critério mais conservativo;
- A cada 10 leituras deveria haver um descanso de 5 minutos.

Os três examinadores realizaram os exames de todos os cortes histológicos, independentemente, um a cada vez. Após o intervalo mínimo de uma semana, os exames foram repetidos.

Depois de todos os examinadores terem realizado os exames pré-calibração, houve uma sessão de calibração, com a participação dos três examinadores e coordenada pela orientadora deste estudo (calibradora), a qual possuía domínio prévio dos critérios e regras adotadas. Foi explicado novamente cada critério e a calibradora juntamente com os três examinadores realizaram, em conjunto, o exame de 40 cortes, escolhidos aleatoriamente, atingindo-se um consenso em relação ao critério a ser atribuído para cada um deles.

Após a sessão de calibração, iniciou-se a fase dos exames Pós-Calibração após intervalo mínimo de uma semana. Os três examinadores realizaram individualmente um novo exame dos cortes histológicos tendo sido os exames repetidos após um intervalo mínimo de sete dias.

5.4 ANÁLISE ESTATÍSTICA:

A reprodutibilidade intra-examinador foi avaliada ao se comparar os resultados de um mesmo examinador no primeiro exame com sua repetição. A reprodutibilidade inter-examinador foi avaliada ao se comparar os resultados dos examinadores entre si, durante o primeiro exame de cada fase. A reprodutibilidade intra e inter-examinadores tanto nos exames pré-calibração quanto nos exames pós-calibração foi avaliada através do Índice de Kappa de Cohen (COHEN, 1960), Índice de Kappa não Ponderado (KINGMAN, 1986) e Índice de Kappa Ponderado (COHEN, 1968).

Por sua natureza dicotômica, ao utilizarmos o Índice de Kappa de Cohen foi necessário estabelecer os pontos de corte nos critérios histológicos. Dessa forma, o ponto de corte esmalte considerou como lesão de cárie todas aquelas que atingissem esmalte ou dentina, excluindo-se apenas o escore “0” (sem desmineralização). O ponto de corte dentina classificou como lesão de cárie aquelas que afetassem a dentina (escores 4, 5 e 6), sendo os demais escores 0, 1, 2 e 3 considerados como “não cárie dentinária”.

Uma questão importante abordada por Cicchetti (1976) e que foi seguida, se relaciona ao tamanho mínimo da amostra. Segundo o autor, este poderia ser calculado através da fórmula abaixo:

$$N \geq 2 k^2 ,$$

onde k seria o número de pontos de critérios de diagnóstico histológico da escala ordinal utilizada.

No presente trabalho, para o qual foi adotada uma escala de sete pontos (tabela 8), o número mínimo da amostra seria de 98 e foi obtida uma boa margem de segurança ao trabalharmos com 123 cortes.

6 RESULTADOS

A tabela 9 apresenta os resultados da reprodutibilidade intra-examinador nos exames pré e pós-calibração, para os três examinadores, nos pontos de corte esmalte e dentina, obtidos através do Índice de Kappa de Cohen e classificados por Landis e Koch (1977).

Tabela 9 – Reprodutibilidade intra-examinador nos exames pré e pós-calibração, nos pontos de corte esmalte e dentina, obtida pelo Índice de Kappa de Cohen e classificada por Landis e Koch (1977).

REPRODUTIBILIDADE INTRA-EXAMINADOR (Índice de Kappa de Cohen)				
	Ponto de Corte Esmalte		Ponto de Corte Dentina	
	Pré-calibração	Pós-calibração	Pré-calibração	Pós-calibração
Examinador A	0,73 (substancial)	0,63 (substancial)	0,79 (substancial)	0,97 (quase perfeita)
Examinador B	0,71 (substancial)	0,61 (substancial)	0,78 (substancial)	0,86 (quase perfeita)
Examinador C	0,58 (moderada)	0,56 (moderada)	0,53 (moderada)	0,87 (quase perfeita)

No ponto de corte esmalte, os valores pré-calibração do examinador experiente A (Ex. A), examinador inexperiente B (Ex. B) e examinador inexperiente C (Ex. C) foram, respectivamente, 0,73, 0,71 e 0,58, classificados por Landis e Koch (1977) como substancial, substancial e moderada. Na etapa pós-calibração a reprodutibilidade intra-examinador foi: 0,63 (Ex. A), 0,61 (Ex. B) e 0,56 (Ex. C), consideradas substancial, substancial e moderada.

No ponto de corte dentina, os valores pré-calibração foram, respectivamente, 0,79 (Ex. A), 0,78 (Ex. B) e 0,53 (Ex. C), classificados como substancial, substancial

e moderada. Na etapa pós-calibração a reprodutibilidade intra-examinador, foi: 0,97 (Ex. A), 0,86 (Ex. B) e 0,87 (Ex. C), todas consideradas quase perfeita.

Para a etapa pré e pós-calibração, no ponto de corte esmalte, o Ex. A e Ex. B foram similares, tendo o Ex. C os resultados inferiores. Quando analisamos o ponto de corte dentina, os Ex. A e B também foram superiores ao Ex. C na pré-calibração, mas na pós-calibração os resultados foram equivalentes para os três examinadores.

Ao analisarmos os resultados individuais dos examinadores percebemos que no ponto de corte esmalte houve uma regularidade dos resultados dos três examinadores do pré para o pós-calibração. Já no ponto de corte dentina, todos os examinadores melhoraram seus resultados na etapa pós-calibração, sendo mais intensa para o Ex. C. Ao compararmos os dois pontos de corte percebemos que na pré-calibração, os três examinadores tiveram resultados semelhantes, tanto no ponto de corte esmalte quanto no ponto de corte dentina. Já na pós-calibração houve melhora para todos os examinadores no ponto de corte dentina, mais evidente, no entanto, para o Ex. C.

A tabela 10 revela os resultados da reprodutibilidade intra-examinador nos exames pré e pós-calibração, para os três examinadores, obtidos através do Índice de Kappa não Ponderado e classificados por Landis e Koch (1977).

Tabela 10 – Reprodutibilidade intra-examinador nos exames pré e pós-calibração obtida através do Índice de Kappa não Ponderado e classificada por Landis e Koch (1977).

REPRODUTIBILIDADE INTRA-EXAMINADOR (Índice de Kappa não Ponderado)		
	Pré-calibração	Pós-calibração
Examinador A	0,56 (moderada)	0,60 (moderada)
Examinador B	0,54 (moderada)	0,53 (moderada)
Examinador C	0,45 (moderada)	0,46 (moderada)

A reprodutibilidade na etapa pré-calibração foi 0,56, 0,54 e 0,45 para o Ex. A, Ex. B e Ex. C e na etapa pós-calibração 0,60, 0,53 e 0,46, resultados considerados por Landis e Koch (1977) como moderada.

A reprodutibilidade intra-examinador, encontrada pelo Kappa não Ponderado, foi considerada semelhante segundo Landis e Koch (1977), individualmente ou quando foram comparados os resultados pré e pós-calibração.

A tabela 11 contém os resultados da reprodutibilidade intra-examinador nos exames pré e pós-calibração, para os três examinadores, obtidos pelo Índice de Kappa Ponderado e classificados por Landis e Koch (1977).

Tabela 11 – Reprodutibilidade intra-examinador, nos exames pré e pós-calibração através do Índice de Kappa Ponderado e classificada por Landis e Koch (1977).

REPRODUTIBILIDADE INTRA-EXAMINADOR (Índice de Kappa Ponderado)		
	Pré-calibração	Pós-calibração
Examinador A	0,85 (quase perfeita)	0,74 (substancial)
Examinador B	0,85 (quase perfeita)	0,85 (quase perfeita)
Examinador C	0,71 (substancial)	0,56 (moderada)

Os valores pré-calibração para o Ex. A, Ex. B e Ex. C foram 0,85, 0,85 e 0,71, respectivamente. Na etapa pós-calibração foram 0,74, 0,85 e 0,56. A reprodutibilidade do Ex. A passou de quase perfeita no pré-calibração para substancial no pós-calibração. Os resultados do Ex. B mantiveram-se no nível quase perfeito e o Ex. C teve uma reprodutibilidade substancial no pré-calibração, passando para moderada no pós-calibração.

Na etapa pré-calibração os Ex. A e B tiveram resultados iguais, sendo superiores ao Ex. C. Já no pós-calibração todos os examinadores foram diferentes em seus resultados. Quando foram analisadas as diferenças individuais dos examinadores nas duas etapas, foi evidenciada uma queda na reprodutibilidade pós-calibração dos Ex A e C, sendo somente o Ex. B estável em seus resultados.

A tabela 12 apresenta os resultados da reprodutibilidade inter-examinador nos exames pré e pós-calibração nos pontos de corte esmalte e dentina, obtidos através do Índice de Kappa de Cohen e classificados por Landis e Koch (1977).

Tabela12 – Reprodutibilidade inter-examinador nos exames pré e pós-calibração nos pontos de corte esmalte e dentina, obtida através do Índice de Kappa de Cohen, e classificada por Landis e Koch (1977)

REPRODUTIBILIDADE INTER-EXAMINADOR (Índice de Kappa de Cohen)				
	Ponto de Corte Esmalte		Ponto de Corte Dentina	
	Pré-calibração	Pós-calibração	Pré-calibração	Pós-calibração
Examinador A/B	0,64 (substancial)	0,47 (moderada)	0,75 (substancial)	0,69 (substancial)
Examinador A/C	0,43 (moderada)	0,38 (sofrível)	0,33 (sofrível)	0,79 (substancial)
Examinador B/C	0,55 (moderada)	0,56 (moderada)	0,38 (sofrível)	0,73 (substancial)

No ponto de corte esmalte, a reprodutibilidade inter-examinador pré-calibração entre Ex. A/B, Ex. A/C e Ex. B/C foi, respectivamente, 0,64, 0,43 e 0,55, classificados por Landis e Koch (1977) como substancial, moderada e moderada. Na etapa pós-calibração a reprodutibilidade inter-examinador, foi 0,47 (Ex. A/B), 0,38 (Ex. A/C) e 0,56 (Ex. B/C), consideradas moderada, sofrível e moderada.

No ponto de corte dentina, a reprodutibilidade inter-examinador pré-calibração foi, respectivamente, 0,75 (Ex. A/B), 0,33 (Ex. A/C) e 0,38 (Ex. B/C), classificadas como substancial, sofrível e sofrível. Na etapa pós-calibração a reprodutibilidade inter-examinador foi 0,69 (Ex. A/B), 0,79 (Ex. A/C) e 0,73 (Ex. B/C), todas consideradas substancial.

Para a etapa pré-calibração e no ponto de corte esmalte, a reprodutibilidade entre Ex. A/B foi superior aos demais, que tiveram resultados similares. Já na pós-calibração, o pior resultado foi obtido entre Ex. A/C, tendo os Ex. A/B e Ex. B/C obtido reprodutibilidade semelhante. Quando foi analisado o ponto de corte dentina, a reprodutibilidade entre Ex. A/B também foi superior aos resultados entre Ex. A/C e

Ex B/C na pré-calibração, mas na pós-calibração os resultados foram equivalentes entre todos os três examinadores.

Em relação a reprodutibilidade inter-examinador, no ponto de corte esmalte houve uma piora dos resultados entre Ex. A/B e A/C do pré para o pós-calibração, tendo a comparação entre Ex. B/C mantido o resultado da pré-calibração. Já no ponto de corte dentina, a reprodutibilidade entre Ex. A/C e B/C foi melhor na etapa pós-calibração, sendo mantida entre os Ex. A/B. Ao se comparar os dois pontos de corte foi notado que na pré-calibração, apenas a reprodutibilidade entre Ex. A/B apresentou resultados semelhantes, tanto no ponto de corte esmalte quanto no ponto de corte dentina, tendo as demais comparações (Ex. A/C e Ex. B/C) apresentado piores resultados no ponto de corte dentina. Já na pós-calibração houve melhora da reprodutibilidade entre todos os examinadores no ponto de corte dentina, mais evidente, no entanto, entre os Ex. A/C.

A tabela 13 apresenta os resultados da reprodutibilidade inter-examinador nos exames pré e pós-calibração obtidos através do Índice de Kappa não Ponderado e classificados por Landis e Koch (1977).

Tabela 13 – Reprodutibilidade inter-examinador nos exames pré e pós-calibração obtida através do Índice de Kappa não Ponderado e classificada por Landis e Koch (1977).

REPRODUTIBILIDADE INTER-EXAMINADOR (Índice de Kappa não Ponderado)		
	Pré-calibração	Pós-calibração
Examinador A/B	0,29 (sofrível)	0,38 (sofrível)
Examinador A/C	0,16 (fraca)	0,29 (sofrível)
Examinador B/C	0,27 (sofrível)	0,40 (sofrível)

Na etapa pré-calibração foi verificada a reprodutibilidade de 0,29, 0,16 e 0,27 entre os Ex. A/B, Ex. A/C e Ex. B/C, resultados considerados por Landis e Koch (1977) como sofrível, fraco e sofrível. Na etapa pós-calibração, os resultados foram, respectivamente, 0,38, 0,29 e 0,40, tidos como sofrível.

A reprodutibilidade inter-examinador foi semelhante entre os Ex. A/B e Ex. B/C, tanto no pré quanto no pós-calibração, com exceção entre os Ex. A/C que na etapa pré-calibração foi inferior aos demais. No entanto, quando se comparou os resultados pré e pós-calibração entre os Ex. A/C foi notada uma pequena melhora na classificação.

A tabela 14 revela os resultados da reprodutibilidade inter-examinador nos exames pré e pós-calibração obtidos através do Índice de Kappa Ponderado e classificados por Landis e Koch (1977).

Tabela 14 – Reprodutibilidade inter-examinador nos exames pré e pós-calibração obtida através do Índice de Kappa Ponderado e classificada por Landis e Koch (1977).

REPRODUTIBILIDADE INTER-EXAMINADOR (Índice de Kappa Ponderado)		
	Pré-calibração	Pós-calibração
Examinador A/B	0,63 (substancial)	0,68 (substancial)
Examinador A/C	0,30 (sofrível)	0,52 (moderada)
Examinador B/C	0,50 (moderada)	0,55 (moderada)

A reprodutibilidade pré-calibração entre os Ex. A/B, Ex. A/C e Ex. B/C foi 0,63, 0,30 e 0,50, respectivamente. Na etapa pós-calibração os valores foram 0,68, 0,52 e 0,55. A reprodutibilidade entre Ex. A/B manteve-se como substancial. Já os resultados entre Ex. A/C passaram de sofrível para moderada e os dos Ex. B/C foram moderada em ambas etapas.

Os resultados da reprodutibilidade entre Ex. A/B foram mais elevados do que entre os Ex. A/C e Ex. B/C tanto na etapa pré quanto pós-calibração. Na pré-calibração a comparação entre Ex. B/C teve melhores resultados do que o Ex. A/C. Já na pós-calibração a reprodutibilidade entre Ex. A/C e Ex. B/C foram similares. Comparando-se o pré e pós-calibração verifica-se que a comparação entre Ex. A/B e Ex. B/C manteve seus resultados e entre Ex. A/C obteve uma melhora na etapa pós-calibração.

7 DISCUSSÃO

Somente através de um correto diagnóstico das lesões de cárie pode-se indicar um tratamento adequado para os pacientes. No entanto, vários autores (WEERHEIJM; van AMERONGEN; EGGINK, 1989; KIDD; RICKETTS; PITTS, 1993; NEWBRUN, 1993; EKSTRAND; RICKETTS; KIDD, 1997) têm demonstrado a dificuldade em se realizar o diagnóstico da cárie oclusal. Na tentativa de se aprimorar o diagnóstico das lesões oclusais novas tecnologias têm surgido. Embora Ekstrand, Ricketts e Kidd (1997) tenham questionado a validade dos estudos laboratoriais sobre métodos diagnósticos de cárie, há uma concordância com a afirmativa de Wenzel e Verdonschot (1994) de que todo sistema diagnóstico deveria ser primeiramente testado *in vitro* antes que pudesse ser utilizado em pacientes. Para se validar tais testes, no entanto, tem sido recomendada a comparação de seus resultados com aqueles estabelecidos por um teste de exatidão máxima, denominado padrão-ouro (PEREIRA, 1995; FERREIRA et al., 1999; ten BOSCH; ANGMAR-MANSSON, 2000).

Padrão-ouro (*gold standard*), por definição, é um método diagnóstico perfeito para se predizer a presença ou ausência de doença (WENZEL; HINTZE, 1999b; HINTZE; WENZEL, 2003). No entanto, a literatura tem demonstrado que até mesmo os padrões-ouro estão sujeitos a variabilidade (WENZEL et al., 1994; WENZEL; VERDONSCHOT, 1994; ten CATE et al., 1999). Wenzel e Verdonschot (1994)

afirmaram que qualquer método de validação estaria sujeito a incertezas desde que dependesse da avaliação subjetiva de um examinador (validador). Este fato fez com que os autores recomendassem o abandono da expressão padrão-ouro por passar a idéia falsa de “um absoluto” e que a terminologia “método de validação ou critério de validação” fosse adotada. Os resultados do presente trabalho também constataram a variabilidade de diagnóstico dos validadores. Isto nos leva a corroborar a adoção das expressões métodos ou critérios de validação, utilizadas por Wenzel e Verdonschot (1994), ao invés da expressão “padrão-ouro”. O presente capítulo se propõe a discutir tanto os aspectos da metodologia utilizada para avaliação dos métodos de validação quanto os resultados encontrados na presente pesquisa.

Existe uma grande preocupação em relação à questão do critério de validação, tendo sido este um dos pontos abordados durante a primeira conferência direcionada à cárie dental do *National Institutes of Health* (NIH – Institutos Nacionais de Saúde) dos Estados Unidos (HOROWITZ, 2004). Um aspecto relatado dizia respeito à validação do próprio método de validação, se deveria ser realizada e como se deveria executá-la. Frequentemente a validação do próprio método de validação não é possível. Assim sendo, a comparação dos resultados obtidos pelo método de validação em diferentes ocasiões e por examinadores distintos faz-se essencial, já que uma das principais características que deve ser apresentada pelo método de validação seria reprodutibilidade (WULFF, 1981 *apud* WENZEL; HINTZE, 1999; HINTZE; WENZEL, 2003; HUYSMANS; LONGBOTTOM, 2004). Esta foi, então, a proposta do presente trabalho.

Vários métodos têm sido utilizados para validação de testes diagnósticos de cárie. Dentre eles são citados inspeção visual, radiografia convencional e digital, preparo cavitário, microscopia de luz polarizada, microrradiografia transversa e

estereomicroscopia (DOWNER, 1975; DOWNER, 1989; LUSSI, 1991; ten BOSCH,1991; WENZEL; LARSEN; FEJERSKOV, 1991; WENZEL; FEJERSKOV,1992; WENZEL; VERDONSCHOT, 1994; WENZEL et al., 1994; van RIJKOM; VERDONSCHOT, 1995; HINTZE; WENZEL; LARSEN, 1995; EKSTRAND et al., 1995; HINTZE; WENZEL, 1996; EKSTRAND; RICKETTS; KIDD, 1997; ASHLEY; BLINKHORN; DAVIES, 1998; TOVO et al., 1998; ten CATE, 1999; WENZEL; HINTZE, 1999; DEERY et al., 2000; PITTS et al., 2001; HINTZE; WENZEL, 2002; HINTZE; WENZEL, 2003; CÔRTEZ; ELLWOOD; EKSTRAND,2003; CASTRO, 2004; ELLWOOD; CÔRTEZ, 2004; HUYSMANS; LONGBOTTOM, 2004; ROCHA et al., 2005; MESTRINER; VINHA; MESTRINER JÚNIOR, 2005; BENGTON et al., 2005).

A inspeção visual das superfícies dentárias falha no preenchimento da primeira característica proposta por Wulff (1981 *apud* WENZEL; HINTZE, 1999) para validação, desde que não se mostra eficientemente reproduzível (EKSTRAND; RICKETTS; KIDD, 1997). Além disso, a inspeção visual não avalia a profundidade da lesão cáriosa com a precisão necessária a um método de validação (HINTZE; WENZEL, 2003; HUYSMANS; LONGBOTTOM, 2004). Já as radiografias são falhas enquanto critério de validação por serem métodos pobremente reproduzíveis e pela possibilidade de diagnósticos falso-positivos nas áreas de desmineralização, desde que restaurações, hipomineralização e alguns artefatos podem ter mesma imagem radiográfica radiolúcida de uma lesão cáriosa (HINTZE; WENZEL, 2002). Wenzel e Verdonschot (1994) mencionaram que a verdadeira validação da cárie somente poderia ser obtida com o preparo cavitário, o que constitui uma afirmativa bastante discutível. Para validação de métodos de diagnóstico de cárie oclusal *in vivo*, o preparo cavitário apresenta-se como uma das poucas validações disponíveis

(CÔRTEZ, 1998). Ela só pode ser realizada, no entanto, em presença de resultado positivo do método de diagnóstico. Este fato impossibilitaria a validação de toda a amostra e a confirmação da ausência de lesão em casos de teste negativo. Soma-se a isto a polêmica da subjetividade da remoção de cárie durante o preparo.

A microscopia de luz polarizada e a microrradiografia transversa têm sido utilizadas principalmente para detecção de zonas de diferentes conteúdos minerais dentro da lesão (HUYSMANS; LONGBOTTOM, 2004). A microscopia de luz polarizada requer embebição em diferentes meios, não melhorando a detecção da profundidade da lesão. Já a microrradiografia transversa utiliza diferentes espessuras de seções dentárias e diferentes exposições à radiação ionizante para esmalte e dentina, tornando inviável o processo de validação conjunta de esmalte e dentina da mesma coroa dentária. A utilização de equipamentos e metodologias mais complexas só parece justificável na medida em que acarrete em uma significativa melhoria dos resultados. Já a inspeção ao estereomicroscópio de cortes dentais tem se apresentado como o mais popular dos métodos de validação *in vitro*, tendo sido relatada desde os primórdios das pesquisas sobre métodos de diagnóstico de cárie (DOWNER, 1975).

A opção no presente estudo pela estereomicroscopia como método a ser avaliado baseou-se na opinião da maioria dos autores que consideraram a estereomicroscopia como o mais confiável dos métodos de validação para testes de diagnóstico de cárie dental que avaliem a profundidade das lesões (DOWNER, 1975; LUSSI, 1991; WENZEL et al., 1994; HINTZE; WENZEL; LARSEN, 1995; HINTZE; WENZEL, 1996; EKSTRAND; RICKETTS; KIDD, 1997; WENZEL; HINTZE, 1999; ten CATE et al., 1999; WENZEL; HINTZE, 1999b; HINTZE; WENZEL, 2002; HINTZE; WENZEL, 2003; HUYSMANS; LONGBOTTOM, 2004). Ela consiste na metodologia

mais simples dentre as que utilizam microscopia e tem sido utilizada na maioria das pesquisas *in vitro* que avaliam métodos de diagnóstico de cárie (HUYSMANS; LONGBOTTOM, 2004). Com frequência a estereomicroscopia é denominada na literatura simplesmente como histologia ou validação histológica.

A histologia de cortes dentais seriados preencheria os critérios considerados fundamentais para um padrão-ouro confiável, segundo Hintze e Wenzel (2002). No entanto, a avaliação microscópica de cortes dentais parece sujeita a intensa variação quanto à sua reprodutibilidade na medida em que depende de uma avaliação subjetiva dos examinadores (WENZEL; VERDONSCHOT, 1994). Huysmans e Longbottom (2004) comentaram que a validação microscópica é baseada em mudanças óticas dos tecidos que poderiam ser facilmente observadas para o esmalte dental, pois uma redução do conteúdo mineral seria caracterizada como uma opacidade. Já na dentina, a relação entre desmineralização e mudanças óticas seria menos direta. Embora muitos estudos tenham relacionado dentina cariada à “descoloração” (FUSAYAMA; OKUSE; HOSODA, 1966; DOWNER, 1975; EKSTRAND; RICKETTS; KIDD, 1997; ASHLEY; BLINKHORN; DAVIES, 1998), os autores ressaltaram que a descoloração da dentina é tempo-dependente e sua relação com desmineralização é variável. Fusayama, Okuse e Hosoda (1966) relataram a possibilidade de cinco colorações diferentes para dentina cariada (preto-amarronzada, marrom, marrom-amarelada, amarelo e amarelo claro), ressaltando que nas cáries crônicas há uma forte descoloração da dentina e em cáries agudas haveria somente uma leve descoloração. Embora este aspecto não tenha sido objetivamente avaliado na presente pesquisa, pôde-se perceber pelo relato dos examinadores durante os exames e durante a calibração que a presença de

descoloração em dentina tendeu a ser acompanhada de maior dúvida no momento de se estabelecer o diagnóstico da profundidade da lesão.

Em revisão de literatura publicada em 2004, Huysmans e Longbottom verificaram que a estereomicroscopia foi o método de validação de escolha em 14 dentre 18 trabalhos *in vitro* sobre diagnóstico de cárie realizados em quatro anos. Embora tenha sido demonstrada uma clara predileção dos pesquisadores por tal metodologia, os autores levantaram uma série de críticas ao método, já que cada pesquisa adota metodologia própria, dificultando a comparação entre estudos. Ponderaram que alguns parâmetros deveriam ser padronizados, como o tipo de microscópio adotado, a magnitude do aumento, o modo de inspeção (cortes secos ou úmidos), o uso ou não de corantes, o número de critérios de diagnóstico utilizados e a espessura das secções histológicas.

A magnitude do aumento apresenta-se bastante variável na literatura (2,5 a 100X) (HINTZE; WENZEL; LARSEN, 1995; EKSTRAND; RICKETTS; KIDD, 1997; HINTZE; WENZEL, 2002; CÔRTEZ; ELLWOOD; EKSTRAND, 2003; CASTRO, 2004). O aumento de 4X foi selecionado na presente pesquisa por permitir a visualização ao microscópio de toda a extensão de esmalte e dentina até o teto da câmara pulpar nas secções histológicas avaliadas sem inclusão da raiz. Também foi realizada a opção pela avaliação da amostra úmida e sem uso de corantes, já que tanto a desidratação dos cortes como a utilização de corantes poderia confundir os examinadores no momento do diagnóstico das lesões.

Outro aspecto discutível no exame de cortes histológicos ao estereomicroscópio diz respeito à composição da amostra. A maioria dos estudos utiliza dentes imaturos, terceiros molares e premolares extraídos, os quais não seriam os dentes mais suscetíveis à cárie (WENZEL et al., 1994; HINTZE; WENZEL;

LARSEN, 1995; EKSTRAND et al., 1995). Assim sendo, a maioria das amostras das pesquisas são compostas por uma maior proporção de dentes saudáveis em relação aos cariados, o que leva a uma amostra inaceitavelmente desbalanceada. Na presente pesquisa foi seguida a orientação de Downer (1975), que considera importante que a amostra abranja dentes com todos os tipos de lesão de cárie e com número balanceado de sítios doentes e saudáveis, refletindo a verdadeira distribuição da doença *in vivo*. Isto, segundo o autor, garantiria uma avaliação segura de sensibilidade e especificidade dos métodos de diagnóstico, além de permitir comparações padronizadas. A amostra da presente pesquisa apresenta uma ampla gama de lesões com uma proporção aproximada entre elas, cuja extensão variava desde a ausência de lesão até cáries profundas em dentina (CÔRTEZ; ELLWOOD; EKSTRAND, 2003; ELLWOOD; CÔRTEZ, 2004). Os sete escores de diagnóstico escolhidos também refletem as profundidades de lesões de cárie relatadas para validação na literatura.

Alguns estudos laboratoriais (WENZEL et al., 1990; EKSTRAND; RICKETTS; KIDD, 1997; TOVO et al., 1998) optaram pelo hemisseccionamento dos dentes para inspeção ao estereomicroscópio. Autores, entretanto, (WENZEL et al. 1994; HINTZE; WENZEL 2003; CASTRO, 2004; HUYSMANS; LONGBOTTOM, 2004), verificaram que lesões cariosas na periferia das fissuras poderiam não ser diagnosticadas durante a validação ao se seccionar os dentes em apenas duas metades, levando a uma menor prevalência de cárie na amostra em estudo. Se uma lesão periférica puder ser detectada pelo teste de diagnóstico em avaliação mas não for detectada pelo critério de validação, o resultado será um registro falso-positivo ao invés de um verdadeiro-positivo, o que levaria a uma avaliação incorreta da acurácia do método testado. O seccionamento seriado dos dentes também está sujeito a

críticas, tendo sido relatado que cerca de 300 μm de material poderia ser perdido durante o seccionamento em série dos dentes com discos de diamante (DOWNER, 1975; HUYSMANS; LONGBOTTOM, 2004). Além disto, a secção analisada ao estereomicroscópio poderia não corresponder exatamente ao sítio examinado pelo método de diagnóstico, ou expressar somente parte dele (HUYSMANS; LONGBOTTOM, 2004). Entretanto, Hintze e Wenzel (2003) demonstraram que pôde ser detectado um maior número de lesões a partir da análise de cortes seriados do que através de hemissecções dentárias. A presente pesquisa optou, então, pela análise de cortes dentais seriados.

Embora a reprodutibilidade do método de validação seja essencial para que este seja considerado confiável, a literatura tem apresentado poucos relatos sobre tal reprodutibilidade (ten BOSCH; ANGMAR-MANSSON, 2000; PITTS et al., 2001; HINTZE; WENZEL, 2003; HUYSMANS; LONGBOTTOM, 2004). Hintze e Wenzel (2003) criticaram a própria metodologia adotada devido ao fato de terem utilizado apenas um examinador ao avaliarem métodos freqüentemente usados como validação dos testes diagnósticos de cárie. Outros trabalhos têm avaliado somente a reprodutibilidade inter-examinador, deixando de lado a variabilidade que poderia existir relacionada ao mesmo validador (WENZEL et al., 1994; HINTZE; WENZEL; LARSEN, 1995; ten CATE et al., 1999). A presente pesquisa optou pela avaliação, então, da reprodutibilidade intra e inter-examinador de vários validadores.

A literatura também nos tem mostrado situações onde o consenso entre examinadores foi utilizado como método de validação (JANHOM et al., 1999; HINTZE; WENZEL, 2002; HINTZE; WENZEL, 2003), embora esta forma de validação, devido às suas limitações, tenha sido considerada como “padrão-prata” (JONES; HUNTER, 1995; JANHOM et al., 1999). Wenzel et al. (1994) lançaram mão

de quatro examinadores para avaliar dois sistemas de validação e concluíram que os validadores, assim como os métodos de validação, poderiam influenciar nos resultados, pela subjetividade das avaliações. Este fato pôde ser observado no presente trabalho, onde houve grande variabilidade na reprodutibilidade tanto intra quanto inter-examinador, mesmo para o examinador mais experiente. Hintze, Wenzel e Larsen (1995) consideraram que a adoção de poucos e bem definidos critérios de diagnóstico seria um fator importante para a reprodutibilidade de um método de validação. No entanto, no presente estudo, mesmo utilizando critérios bem definidos, que abrangeram desde a ausência total de doença até casos onde a lesão se estendia até além do terço interno da dentina, e ainda após a calibração dos examinadores, verificou-se que houve muitas contradições no momento de se estabelecer o diagnóstico.

Além dos reduzidos estudos disponíveis na literatura que se propuseram a analisar a reprodutibilidade da estereomicroscopia enquanto critério de validação para os métodos de diagnóstico de cárie (WENZEL et al., 1994; HINTZE; WENZEL; LARSEN, 1995; EKSTRAND et al., 1995; PITTS et al., 2001; HINTZE; WENZEL,2003), existe o fato de cada um destes trabalhos ter adotado metodologia própria, dificultando a comparação entre estudos. O único trabalho que pudemos efetivamente comparar com os nossos resultados foi o de Pitts et al. (2001). Os trabalhos de Ekstrand et al. (1995) e Hintze e Wenzel (2003) chamam atenção pelos elevados valores de reprodutibilidade encontrados, mas os dados fornecidos geram dificuldade de comparação, já que não foram relatados os pontos de corte ou o tipo de Kappa adotados. Hintze e Wenzel (2003) utilizaram apenas um examinador em seu estudo, tendo obtido reprodutibilidade quase perfeita frente ao exame de cortes dentais seriados ao estereomicroscópio. Ekstrand et al. (1995) trabalharam com dois

examinadores, tendo ambos obtido quase perfeita reprodutibilidade intra e inter-examinadores. Wenzel et al. (1994) discutiram a influência que o validador e o método de validação poderia exercer sobre o resultado de testes de diagnóstico de cárie oclusal. Para tal utilizaram quatro validadores e em relação à estereomicroscopia obtiveram valores de reprodutibilidade inter-examinador moderada. No entanto não foi analisada a reprodutibilidade intra-examinador, não havendo também relato quanto aos pontos de corte ou tipo de Kappa de escolha. Hintze, Wenzel e Larsen (1995) tiveram como objetivo comparar a acurácia de quatro métodos de validação. Embora tenham adotado o Índice de Kappa de Cohen para o cálculo da reprodutibilidade dos três examinadores, não foram mencionados os pontos de corte utilizados e reportaram somente a média da reprodutibilidade entre os examinadores (substancial), sem haver relato quanto aos valores individuais.

A análise estatística da reprodutibilidade dos examinadores do presente trabalho foi feita utilizando-se os Índices de Kappa de Cohen, Kappa não Ponderado e Kappa Ponderado. A estatística de Kappa expressa a concordância percentual entre dois exames ou entre dois examinadores, excluindo-se o que acontece por mero acaso. A estatística Kappa é freqüentemente utilizada para avaliação da reprodutibilidade dos examinadores dos métodos de diagnóstico de cárie, por ser mais fidedigna do que a simples concordância percentual. O Índice de Kappa de Cohen (COHEN, 1960) é aplicado pela maioria dos autores para analisar o grau de concordância intra-examinador e entre examinadores de métodos de diagnóstico de cárie (LUSSI, 1991; HINTZE; WENZEL; LARSEN, 1995; EKSTRAND et al., 1995; ten CATE et al., 1999; DEERY et al., 2000; PITTS et al., 2001; HINTZE; WENZEL, 2003). Entretanto, pela sua natureza dicotômica, o Kappa de Cohen não

avalia adequadamente a reprodutibilidade em metodologias que envolvam múltiplos critérios de diagnóstico de cárie. Para utilizá-lo é necessário estipular pontos de corte nos critérios adotados e a partir destes definir o que seria “saúde” e “doença”. Tradicionalmente a literatura relata pontos de corte em esmalte e dentina. O ponto de corte esmalte considera como cárie (“doença”) todas as lesões que afetem esmalte ou esmalte e dentina. O ponto de corte em dentina classifica como cárie (“doença”) todas as lesões que atinjam somente dentina e como “saudável” tanto a ausência de lesões como todas as lesões em esmalte. Ao se analisar o ponto de corte esmalte, por exemplo, se um dos validadores classificasse um dos sítios como cárie atingindo metade externa do esmalte e o outro examinador como cárie atingindo terço interno da dentina, isto seria considerado um resultado concordante, ambos classificados como “doença”. Assim sendo, grandes discordâncias de escores intra e entre examinadores são consideradas concordâncias e o Kappa de Cohen sobrestima a reprodutibilidade em relação ao Kappa não Ponderado e Kappa Ponderado. De acordo com Bulman e Osborn (1989), pela característica dicotômica do Índice de Kappa de Cohen, pode-se encontrar dificuldades em se medir a severidade da doença ou classificar a profundidade de uma lesão de cárie. A utilização deste tipo de teste estatístico poderia levar a conclusões indevidas, na medida em que implica na sobrestimação dos valores de reprodutibilidade. Tal fato foi também verificado no presente trabalho, já que com o Índice de Kappa de Cohen foram obtidos sempre melhores resultados de reprodutibilidade do que com os Índices Kappa Ponderado e não Ponderado.

O Índice de Kappa não Ponderado (KINGMAN, 1986) considera somente o percentual de total concordância entre cada um dos critérios de diagnóstico e representa o resultado mais fiel da reprodutibilidade dos examinadores no caso de

múltiplos critérios. Já o Índice de Kappa Ponderado (COHEN, 1968; CICHETTI, 1976) também avalia o percentual total de concordância entre múltiplos critérios, porém atribui pesos diferentes às discordâncias de diagnóstico. Dessa forma passa a ser mais grave (com resultados piores em termos de reprodutibilidade), por exemplo, se um examinador classificar um sítio como cárie até metade externa do esmalte e o segundo examinador como cárie até terço externo da dentina do que se o segundo examinador atribuir escore cárie até metade interna do esmalte. Assim, em trabalhos envolvendo múltiplos escores, o uso do Kappa Ponderado e Kappa não Ponderado para avaliação da reprodutibilidade dos examinadores se mostra mais adequado. Onde seja importante ressaltar o peso dos erros, parece mais apropriado adotar o Índice de Kappa Ponderado para a análise da reprodutibilidade dos examinadores. Apesar deste fato, foram identificados poucos estudos sobre cárie que utilizam o Kappa não Ponderado (EKSTRAND; RICKETTS; KIDD, 1997; PERES; TRAEBERT; MARCENES, 2001) e nenhum que reporte o Kappa Ponderado. O Kappa não Ponderado, devido ao rigor em sua avaliação entre as concordâncias do examinador, deveria ser a estatística de escolha para a avaliação da reprodutibilidade dos métodos de validação de cárie.

Os examinadores que participaram do presente estudo apresentavam experiência prévia diferente em relação ao exame das secções dentárias ao estereomicroscópio. Havia a hipótese de que o validador experiente pudesse apresentar reprodutibilidade superior aos inexperientes. Analisando-se a reprodutibilidade intra-examinador através de quaisquer dos Índices de Kappa adotados (Cohen, não Ponderado e Ponderado) tanto nas etapas pré quanto pós-calibração (tabelas 9, 10 e 11), verificamos que o examinador experiente (Ex. A) apresentou resultados semelhantes a um examinador inexperiente (Ex. B) e superior

ao outro examinador inexperiente (Ex. C). Na avaliação após calibração pelo Kappa Ponderado, a reprodutibilidade intra-examinador de um validador inexperiente (Ex B) foi superior (quase perfeita) à do examinador experiente (Ex. A, substancial). A hipótese de que o validador experiente apresentaria como regra uma reprodutibilidade superior aos demais não se confirmou no presente trabalho.

Autores (HINTZE; WENZEL; LARSEN, 1995; PITTS et al., 2001; HUYSMANS; LONGBOTTOM, 2004) têm ressaltado a importância da calibração dos examinadores quando se avalia determinado método de validação, o que segundo Hintze, Wenzel e Larsen (1995) acarretaria em melhoria na reprodutibilidade. A Organização Mundial de Saúde (OMS), segundo Peres, Traebert e Marcenes (2001), definiu calibração como conjunto de procedimentos realizados para diminuir as discrepâncias de interpretação nos diagnósticos, o que pode ser conseguido pela repetição de exames na mesma amostra pelos mesmos examinadores ou pelo mesmo examinador em tempos diferentes. Ainda de acordo com a OMS, os objetivos da calibração seriam: (1) assegurar uniformidade de interpretação, entendimento e aplicação dos critérios das várias doenças e condições a serem observadas e registradas; (2) assegurar que cada um dos examinadores pudesse trabalhar consistentemente com o padrão adotado e (3) minimizar as variações entre diferentes examinadores. Baseado em tais afirmativas, o presente trabalho se propôs a verificar o impacto da calibração na reprodutibilidade dos examinadores do método de validação. A hipótese inicial era que a calibração pudesse melhorar principalmente os resultados da reprodutibilidade entre os diferentes examinador. Porém, de modo geral, isto não pôde ser comprovado, como explicitado a seguir.

O Índice de Kappa de Cohen mostrou uma manutenção dos resultados de reprodutibilidade intra-examinador nas etapas pré e pós-calibração, tanto para o

examinador experiente (Ex. A) quanto para os examinadores inexperientes (Ex. B e C), no ponto de corte esmalte (tabela 9). Houve uma maior diferença entre os resultados do Ex. C (inexperiente) quando comparado aos demais, tendo sido considerado moderada antes e após calibração para o ponto de corte esmalte, enquanto os examinadores A (experiente) e B (inexperiente) tiveram reprodutibilidade substancial. Entretanto, ao ser analisado o ponto de corte dentina houve melhora para todos os examinadores na etapa pós-calibração, independente da experiência prévia apresentada por eles. O examinador experiente (Ex. A), assim como um dos examinadores inexperientes (Ex. B) passaram de substancial para quase perfeita, tendo o outro examinador inexperiente (Ex. C) experimentado melhora mais acentuada, passando de moderada para quase perfeita. O Índice de Kappa de Cohen evidenciou que a calibração melhorou os resultados dos examinadores apenas no ponto de corte dentina, mantendo-se inalterada para o ponto de corte esmalte.

Pitts et al. (2001) avaliaram a validação histológica no diagnóstico da cárie oclusal e a reprodutibilidade dos examinadores da validação. Foram avaliados três examinadores e a análise estatística baseou-se no Índice de Kappa de Cohen, adotando os pontos de corte esmalte e dentina. No entanto, não houve relato referente a calibração e experiência prévia dos validadores. Os referidos autores reportaram uma reprodutibilidade intra-examinador variando de 0,58 a 0,82 (moderada a quase perfeita) no ponto de corte esmalte e de 0,37 a 0,88 (sofrível a quase perfeita) no ponto de corte em dentina, enquanto no presente estudo a variação foi de 0,58 a 0,73 (moderada a substancial) para o ponto de corte esmalte e de 0,53 a 0,79 (moderada a substancial) para o ponto de corte em dentina. Pode-se perceber que, assim como no presente trabalho, Pitts et al. (2001) também relataram

resultados insatisfatórios quanto à reprodutibilidade de seus examinadores da validação. É importante ressaltar que foi possível a comparação apenas dos resultados obtidos através do Índice de Kappa de Cohen, que implica em valores mais elevados de reprodutibilidade em relação aos Kappa Ponderado e não Ponderado.

Partindo-se do princípio que estes testes validam métodos de diagnóstico de cárie para serem usados *in vivo*, a confiança em seus resultados é fundamental. Para tal, torna-se inaceitável um método de validação cuja reprodutibilidade seja inferior à quase perfeita (Kappa 0,80 a 1,00). Confiar nos resultados das validações feitas com tais testes, tidos como “padrões-ouro”, aumenta a probabilidade de realização de preparos cavitários desnecessários ou até mesmo de não se restaurar elementos que efetivamente apresentem-se cariados.

No presente estudo, quando a reprodutibilidade intra-examinador foi avaliada pelo Índice de Kappa não Ponderado (tabela 10), houve uma maior igualdade dos resultados para os três examinadores, tanto na fase pré quanto na pós-calibração, tendo sido considerados como moderada. A experiência do examinador, assim como a calibração, não exerceu influência sobre os resultados. Com este índice foram consideradas apenas as áreas de total concordância entre todos os critérios de diagnóstico, tendo por isso havido uma grande redução dos resultados da reprodutibilidade quando comparado ao Kappa de Cohen.

Ao aplicarmos o Índice de Kappa Ponderado houve uma melhora dos resultados para a reprodutibilidade intra-examinador em relação ao Kappa não Ponderado (tabela 11), já que esta avaliação aplica pesos aos erros dos examinadores, o que demonstrou que os erros ocorreram primordialmente entre critérios próximos. A reprodutibilidade intra-examinador na etapa pré-calibração

variou de substancial a quase perfeita e após calibração de moderada a quase perfeita. Na etapa pré-calibração o examinador inexperiente (Ex. C) continuou apresentando resultados inferiores (substancial) aos demais, que tiveram reprodutibilidade idêntica (quase perfeita), independentemente de sua experiência. Na fase pós-calibração todos os examinadores tiveram resultados diferentes, tendo sido inesperado constatar piora da reprodutibilidade do examinador inexperiente (Ex. C), de substancial para moderada após calibração e principalmente do examinador experiente (Ex. A) passando de quase perfeita para substancial. A melhoria dos resultados após calibração seria mais esperado. Estes resultados aparentemente conflitantes reforçam a conclusão que, de uma maneira geral, nem experiência prévia nem calibração dos examinadores pode assegurar uma reprodutibilidade quase perfeita intra-examinador para o método de validação por estereomicroscopia.

Apesar da literatura ter considerado o exame de cortes dentais ao estereomicroscópio como o melhor critério de validação para testes de diagnóstico de cárie, pudemos comprovar pelos resultados da reprodutibilidade intra-examinador do presente trabalho que ele é sujeito a intensa variabilidade por depender de uma avaliação subjetiva dos examinadores.

Quando analisada pelo Índice de Kappa de Cohen, a reprodutibilidade entre os examinadores experiente e inexperientes A/B e A/C apresentou piores resultados na etapa pós-calibração, no ponto de corte esmalte, variando de substancial a moderada e de moderada a sofrível, respectivamente (tabela 12), fato este surpreendente. Apenas a comparação da reprodutibilidade entre os examinadores inexperientes B/C, para o ponto de corte esmalte, permaneceu como moderada. No entanto, para o ponto de corte dentina, a calibração melhorou os resultados entre experiente/inexperiente A/C e inexperiente/inexperiente B/C, passando ambos, de

sofrível a substancial. A reprodutibilidade entre A/B manteve-se como substancial nas duas etapas no ponto de corte dentina. Assim como para a reprodutibilidade intra-examinador, o Índice de Kappa de Cohen evidenciou que a calibração foi capaz de melhorar os resultados apenas entre alguns examinadores, além disto ter ocorrido somente no ponto de corte dentina, com resultados inalterados no ponto de corte esmalte.

Comparando-se os resultados referentes à reprodutibilidade entre examinadores do presente estudo com os relatados por Pitts et al. (2001), ambos apresentaram valores semelhantes pelo Kappa de Cohen. Pitts et al. (2001) obtiveram reprodutibilidade para os pontos de corte esmalte e dentina variando de 0,46 a 0,63 (moderada a substancial) e de 0,44 a 0,64 (moderada a substancial), respectivamente, enquanto no presente estudo a variação foi de 0,43 a 0,64 (moderada a substancial) e de 0,33 a 0,75 (sofrível a substancial).

Quando a reprodutibilidade entre examinadores foi avaliada pelo Índice de Kappa não Ponderado (concordância total) houve uma queda acentuada nos resultados para todos os examinadores em relação ao Kappa de Cohen, tanto na fase pré quanto na pós-calibração, tendo a reprodutibilidade variado de fraca a sofrível (tabela 13). Uma reprodutibilidade sofrível significa que houve uma concordância de no máximo 40% entre os examinadores, antes ou após a calibração. Este é um resultado inadmissível para um método de validação.

Ao aplicarmos o Índice de Kappa Ponderado houve uma melhora dos resultados para a reprodutibilidade inter-examinador (tabela 14) quando comparado ao Kappa não Ponderado, demonstrando novamente que os erros ocorreram em critérios próximos. A reprodutibilidade entre os examinadores antes da calibração variou de sofrível a substancial e de moderada a substancial após a calibração. A

calibração somente acarretou melhora nos resultados na comparação entre os examinadores experiente/inexperiente A/C, passando de sofrível para moderada. Na etapa pré-calibração houve uma variação maior na reprodutibilidade entre examinadores do que após a calibração. Embora tenha havido uma melhora dos resultados em relação àqueles apresentados pelo Kappa não Ponderado, esta reprodutibilidade permaneceu inaceitável para um método de validação.

Os resultados referentes à reprodutibilidade entre examinadores foram geralmente piores na comparação entre os examinadores A (experiente) e C (inexperiente) pelos três tipos de Kappa analisados.

A partir da reprodutibilidade obtida entre os examinadores, mais uma vez pôde-se verificar que, de modo geral, nem a calibração ou experiência prévia do examinador contribuiu para melhoria dos resultados.

Segundo Fletcher, R.H., Fletcher, S.W. e Wagner (1996) o “padrão-ouro” é freqüentemente difícil de ser encontrado, sendo muitas vezes necessário escolher como método de validação um teste que, mesmo não sendo perfeito, seja o melhor disponível. Nas conclusões do “Workshop de Colaboração Internacional de Estudos Clínicos sobre Cárie”, apresentadas por Pitts e Stamm (2002), foi abordada a deficiência dos métodos de validação disponíveis, fato que tem sido discutido por diversos autores (WENZEL; VERDONSCHOT, 1994; ten BOSCH; ANGMAR-MANSSON, 2000; PITTS, 2001; HUYSMANS; LONGBOTOM, 2004). Neste evento foi lançado o desafio de se desenvolver validações que possam acompanhar o avanço dos novos métodos de diagnóstico de cárie, incluindo avaliação de lesões não cavitadas e até mesmo de parâmetros da doença que não têm padrão-ouro histológico, como atividade e infecção. ten Bosch e Angmar-Mansson (2000) já haviam ressaltado que o progresso da ciência traria métodos diagnósticos melhores

do que os já estabelecidos padrões-ouro. Isto entra em conflito com a idéia de que qualquer novo método deve ser validado contra os atuais métodos de validação. Wenzel e Verdonshot (1994) afirmaram que confrontar os resultados de novos métodos de diagnóstico de cárie com critérios de validação que já se têm demonstrado apresentar grande variabilidade faz com que o novo método de diagnóstico sob investigação nunca seja capaz de demonstrar maior acurácia do que o critério de validação. Na verdade, a subestimação da validade do novo método poderia ser um erro da técnica de validação.

As características da cárie dental, graças à ação dos fluoretos, têm mudado bastante nas últimas décadas. Este fato tem obrigado os profissionais a lançarem mão de meios diagnósticos cada vez mais sensíveis para que o correto diagnóstico possa ser estabelecido e o tratamento mais adequado para o paciente seja implantado. Para que tal objetivo seja alcançado, novos métodos de diagnóstico têm surgido, necessitando de uma validação para que se possa utilizá-los, com segurança, na clínica diária. A afirmação tradicional que o exame histológico é o padrão-ouro para avaliação dos métodos de diagnóstico de cáries deve ser considerada com cautela. Os baixos valores de reprodutibilidade e a grande variabilidade dos resultados obtidos no presente estudo fazem pensar na necessidade de que se revejam os conceitos a respeito dos critérios de validação atualmente disponíveis. Seria no mínimo arriscado confiar em um método de diagnóstico que tenha sido validado por um outro método sujeito a limitações. Futuros estudos deverão se preocupar em desenvolver métodos de validação mais sensíveis e precisos, onde sua reprodutibilidade seja avaliada adequadamente.

8 CONCLUSÕES

A partir dos resultados do presente trabalho e de acordo com a metodologia aplicada pudemos concluir que:

1. O exame de cortes dentais seriados ao estereomicroscópio apresentou valores de reprodutibilidade inaceitáveis e com grande variabilidade dos resultados tanto intra quanto inter-examinadores;
2. O exercício de calibração não melhorou a reprodutibilidade intra e inter-examinadores;
3. A experiência prévia não exerceu influência sobre a reprodutibilidade intra e inter-examinadores;
4. O Índice de Kappa de Cohen não se mostrou adequado para a análise da reprodutibilidade em casos de múltiplos critérios diagnósticos. A análise estatística de estudos sobre critérios de validação deveria recair sobre o Índice de Kappa não Ponderado, por ser o mais rigoroso em sua avaliação;
5. É essencial que estudos sobre métodos de diagnóstico de cárie reportem dados sobre a reprodutibilidade dos examinadores dos métodos de validação.

9 REFERÊNCIAS BIBLIOGRÁFICAS

1. ASHLEY, P.F.; BLINKHORN, A.S.; DAVIES, R.M. Occlusal caries diagnosis: an in vitro histological validation of the Electronic Caries Monitor (ECM) and others methods. **J Dent**, v. 26, p. 83-88, 1998.
2. BENGTON, A.L. et al. Influence of examiner's clinical experience in detecting occlusal caries in primary teeth. **Pediatr Dent**, v. 27, p. 238-243, 2005.
3. BERKEY, C.S. et al. Statistical methods for comparing dental diagnostic procedures. **Community Dent Oral Epidemiol**, v. 18, p. 169-176, 1990.
4. BULMAN, J.S.; OSBORN, J.F. Measuring diagnostic consistency. **Br Dent J**, v. 166, p. 377-381, 1989.
5. CASTRO, Fernanda Carvalho de. **Avaliação dos métodos clínico, clínico-radiográficos e laser diodo na detecção de lesões de cárie adjacentes a restaurações de amálgama**: estudo *in vitro*. Belo Horizonte, MG, 2004: Dissertação de Mestrado. Faculdade de Odontologia da UFMG, 2004.
6. CICHETTI, D.V. Assessing inter-rater reliability for rating scales: resolving some basic issues. **Brit J Psychiat**, v. 129, p. 452-456, 1976.
7. COHEN, J. A coefficient of agreement for nominal scales. **Educ Psychol Measurement**, v.20, p. 37-46, 1960.
8. COHEN, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. **Psychological Bulletin**, v. 70, p. 213-220, 1968.
9. CONSOLARO, A. **Cárie Dentária – Histopatologia e Correlações Clínico-Radiográficas – Bauru: FOB – USP, 1996, 48p.**

10. CÔRTEZ, Denise Fonseca. **Validação *in vivo* do diagnóstico de cárie oclusal e proximal em dentes posteriores por meio de exame clínico visual, transiluminação por fibra ótica (FOTI) e radiografias interproximais.** Bauru, SP, 1998. Tese de Doutorado. Faculdade de Odontologia de Bauru, Universidade de São Paulo, 1998.
11. CÔRTEZ D.F.; ELLWOOD, R.P.; EKSTRAND, K.R. An in vitro comparison of a Combined FOTI/Visual examination of occlusal caries with other caries diagnostic methods and the effect of stain on their diagnostic performance. **Caries Res**, v. 37, p. 8-16, 2003.
12. DEERY, C. et al. Effect of diagnostic criteria and validation system on the interpretation of caries diagnostic results. **Caries Res**, v. 34, p. 322-323, 2000.
13. DOOLAND, M.; SMALES, R. The diagnosis of fissure caries in permanent molar teeth. **J Dent Child**, v. 49, p. 181-185, May-Jun. 1982.
14. DOUGLASS, C. W. Evaluating diagnostic testes. **Adv Dent Res**, v. 7, p. 66-69, 1993.
15. DOUGLASS, C.W.; McNEIL, B.J. Clinical decision analysis methods applied to diagnostic tests in dentistry. **J Dental Educ.**, v. 47, p. 708-712, 1983.
16. DOWNER, M.C. Concurrent validity of an epidemiological diagnostic system for caries with the histological appearance of extracted teeth as validating criterion. **Caries Res.**, v. 9, p. 231-246, 1975.
17. DOWNER, M.C. Validation of methods used in dental caries diagnosis. **Int Dent J**, v. 39, p. 241-246, 1989.
18. EKSTRAND, K.R. et al. Relationship between external and histologic features of progressive stage of caries in the occlusal fossa. **Caries Res**, v. 29, p. 243-250, 1995.
19. EKSTRAND, K.R.; RICKETTS, D.N.J; KIDD, E.A.M. Reproducibility and accuracy of three methods for assessment of demineralization depth on the occlusal surface: An in vitro examination. **Caries Res**, v. 31, p. 224-231, 1997.

20. ELDERTON, R.J. Assessment and clinical management of early caries in young adults: invasive versus non invasive methods. **Br Dent J**, v. 158, p. 440-444, Jun. 1985.
21. ELLWOOD, R.P. CÔRTEZ, D.F. In vitro assessment of methods of applying the Electrical Caries Monitor for the detection of occlusal caries. **Caries Res**, v. 38, p. 45-53, 2004.
22. ESPELID, I.; TVEIT, A.B.; TJELTVEIT, A. Variations among dentists in radiographic detection of occlusal caries. **Caries Res**, v.28, p. 169-175, May/Jun. 1994.
23. FERREIRA, E.F. et al. Testes de diagnóstico: Conhecendo suas propriedades e adequando sua utilização em lesões cariosas. **Revista do CROMG**, v. 5, jan./abr.,1999.
24. FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, E.H. Diagnóstico. In: _____. **Epidemiologia clínica**. 3^a ed. Porto Alegre: Artes Médicas, 1996. cap. 3, p. 52-83.
25. FUSAYAMA, T.; OKUSE, K.; HOSODA, H. Relationship between hardness, discoloration, and microbial invasion in carious dentin. **J Dent Res**, v. 45, p. 1033-1046, 1966.
26. HINTZE, H.; WENZEL, A. Clinical and laboratory radiographic caries diagnosis. A study of the same teeth. **Dentomaxillofac Radiol**, v. 25, p. 115-118, 1996.
27. HINTZE, H.; WENZEL, A. Diagnostic outcome of methods frequently used for caries validation: a comparison of clinical examination, radiography and histology following hemisectioning and serial tooth sectioning. **Caries Res**, v. 37, p. 115-124, 2003.
28. HINTZE, H.; WENZEL, A. Influence of the validation method on diagnostic accuracy for caries. A comparison of six digital and two conventional radiographic systems. **Dentomaxillofac Radiol**, v. 31, p. 44-49, 2002.
29. HINTZE, H.; WENZEL, A; LARSEN, M.J. Stereomicroscopy, film radiography, microradiography, and naked-eye inspection of tooth sections as validation for occlusal caries diagnosis. **Caries Res**, v. 29, p. 359-363, 1995.

30. HOFFMAN, S. Histopatologia das lesões de cárie. In: _____. **Cáries dentárias: bases biológicas**. Rio de Janeiro: Guanabara Koogan, 1984, cap 9, p. 200-218.
31. HOLMEN, L.; THYLSTRUP, A.; ARTUN, J. Surface changes during the arrest of active enamel caries lesions in vivo. A scanning electron microscope study. **Acta Odontol Scand**, v. 45, p. 383-390, 1987.
32. HOROWITZ, A.M. A report of the NIH consensus development conference on diagnosis and management of dental caries throughout life. **J Dent Res** (Spec Iss C), v. 83, p. 15-17, 2004.
33. HUYSMANS, M.C.; LONGBOTTOM, C. The challenges of validating diagnostic methods and selecting appropriate gold standards. **J Dent Res** (Spec Iss C), v. 83, p. 48-52, 2004.
34. JANHOM, A. et al. Effect of noise on the compressibility and diagnostic accuracy for caries detection of digital bitewing radiographs. **Dentomaxillofac Radiol**, v. 28, p. 6-12, 1999.
35. JONES, J.; HUNTER, D. Consensus methods for medical and health services research (review). **Br Med J**, v. 311, p. 376-380, 1995.
36. JUHL, M. Localization of carious lesions in occlusal pits and fissures of human premolars. **Scand J Dent Res**, v. 91, p. 251-255, 1983.
37. KIDD, E.A.M.; RICKETTS, D.N.J.; PITTS, N.B. Occlusal caries diagnosis: a changing challenge for clinicians and epidemiologists. **J Dent**, v. 21, p. 323-331, 1993.
38. KINGMAN, A. A procedure for evaluating the reability of a gingivitis index. **J Clin Periodontol**, v. 13, p. 385-391, 1986.
39. KÖNIG, K.G. Dental morphology in relation to caries resistance with special reference to fissures as susceptible areas. **J Dent Res**, v. 42, p. 461-476, 1963.
40. LANDIS, R.J.; KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159-174, 1977.

41. LUSI, A. Validity of diagnostic and treatment decisions of fissure caries. **Caries Res**, v. 25, p. 296-303, Jul./Aug. 1991.
42. MESTRINER, S.F.; VINHA, D.; MESTRINER JÚNIOR, W. Comparison of different methods for the occlusal dentine caries diagnosis. **J Appl Oral Sci**, v. 13, p. 28-34, 2005.
43. MITROPOULUS, C.M.; DOWNER, M.C. Inter-examiner agreement in the diagnosis of dental caries among officers of the Reference Service. **B Dent J**, v. 21, p. 227-228, 1987.
44. MOTA, L.Q. et al. Diagnóstico da cárie oclusal incipiente. **J Bras Odontopediatr Odontol Bebê**, v. 5, p. 3210327, 2002.
45. NAMEN, Fátima. **Elaboração de Teses e Dissertações**. Rio de Janeiro: Rubio, 2006, 96p.
46. NEWBRUN, E. Problems in caries diagnosis. **Int Dent J**, v. 43, p. 133-142, 1993.
47. NYVAD, B. Diagnosis versus detection of caries. **Caries Res** v. 38, p. 192-198, 2004.
48. NYVAD, B.; MACHIULSKIENE, V.; BAELUM, V. Construct and predictive validity of clinical caries diagnostic criteria assessing lesion activity. **J Dent Res**, v. 82, p. 117-122, 2003.
49. OPTOVAC. Disponível em:
<<http://www.funbec.com.br/bmSpecMicroEst1.html>> . acesso em 28 fevereiro 2006.
50. PEREIRA, M.G. Aferição dos Eventos. In: _____. **Epidemiologia: teoria e prática**. Rio de Janeiro: Guanabara-Koogan, 1995. cap.17:, p. 358-376.
51. PERES, M.A.; TRAEBERT, J.; MARCENES, W. Calibração de examinadores para estudos epidemiológicos de cárie dentária. **Cad. Saúde Pública**, Rio de Janeiro, v. 17, p. 153-159, jan-fev, 2001.

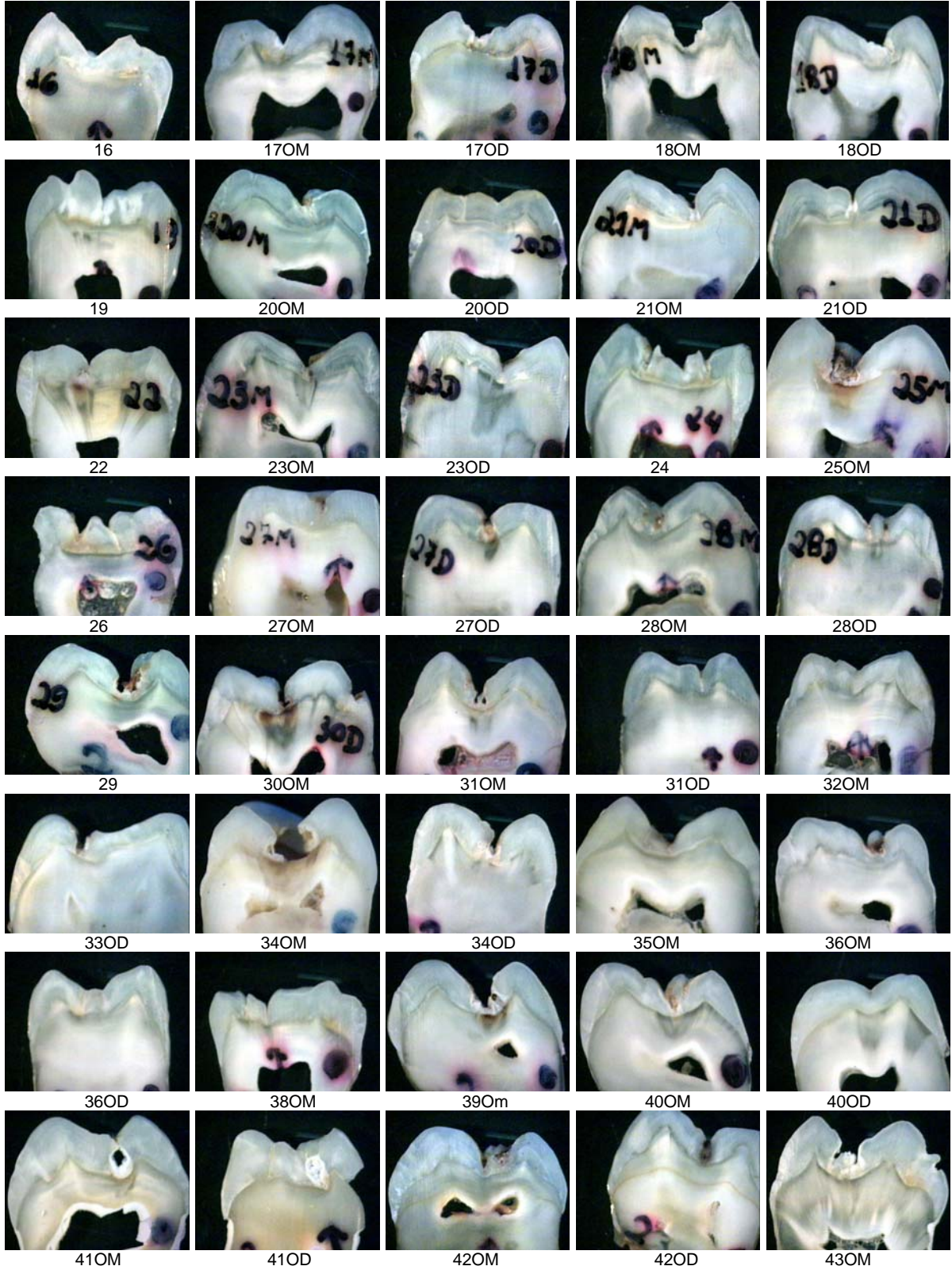
52. PITTS, N.B. et al. Histological validation of caries diagnosis – intra and inter assessor variation and its implications. **Caries Res**, v. 35, p. 278-279, 2001.
53. PITTS, N.B.; STAMM, J.W. Caries research implications on the International Collaborative Workshop on Caries Clinical Trials. **Caries Res**, v. 36, p. 208-209, 2002.
54. PITTS, N.B.; STAMM, J.W. International consensus workshop on caries clinical trials (ICW-CCT) – Final consensus statements: agreeing where the evidence leads. **J Dent Res**, v. 83 (Spec Iss C), p. 125-128, 2004.
55. ROCHA, A.S.P.S. et al. Interexaminer agreement in caries radiographic diagnosis by conventional and digital radiographs. **J Appl Oral Sci**, v. 13, p. 329-333, 2005.
56. SAWLE, R.F.; ANDLAW, R.J. Has occlusal caries become more difficult to diagnose? A study comparing clinically undetected lesions in molar teeth of 14-16 year old children in 1974 and 1992. **Br Dent J**, v. 164, p. 209-211, 1988.
57. ten BOSCH, J.J.; ANGMAR-MANSSON, B. A review of quantitative methods for studies of mineral content of intra-oral incipient caries lesions. **J Dent Res**, v. 70, p. 2-14, 1991.
58. ten BOSCH, J.J.; ANGMAR-MANSSON, B. Characterization and validation of diagnosis methods. **Monogr Oral Sci**, v. 17, p. 174-189, 2000.
59. ten CATE, B. et al. In vitro validation studies of quantitative light-induced fluorescence. In: _____. **Proceedings of the 4th Annual Indiana Conference** – Early detection of dental caries II. Indianapolis: George K. Stookey (ed.), Indiana University School of Dentistry, 1999.
60. THYLSTRUP, A.; FEJERSKOV, O. **Cariologia Clínica**. 2^a ed. São Paulo: Santos, 1988. 388 p.
61. TOVO, M.F. et al. Avaliação do exame visual, visual-tátil e radiográfico no diagnóstico de lesões de cáries dentinárias em superfície oclusal de molares decíduos. **Rev. FOB**, v. 6, p. 39-45, abr./jun. 1998.

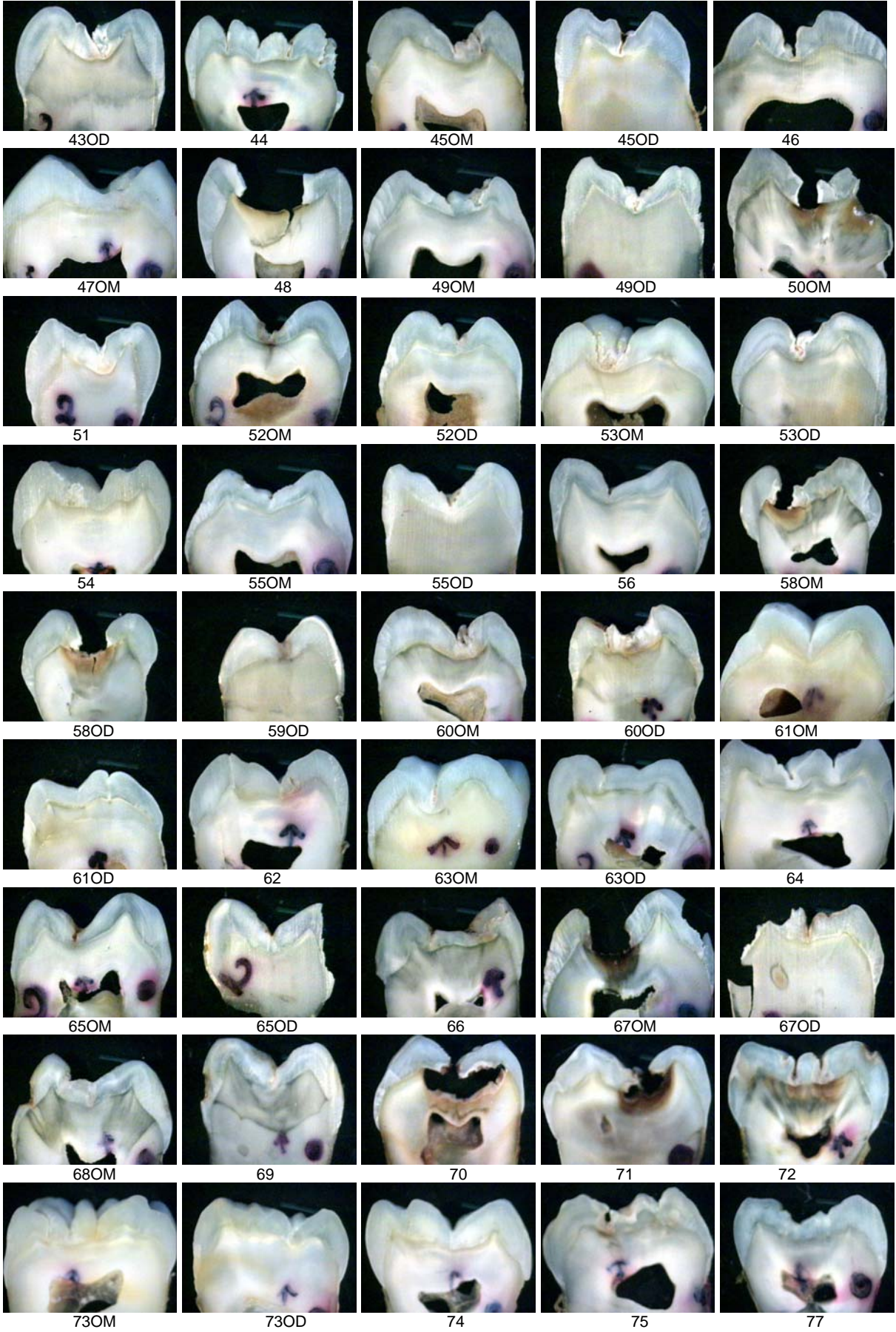
62. van RIJKOM, H.M.; VERDONSCHOT, E.H. Factors involved in validity measurements of diagnostic testes for approximal caries – a meta-analysis. **Caries Res**, v. 29, p. 364-370, 1995.
63. VERDONSCHOT, E.H. et al. Effects of digital grey-scale modification on the diagnosis of small approximal carious lesions. **J Dent**, v. 20, p. 44-49, 1992.
64. VERDONSCHOT, E.H.; WENZEL, A.; BRONKHORST, E.M. Applicability of receiver operating characteristic (ROC) analysis on discrete caries depth ratings. **Community Dent Oral Epidemiol**, v. 21, p. 203-208, 1993a.
65. VERDONSCHOT, E.H.; WENZEL, A.; BRONKHORST, E.M. Assessment of diagnostic accuracy in caries detection: an analysis of two methods. **Community Dent Oral Epidemiol**, v. 21, p. 269-272, 1993b.
66. WEERHEIJM, K.L.; GRUJTHUYSEN, R.J.M.; van AMERONGEN, W.E. Prevalence of hidden caries. **J Dent Child**, v. 59, p. 408-412, 1992.
67. WEERHEIJM, K.L.; van AMERONGEN, W.E.; EGGINK, C.O. The clinical diagnosis of occlusal caries: a problem. **J Dent Child**, v. 56, p. 196-200, 1989.
68. WEFEL, J.S.; CLARKSON, B.H.; HEILMAN, J.R. Natural root caries: a histologic and microradiographic evaluation. **J Oral Pathol**, v. 14, p. 615-623, 1985.
69. WENZEL, A. et al. Depth of occlusal caries assessed clinically, by conventional film radiographs, and by digitized, processed radiographs. **Caries Res**, v. 24, p. 327-333, 1990.
70. WENZEL, A. et al. Impact of the validator and the validation method on the outcome of occlusal caries diagnosis. **Caries Res**, v. 28, p. 373-377, 1994.
71. WENZEL, A.; FEJERSKOV, O. Validity of diagnosis of questionable caries lesions in occlusal surfaces of extracted third molars. **Caries Res**, v. 26, p. 188-194, 1992.
72. WENZEL, A.; HINTZE, H. Comparison of microscopy and radiography as gold standards in radiographic caries diagnosis. **Dentomaxillofac Radiol**, v. 28, p. 182-185, 1999a.

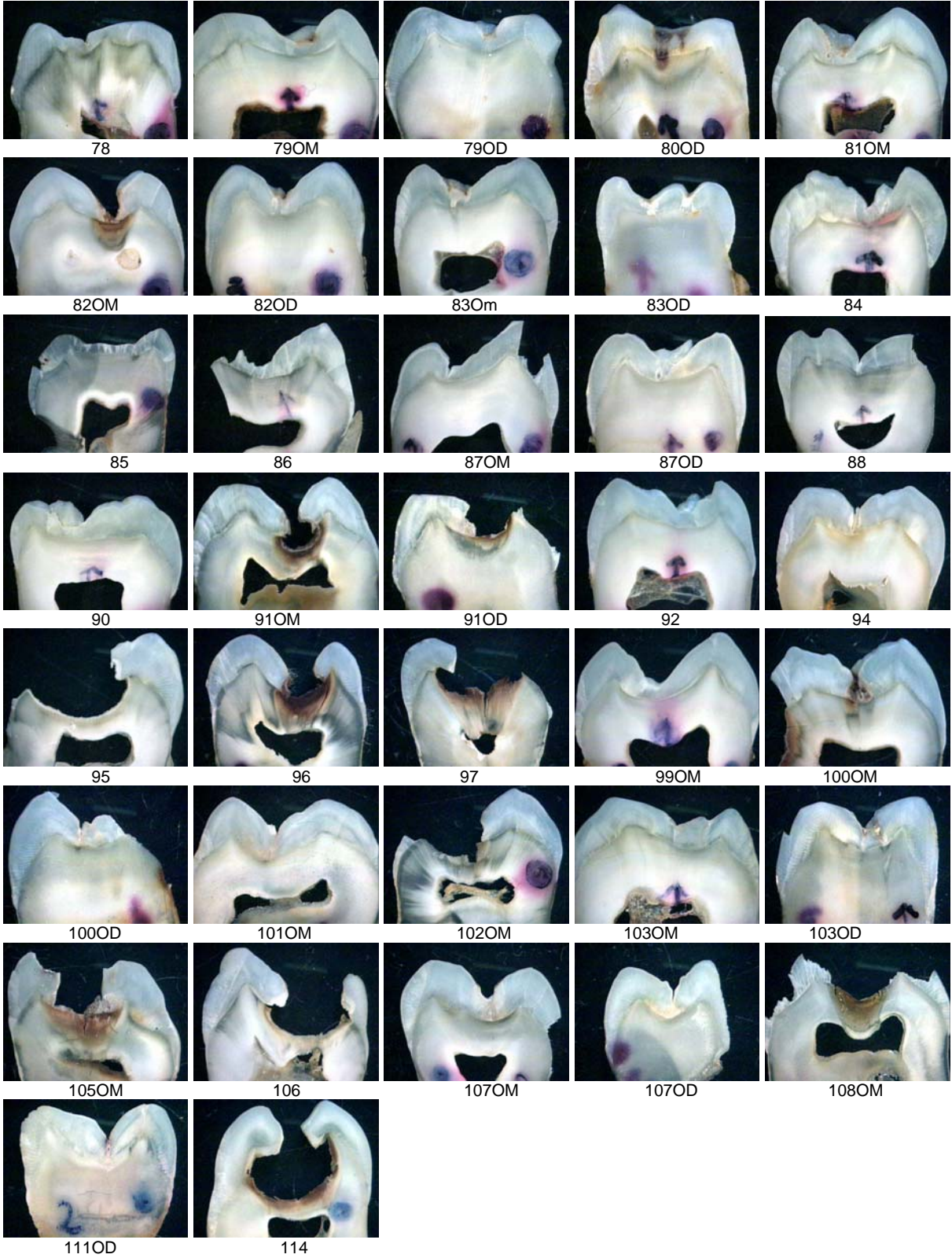
73. WENZEL, A.; HINTZE, H. The choice of gold standard for evaluating tests for caries diagnosis. **Dentomaxillofac Radiol**, v. 28, p. 132-136, 1999b.
74. WENZEL, A.; LARSEN, M.J.; FEJERSKOV, O. Detection occlusal caries without cavitation by visual inspection, film radiographs, xeroradiographs and digitized radiographs. **Caries Res**, v. 25, p. 365-371, 1991.
75. WENZEL, A.; VERDONSCHOT, E.H. Some considerations in the evaluation of diagnostic tests in dentistry. **Dentomaxillofac Radiol**, v. 23, p. 179-182, 1994.
76. WHO (World Health Organization). **Calibration of Examiners for Oral Health Epidemiology Surveys**. Technical Report. Geneva: WHO, 1993

ANEXOS

ANEXO I: Imagens dos cortes histológicos utilizados no estudo







ANEXO II: Resultados obtidos pelos examinadores sem calibração

A (sem calibração)		
secção	1o ex.	2o ex.
16	0	0
17 OM	1	1
17 OD	3	3
18 OM	3	3
18 OD	3	3
19	3	3
20 OM	2	2
20 OD	2	1
21 OM	3	3
21 OD	3	3
22	3	3
23 OM	4	3
23 OD	2	3
24	2	2
25 OM	4	4
26	2	1
27 OM	1	1
27 OD	4	3
28 OM	2	2
28 OD	2	3
29	4	3
30 OD	4	4
31 OM	3	3
31 OD	2	2
32 OM	0	0
33 OD	3	3
34 OM	4	4
34 OD	3	3
35 OM	1	2
36 OM	1	1
36 OD	1	3
38 OM	2	2
39 OM	4	4
40 OM	3	3
40 OD	0	0
41 OM	2	4
41 OD	4	4
42 OM	2	2
42 OD	3	3
43 OM	3	3
43 OD	3	3
44	2	1
45 OM	3	3
45 OD	3	3

B (sem calibração)		
secção	1o ex.	2o ex.
16	0	0
17 OM	1	1
17 OD	3	3
18 OM	4	3
18 OD	4	2
19	3	0
20 OM	3	2
20 OD	4	0
21 OM	2	3
21 OD	2	1
22	6	4
23 OM	4	4
23 OD	5	1
24	2	2
25 OM	5	4
26	2	2
27 OM	1	1
27 OD	4	4
28 OM	2	2
28 OD	4	1
29	6	2
30 OD	6	5
31 OM	5	5
31 OD	1	1
32 OM	0	0
33 OD	2	3
34 OM	4	4
34 OD	2	2
35 OM	1	2
36 OM	2	2
36 OD	0	1
38 OM	3	3
39 OM	5	5
40 OM	4	4
40 OD	0	0
41 OM	2	2
41 OD	2	3
42 OM	2	2
42 OD	3	3
43 OM	2	4
43 OD	1	2
44	1	1
45 OM	2	3
45 OD	2	2

C (sem calibração)		
secção	1o ex.	2o ex.
16	0	4
17 OM	2	4
17 OD	2	3
18 OM	3	4
18 OD	4	3
19	0	4
20 OM	1	1
20 OD	0	1
21 OM	3	3
21 OD	4	3
22	0	5
23 OM	6	5
23 OD	6	6
24	2	2
25 OM	5	5
26	1	1
27 OM	2	2
27 OD	5	5
28 OM	4	4
28 OD	2	4
29	2	3
30 OD	6	6
31 OM	2	3
31 OD	0	1
32 OM	0	0
33 OD	3	2
34 OM	6	6
34 OD	3	4
35 OM	1	1
36 OM	1	1
36 OD	0	0
38 OM	2	4
39 OM	5	5
40 OM	6	6
40 OD	1	1
41 OM	4	4
41 OD	0	3
42 OM	2	2
42 OD	3	2
43 OM	1	1
43 OD	2	2
44	4	3
45 OM	3	4
45 OD	3	2

46	3	3
47 OM	0	0
48	6	6
49 OM	2	1
49 OD	3	3
50 OM	4	5
51	1	1
52 OM	3	4
52 OD	2	2
53 OM	2	3
53 OD	3	3
54	0	0
55 OM	1	2
55 OD	1	2
56	2	3
58 OM	4	4
58 OD	5	6
59 OD	2	2
60 OM	2	2
60 OD	3	2
61 OM	0	2
61 OD	0	0
62	2	1
63 OM	1	2
63 OD	1	1
64	2	2
65 OM	2	3
65 OD	1	1
66	3	3
67 OD	3	3
68 OM	2	2
68 OD	2	1
69	5	5
70	5	4
71	3	4
72	6	6
73 OM	2	2
73 OD	3	3
74	0	0
75	3	4
77	1	3
78	1	2
79 OM	1	0
79 OD	1	1
80 OD	3	4
81 OM	3	1
82 OM	4	4
82 OD	2	3
83 OM	3	3
83 OD	0	3
84	0	1
85	3	0

46	2	3
47 OM	0	0
48	6	6
49 OM	1	1
49 OD	3	2
50 OM	5	5
51	0	0
52 OM	3	4
52 OD	1	1
53 OM	1	2
53 OD	2	3
54	0	0
55 OM	1	1
55 OD	1	1
56	1	1
58 OM	5	5
58 OD	4	5
59 OD	2	2
60 OM	3	2
60 OD	2	4
61 OM	0	0
61 OD	0	0
62	0	1
63 OM	1	1
63 OD	0	0
64	1	0
65 OM	3	2
65 OD	1	1
66	2	4
67 OD	3	2
68 OM	2	2
68 OD	1	1
69	6	6
70	5	5
71	6	6
72	6	6
73 OM	0	0
73 OD	0	0
74	0	0
75	3	4
77	0	1
78	3	1
79 OM	1	1
79 OD	1	1
80 OD	4	4
81 OM	2	1
82 OM	4	5
82 OD	1	1
83 OM	2	3
83 OD	0	0
84	0	2
85	2	3

46	2	3
47 OM	0	0
48	6	6
49 OM	4	2
49 OD	3	2
50 OM	6	6
51	1	1
52 OM	5	5
52 OD	1	1
53 OM	3	2
53 OD	2	2
54	0	0
55 OM	1	1
55 OD	2	2
56	4	1
58 OM	5	5
58 OD	6	6
59 OD	3	4
60 OM	1	2
60 OD	4	4
61 OM	0	0
61 OD	0	3
62	2	4
63 OM	4	3
63 OD	0	1
64	3	3
65 OM	4	3
65 OD	1	1
66	6	2
67 OD	2	3
68 OM	5	2
68 OD	5	6
69	6	6
70	5	6
71	6	6
72	6	6
73 OM	0	0
73 OD	0	3
74	0	1
75	6	6
77	5	6
78	6	6
79 OM	4	4
79 OD	1	1
80 OD	5	5
81 OM	3	2
82 OM	6	6
82 OD	3	2
83 OM	4	3
83 OD	0	0
84	2	3
85	5	4

86	0	0
87 OM	2	2
87 OD	0	3
88	3	0
90	0	0
91 OM	5	5
91 OD	4	4
92	0	0
94	3	3
95	5	6
96	6	6
97	5	6
99 OM	1	0
100 OM	3	4
100 OD	3	3
101 OM	3	1
102 OM	5	5
103 OM	1	1
103 OD	3	3
105 OM	5	6
106	6	6
107 OM	3	3
107 OD	3	3
108 OM	5	5
108 OD	3	2
111 OD	0	0
114	3	3

86	0	0
87 OM	2	2
87 OD	0	1
88	0	2
90	1	1
91 OM	6	6
91 OD	4	4
92	0	1
94	1	1
95	6	6
96	6	6
97	6	6
99 OM	0	0
100 OM	5	5
100 OD	1	2
101 OM	0	0
102 OM	6	6
103 OM	0	0
103 OD	1	2
105 OM	6	6
106	6	6
107 OM	1	2
107 OD	2	2
108 OM	6	6
108 OD	2	2
111 OD	0	1
114	2	2

86	6	3
87 OM	3	3
87 OD	5	4
88	5	3
90	5	5
91 OM	6	6
91 OD	4	4
92	5	2
94	4	3
95	6	6
96	6	6
97	6	6
99 OM	4	4
100 OM	5	5
100 OD	4	5
101 OM	0	1
102 OM	6	6
103 OM	4	3
103 OD	4	4
105 OM	6	6
106	6	6
107 OM	5	5
107 OD	1	4
108 OM	6	6
108 OD	4	2
111 OD	0	0
114	4	3

ANEXO III: Resultados obtidos pelos examinadores após calibração

A (com calibração)		
secção	1o ex.	2o ex.
16	0	0
17 OM	1	1
17 OD	3	3
18 OM	3	3
18 OD	1	2
19	3	3
20 OM	2	2
20 OD	1	1
21 OM	3	3
21 OD	3	2
22	3	3
23 OM	3	3
23 OD	3	2
24	0	1
25 OM	4	4
26	0	0
27 OM	2	1
27 OD	4	4
28 OM	2	2
28 OD	3	3
29	3	3
30 OD	4	4
31 OM	3	3
31 OD	2	2
32 OM	0	0
33 OD	3	3
34 OM	5	5
34 OD	3	3
35 OM	2	1
36 OM	1	1
36 OD	3	0
38 OM	0	2
39 OM	4	4
40 OM	3	3
40 OD	0	0
41 OM	4	4
41 OD	4	4
42 OM	1	1
42 OD	3	2
43 OM	1	2
43 OD	1	2
44	0	0
45 OM	2	2
45 OD	3	3
46	3	3
47 OM	0	0

B (com calibração)		
secção	1o ex.	2o ex.
16	0	0
17 OM	1	1
17 OD	3	2
18 OM	4	3
18 OD	2	1
19	4	2
20 OM	2	2
20 OD	0	1
21 OM	3	3
21 OD	2	2
22	4	4
23 OM	1	2
23 OD	1	1
24	2	2
25 OM	4	4
26	1	1
27 OM	1	2
27 OD	3	3
28 OM	2	2
28 OD	2	2
29	4	3
30 OD	4	4
31 OM	3	3
31 OD	0	1
32 OM	0	0
33 OD	2	3
34 OM	5	4
34 OD	3	2
35 OM	2	1
36 OM	1	2
36 OD	0	2
38 OM	3	3
39 OM	4	4
40 OM	2	3
40 OD	0	0
41 OM	2	3
41 OD	3	3
42 OM	2	2
42 OD	2	2
43 OM	2	2
43 OD	2	3
44	1	2
45 OM	2	3
45 OD	3	3
46	2	3
47 OM	0	0

C (com calibração)		
secção	1o ex.	2o ex.
16	0	0
17 OM	3	3
17 OD	1	2
18 OM	2	2
18 OD	1	1
19	0	2
20 OM	1	1
20 OD	0	1
21 OM	3	2
21 OD	2	1
22	0	0
23 OM	2	2
23 OD	1	1
24	2	2
25 OM	3	3
26	1	1
27 OM	1	1
27 OD	4	2
28 OM	2	2
28 OD	2	0
29	2	2
30 OD	4	2
31 OM	2	2
31 OD	1	1
32 OM	0	0
33 OD	3	2
34 OM	4	4
34 OD	2	2
35 OM	2	1
36 OM	1	1
36 OD	1	0
38 OM	2	2
39 OM	4	4
40 OM	2	2
40 OD	0	0
41 OM	1	2
41 OD	0	0
42 OM	2	2
42 OD	2	1
43 OM	1	1
43 OD	2	2
44	3	3
45 OM	2	1
45 OD	2	2
46	2	2
47 OM	0	0

48	4	5
49 OM	0	0
49 OD	3	3
50 OM	4	4
51	0	0
52 OM	3	3
52 OD	2	1
53 OM	0	1
53 OD	3	2
54	0	0
55 OM	0	0
55 OD	0	0
56	0	0
58 OM	4	4
58 OD	5	5
59 OD	2	2
60 OM	2	2
60 OD	3	3
61 OM	0	0
61 OD	0	0
62	1	0
63 OM	1	1
63 OD	1	0
64	2	2
65 OM	2	2
65 OD	1	1
66	1	1
67 OD	2	3
68 OM	2	0
68 OD	1	2
69	6	6
70	4	5
71	3	3
72	5	6
73 OM	0	0
73 OD	1	0
74	0	0
75	4	3
77	1	0
78	0	0
79 OM	0	0
79 OD	1	1
80 OD	1	1
81 OM	4	4
82 OM	4	5
82 OD	2	1
83 OM	3	0
83 OD	0	3
84	0	0
85	0	0
86	0	0
87 OM	0	1

48	4	4
49 OM	1	2
49 OD	2	3
50 OM	4	4
51	1	3
52 OM	4	4
52 OD	0	1
53 OM	2	2
53 OD	2	2
54	0	0
55 OM	1	1
55 OD	1	1
56	1	1
58 OM	4	4
58 OD	4	4
59 OD	2	1
60 OM	2	2
60 OD	3	2
61 OM	0	1
61 OD	0	0
62	0	1
63 OM	1	2
63 OD	0	0
64	2	2
65 OM	2	2
65 OD	1	1
66	3	2
67 OD	3	2
68 OM	2	2
68 OD	3	3
69	6	6
70	5	5
71	4	4
72	6	6
73 OM	0	0
73 OD	1	2
74	0	0
75	3	4
77	1	1
78	1	3
79 OM	1	1
79 OD	1	1
80 OD	4	1
81 OM	1	4
82 OM	5	5
82 OD	3	2
83 OM	3	3
83 OD	0	0
84	0	1
85	0	0
86	0	0
87 OM	0	1

48	4	4
49 OM	1	1
49 OD	3	2
50 OM	4	4
51	1	0
52 OM	3	2
52 OD	1	0
53 OM	2	2
53 OD	3	2
54	0	0
55 OM	1	1
55 OD	2	1
56	1	1
58 OM	4	4
58 OD	4	4
59 OD	2	2
60 OM	1	1
60 OD	3	2
61 OM	0	0
61 OD	3	0
62	0	0
63 OM	2	2
63 OD	0	0
64	0	1
65 OM	2	2
65 OD	1	0
66	2	1
67 OD	2	2
68 OM	0	1
68 OD	5	4
69	5	5
70	4	4
71	2	4
72	5	5
73 OM	2	0
73 OD	3	1
74	0	0
75	3	2
77	0	1
78	0	3
79 OM	1	1
79 OD	0	0
80 OD	4	1
81 OM	1	3
82 OM	4	4
82 OD	1	1
83 OM	3	1
83 OD	0	1
84	0	0
85	0	0
86	0	0
87 OM	2	2

87 OD	0	1
88	0	1
90	2	0
91 OM	5	6
91 OD	4	4
92	0	1
94	3	0
95	6	6
96	5	6
97	5	6
99 OM	3	3
100 OM	0	0
100 OD	3	3
101 OM	0	0
102 OM	5	5
103 OM	1	0
103 OD	3	3
105 OM	5	5
106	6	6
107 OM	2	2
107 OD	1	2
108 OM	5	6
108 OD	2	2
111 OD	0	0
114	2	2

87 OD	2	2
88	0	1
90	0	1
91 OM	5	5
91 OD	4	4
92	0	0
94	2	2
95	6	6
96	5	5
97	5	5
99 OM	0	2
100 OM	4	4
100 OD	3	3
101 OM	0	0
102 OM	6	5
103 OM	0	1
103 OD	0	3
105 OM	5	5
106	6	6
107 OM	0	2
107 OD	3	3
108 OM	6	5
108 OD	2	2
111 OD	0	0
114	1	2

87 OD	0	2
88	3	2
90	2	2
91 OM	5	4
91 OD	4	4
92	0	0
94	2	0
95	6	5
96	5	5
97	4	5
99 OM	0	3
100 OM	3	0
100 OD	2	2
101 OM	0	0
102 OM	5	4
103 OM	0	0
103 OD	2	0
105 OM	5	5
106	4	5
107 OM	4	1
107 OD	3	1
108 OM	5	5
108 OD	3	2
111 OD	0	0
114	1	1

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)