

**RODRIGO NORREMOSE COSTA**

**NECESSIDADE DE TRATAMENTO ORTODÔNTICO:  
VALIDAÇÃO DO DAI (ÍNDICE DE ESTÉTICA DENTAL) E DO ICON  
(ÍNDICE DE COMPLEXIDADE, RESULTADO E NECESSIDADE)  
PARA A REGIÃO DE BELO HORIZONTE**

**BELO HORIZONTE  
Faculdade de Odontologia  
UNIVERSIDADE FEDERAL DE MINAS GERAIS  
2007**

**RODRIGO NORREMOSE COSTA**

**NECESSIDADE DE TRATAMENTO ORTODÔNTICO:  
VALIDAÇÃO DO DAI (ÍNDICE DE ESTÉTICA DENTAL) E DO ICON  
(ÍNDICE DE COMPLEXIDADE, RESULTADO E NECESSIDADE)  
PARA A REGIÃO DE BELO HORIZONTE**

Dissertação apresentada ao Programa de Pós Graduação da Faculdade de Odontologia da UFMG como parte dos requisitos para obtenção do título de Mestre em Odontologia.

Área de concentração: Clínica Odontológica

**ORIENTADOR:**

Prof. Dr. Allyson Nogueira Moreira  
Faculdade de Odontologia da UFMG

**CO-ORIENTADORA:**

Profa. Dra. Cláudia Silami de Magalhães  
Faculdade de Odontologia da UFMG

Belo Horizonte

Faculdade de Odontologia - Universidade Federal de Minas Gerais

2007

C837n  
2007  
T

Costa, Rodrigo Norremose

Necessidade de tratamento ortodôntico: validação do DAI (índice de estética dental) e do ICON (índice de complexidade, resultado e necessidade) para a região de Belo Horizonte / Rodrigo Norremose Costa, 2007.

139 fls.:il.

Orientador: Allyson Nogueira Moreira

Co-orientadora: Cláudia Silami de Magalhães

Dissertação (Mestrado) - Universidade Federal de Minas Gerais, Faculdade de Odontologia.


1. Má oclusão – Diagnóstico – Teses. 2. Índice de gravidade de doença – Teses. 3. Validade dos testes – Teses. I. Moreira, Allyson Nogueira. II. Magalhães, Cláudia Silami de. III. Universidade Federal de Minas Gerais, Faculdade de Odontologia. III. Título.

BLACK D047




**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**FACULDADE DE DONTOLOGIA**  
**Programa de Pós-Graduação em Odontologia**

Dissertação intitulada "**Necessidade de tratamento ortodôntico: validação do ICON (Índice de complexidade, resultado e necessidade) e o DAI (Índice de estética dental) para a cidade de Belo Horizonte**", área de concentração em Clínica Odontológica, apresentada pelo candidato **Rodrigo Norremose Costa**, para obtenção do grau de Mestre em Odontologia, aprovada pela banca examinadora constituída pelos seguintes professores:

  
Prof. Dr. Allyson Nogueira Moreira  
Orientador – FO-UFMG

  
Profa. Dra. Cláudia Silami de Magalhães  
Co-Orientadora – FO-UFMG

  
Profa. Dra. Branca Heloisa de Oliveira Martins Vieira  
FO-UERJ

  
Prof. Dr. Saul Martins de Paiva  
FO-UFMG

  
Prof. Dr. Saul Martins de Paiva  
Coordenador *pro tempore* do Colegiado do  
Programa de Pós-Graduação em Odontologia

Belo Horizonte, 31 de janeiro de 2007.

## **Agradecimentos**

**À UFMG e à FO-UFMG,**

*pela infra-estrutura e possibilidade de acesso a disciplinas e professores que contribuíram para o meu crescimento profissional e acadêmico.*

**Ao CPGO,**

*pelo apoio e por zelar pela crescente qualidade da nossa pós-graduação.*

**À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES),**

*pelo apoio financeiro.*

**À PUCMG,**

*pelo apoio e disponibilidade em contribuir para a realização deste trabalho.*

**Aos Coordenadores do CPGO, Profa. Dra. Isabela Almeida Pordeus e Prof. Dr. Saul Martins de Paiva,**

*pelo apoio, competência, disponibilidade e amizade.*

**Aos meus orientadores, Prof. Dr. Allyson Nogueira Moreira e Profa. Dra. Cláudia Silami de Magalhães,**

*pelos ensinamentos, atenção, confiança, solidariedade, amizade e pela oportunidade de realização deste trabalho.*

**À Profa. Dra. Maria de Lourdes Massara,**

*pela atenção e disponibilidade.*

**Aos Professores da FO-UFMG,**

*pelo prazer de rever e conviver com velhos e novos amigos.*

**Aos Professores da Ortodontia – FO-UFMG,**

*pela receptividade, atenção, amizade e disponibilidade em contribuir ativamente para a realização deste trabalho.*

**Ao Prof. Eustáquio Afonso Araújo, Magnífico Reitor da PUC-MG,**

*pela disponibilidade, atenção, e apoio ao possibilitar a colaboração necessária para a realização deste trabalho.*

**Ao Prof. Tarcísio Junqueira Pereira, Diretor da FO-PUC,**

*pela receptividade, atenção, apoio e colaboração.*

**Ao Prof. Dr. Ênio Tonani Mazzeiro, Coordenador do Curso de Mestrado em Ortodontia da PUC,**

*pela receptividade, apoio e colaboração.*

**Aos Professores do Curso de Mestrado em Ortodontia da PUC,**

*pela receptividade, atenção, amizade e disponibilidade em colaborar na realização deste trabalho.*

**Aos Alunos do curso de Mestrado em Ortodontia da PUC,**

*pela convivência, alegria, amizade, receptividade e disponibilidade em dividir o espaço de trabalho.*

**À Profa. Dra. Ângela Maria Quintão Lana e ao Prof. Dr. João Paulo Amaral Haddad,**

*pela iniciação na Bioestatística, ponto fundamental na compreensão e execução deste trabalho e de meu crescimento científico.*

**Aos meus colegas, amigos e colaboradores: Antônio, Gisele, Maria Helena, Sérgio e Vera Christina,**

*pela disponibilidade, amizade, incentivo e participação voluntária, desde o início, como examinadores colaboradores, e a quem devo muito do sucesso no treinamento de calibração do examinador e da elaboração da metodologia deste trabalho.*

**Aos colegas do Mestrado,**

*pela convivência, alegria, solidariedade, cumplicidade, incentivo, companheirismo e amizade.*

**A todos os funcionários do CPGO, da Biblioteca, da FO-UFMG da FO-PUC e da Copiadora,**

*que contribuíram direta e indiretamente para a realização deste trabalho.*

**Aos amigos e colegas Juliana e Luis Otávio,**

*pela amizade, disponibilidade, compreensão, apoio e colaboração.*

**À minha família, em especial aos meus pais e irmãos,**

*pelo apoio, dedicação, compreensão e paciência na convivência “semi-ausente” diária nesses dois anos do mestrado, e por tudo o que fizeram e ainda fazem por mim.*

**Aos meus avós Paulina (*in memoriam*) e Hans,**

*pela convivência, interesse, carinho, exemplo, e incentivo em todos os momentos.*

**Aos amigos,**

*especialmente ao Pedro (*in memoriam*), pela amizade incondicional, aventuras, bons e maus momentos juntos, ouvidos e ombros sempre prontos, compreensão e por aguardar, sem reclamar demais, o retorno à convivência normal.*

**Ao Dalai Lama,**

*cujos textos iluminados me trouxeram compreensão, tranquilidade e paz espiritual, principalmente em momentos difíceis.*

## RESUMO

Índices oclusais como o índice de estética dental (DAI) e o índice de complexidade, resultado e necessidade (ICON), são utilizados para determinar a necessidade ou prioridade de tratamento ortodôntico em serviços públicos de atenção odontológica e, no Brasil, o DAI tem sido usado em levantamentos epidemiológicos. O objetivo deste estudo foi verificar a validade dos dois índices, para necessidade de tratamento ortodôntico para a região de Belo Horizonte, Minas Gerais. Os índices foram comparados à opinião agregada de um painel formado por vinte ortodontistas experientes sobre a necessidade de tratamento de um conjunto de 108 pares de modelos de estudo. Os modelos foram escolhidos de forma a representar uma grande variedade de tipos e gravidades da má oclusão. Um examinador, calibrado para os dois índices, mediu os modelos segundo ambos os critérios. A confiabilidade do examinador e do painel de especialistas foi verificada. Valores de ICC indicaram níveis excelentes de confiabilidade do examinador para o DAI ( $r_1 \geq 0,969$ ) e para o ICON ( $r_1 \geq 0,952$ ). Os níveis de concordância interexaminadores do painel de especialistas variaram de pobre a bom e os de concordância intra-examinadores variaram de pobre a excelente. A acurácia, medida pela área sob a curva ROC, foi igual a 81,83% para o DAI e 88,75% para o ICON. A diferença entre as áreas sob as curvas ROC foi significativa ( $p=0,0435$ ). Considerando as características desejáveis para um teste de rastreamento, relativas a condições de baixa prevalência (26,4%), os pontos de corte originais para o DAI (31) e para o ICON (43) foram considerados adequados. Os valores de especificidade para o DAI (86,67%) e para o ICON (93,33%) resultaram em valores preditivos positivos iguais a 62,38% e 79,89%, respectivamente. De acordo com a opinião do painel, 86% dos casos apresentaram necessidade de tratamento. Na amostra estudada, com alta prevalência de necessidade de tratamento, o DAI classificou corretamente 70,37% dos casos. O ICON concordou em 76,85% com as decisões do painel. Os bons níveis de acurácia indicam que os índices podem ser utilizados de forma a refletir a opinião de um painel de especialistas de Belo Horizonte.

**Palavras-chave:** má oclusão, índice de gravidade de doença, sensibilidade e especificidade, curva ROC, variações dependentes do observador, reprodutibilidade dos testes, validade dos testes.

## ABSTRACT

### **Orthodontic treatment need: validation of the DAI (dental aesthetic index) and the ICON (index of complexity, outcome and need) for the area of Belo Horizonte, Brazil**

Occlusal indices, such as the Dental Aesthetic Index (DAI) and the Index of Complexity, Outcome and Need (ICON) are used to determine eligibility for orthodontic treatment in public funded services. In Brazil, epidemiological surveys have assessed orthodontic treatment need prevalence by means of the DAI. The aim of this study was to investigate the validity of both indices for orthodontic treatment need in the area of Belo Horizonte, Minas Gerais. The DAI and ICON scores were compared to an aggregate opinion on treatment need formed by an expert panel of twenty specialists. A calibrated examiner in the DAI and ICON scored 108 study casts using the indices. The examiner's and the panel's reliability were assessed. Excellent ICC values were obtained for the DAI ( $r_1 \geq 0,969$ ) and ICON ( $r_1 \geq 0,952$ ). The the panel had interrater agreement levels that ranged from poor to good and intrarater levels ranging from poor to excellent. The accuracy measured by the area under ROC curve figured 81,83% for the DAI and 88,75% for the ICON. The difference between the areas under the ROC curves was significant ( $p=0,0435$ ). The original cutoff points for the DAI (31) and ICON (43) were considered adequate when the desirable characteristics of a screening test for a low prevalence condition are taken into account. The specificity figures for DAI (86,67%) and ICON (93,33%) resulted in positive predictive values of 62,38% and 79,89%, respectively. The panel assigned 86% of the casts into treatment category. Using this high prevalence sample, the DAI classified 70,37% correctly. The ICON scores agreed with the panel 76,85% of the cases. The good accuracy levels achieved by both indices suggest DAI and ICON may be used so as to reflect the opinion of specialists from Belo Horizonte area.

**Keywords:** malocclusion, severity of illness index, sensitivity and specificity, ROC curve, observer variation, reproducibility of results, validity of tests.



## LISTA DE ABREVIATURAS E SIGLAS

|         |  |
|---------|--|
| CV      | - Coeficiente de variação  |
| DAI     | - Índice de estética dental ( <i>dental aesthetic index</i> )  |
| E       | - Especificidade   |
| EUA     | - Estados Unidos da América  |
| FO-UFMG | - Faculdade de Odontologia da Universidade Federal de Minas Gerais   |
| ICC     | - Coeficiente de correlação intraclasse ( <i>intra-class correlation coefficient</i> )                                       |
| ICON    | - Índice de complexidade, resultado e necessidade ( <i>index of complexity outcome and need</i> )                            |
| IOTN    | - Índice de Necessidade de Tratamento Ortodôntico ( <i>index of orthodontic treatment need</i> )                             |
| LILACS  | - Literatura Latino Americana e do Caribe em Ciências da Saúde   |
| MEDLINE | - Sistema virtual de busca e análise de literatura médica ( <i>Medical Literature Analysis and Retrieval System Online</i> ) |
| NCHS    | - Centro Nacional de Estatísticas de Saúde dos Estados Unidos da América ( <i>National Center for Health Statistics</i> )    |
| OMS     | - Organização Mundial de Saúde   |
| PAR     | - Pontuação da Avaliação de Pares ( <i>peer assessment rating</i> )  |
| ROC     | - Característica de operação do receptor ( <i>receiver operating characteristic</i> )  |
| RV      | - Razão de verossimilhança   |
| S       | - Sensibilidade  |
| SCAN    | - Escala contínua padronizada de necessidade estética ( <i>standardized continuum of aesthetic need</i> )                    |
| SciELO  | - Biblioteca eletrônica científica virtual ( <i>Scientific Electronic Library Online</i> )                                   |
| UFMG    | - Universidade Federal de Minas Gerais   |
| VPN     | - Valor preditivo negativo   |
| VPP     | - Valor preditivo positivo   |
| WHO     | - Organização Mundial de Saúde ( <i>World Health Organization</i> )  |

## LISTA DE FIGURAS

|           |   |     |
|-----------|---|-----|
| FIGURA 1  | - Fluxograma da metodologia da obtenção do padrão de referência ..... | 49  |
| FIGURA 2  | - Fluxograma da metodologia da calibração .....                       | 50  |
| FIGURA 3  | - Fluxograma da metodologia da validação.....                         | 58  |
| FIGURA 4  | - Sondas periodontais.....  | 119 |
| FIGURA 5  | - Região para avaliação de dentes ausentes .....                      | 120 |
| FIGURA 6  | - Diastema na linha mediana.....                                      | 121 |
| FIGURA 7  | - Maior irregularidade anterior na maxila .....                       | 122 |
| FIGURA 8  | - Trespasse horizontal positivo.....                                  | 122 |
| FIGURA 9  | - Trespasse horizontal negativo .....                                 | 123 |
| FIGURA 10 | - Medição da mordida aberta anterior .....                            | 123 |
| FIGURA 11 | - Intercuspidação nos segmentos posteriores.....                      | 124 |
| FIGURA 12 | - Escala SCAN (Componente Estético).....                              | 126 |

## LISTA DE GRÁFICOS

|            |   |    |
|------------|---|----|
| GRÁFICO 1  | - Gráficos de Bland-Altman das diferenças versus as médias entre os pares de examinadores para o índice DAI em vinte pares de modelos.....                                      | 68 |
| GRÁFICO 2  | - Histograma das diferenças das medidas entre o pesquisador e o padrão de referência para o índice DAI.....   | 71 |
| GRÁFICO 3  | - Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas pelo pesquisador e o padrão de referência para o índice DAI, em trinta pares de modelos.....   | 72 |
| GRÁFICO 4  | - Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice DAI, em trinta pares de modelos.....  | 73 |
| GRÁFICO 5  | - Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice DAI, em trinta pares de modelos.....   | 74 |
| GRÁFICO 6  | - Gráficos de Bland-Altman das diferenças versus as médias entre os pares de examinadores para o índice ICON em vinte pares de modelos.....                                     | 77 |
| GRÁFICO 7  | - Histograma das diferenças das medidas entre o pesquisador e o padrão de referência para o índice ICON, em trinta pares de modelos.....  | 79 |
| GRÁFICO 8  | - Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas pelo pesquisador e o padrão de referência para o índice ICON, em trinta pares de modelos ..... | 80 |
| GRÁFICO 9  | - Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice ICON, em trinta pares de modelos .....  | 81 |
| GRÁFICO 10 | - Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice ICON, em trinta pares de modelos ..... | 82 |
| GRÁFICO 11 | - Histograma da distribuição das médias das opiniões dos 20 especialistas para o conjunto de 108 modelos, numa escala de 1 a 7 pontos.....                                      | 85 |
| GRÁFICO 12 | - Distribuição dos valores do DAI medidos no conjunto de 108 pares de modelos.....  | 89 |

|            |  |     |
|------------|--|-----|
| GRÁFICO 13 | - Distribuição dos valores do ICON medidos no conjunto de 108 pares de modelos.....  | 91  |
| GRÁFICO 14 | - Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice DAI, em quarenta pares de modelos.....   | 93  |
| GRÁFICO 15 | - Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice DAI, em quarenta pares de modelos.....  | 94  |
| GRÁFICO 16 | - Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice ICON em quarenta pares de modelos. ....  | 95  |
| GRÁFICO 17 | - Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice ICON, em quarenta pares de modelos..... | 96  |
| GRÁFICO 18 | - Curva ROC para o DAI (n=108).....  | 98  |
| GRÁFICO 19 | - Curva ROC para o ICON (n=108) .....  | 102 |
| GRÁFICO 20 | - Comparação entre as áreas sob as curvas ROC para o ICON e para o DAI.....  | 106 |
| GRÁFICO 21 | - Histograma da distribuição das categorias da escala SCAN no conjunto de 108 modelos.....   | 136 |

## LISTA DE TABELAS

|           |   |    |
|-----------|---|----|
| TABELA 1  | - Resultados dos níveis de concordância dos pares de examinadores para os valores do DAI em vinte pares de modelos.....                           | 66 |
| TABELA 2  | - Resultados do erro sistemático e seus limites de confiança e dos limites de concordância para os valores do DAI em vinte pares de modelos.....  | 67 |
| TABELA 3  | - Concordância dos pares de examinadores para os componentes do DAI em vinte pares de modelos.....  | 69 |
| TABELA 4  | - Concordância entre o pesquisador e o padrão de referência do DAI em trinta pares de modelos de estudo.....                                      | 70 |
| TABELA 5  | - Concordância intra-examinador do pesquisador para o índice DAI em trinta pares de modelos.....  | 72 |
| TABELA 6  | - Resultados dos níveis de concordância dos pares de examinadores para os valores do ICON em vinte pares de modelos.....                          | 75 |
| TABELA 7  | - Resultados do erro sistemático e seus limites de confiança e dos limites de concordância para os valores do ICON em vinte pares de modelos..... | 75 |
| TABELA 8  | - Concordância entre os pares de examinadores para os componentes do ICON.....  | 76 |
| TABELA 9  | - Concordância entre o pesquisador e o padrão de referência do ICON.....  | 79 |
| TABELA 10 | - Concordância intra-examinador do pesquisador para o índice ICON, em trinta modelos.....   | 81 |
| TABELA 11 | - Médias dos valores atribuídos ao mesmo conjunto de 108 pares de modelos por cada um dos especialistas.....                                      | 84 |
| TABELA 12 | - Distribuição das médias de opinião para os 108 modelos.....   | 84 |
| TABELA 13 | - Ponto de tratamento indicado - PTI.....   | 86 |
| TABELA 14 | - Distribuição da decisão dicotômica do painel.....   | 86 |
| TABELA 15 | - Concordância interexaminadores por pares.....   | 87 |

|           |   |     |
|-----------|---|-----|
| TABELA 16 | - Concordância intra-examinadores para os valores da opinião .....  | 87  |
| TABELA 17 | - Valores obtidos pelos 108 modelos após remoção de sete especialistas discordantes e que apresentaram valores extremos .....   | 88  |
| TABELA 18 | - Valores do DAI em 108 modelos de estudo.....  | 90  |
| TABELA 19 | - Distribuição dos resultados do DAI, com ponto de corte igual a 31, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos.....            | 90  |
| TABELA 20 | - Valores do ICON em 108 modelos de estudo.....   | 91  |
| TABELA 21 | - Distribuição dos resultados do ICON, com ponto de corte igual a 43, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos.....           | 92  |
| TABELA 22 | - Concordância intra-examinador do pesquisador para o índice DAI, em quarenta pares de modelos.....   | 92  |
| TABELA 23 | - Concordância intra-examinador do pesquisador para o índice ICON, em quarenta modelos.....   | 94  |
| TABELA 24 | - Sensibilidade e especificidade de pontos de corte do DAI em relação ao padrão-ouro.....   | 99  |
| TABELA 25 | - Distribuição dos resultados do DAI, com ponto de corte igual a 30,4, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos.....          | 99  |
| TABELA 26 | - Comparação das características de desempenho diagnóstico do DAI utilizando os pontos de corte original e ajustado na sua aplicação em 108 pares de modelos de estudo .....  | 100 |
| TABELA 27 | - Sensibilidade e especificidade de pontos de corte do ICON em relação ao padrão-ouro.....  | 103 |
| TABELA 28 | - Distribuição dos resultados do ICON, com ponto de corte igual a 40, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos.....           | 103 |
| TABELA 29 | - Comparação das características de desempenho diagnóstico do ICON utilizando os pontos de corte original e ajustado na sua aplicação em 108 pares de modelos de estudo ..... | 104 |

|           |  |     |
|-----------|--|-----|
| TABELA 30 | - Desempenho diagnóstico do DAI e do ICON utilizando os pontos de corte originais na sua aplicação em 108 pares de modelos de estudo ..... | 106 |
| TABELA 31 | - Comparação entre porcentagens das áreas sob as curvas ROC do DAI e do ICON .....   | 106 |
| TABELA 32 | - Frequência de classificações de má oclusão em relação à classificação pela escala SCAN no conjunto de 211 modelos da FO-UFMG .....       | 130 |
| TABELA 33 | - Conjunto de vinte pares de modelos utilizados na aplicação e avaliação dos critérios .....   | 131 |
| TABELA 34 | - Frequência de classificações de má oclusão em relação à escala SCAN no conjunto de 30 modelos .....                                      | 132 |
| TABELA 35 | - Frequência de classificações de má oclusão em relação à classificação pela escala SCAN no conjunto de 445 modelos da FO-UFMG .....       | 135 |
| TABELA 36 | - Frequência de classificações de má oclusão em relação à classificação pela escala SCAN no conjunto de 108 modelos .....                  | 136 |

## LISTA DE EQUAÇÕES

|            |  |    |
|------------|--|----|
| EQUAÇÃO 1  | - Modelo de mensuração.....                                    | 34 |
| EQUAÇÃO 2  | - Variância do valor observado.....                            | 34 |
| EQUAÇÃO 3  | - Coeficiente de confiabilidade de correlação intraclassa..... | 34 |
| EQUAÇÃO 4  | - Raiz da média de quadrados do erro.....                      | 37 |
| EQUAÇÃO 5  | - Coeficiente de correlação intraclassa – ICC.....             | 37 |
| EQUAÇÃO 6  | - Limite inferior do intervalo de confiança do ICC.....        | 37 |
| EQUAÇÃO 7  | - Estatística kappa.....                                       | 38 |
| EQUAÇÃO 8  | - Variância aproximada do kappa simples.....                   | 39 |
| EQUAÇÃO 9  | - Variância aproximada do kappa ponderado.....                 | 39 |
| EQUAÇÃO 10 | - Valor preditivo positivo.....                                | 43 |
| EQUAÇÃO 11 | - Valor preditivo negativo.....                                | 43 |
| EQUAÇÃO 12 | - Razão de verossimilhança.....                                | 46 |
| EQUAÇÃO 13 | - Cálculo do tamanho amostral.....                             | 59 |



## LISTA DE QUADROS

|          |   |   |     |
|----------|---|---|-----|
| QUADRO 1 | - | Significância clínica para valores de kappa, ICC e concordância observada .....           | 63  |
| QUADRO 2 | - | Componentes da equação de regressão do DAI com seus respectivos coeficientes (pesos)..... | 118 |
| QUADRO 3 | - | Categorização do DAI a partir dos pontos de corte .....                                   | 118 |
| QUADRO 4 | - | Componentes do ICON.....  | 125 |
| QUADRO 5 | - | Componente avaliação de espaços.....  | 127 |
| QUADRO 6 | - | Componente vertical anterior.....   | 128 |
| QUADRO 7 | - | Resumo do protocolo da avaliação dos componentes do ICON .....                            | 128 |

# SUMÁRIO

|          |  |            |
|----------|--|------------|
| <b>1</b> | <b>INTRODUÇÃO</b>  | <b>17</b>  |
| <b>2</b> | <b>REVISÃO DA LITERATURA</b>   | <b>19</b>  |
| 2.1      | Índices oclusais   | 19         |
| 2.2      | DAI  | 25         |
| 2.3      | ICON   | 30         |
| 2.4      | Confiabilidade   | 33         |
| 2.5      | Validade   | 40         |
| 2.5.1    | Sensibilidade e especificidade   | 42         |
| <b>3</b> | <b>OBJETIVOS</b>   | <b>47</b>  |
| 3.1      | Objetivo geral   | 47         |
| 3.2      | Objetivos específicos  | 47         |
| <b>4</b> | <b>METODOLOGIA</b>   | <b>48</b>  |
| 4.1      | <b>PARTE I: estudo de confiabilidade para calibração do examinador</b> | <b>48</b>  |
| 4.1.1    | Estudo dos critérios   | 50         |
| 4.1.2    | Aplicação e avaliação dos critérios                                    | 53         |
| 4.1.3    | Estabelecimento do consenso para os critérios de medida                | 55         |
| 4.1.4    | Calibração   | 56         |
| 4.2      | <b>PARTE II: Validação dos índices</b>                                 | <b>56</b>  |
| 4.2.1    | Materiais  | 57         |
| 4.2.2    | Seleção do conjunto de modelos de estudo                               | 59         |
| 4.2.3    | Obtenção do padrão-ouro  | 60         |
| 4.2.4    | Aplicação dos índices pelo pesquisador                                 | 62         |
| 4.3      | Análise de dados   | 62         |
| 4.4      | Aspectos éticos  | 65         |
| <b>5</b> | <b>RESULTADOS E DISCUSSÃO</b>  | <b>66</b>  |
| 5.1      | <b>Estudo de confiabilidade para calibração do examinador</b>          | <b>66</b>  |
| 5.1.1    | DAI  | 66         |
| 5.1.2    | ICON   | 75         |
| 5.2      | <b>Validação dos índices</b>   | <b>83</b>  |
| 5.2.1    | Padrão-ouro: a opinião dos especialistas                               | 83         |
| 5.2.2    | Valores dos índices  | 89         |
| 5.2.3    | Sensibilidade, especificidade e acurácia                               | 97         |
| <b>6</b> | <b>CONCLUSÃO</b>   | <b>111</b> |
|          | <b>REFERÊNCIAS</b>   | <b>112</b> |
|          | <b>APÊNCICES</b>   | <b>118</b> |
|          | <b>ANEXOS</b>  | <b>138</b> |

# 1 INTRODUÇÃO

A partir da perspectiva do diagnóstico, diferentemente dos outros problemas bucais, a má oclusão não é classificada como uma doença e é difícil de ser definida (NELSON, 1999; RICHMOND *et al.*, 1994; SHEIHAM, 2000). É um termo genérico usado para designar qualquer alteração ou desvio da normalidade nas relações entre as estruturas dento-faciais (STEDMAN, 1995), subdividido em classificações didáticas de acordo com o plano do espaço ou a parte anatômica envolvidos (NIELSEN, 1992).

Portanto, a má oclusão não é uma entidade única e as diversas classificações ou tipos de má oclusão podem estar presentes de forma isolada ou combinada. O grau de morbidade, e seu conseqüente impacto na qualidade de vida, dependem não só do tipo de desvio, mas também de sua gravidade e da interação com outros desvios. Essas interações podem ocorrer no sentido de um tipo de desvio agravar ou compensar outro desvio presente (NIELSEN, 1991).

Conseqüentemente, a ortodontia utiliza como meta de tratamento o conceito de “oclusão ideal”, em oposição ao conceito de oclusão “normal” ou “saudável”. Esse ideal arbitrário torna evidente a ausência de uma definição objetiva de má oclusão, resultando na crença que a necessidade de tratamento ortodôntico é uma das condições mais difíceis de se medir (COHEN & JAGO, 1976; McLAIN & PROFFITT, 1985).

Vários métodos de registro de más oclusões, inicialmente qualitativos e posteriormente quantitativos, foram propostos. Os métodos qualitativos apresentam a desvantagem de não serem capazes de definir objetivamente as situações que requerem ou não o tratamento. Posteriormente ao surgimento dos métodos qualitativos, métodos quantitativos passaram a ser desenvolvidos desde a década de 1960 (BRESOLIN, 2000; BROOK & SHAW, 1989; CONS *et al.*, 1989; DRAKER, 1960; JÄRVINEN, 2001; LINDER-ARONSON, 1974; SALZMANN, 1968; SHAW *et al.*, 1991b; SUMMERS, 1971).

Embora haja críticas à utilização de índices oclusais, o desenvolvimento de índices com características próximas do ideal seria de grande importância para a correta avaliação da situação da saúde bucal e implementação das ações adequadas (YOUNIS *et al.*, 1997). Alguns autores descreveram os critérios de um índice oclusal ideal (DRAKER, 1960; SHAW *et al.*, 1991b; SUMMERS, 1971), o qual deve ser: clinicamente válido para a região, válido no decorrer do tempo, confiável, objetivo e passível de análise quantitativa, aplicável

em paciente e em modelos de estudo, utilizável por clínicos e técnicos, de fácil aprendizado e rápida aplicação e aceitável pelos especialistas e pelo paciente.

Requisitos básicos de qualquer índice ou sistema de medida são a sua **validade** e **confiabilidade**. Enquanto a validade é a capacidade do índice medir o que realmente se propõe a medir, a confiabilidade é a capacidade da mesma pontuação ou resultado ser obtido num reexame feito pelo mesmo ou por outro examinador (CARLOS, 1970; GORDIS, 2004).

A má oclusão apresenta alta prevalência na população, com valores nacionais de 58,14% para a faixa etária de doze anos (BRASIL, 2004), e locais, para Belo Horizonte, de 61,9% para idades entre dez e quatorze anos (MARQUES *et al.*, 2005). Diante da sua crescente importância em saúde pública, da demanda por levantamentos epidemiológicos mais confiáveis, da limitação de recursos para tratamento e da dificuldade para o clínico em diagnosticá-la e avaliar corretamente a sua gravidade, torna necessária a utilização de um instrumento válido para medir objetivamente os critérios de indicação e priorização do tratamento ortodôntico.

## 2 REVISÃO DA LITERATURA

### 2.1 ÍNDICES OCLUSAIS

A necessidade de tratamento deve ser determinada por três tipos de informação que incluem sinais objetivos, sintomas subjetivos e adequação social. Os sinais objetivos são desvios de normas consideradas ideais. Tais referências somente devem ser utilizadas se estiver comprovado que o grupo de indivíduos no qual as normas foram embasadas apresenta melhor saúde, menos defeitos ou que se adapta melhor que outros grupos. Os sintomas subjetivos incluem o reconhecimento, por parte do paciente ou de seus pais, no caso de crianças, do desvio como sendo um problema que demanda tratamento. Já a adequação social é o reconhecimento, por parte da sociedade, de que a má oclusão cria um problema para o paciente (PRAHL-ANDERSEN, 1978).

Os principais fatores que justificariam o tratamento ortodôntico são a melhora na estética dental, no ajuste psicossocial e na função oclusal. Mas as evidências que suportam a intervenção baseada na melhora funcional, de uma forma geral, são fracas. Há pouca evidência para sugerir que a oclusão, de forma geral, está associada à auto-estima e à qualidade de vida, com vantagens para a saúde bucal. O entendimento sobre as conseqüências da má oclusão sobre a saúde bucal não é seguro nem consensual. A ciência que justifica algumas características oclusais específicas como indicação para o tratamento ortodôntico tem se desenvolvido nos últimos anos e, atualmente, não é mais claro se o maior risco de cárie, doença periodontal e problemas nas articulações temporo-mandibulares seriam justificativas válidas para a realização do tratamento ortodôntico. Apenas um número limitado de características da oclusão está associado a um aumento mensurável na auto-estima, realização plena na vida e melhora na saúde bucal. Essas características são: fissura labial e palatina, mordida cruzada posterior, trespasse horizontal excessivo, trespasse horizontal negativo, impactação dentária, mordida aberta anterior, hipodontia, sobremordida profunda, apinhamento dental e espaçamento dental (DANIELS & RICHMOND, 2000; LUTHER, 1998; McLAIN & PROFFITT, 1985; SHAW *et al.*, 1991a).

A importância da estética facial e das conseqüências psicossociais da má oclusão tem sido amplamente discutida recentemente (BRESOLIN, 2000; DE OLIVEIRA & SHEIHAM, 2004; SOLOW, 1995). Estudos sugerem que a melhora estética é o principal motivo para os pacientes que buscam tratamento ortodôntico (GOCHMAN, 1975;

MCKIERNAN *et al.*, 1992; SHAW, 1981; TULLOCH *et al.*, 1984). Em muitos casos, essa pode ser a principal indicação para o tratamento ortodôntico (EVANS & SHAW, 1987; GRABER & LUCKER, 1980; HOLMES, 1992a, b; JÄRVINEN, 2001; LEWITT & VIROLAINEN, 1968; LINDSEY & HODGKINS, 1983; SHAW *et al.*, 1975, 1991a).

A motivação para procurar tratamento ortodôntico parece não estar correlacionada com a saúde dental nem com as necessidades estéticas do paciente. Isso evidencia a dificuldade na percepção da má oclusão e enfatiza a necessidade de se estabelecer exatamente a característica que desagrade o paciente em relação aos seus dentes (HOLMES 1992b). A auto-percepção da aparência dental pelo paciente, ou por seus pais, no caso de crianças, é importante na determinação da demanda pelo tratamento. (SHAW *et al.*, 1975; SHAW, 1981). A percepção da própria condição oclusal e necessidade de tratamento é difícil de ser medida e as diferentes metodologias adotadas na sua avaliação dificultam a comparação entre os estudos (ESPELAND, 1992).

Tem sido sugerido que há uma variedade de más oclusões que são consideradas aceitáveis e que sua discriminação pode variar entre indivíduos e entre gêneros (ESPELAND, 1992). Lewitt e Virolainen (1968), Shaw (1981) e Holmes (1992b) observaram pacientes insatisfeitos com dentições esteticamente agradáveis e, por outro lado, pacientes satisfeitos com problemas estéticos aparentemente graves. A capacidade de reconhecer os sintomas subjetivos da própria má oclusão, em adultos, pode estar associada ao perfil psicológico. Enquanto pacientes com tendências neuróticas apresentam percepção próxima à do ortodontista, os perfis psicológicos normais tendem a subestimar a gravidade da má oclusão (MCKIERNAN *et al.*, 1992). Dois estudos com metodologia semelhante investigaram a necessidade de tratamento ortodôntico e sua demanda subjetiva em escolares britânicos com idades entre nove e doze anos (SHAW, 1981; HOLMES, 1992b). Seus resultados sugerem que a capacidade de autopercepção das características da oclusão também estão associados ao gênero, idade, auto-estima, padrão de atendimento odontológico e histórico de discriminação devido a características dentais (apelidos e gozações na escola). A insatisfação com a aparência dental em crianças com idades entre nove e doze anos foi maior entre as meninas (SHAW, 1981; HOLMES, 1992b) e pode estar associada ao aumento da idade (SHAW, 1981), apesar de não ter sido observada maior necessidade objetiva de tratamento nesse grupo (HOLMES, 1992b). A escala SCAN, em relação ao questionário estruturado, correlacionou melhor os níveis de percepção da atratividade e de satisfação com a aparência dental, quando comparados às observações feitas pelo ortodontista (HOLMES, 1992b).

O conceito do paciente sobre a própria má oclusão pode apresentar grande variação em relação à avaliação profissional (HOLMES 1992b). É provável que dentistas, e particularmente ortodontistas, tenham uma visão mais crítica da escala de estética dental do que a sociedade em geral (PRAHL-ANDERSEN, 1978). Observou-se que a percepção sobre a má oclusão e a aceitabilidade da aparência dental é diferente entre ortodontistas, crianças e seus pais (SHAW, 1975; 1981). Adultos leigos, crianças, estudantes de odontologia e dentistas identificaram menos más oclusões em filmes do que ortodontistas (SHAW 1975). A maioria das crianças e seus pais apresentaram dificuldade em descrever as características da oclusão na região anterior (SHAW, 1981). Lindsay e Hodgkins (1983) observaram que crianças e seus pais subestimaram igualmente a gravidade das más oclusões em relação às avaliações realizadas por ortodontistas. Num estudo realizado em 104 escolares com idades entre 9,8 a 11,6 anos, Espeland *et al.*(1992) observaram associação significativa ( $p < 0,01$ ) entre a percepção da má oclusão pela criança e por seus pais, enquanto a sua concordância com os profissionais variou para características oclusais distintas. Petersen e Dahlström (1998) avaliaram a percepção de dentistas, ortodontistas e pacientes sobre a necessidade de tratamento em fotografias intra-orais. Observaram que os pacientes avaliaram os casos com mais rigor e sua confiabilidade intra e interexaminadores foi menor do que a dos profissionais. Os dentistas e ortodontistas fizeram uma avaliação menos rigorosa quanto à necessidade estimada e a sua confiabilidade intra e interexaminadores foi maior, mesmo sem nenhuma calibração prévia. Não foi encontrada diferença entre os gêneros, sugerindo que essa diferença possa não existir quando a dentição analisada não é a própria.

Numa amostra de 315 escolares de Florianópolis, Brasil, com idades entre 14 e 18 anos, Peres *et al.* (2002) verificaram o impacto de características biométricas da oclusão na autopercepção da necessidade de tratamento ortodôntico e na satisfação com a aparência estética. Foi observado que a presença de apinhamento incisal e trespasse horizontal, medidos conforme os critérios do DAI, estavam associados à insatisfação com a aparência. A necessidade de tratamento ortodôntico foi melhor percebida pelos adolescentes que apresentaram irregularidade anterior na mandíbula, trespasse horizontal e diastema anterior. Determinadas características da oclusão e a gravidade do seu desvio podem ser aceitáveis pela população e podem assim interferir na autopercepção da necessidade de tratamento.

Nem toda criança portadora de má oclusão está consciente sobre a mesma, pois há uma variabilidade considerável de ajustamento à irregularidade, desde a total desatenção até a preocupação profunda. Portanto, tentativas de se medir objetivamente a má oclusão através de normas biométricas nem sempre avaliarão o problema de forma significativa para o paciente (SHAW *et al.*, 1975).

Num estudo realizado em 1.675 adolescentes brasileiros, De Oliveira e Sheiham (2004) observaram que o tratamento ortodôntico reduziu de forma significativa os impactos sobre a saúde, percebidos pelos adolescentes, relacionados à própria saúde bucal. A necessidade de tratamento foi medida pelo componente dental do índice de necessidade de tratamento ortodôntico (*index of orthodontic treatment need – IOTN*) e duas medidas de qualidade de vida relacionadas à saúde bucal foram utilizadas: Impacto Oral na Performance Diária (*oral impacts on daily performance – OIDIP*) e a versão resumida do Perfil de Impacto da Saúde Oral (*oral health impacts profile – OHIP-14*). Foi observado que a necessidade de tratamento, avaliada de forma normativa através do IOTN, não registrou dimensões psicossociais importantes da qualidade de vida relacionadas à saúde bucal. O tratamento da má oclusão, que possui um grande componente psicossocial, requer a utilização de medidas de qualidade de vida relacionadas à saúde bucal.

Para que um tratamento ortodôntico seja bem indicado, os benefícios reais à saúde do paciente trazidos pela correção da má oclusão devem ser confrontados com os fatores relacionados aos riscos inerentes ao próprio tratamento. Não devem ser subestimados os riscos de iatrogenia, insucesso parcial ou total e falta de estabilidade dos resultados a longo prazo (SHAW *et al.*, 1991a, 1995). São poucos os casos em que os objetivos do tratamento ortodôntico são atingidos (KATTNER & SCHNEIDER, 1993). O ponto no qual os riscos potenciais do tratamento ortodôntico superam os seus benefícios é um problema controverso e depende de análise individualizada caso a caso. Além disso, fatores não clínicos podem afetar a decisão do especialista sobre necessidade de tratamento em casos limítrofes. A prescrição real de tratamento pode também ser modificada pelas demandas individuais do paciente (DANIELS & RICHMOND, 2000; RICHMOND & DANIELS, 1998).

Num estudo para verificar a percepção de dentistas sobre a necessidade de tratamento ortodôntico no País de Gales e na Inglaterra, Richmond *et al.* (1994) observaram que os critérios utilizados nesse julgamento não são claros. A opinião sobre necessidade de tratamento sobre 320 modelos de estudo, baseada em critérios estéticos e



relacionados à saúde, foi registrada por 74 dentistas. Entre esses, 44 eram ortodontistas (22 consultores e 22 clínicos), 26 generalistas (15 clínicos e 11 comunitários), mais quatro dentistas não classificados nesses grupos. A análise de concordância intra-examinadores, realizada em 16 modelos, obteve valores de kappa que variaram de 0,54 a 0,97 para necessidade estética e entre 0,12 e 0,89 para necessidade baseada na saúde dental. A concordância interexaminadores variou de forma semelhante. A necessidade estética apresentou níveis bons de concordância ( $\kappa = 0,68$  e  $\kappa = 0,70$ ) dentro dos grupos de ortodontistas enquanto os generalistas alcançaram níveis razoáveis ( $\kappa = 0,58$  e  $\kappa = 0,59$ ). A necessidade baseada em critérios de saúde bucal apresentou valores mais baixos para concordância. Os grupos de ortodontistas obtiveram valores razoáveis ( $\kappa = 0,50$  e  $\kappa = 0,52$ ) e os generalistas apresentaram valores pobres ( $\kappa = 0,36$  e  $\kappa = 0,36$ ).

Richmond e Daniels (1998) obtiveram valores de concordância que apresentaram alguma variação em relação à opinião de 97 ortodontistas de nove países sobre necessidade de tratamento ortodôntico de 240 modelos de estudo. Ao avaliar a necessidade relacionada à estética dental, obtiveram valores para a concordância intra-examinadores de  $\kappa_w = 0,60$  (0,58). Para a necessidade relacionada à saúde dental, os valores observados foram de  $\kappa_w = 0,57$  (0,55).

Um índice numérico pode medir a ocorrência de um problema ou estabelecer seu nível de gravidade. Os índices oclusais são expressões matemáticas que combinam características oclusais medidas por um sistema numérico, ponderado ou não, com o objetivo de quantificar o grau de gravidade da má oclusão. Os diversos desvios da normalidade nas arcadas dentárias e no esqueleto facial têm reflexos variados nas funções do sistema estomatognático, na aparência e auto-estima dos indivíduos afetados. A complexidade dessas variáveis constitui um desafio na criação de índices que representem os diferentes graus de gravidade desses problemas (BRESOLIN, 2000).

Com a principal finalidade de detectar aqueles pacientes com maior necessidade de tratamento ortodôntico, índices foram desenvolvidos desde a década de 1960. Índices como o “índice de desvio labio-lingual incapacitante”, desenvolvido por DRAKER (1960), medidos a partir de exames clínicos ou modelos de estudo, tinham como vantagem não necessitarem de radiografias e poderem ser realizados por examinadores treinados. Essas características eram essenciais para levantamentos epidemiológicos de grandes grupos populacionais. Entretanto, essas avaliações continham apenas os sinais objetivos, ou parte

deles, considerando a ausência de radiografias. Assim, sua utilidade em avaliar a real necessidade de tratamento era mínima (MCLAIN & PROFFITT, 1985).

As medidas oclusais e os valores obtidos pelos índices são convencionais, refletindo as opiniões sobre as desvantagens e riscos das más oclusões para a saúde, na época em que são desenvolvidos (LUTHER, 1998; MCLAIN & PROFFITT, 1985).

A maioria dos índices oclusais atuais contempla três componentes principais da necessidade de tratamento: problemas psicossociais, função oral e a relação com o risco de traumatismo dental e doenças bucais. Sua principal finalidade é selecionar os portadores de más oclusões mais graves para o tratamento subsidiado por programas do governo (BENTELE *et al.*, 2002). Em situações nas quais o provedor de tratamento ortodôntico é o serviço público, com recursos limitados para o tratamento da população, o índice de necessidade de tratamento pode ter seu ponto de corte ajustado de forma a priorizar os casos mais graves, ou seja, os que mais se beneficiariam com o tratamento (SALZMANN, 1968). Apesar da utilização de índices de necessidade de tratamento ser praticamente limitada a países que têm programas governamentais de atenção à saúde odontológica, sua aplicação tem também grande importância no registro da prevalência e gravidade de más oclusões em estudos epidemiológicos, auditoria e referenciamento de pacientes. (BEGLIN *et al.*, 2001; BRASIL, 2004; JÄRVINEN, 2001; LIEPA *et al.*, 2003; YOUNIS *et al.*, 1997).

A maior vantagem em se utilizar índices é maximizar a concordância inter e intra-examinadores, além de possibilitar um diagnóstico mais objetivo, seguro e que possa ser utilizado pelo clínico geral (DANIELS & RICHMOND, 2000; RICHMOND & DANIELS, 1998). Os índices oclusais que avaliam a gravidade da má oclusão e os que determinam a necessidade de tratamento podem ser utilizados por cirurgiões dentistas ou técnicos calibrados (RICHMOND *et al.*, 1995). São de fácil aprendizado e aplicação. Porém, para que possam ser utilizados adequadamente, é necessário que sua validação seja verificada para a região onde se pretende utilizá-los (POLEDNAK, 2000; RICHMOND & DANIELS, 1998; WILT, *et al.*, 1999).

Vários índices oclusais e de necessidade de tratamento são utilizados mundialmente. Entre eles figuram o índice de estética dental (*dental aesthetic index* – DAI), o índice de necessidade de tratamento ortodôntico (*index of orthodontic treatment need* – IOTN) e a pontuação da avaliação de pares (*peer assessment rating* – PAR). O DAI é o índice oclusal de rastreamento (*screening*) recomendado pela Organização Mundial de Saúde – OMS (1999) e, no Brasil, vem sendo utilizado em levantamentos epidemiológicos realizados pelo

Ministério da Saúde (BRASIL, 2004; BRESOLIN, 2000) e em estudos locais (MARQUES *et al.*, 2005; PERES *et al.*, 2002). O IOTN e o PAR são os mais utilizados, principalmente no Reino Unido, para pesquisa, auditoria e gestão de serviços de atendimento clínico. Medem características complementares e são geralmente utilizados em conjunto. Enquanto o primeiro mede a necessidade de tratamento, o segundo tem o objetivo de medir a melhora obtida com o tratamento realizado (BROOK & SHAW, 1989; FOX *et al.*, 2002; RICHMOND *et al.*, 1992).

Um estudo examinou a confiabilidade da informação obtida com o IOTN em situação clínica comparada com modelos de estudo e fotografias. Para o componente de saúde dental, foi encontrada concordância substancial (kappa ponderado de 0,62 a 0,70) entre as observações medidas clinicamente e em modelos de estudo. O componente estético obteve concordância substancial (kappa ponderado de 0,69 a 0,79) quando as medidas clínicas e as realizadas em modelos foram comparadas. Se consideradas as avaliações em fotografias, a sua comparação com as medidas clínicas e em modelos obtiveram apenas concordância de razoável a moderada (kappa ponderado de 0,22 a 0,43).

Embora, na prática, índices oclusais e índices de necessidade de tratamento sejam utilizados de forma associada, até recentemente nenhum índice unificado para avaliação tanto da necessidade de tratamento como da gravidade do desvio da oclusão normal/ideal havia sido desenvolvido ou validado. Recentemente, Daniels e Richmond (2000) propuseram um novo índice unificado, o índice de complexidade, resultado e necessidade de tratamento (*index of complexity, outcome and need* – ICON), com o objetivo de superar as limitações dos índices anteriores e reunir nele, de forma simplificada, a avaliação de características antes realizadas somente por índices distintos (FOX *et al.*, 2002).

## 2.2 DAI

O índice de estética dental é um índice quantitativo, proposto, em 1986 por Jenny e Cons, para avaliar o grau de impacto estético da dentição. Tem sido adotado pela OMS como um índice de rastreamento trans-cultural (OMS, 1999). O DAI é uma equação de regressão logística que relaciona, matematicamente, a percepção do público sobre a estética dental com medidas físicas objetivas de características oclusais associadas às más oclusões (BEGLIN *et al.*, 2001; JENNY & CONS, 1988, 1996a).

Os conceitos teóricos que embasam o DAI permitem afirmar que quanto mais o aspecto dentário de uma pessoa é desviado das normas sociais, maior a probabilidade dela apresentar limitações sociais e, conseqüentemente, necessidade de tratamento ortodôntico. Em epidemiologia, o DAI pode ser utilizado como índice de gravidade da má oclusão e como índice de necessidade de tratamento ortodôntico (JENNY & CONS, 1988, 1989, 1996a, b).

Para se obter o valor do DAI, são avaliadas dez características oclusais conforme padrões socialmente definidos para estética dental: dentes ausentes, apinhamento anterior, espaçamento anterior, diastema na linha média, maior irregularidade anterior na maxila, maior irregularidade anterior na mandíbula, trespasse horizontal maxilar, trespasse horizontal negativo, mordida aberta anterior e relação intermolar ântero-posterior (BEGLIN *et al.*, 2001; JENNY & CONS, 1988, 1996b).

Tais características são medidas e os valores a elas atribuídos são multiplicados pelos seus respectivos pesos e totalizados com a adição de uma constante. O resultado, um valor entre treze a aproximadamente oitenta, pode ser considerado uma escala de intervalo, que pode ser categorizado através de pontos de corte. O protocolo dos critérios e métodos de medição para cada componente foi descrito detalhadamente no APÊNDICE A (JENNY & CONS, 1996b; OMS, 1999).

Em sua concepção original, a distribuição da pontuação do DAI, em percentis, numa amostra grande da população, permitia priorizar a necessidade de tratamento conforme a gravidade relativa da má oclusão. Dado um determinado valor medido pelo índice, através de uma tabela de percentis, era possível calcular qual a parcela da população local teria resultados maiores ou menores que o apresentado. Dessa forma, administradores públicos podiam definir um ponto de corte conveniente para que o índice selecionasse um número máximo de pessoas a serem atendidas, conforme a quantidade de recursos disponíveis (JENNY & CONS, 1988).

Estudos mais recentes estabeleceram pontos de corte para definir quatro categorias de gravidade da má oclusão e de necessidade de tratamento ortodôntico. O ponto de corte que estabelece a categoria de más oclusões incapacitantes foi definido a partir da correlação entre a os valores do DAI e a opinião de ortodontistas experientes em triagem de más oclusões incapacitantes, em programas públicos. Esses valores foram medidos numa amostra com 1.306 pares de modelos de estudo representativa de meio milhão de estudantes com faixa etária de quinze a dezoito anos. (JENNY *et al.*, 1993).

Num estudo posterior, Jenny e Cons (1996a) determinaram os pontos de corte para as categorias de má oclusão com menor gravidade. Utilizou-se a distribuição da frequência dos valores do DAI nos mesmos 1.306 modelos utilizados no estudo prévio e os dados do relatório do Centro Nacional de Estatísticas de Saúde dos Estados Unidos da América (*National Center for Health Statistics* – NCHS) sobre a distribuição da má oclusão em categorias de gravidade. Os pontos de corte foram definidos através da comparação da distribuição percentual, por gravidade da má oclusão, relatados pelo NCHS e a porcentagem acumulada dos valores do DAI. Resultados iguais ou menores que 25 representam oclusão normal ou má oclusão mínima, requerendo nenhum tratamento ou apenas um tratamento leve. Pontuações entre 26 e 30 representam má oclusão definida, sugerindo tratamento eletivo. Valores entre 31 e 35 representam má oclusão grave com tratamento altamente desejável. Pontuações iguais ou acima de 36 pontos representam má oclusão muito grave ou incapacitante com tratamento obrigatório. Para categorização dicotômica em “com necessidade” e “sem necessidade” de tratamento ortodôntico, o valor do ponto de corte recomendado é igual a 31. Assim, valores iguais ou maiores que 31 determinam necessidade de tratamento ortodôntico (BEGLIN *et al.*, 2001; JENNY *et al.*, 1993; JENNY & CONS, 1996a; 1996b).

A medição e o registro dos valores referentes aos dez componentes do DAI são realizadas em aproximadamente três minutos. O DAI foi desenvolvido para ser utilizado na dentição permanente, em indivíduos com idade a partir dos doze anos. Porém pode ser facilmente adaptado para a utilização na dentadura mista. Em estudos de calibração, recomenda-se que os limites de concordância não ultrapassem a amplitude de -4,0 a 4,0 na escala de valores do DAI (JENNY & CONS, 1988; 1996b; OMS, 1999).

Cons *et al.* (1989) avaliaram a percepção da estética dental por estudantes da Tailândia, Austrália e da República Democrática da Alemanha. Foi observado que esses estudantes concordavam com o que os nativos de reservas indígenas e estudantes norte-americanos consideravam como características oclusais socialmente aceitáveis ou não, numa escala ordinal de seis pontos. As opiniões registradas para um conjunto de 25 modelos de estudos foram ordenadas em postos, e a correlação por grupos de estudantes foi calculada através da correlação de Spearman, cujos valores variaram entre 0,81 a 0,96. Por ser a Tailândia um país em desenvolvimento, foi generalizado que o DAI pode ser utilizado sem modificações, tanto em países industrializados como em países em desenvolvimento, para avaliar a necessidade de tratamento ortodôntico.

O DAI avalia a gravidade do impacto estético da má oclusão, sendo a necessidade de tratamento uma conclusão subjetiva inferida dessa classificação (BRESOLIN, 2000). Uma limitação na sua utilização como índice de necessidade de tratamento ortodôntico, é desconsiderar que alterações em diferentes locais da cavidade bucal produzem seu efeito apenas naquele local. O fato de somar valores atribuídos a várias características de locais diferentes e não ponderados por grau de morbidade, pode resultar num valor final aumentado que não corresponda exatamente à gravidade da má oclusão. Assim, pequenas alterações oclusais pouco importantes em relação ao seu impacto na saúde bucal, se somadas, podem acumular uma alta pontuação (SHAW, 1991b).

A medição direta de alguns componentes em escala milimétrica pode fazer com que pequenos erros tenham efeito exagerado após a multiplicação desses valores pelos coeficientes. O DAI não considera a mordida cruzada nos segmentos bucais, os desvios na linha média nem a mordida profunda excessiva. Esses fatores podem ter um impacto considerável na gravidade da má oclusão e, por desconsiderá-los, o índice perde força. (OTUYEMI & NOAR 1996a). Algumas características relacionadas à necessidade de tratamento ortodôntico não são incluídas no cálculo do DAI: molares ausentes, dentes impactados, mordidas cruzadas posteriores e discrepâncias da linha média. A relevância de um índice como o DAI é controversa, porque a necessidade de tratamento ortodôntico depende de outros fatores além da estética (DANYLUK et al., 1999).

Um fator que pode afetar a validade e a reprodutibilidade de um índice é o seu desempenho com o decorrer do tempo (SUMMERS, 1971). Foi avaliada uma amostra com 45 indivíduos, sendo que 29 casos estavam na fase final da dentadura mista e dezesseis na fase inicial da dentição permanente. Os casos foram medidos com o DAI, acompanhados dos 12 aos 16 anos de idade e medidos novamente. Os valores do DAI mostraram que, nesse período do desenvolvimento, pode haver alterações oclusais com potencial de alterar os resultados dos índices. Após esses quatro anos, os valores do DAI reduziram, em média, 5,1 pontos em toda a amostra. No grupo da dentadura mista, a redução foi significativamente maior ( $p < 0,01$ ). Os componentes do DAI, que mais contribuíram para as alterações temporais em seus valores, foram o trespassse horizontal maxilar, apinhamento e espaçamento anteriores, diastema na linha média e relacionamento intermolares (TARVIT & FREER, 1998).

Keay *et al.* (1993) verificaram a validade do DAI em uma amostra ( $n = 105$ ), com idade média de doze anos e 1,3 meses, composta por setenta indivíduos com

dentições na fase da dentadura mista e 35 na fase da dentição permanente. As avaliações ortodônticas geralmente devem ser realizadas na dentadura mista, de forma que um possível tratamento possa ser planejado em idade adequada. Os valores do DAI para os setenta casos foram comparados com a opinião subjetiva de um painel de quatro especialistas sobre a gravidade da má oclusão desses mesmos casos. O ponto de corte utilizado foi igual a 32,5, pois não havia sido publicado, até então, uma definição nesse sentido. Os valores de sensibilidade (S) e especificidade (E) foram de 0,86 e 0,60, respectivamente. Os valores preditivos positivos (VPP) e negativos (VPN) calculados foram: VPP = 0,54 e VPN = 0,89. A comparação dos valores de sensibilidade e especificidade da dentadura mista (S = 0,83; E = 0,59) e da dentição permanente (S = 0,92; E = 0,60) mostrou que a primeira apresentou valores menores. O DAI pode ser adequado como um índice de rastreamento, embora tenha classificado como necessidade de tratamento uma parcela de falsos-positivos (0,40).

Um estudo de validação do DAI para necessidade de tratamento, com um grande número de ortodontistas como painel validador, foi realizado em 2001 (BEGLIN *et al.*, 2001). Os valores do índice, medidos por um examinador calibrado, em 156 pares de modelos de estudo, foram confrontados com a opinião subjetiva do painel validador (n = 15). A concordância intra-examinador do pesquisador calibrado, obtida pela estatística kappa (limite inferior do intervalo de confiança de 95% em parênteses), foi igual a 0,96 (0,93). Para o painel de especialistas, a concordância interexaminadores foi  $\kappa_w = 0,81$  (0,81), e a concordância intra-examinadores foi  $\kappa_w = 0,92$  (0,90). A opinião do painel classificou 64% dos casos na categoria de necessidade de tratamento enquanto o DAI classificou 65% na mesma categoria. Os valores de sensibilidade e especificidade para o ponto de corte recomendado, igual a 31, foram: S = 0,78 e E = 0,92. Se for considerado o ponto de corte otimizado, igual a 28, esses valores são alterados para S = 0,93 e E = 0,84. O DAI parece ser uma medida válida da necessidade de tratamento conforme percebida por ortodontistas, sendo a área sob a curva ROC igual a 0,95 (BEGLIN *et al.*, 2001).

Não foram encontrados relatos da validação do DAI no Brasil.

O DAI tem sido utilizado para avaliar a gravidade da má oclusão e a prevalência da necessidade de tratamento ortodôntico em vários estudos nacionais (BRASIL, 2004; MARQUES *et al.*, 2005, 2006). O relatório dos resultados principais do levantamento epidemiológico Projeto SB Brasil 2003 apresentou, para a região sudeste, numa amostra de adolescentes com 12 anos de idade (n = 8.052), a prevalência de 42,28% com necessidade

de tratamento ortodôntico definida pelo ponto de corte igual a 31 (BRASIL, 2004). Marques *et al.* (2005) observaram uma prevalência de necessidade de tratamento ortodôntico de 26,4% numa amostra (n = 333) representativa da população de adolescentes com idades entre 10 e 14 anos em Belo Horizonte, considerando o ponto de corte para o DAI igual a 31.

### 2.3 ICON

O ICON é um índice quantitativo multifuncional que pode ser utilizado para medir simultaneamente a necessidade de tratamento ortodôntico, sua complexidade e a aceitabilidade do seu resultado, dependendo da definição do ponto de corte. No seu desenvolvimento, uma amostra de 338 modelos de estudo (240 iniciais e 98 tratados) foi avaliada por um painel internacional composto de 97 ortodontistas de nove países (Alemanha, Grécia, Hungria, Itália, Países Baixos, Noruega, Espanha, Reino Unido e Estados Unidos). O painel deu sua opinião subjetiva sobre a complexidade da má oclusão, necessidade de tratamento, melhora obtida pelo tratamento e sua aceitabilidade. Para cada modelo, valores objetivos foram atribuídos de forma abrangente às diversas características oclusais, de acordo com um protocolo numérico definido para: alinhamento nos segmentos anteriores superior e inferior, relação vertical anterior, linha média, presença de dentes impactados, alinhamento nos segmentos bucais superiores e inferiores, relacionamento ântero-posterior e vertical nos segmentos bucais, mordida cruzada, dentes ausentes por qualquer razão (exceto terceiros molares), avaliação estética baseada na escala contínua padronizada de necessidade estética (*standardized continuum of aesthetic need – SCAN*), trespasse horizontal positivo ou negativo (em milímetros), inclinação dos incisivos superiores relativa ao plano oclusal, discrepância alveolar generalizada nos arcos superior e inferior e competência labial. Na avaliação dos segmentos bucais, é considerada a soma dos lados direito e esquerdo (DANIELS & RICHMOND, 2000; RICHMOND *et al.*, 1998).

Através de um modelo de regressão logística, cinco variáveis altamente preditivas foram identificadas para necessidade de tratamento: escala SCAN, discrepância dento-alveolar no arco superior, mordida cruzada, relacionamento vertical anterior, relacionamento ântero-posterior do segmento bucal. A equação do ICON integra as variáveis que avaliam as características estéticas e aquelas relacionadas à saúde bucal (DANIELS & RICHMOND, 2000; FOX *et al.*, 2002; RICHMOND *et al.*, 1998).



Desenvolvida por Evans e Shaw (1987), a escala SCAN consiste numa escala ilustrada por uma série de dez fotografias numeradas. A aparência das arcadas do paciente é comparada às fotografias que representam níveis de necessidade de tratamento baseados na estética. Essas imagens foram pontuadas por pessoas leigas, de acordo com o grau de atratividade estética, e posicionadas com espaços equidistantes numa série de dez pontos. Um valor é atribuído à atratividade geral condizente com as categorias ilustradas pelas fotografias, indicando a necessidade do tratamento baseada no prejuízo estético que, por inferência, reflete a necessidade psicossocial para o tratamento ortodôntico. A escala pode ser categorizada em três níveis de necessidade de tratamento. As fotos de 1 a 4 representam a categoria sem indicação para tratamento, as de 5 a 7 o limite entre não necessidade e a necessidade, e as de 8 a 10 indicam a necessidade de tratamento (SHAW *et al.*, 1991b).

A utilização da escala SCAN pode ser criticada por apresentar uma visão bidimensional e limitada de exemplos da aparência dental. O julgamento da impressão estética geral, em vez da semelhança morfológica, é problemático nos casos cujas características principais não são mostradas nas fotografias. Esse é o caso da mordida aberta anterior e da mordida cruzada anterior, que não representadas na escala (OTUYEMI & NOAR, 1996b).

O desempenho da escala SCAN, no decorrer do tempo, foi verificado por Tarvit e Freer (1999). A amostra era composta por 45 indivíduos com idades variando de doze a dezesseis anos, sendo que 29 casos estavam na fase final da dentadura mista e dezesseis na fase inicial da dentição permanente. Após quatros anos de acompanhamento sem tratamento ortodôntico, novas medidas foram realizadas e comparadas. Os valores da escala foram significativamente reduzidos ( $p < 0,05$ ) no grupo da dentadura mista, enquanto sua redução no total da amostra não foi significativa, após o período avaliado. O componente do IOTN que mede o impacto da má oclusão na saúde bucal, também avaliado, não foi alterado significativamente no decorrer do período observado.

Para se obter o valor do ICON, cada característica é medida e categorizada conforme um protocolo que define a pontuação para cada item. Os valores atribuídos às cinco características (quatro componentes oclusais e um estético) são multiplicados por seus respectivos coeficientes (pesos) e somados. O valor total é uma pontuação única, situada numa escala de intervalo, que pode variar de sete a 128 pontos (DANIELS & RICHMOND, 2000; FOX *et al.*, 2002). O protocolo dos critérios e métodos de medição para cada componente do ICON foi descrito detalhadamente no APÊNDICE B.

O ponto de corte recomendado pelos autores para a avaliação da necessidade de tratamento é igual a 43. Se o valor do índice for maior que 43, o tratamento ortodôntico é indicado. Esse valor foi determinado pelo traçado da especificidade, sensibilidade e acurácia geral. A validade do índice para decisões de necessidade de tratamento obteve valores de especificidade de 0,84, de sensibilidade de 0,85 e de acurácia geral (porcentagem de classificações corretas) de 0,85 (DANIELS & RICHMOND, 2000).

O ICON tem como vantagem a simplicidade na sua utilização, pois possui poucas características a serem medidas, além de integrar na mesma equação tanto o componente estético assim como os relacionados à morbidade da má oclusão. Sua aplicação é relativamente rápida, de aproximadamente um minuto para cada caso, até o máximo de três minutos em casos mais complexos. Outra vantagem é a ampliação da faixa etária, tornando viável uma avaliação na fase tardia da dentadura mista, a partir do período intertransicional. Pode ser aplicado em modelos de estudos sem necessidade de adaptação e não exige instrumentos de medida, exceto uma régua milimetrada e a escala SCAN (DANIELS & RICHMOND, 2000). Em estudos de calibração para o ICON, recomenda-se que os limites de concordância não ultrapassem a amplitude de -18,0 a 18,0 (BROWN & RICHMOND, 2005).

A validade do ICON, para determinar a necessidade de tratamento ortodôntico, conforme a percepção de necessidade por ortodontistas norte-americanos, foi verificada por Firestone *et al.* (2002a). Os valores do ICON foram medidos em 170 pares de modelos e comparados à opinião subjetiva de um painel constituído por quinze ortodontistas da região central do estado de Ohio, Estados Unidos da América – EUA. O examinador calibrado para o índice demonstrou concordância intra-examinadores de  $\kappa_p = 0,89$  (0,74). Para o painel de especialistas, a concordância interexaminadores foi de  $\kappa_p = 0,81$  (0,81) e a intra-examinadores de  $\kappa_p = 0,92$  (0,90). A opinião do painel classificou 64% dos casos na categoria de necessidade de tratamento enquanto o ICON classificou 65% na mesma categoria. Utilizando um ponto de corte igual a 42, os valores de sensibilidade e especificidade foram  $S = 0,94$  e  $E = 0,86$ . Valores preditivos positivos e negativos para esse ponto de corte foram:  $VPP = 0,92$ ; e  $VPN = 0,90$ . Um ponto de corte igual a 52 obteve um maior equilíbrio entre os valores de sensibilidade e especificidade:  $S = 0,92$  e  $E = 0,93$ . Os novos valores preditivos foram:  $VPP = 0,96$ ; e  $0,87$ . A área sob a curva ROC obtida foi igual a 0,97, indicando alta validade do índice.

Cooke (2006) verificou a validade do ICON contra a opinião de quinze ortodontistas do estado da Califórnia, EUA. A opinião sobre necessidade de tratamento, baseada numa escala com doze categorias, foi atribuída a 153 pares de modelos de estudo. Observando-se a distribuição da classificação de um a doze pontos, atribuída ao mesmo conjunto de modelos, a menor média foi igual a 6,03 e a maior foi igual a 9,95. Evidenciou-se alguma variação na forma como os especialistas julgaram diferentemente os mesmos modelos utilizando a mesma escala. A remoção de dois ortodontistas que não avaliaram um número suficiente de casos reduziu a variância observada. A utilização de um ponto de corte igual a 37,5 refletiu melhor a opinião dos especialistas do que o recomendado originalmente, igual a 43.

A validação do ICON, para necessidade de tratamento, na Holanda, foi realizada através da comparação dos valores do índice, medidos em 102 pares de modelos de estudo, por um examinador calibrado, com a opinião de sete ortodontistas daquele país. O painel foi originalmente constituído por nove especialistas e dois foram excluídos devido aos baixos valores de concordância intra-examinadores e interexaminadores. A concordância intra-examinador foi de  $r_1 = 0,89$ , medida com o coeficiente de correlação intraclasse – ICC. O painel de especialistas obteve valores de concordância interexaminadores de  $r_1 = 0,90$  e intra-examinadores entre  $r_1 = 0,60$  e  $r_1 = 0,86$ . Os valores de sensibilidade e especificidade foram iguais a 1,0 e 0,36, respectivamente. A escolha de um novo ponto de corte, igual a 52, melhorou o equilíbrio entre sensibilidade e especificidade, assumindo os respectivos valores:  $S = 0,91$  e  $E = 0,84$ . O componente estético (escala SCAN) foi o que obteve maior discordância entre um examinador calibrado e outro não calibrado. Durante a calibração do examinador, esse foi o componente que apresentou maior dificuldade para seu aprendizado (LOUWERSE *et al.*, 2006).

## 2.4 CONFIABILIDADE

Os índices oclusais são métodos de medida padronizados. Todo processo de mensuração pode conter erros e esses podem originar-se no observador, no instrumento ou método, ou mesmo em alguma circunstância que afete o momento da medida. Somados à variabilidade do objeto estudado, além da possibilidade de eventos desconhecidos que acrescentem mais erro à medida, os erros afetam a precisão da representação do fenômeno estudado (LUIZ, 2005). Conforme o modelo de mensuração, quando o examinador realiza

determinada medida, o valor observado ( $X$ ) é a somatória do valor real ( $T$ ), do erro sistemático ( $r$ ) e do erro aleatório ( $e$ ) (EQUAÇÃO 1) (FLEISS, 1986; ROBERTS & RICHMOND, 1997).

$$X = T + (r + e) \quad (\text{EQUAÇÃO 1})$$

Um modelo estatístico da confiabilidade pode ser desenvolvido a partir do modelo de mensuração. Seja  $X$  o valor observado de uma determinada variável. O valor foi obtido a partir de medidas (exame físico, questionário, exame laboratorial, etc.) confiáveis em condições controladas e, mesmo assim, cada nova medida difere da anterior em alguma extensão. Se medidas repetidas dessa variável de interesse forem realizadas em condições com menor variabilidade possível,  $T$  representará a média dessas muitas medidas hipotéticas repetidas. Uma medida isolada  $X$  será diferente de  $T$  por várias razões: erro aleatório, mau entendimento da pergunta do entrevistador ou da resposta do entrevistado (em caso de questionário), dificuldades inerentes à própria variável, ou calibração imperfeita do instrumento ou método de medida. Se  $(r+e)$  representam a diferença entre uma observação única  $X$  e seu valor médio  $T$ , o modelo linear clássico de um valor observado pode ser representado pelo modelo de mensuração.

Numa população, o valor sem erro  $T$  varia em torno de um valor médio  $\mu$  com uma variância de  $\sigma_T^2$ . Para uma medida única, o erro aleatório varia em torno de uma média zero. Considerando que a distribuição dos erros seja independente do valor de  $T$ ,  $(r+e)$  tem uma variância de  $\sigma_{(r+e)}^2$ , não importa o valor de  $T$ , sendo então a variância de  $X$ :

$$\sigma_x^2 = \sigma_T^2 + \sigma_{(r+e)}^2 \quad (\text{EQUAÇÃO 2})$$

Há dois componentes para a variabilidade dentro de uma série de medidas em diferentes objetos: a variabilidade entre os seus valores reais mais a variabilidade dos seus erros. Um coeficiente único que expressa a magnitude relativa dos dois componentes da variância de  $X$  é o coeficiente de confiabilidade de correlação intraclasse ( $r_I$ , ICC) (FLEISS, 1986).

$$r_I = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{(r+e)}^2} \quad (\text{EQUAÇÃO 3})$$

O erro aleatório reduz a precisão, enquanto o erro sistemático afeta sua validade. O viés de informação e os erros aleatórios de mensuração comprometem respectivamente a validade e a precisão das medidas em questão (COSTA e NADANOVSKY, 2005a).

O erro sistemático afeta a comparação de medidas realizadas por examinadores diferentes e, no caso de vários examinadores, aumenta a variância. Já o erro aleatório não altera a média das medidas, mas afeta a variância de forma a dificultar a definição de diferenças estatisticamente significativas. Reduz também a correlação entre variáveis, fato conhecido como atenuação, levando conseqüentemente à necessidade de uma amostra maior (ROBERTS & RICHMOND, 1997).

A utilização de índices requer o treinamento de examinadores para sua calibração, de forma a minimizar os erros de mensuração. Esses erros podem ser estatisticamente avaliados através dos estudos de confiabilidade ou concordância. No caso de haver uma medida considerada como correta, o padrão-ouro, contra a qual as medidas serão comparadas, os estudos avaliam a validade (acurácia ou conformidade) da mensuração. Os estudos conhecidos como de consistência têm como objetivo avaliar a reprodutibilidade ou repetibilidade das medidas. Geralmente os estudos de confiabilidade agregam avaliações de conformidade e consistência (LUIZ, 2005).

Estudos de confiabilidade das medidas realizadas por um examinador, devem verificar tanto a sua validade em relação a um padrão de referência aceito como correto, como sua repetibilidade. Embora o valor “correto” talvez nunca possa ser conhecido, é importante que haja um padrão de referência através do qual se possa avaliar o erro sistemático do examinador e assegurar uma padronização dos critérios de medida. Isso torna possível a comparação entre medidas realizadas por examinadores diferentes (RICHMOND *et al.*, 1995).

O valor adotado como padrão de referência corresponde ao  $T$  no modelo de mensuração (EQUAÇÃO 1). Presume-se que este valor seja confiável (válido e consistente) ou, pelo menos, que a possibilidade de erro seja menor do que a do examinador em treinamento. Quando nenhum examinador pode ser considerado infalível, esse padrão de referência pode ser obtido através de um consenso entre examinadores treinados. Após a adoção de um padrão de referência aceito como confiável, embora na prática isso possa não ser verdadeiro, todo o erro aleatório e sistemático será atribuído ao examinador em treinamento (ROBERTS & RICHMOND, 1997).

Há diversas formas para avaliar o grau de concordância entre valores medidos. Testes estatísticos, coeficientes e abordagem gráfica podem ser utilizados de forma complementar para verificar melhor a confiabilidade de variáveis numéricas e categóricas, ordinais e dicotômicas (LUIZ, 2005; ROBERTS & RICHMOND, 1997).

Os valores obtidos pelos coeficientes de concordância estão localizados no centro de um intervalo de probabilidade de obtenção do resultado. A obtenção do intervalo de confiança significa que se a medida em questão fosse repetida cem vezes, em condições semelhantes, haveria uma faixa de valores como possíveis resultados para o coeficiente de concordância, dada a variabilidade das medidas. Nesse caso, pode ser considerado o limite inferior do intervalo de confiança como o menor valor provável obtido pelo coeficiente, pois há interesse apenas em estabelecer, com um grau de certeza, se o valor obtido pelo coeficiente está acima de um valor mínimo específico. Quanto menor a variabilidade das medidas, menor a amplitude do intervalo de confiança (RICHMOND *et al.*, 1995; ROBERTS & RICHMOND, 1997).

Os valores dos índices DAI e ICON são obtidos através da soma de componentes categóricos ponderados e totalizados. O resultado é uma escala de intervalo que, para efeito da avaliação de sua confiabilidade, pode ser tratada como variável numérica (LUIZ, 2005; ROBERTS & RICHMOND, 1997).

Alguns testes estatísticos e coeficientes são utilizados tradicionalmente para avaliar a concordância de variáveis numéricas. O coeficiente da correlação linear de Pearson ( $r$ ) mede a força do relacionamento linear. A correlação linear tem sido utilizada equivocadamente no sentido de avaliar a concordância. É insuficiente que as medidas tenham uma correlação elevada, pois medidas com correlação perfeita ( $r = 1$ ) podem estar, num gráfico de dispersão, numa linha bem diferente da linha de igualdade de 45 graus (BLAND & ALTMAN, 1986, 1999; LUIZ, 2005).

O teste- $t$  pareado pode ser uma alternativa estatística para avaliar se o erro sistemático é diferente de zero. Porém esta abordagem pode falhar quando o erro aleatório é elevado, tornando o desvio-padrão das diferenças maior e, conseqüentemente,  $t$  menor. Neste caso, aumenta-se a chance de que a hipótese nula ( $H_0$ : erro  $r = 0$ ) não seja rejeitada. Assim, torna-se menos provável que o teste- $t$  aponte que o erro sistemático seja estatisticamente significativo (ROBERTS & RICHMOND, 1997).

A raiz da média de quadrados do erro “intragrupo” (*root mean square error – RMSE*), obtido a partir da análise de variância, pode ser utilizado para quantificar o erro aleatório. Estima o desvio-padrão que representa o erro de mensuração de cada medida, desde que não haja erro sistemático. Na presença de erro sistemático, o desvio-padrão é superestimado pela inclusão dos dois tipos de erro. No caso da comparação entre duas

medidas, pode ser calculado conforme a fórmula abaixo, onde  $d_i$  é a diferença entre as duas medidas (FLEISS, 1981, 1986; ROBERTS & RICHMOND, 1997).

$$RMSE = \sqrt{\frac{\left(\sum_{i=1}^n d_i^2\right)}{2n}} \quad (\text{EQUAÇÃO 4})$$

O coeficiente de correlação intraclasse (*intraclass correlation coefficient* – ICC –  $r_I$ ) é um coeficiente que avalia o grau de concordância desconsiderando aquela obtida apenas pela chance. Relaciona a magnitude do erro de medida à variabilidade da amostra estudada (ROBERTS & RICHMOND, 1997). Seus resultados são valores que podem variar de zero à unidade. O valor de  $r_I$  igual a zero significa que as medidas são tão pouco confiáveis que as diferenças entre as medidas são creditadas exclusivamente aos erros aleatórios (FLEISS, 1986).

Um valor menor que 0,4 para o limite inferior do seu intervalo de confiança de 95% constitui baixa concordância. Valores entre 0,4 e 0,75 são considerados razoáveis ou bons, enquanto acima de 0,75 são excelentes (FLEISS, 1981).

Para estudos de confiabilidade com repetição simples, Fleiss (1986) sugere a utilização das seguintes fórmulas para a estimativa do ICC e do limite inferior do intervalo de confiança:

$$r_I = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_{\sigma^2}} = \frac{BMS - WMS}{BMS + (k_0 - 1)WMS} \quad (\text{EQUAÇÃO 5})$$

$$r_I \geq \frac{\frac{BMS}{WMS} - F_{N-1, K-N, \alpha}}{\frac{BMS}{WMS} + (k_0 - 1)F_{N-1, K-N, \alpha}} \quad (\text{EQUAÇÃO 6})$$

Na estimativa do ICC, a partir da análise de variância,  $BMS$  é o quadrado médio entre os objetos medidos,  $WMS$  é o quadrado médio intra-objetos medidos e  $k_0$  é o número de medidas realizadas em cada objeto. A variância das médias e a média das variâncias são representadas por  $\sigma_x^2$  e  $\sigma_{\sigma^2}$ , respectivamente. No cálculo do intervalo de confiança unicaudal aproximado de  $100(1-\alpha)\%$  para  $r_I$ ,  $N-1$  e  $K-N$  são os graus de liberdade da distribuição  $F$  com corte na proporção  $\alpha$  na cauda superior.

Bland e Altman (1986) propuseram um método gráfico para medir a concordância de variáveis numéricas através do traçado de um diagrama de dispersão das diferenças entre as medidas contra as médias das mesmas medidas, para cada observação.

Essa abordagem possibilita a quantificação das discordâncias entre medidas individuais. Além disso, permite a visualização da magnitude dos erros sistemático e

aleatório, fornecendo também um intervalo de concordância para as medidas. Quando se avalia a diferença entre medidas realizadas por dois métodos ou examinadores num mesmo objeto, pode haver uma tendência consistente de um método ou examinador superar o outro. Esse viés é o erro sistemático e é estimado pela média das diferenças. Haverá também variação em torno dessa média, a qual pode ser estimada através do desvio-padrão das diferenças. Estas estimativas somente são significativas se o erro sistemático (viés) e a variabilidade (erro aleatório) forem distribuídos uniformemente através da faixa de valores medidos. O pressuposto que os erros aleatórios apresentem distribuição normal, e que a magnitude das diferenças seja independente das médias, é uma limitação na abordagem gráfica original de Bland-Altman (BLAND & ALTMAN, 1986, 1999; BROWN & RICHMOND, 2005; LUIZ, 2005; ROBERTS & RICHMOND, 1997).

Numa publicação mais recente, os autores fizeram uma atualização, de forma a resolver os casos em que há heteroscedasticidade ou uma correlação entre diferenças e médias. Torna-se possível assim uma avaliação mais precisa da concordância, pois o cálculo dos limites de concordância é baseado numa simples transformação logarítmica ou na análise de regressão (BLAND & ALTMAN, 1999; BROWN & RICHMOND, 2005).

No caso de variáveis categóricas, dicotômicas ou com mais categorias, uma forma simples de se calcular a concordância é através da concordância percentual. Apesar da grande simplicidade de sua utilização, esse tipo de cálculo apresenta como desvantagens a dependência das freqüências das categorias da variável e a desconsideração das concordâncias que podem ocorrer pela chance (LUIZ, 2005).

Um cálculo de concordância para variáveis categóricas que corrige a desvantagem de contabilizar a concordância ocorrida ao acaso é o coeficiente kappa ( $\kappa$ ) (COHEN, 1960). Em publicação posterior foi generalizado para medidas categóricas ordinais, situação na qual a importância relativa de cada possível discordância necessita ser quantificada (COHEN, 1968).

A estatística kappa pode ser explicada a partir da concordância atribuível, ou seja, a concordância observada menos aquela obtida pela chance ( $p_o - p_e$ ). É a proporção da concordância atribuível em relação ao seu valor máximo possível ( $1 - p_e$ ).

$$\kappa_w = \frac{p_o(w) - p_e(w)}{1 - p_e(w)} \quad (\text{EQUAÇÃO 7})$$

Pesos ( $w$ ) podem ser atribuídos, baseados em critérios racionais ou clínicos, aos diferentes níveis de concordância dentro da escala ordinal (CICCHETTI, 1976; FLEISS, 1981). Considerando que o sistema de ponderação altera os valores de kappa ponderado, a



sua escolha deve ser feita *a priori*, em função do tipo e do número de categorias. Há dois sistemas padronizados de ponderação que podem ser utilizados na maioria das situações: o linear e o quadrático (CICCHETTI, 1976; ROBERTS & RICHMOND, 1997). A utilização da ponderação quadrática torna o kappa ponderado equivalente ao ICC (FLEISS & COHEN, 1973). Isso sugere sua utilização em escalas que possuam grande número de categorias, ou nas que possam ser interpretadas como contínuas, tal como a escala SCAN (ROBERTS & RICHMOND, 1997). Já ponderação linear, parece mais adequada nos casos em que o número de categorias é pequeno e tende a ter diferenças qualitativas (CICCHETTI, 1976; ROBERTS & RICHMOND, 1997).

O kappa simples é um caso especial do kappa ponderado quando os pesos são iguais a zero em todas as células da tabela de contingência, exceto as da diagonal principal de concordância completa, para as quais o peso é igual a um (COHEN, 1968).

Os valores do kappa simples, kappa ponderado quadrático e kappa ponderado linear não são comparáveis. O número de categorias da variável analisada também afeta os valores de kappa. Portanto, não é possível comparar valores de kappa entre duas medições distintas cujas variáveis apresentem números de categorias diferentes, mesmo que a matriz de ponderação escolhida seja a mesma (CICCHETTI, 1976).

Os valores de kappa podem variar de 1 a -1. No caso de concordância completa,  $\kappa = +1$ . Se a concordância observada é maior ou igual à esperada pela chance,  $\kappa \geq 0$ . Se a concordância observada for menor ou igual à esperada pela chance,  $\kappa \leq 0$ . O valor mínimo de  $\kappa$  depende das proporções marginais da tabela de contingência das freqüências de concordância observadas, a partir das quais a concordância esperada pela chance é calculada. As fórmulas para o cálculo do erro padrão do kappa simples e do kappa ponderado para amostras grandes foram publicadas originalmente com erro que superestimava a variância de kappa. Em publicação posterior foram apresentadas novas fórmulas válidas para variância aproximada do kappa simples ( $\kappa$ ) e ponderado ( $\kappa_w$ ) para grandes amostras (FLEISS *et al.*, 1969):

$$\sigma^2(\kappa) = \frac{1}{N(1-p_e)^4} \left\{ \sum_{i=1}^k p_{ii} \times [(1-p_e) - (p_i + p_i)(1-p_o)]^2 + (1-p_o)^2 \sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k p_{ij} (p_i + p_j)^2 - (p_o p_e - 2p_e + p_o)^2 \right\} \quad (\text{EQUAÇÃO 8})$$

$$\sigma^2(\kappa_w) = \frac{1}{N(1-p_e)^4} \left\{ \sum_{i=1}^k \sum_{j=1}^k p_{ij} [w_{ij}(1-p_e) - (\bar{w}_{i.} + \bar{w}_{.j})(1-p_o)]^2 - (p_o p_e - 2p_e + p_o)^2 \right\} \quad (\text{EQUAÇÃO 9})$$

Não é simples avaliar o nível de significância clínica ou prática de um determinado valor de kappa, especialmente no que se refere à comparação entre os níveis de concordância de variáveis com um número diferente de categorias. Landis e Koch (1977) sugeriram a utilização de uma nomenclatura padronizada, embora as divisões fossem arbitrárias, para descrever a força da concordância associada aos valores de kappa. Nessa classificação, valores menores que zero apresentam concordância pobre; entre zero e 0,2, leve; entre 0,21 e 0,4, razoável; entre 0,41 e 0,6, moderada; entre 0,61 e 0,8, substancial ou expressiva e; entre 0,81 e 1,0, boa ou quase perfeita.

Fleiss (1981) sugeriu que os valores de kappa fossem interpretados da mesma forma como os do ICC. De forma consistente com esta orientação, Cicchetti *et al.* (1992) propuseram que os valores de kappa fossem interpretados em quatro níveis de significância clínica, de forma relacionada à avaliação dos níveis de acurácia no modelo de sensibilidade e especificidade. Para valores menores que 0,4, a significância clínica é considerada pobre; entre 0,4 e 0,59, razoável; entre 0,6 e 0,74, boa e; entre 0,75 e 1,0, excelente.

## 2.5 VALIDADE

A validade de índices oclusais é um tipo de validade de critério, tais como a de testes psicométricos, classificada como validade concorrente (YOUNIS *et al.*, 1997).

É uma medida de concordância entre os resultados obtidos pelo instrumento e os obtidos por uma medida “mais objetiva” da mesma população. A medida “objetiva”, o padrão-ouro, é obtida por meio de um instrumento reconhecido e bem estabelecido (STATISTICS.COM, 2005; YOUNIS *et al.*, 1997).

Considerando a dificuldade na identificação e quantificação dos sinais objetivos da má oclusão e da necessidade de tratamento ortodôntico, adquiriu-se crença de que a opinião dos especialistas é o melhor determinante da necessidade de tratamento (RICHMOND *et al.*, 1992; SUMMERS, 1971; YOUNIS *et al.*, 1997). Portanto, não existe um padrão-ouro ou “verdade” definitiva sobre a decisão da necessidade de tratamento. A decisão agregada de especialistas em ortodontia é geralmente considerada como o padrão-ouro contra o qual qualquer índice oclusal deve ser validado (B EGLIN *et al.*, 2001; DE GUZMAN *et al.*, 1995; FIRESTONE *et al.*, 2002a, b; YOUNIS *et al.*, 1997).

Nesse contexto, a avaliação da validade de um índice consiste em investigar se o índice realmente mede o que ele se propõe a medir. Quando utilizados para detectar

indivíduos afetados ou não pela condição medida, os índices oclusais são perfeitamente análogos aos muitos testes diagnósticos usados na pesquisa e na clínica médica (CARLOS, 1970; SUMMERS, 1971).

Os testes diagnósticos, de forma geral, são procedimentos para identificar indivíduos doentes. Podem compreender a anamnese, exames físicos, e exames complementares. De forma ideal, não deveriam errar na classificação de indivíduos afetados e não afetados. Porém várias fontes de erro podem existir, como a variabilidade populacional dos processos fisiopatológicos, ou mesmo devido às características e limitações do próprio teste utilizado. Portanto, o erro de diagnóstico é um tipo de erro de mensuração (COSTA & NADANOVSKY, 2005b).

A validade de um teste de diagnóstico deve ser investigada conforme sua capacidade de discriminar corretamente os casos positivos e negativos. Duas propriedades são avaliadas: a sensibilidade e a especificidade (COSTA & NADANOVSKY, 2005b).

Há evidência de que a determinação da necessidade de tratamento e de seu planejamento pelo especialista apresenta grande variação entre profissionais e pode ser afetada pela localização geográfica. Na medicina também há grande evidência de que padrões de tratamento variam geograficamente, mesmo que a prevalência das doenças relacionadas não varie entre as regiões (POLEDNAK, 2000; RICHMOND *et al.*, 1994, 1998; SHEIHAM, 2000; WILT *et al.*, 1999; YOUNIS, 1997).

Há diferenças nas validades de índices distintos, embora o conhecimento atual não permita avaliar o seu grau (JÄRNIVEN, 2001). Índices epidemiológicos podem ser desenvolvidos e aplicados em determinada situação. Para que sejam utilizados em outras circunstâncias, pode ser apropriado reavaliar se os valores da medida são adequados para a nova situação (ROBERTS & RICHMOND, 1997; LOUWERSE, 2006). Portanto, o limiar de decisão, ou ponto de corte de um índice pode requerer um ajuste para padrões nacionais diferentes, antes de sua aplicação numa nova situação e, em países com dimensões continentais, pode haver diferenças regionais menores que podem influenciar a decisão sobre necessidade de tratamento (B EGLIN *et al.*, 2001).

Testes de validação e confiabilidade têm sido realizados por um grande grupo de ortodontistas para um número limitado de índices e têm demonstrado que vários deles refletem com exatidão a opinião dos especialistas que, por sua vez, apresentaram altos níveis de concordância interexaminadores (B EGLIN *et al.*, 2001; DE GUZMAN *et al.*, 1995; FIRESTONE *et al.*, 2002a; LOUWERSE *et al.*, 2006; RICHMOND *et al.*, 1992; YOUNIS *et al.*,

1997). É possível que a convergência de opiniões entre especialistas norte-americanos seja devida à padronização do processo educacional, cujo currículo clínico é direcionado por padrões de acreditação e critérios definidos pela Associação Americana de Ortodontistas, o que levou a uma noção comum sobre o que constitui uma oclusão ideal (YOUNIS *et al.*, 1997).

Younis *et al.* (1997) utilizaram 160 pares de modelos para validação de índices oclusais, enquanto Beglin *et al.* (2001) e Firestone *et al.* (2002a, b) utilizaram 170, todos baseados no cálculo amostral realizado por Richmond *et al.* (1992) em seu estudo de validação do índice oclusal PAR, cuja concordância encontrada, avaliada pela estatística kappa foi  $\kappa = 0,57$ . Esse estudo recomenda a utilização de um tamanho mínimo de 160 pares de modelos, empregando um poder de teste de 80% e erro alfa de 5%.

### 2.5.1 SENSIBILIDADE E ESPECIFICIDADE

A sensibilidade é a capacidade do teste em identificar como positivos os casos realmente positivos. A especificidade é a habilidade do teste em classificar como negativos os casos realmente negativos. A partir dos valores de sensibilidade e especificidade podem ser deduzidas as porcentagens de falso-negativos e falso-positivos, seus respectivos complementos (COSTA e NADANOVSKY, 2005b; GORDIS, 2004; PEREIRA, 1995).

#### 2.5.1.1 VALORES PREDITIVOS

Na prática clínica é muito importante avaliar qual a chance de um caso específico dado como positivo ou negativo pelo teste diagnóstico ser realmente positivo ou negativo, respectivamente. O valor preditivo positivo é a proporção de indivíduos que apresentam a condição testada entre os casos com teste positivo, enquanto o valor preditivo negativo é a proporção de indivíduos não afetados pela condição entre os que apresentam teste negativo. A interpretação dos valores preditivos pode ser resumida como a probabilidade do indivíduo apresentar realmente a condição dada como resultado do teste. Sua interpretação deve ser cuidadosa, pois trata-se de uma probabilidade que é influenciada pela prevalência da condição estudada e pela especificidade do teste quando a prevalência é baixa na população testada. Os valores preditivos positivos aumentam com o incremento da prevalência, enquanto os negativos diminuem (COSTA e NADANOVSKY, 2005b; PEREIRA, 1995). Alterações no valor da especificidade causam grande impacto nos

valores preditivos positivo e negativo. Um valor maior para especificidade reflete um grande aumento no valor preditivo positivo, especialmente em situação de baixa prevalência (COSTA e NADANOVSKY, 2005b; GORDIS, 2004).

O cálculo dos valores preditivos positivo e negativo a partir dos valores de sensibilidade, especificidade e prevalência pode ser descrito conforme as expressões abaixo (COSTA e NADANOVSKY, 2005b):

$$VPP = \frac{S \times P}{(S \times P) + (1 - E) \times (1 - P)} \quad (\text{EQUAÇÃO 10})$$

$$VPN = \frac{E \times (1 - P)}{(1 - S) \times P + E \times (1 - P)} \quad (\text{EQUAÇÃO 11})$$

### 2.5.1.2 DEFINIÇÃO DO PONTO DE CORTE IDEAL

O ponto de corte ideal para um índice com resultado numérico seria aquele no qual a sensibilidade e a especificidade fossem iguais a 100%. Na prática isso é raramente possível, pois elas apresentam correlação inversamente proporcional. Qualquer tentativa de melhorar a sensibilidade implica na piora da especificidade e vice-versa (PEREIRA, 1995). A escolha do valor para o ponto de corte dependerá do valor relativo que representam os falsos positivos e os falsos negativos. Os resultados falsos positivos, principalmente, apresentam impactos indesejáveis. Sob o ponto de vista do indivíduo, podem acarretar prejuízo emocional por rotular erroneamente o indivíduo com um resultado positivo que pode ser socialmente estigmatizante. No aspecto da saúde pública, acarretam aumento nos custos econômicos, pois traz a necessidade de realizar novos testes para discriminar os indivíduos que realmente apresentam a condição testada. Os falsos negativos têm sua importância dependente da natureza e da gravidade da alteração, além da eficiência das medidas de intervenção disponíveis, considerando se essa eficiência é maior se a intervenção for realizada no início do desenvolvimento da alteração em questão. Os falsos negativos, por acreditarem estar livres da condição estudada, podem seguir sem o acompanhamento ou tratamento adequado (GORDIS, 2004).

### 2.5.1.3 CURVA ROC

Como indicador global do desempenho de um teste de diagnóstico cujo resultado é uma escala numérica, a curva ROC (*receiver/relative operating characteristic* – característica de operação relativa/do receptor) do teste diagnóstico foi inicialmente desenvolvida para testar a detecção e interpretação de sinais de radar contra um ruído de fundo. Posteriormente foi aplicada no campo da radiologia para o diagnóstico por imagens (BECK & SCHULTZ, 1986). É uma descrição simples e completa dos efeitos de qualquer ponto de corte possível sobre a sensibilidade e especificidade do teste. A curva ROC informa sobre a habilidade do teste em refletir a real situação, ou a sua capacidade preditiva de forma precisa e válida numa escala padronizada e fácil de interpretar (SWETS, 1988). Descreve as *características* de detecção inerentes ao teste e o *receptor* da sua informação pode *operar* em qualquer ponto da curva por meio de um ponto de corte apropriado (METZ, 1978).

A questão do teste diagnóstico baseia-se nas conseqüências da tomada de decisão em relação à definição do ponto de corte. Um teste de diagnóstico cujos resultados são valores contínuos tem a distribuição dos mesmos de forma sobreposta entre os indivíduos realmente afetados e não afetados. Diferentes escolhas para um ponto de corte têm como conseqüência vários tipos de decisões corretas e incorretas. A escolha de um determinado ponto de corte classifica arbitrariamente os indivíduos com resultados abaixo desse valor como negativos e os com resultados mais altos como positivos. Valores mais altos para o ponto de corte tornam menores as freqüências de falsos positivos e verdadeiros positivos e, por outro lado, aumentam as freqüências de decisões verdadeiras negativas e falsas negativas. Um ponto de corte adequado deve ser escolhido de forma a contemplar um compromisso entre as conseqüentes vantagens e desvantagens (METZ, 1978).

Vários pontos de corte possíveis estão associados a valores correspondentes de sensibilidade e especificidade. Ao traçar um gráfico da sensibilidade e especificidade no eixo das ordenadas em função dos valores de probabilidade prevista para cada ponto de corte plausível, é possível visualizar a localização de um ponto de corte adequado (DOHOO *et al.*, 2003).

O gráfico da curva ROC convencional é traçado com a sensibilidade no eixo das ordenadas contra o complemento da especificidade (1-E). A curva passa inevitavelmente pelo canto inferior esquerdo do gráfico ( $S = 0, 1-E = 0$ ), porque é possível classificar todos os resultados como negativos, e pelo canto superior direito ( $S = 1, 1-E = 1$ ) porque também é possível classificar todos os resultados como positivos. Assim, se o

teste fornece alguma informação na tomada de decisão, os pontos intermediários da curva devem estar localizados acima da diagonal principal da área ROC (METZ, 1978).

Como os valores de sensibilidade e especificidade são independentes da prevalência, a curva ROC fornece informação sobre a capacidade de detecção do teste de forma independente da prevalência da condição testada e do ponto de corte escolhido (METZ, 1978; SWETS, 1988).

Além das avaliações qualitativas a partir da curva ROC, técnicas foram desenvolvidas e propostas para avaliar as suas propriedades estatísticas desde o início da década de 1980. O indicador estatístico mais popular para descrever a curva ROC é a estimativa da área sob a curva adequada e suavizada, que varia de 0,5 (nenhuma acurácia aparente) a 1,0 (acurácia perfeita), à medida que a curva pode se estender em direção ao canto superior esquerdo da área do gráfico. A área sob a curva tem sido descrita como a representação da probabilidade de um caso com necessidade de tratamento, escolhido aleatoriamente, ser classificado corretamente ou com necessidade maior do que um caso sem necessidade de tratamento, também selecionado aleatoriamente (HANLEY & McNEIL, 1982). Tem sido proposto que a curva ROC seja uma medida mais significativa do valor de acurácia do teste de diagnóstico do que a “acurácia geral” descrita como a porcentagem de casos classificados corretamente. Isto porque, diferentemente da acurácia geral, a curva ROC não é dependente da prevalência de uma doença na população. Além disso, dois testes com a mesma porcentagem de casos corretamente classificados, mas com sensibilidade e especificidade diferentes não resultarão na mesma curva ROC (METZ, 1978).

Testes de diagnóstico cujos resultados são escalas numéricas podem ter seus desempenhos comparados através da análise estatística da diferença entre as áreas das respectivas curvas ROC. Quando essas curvas são baseadas em testes realizados nos mesmos indivíduos, a natureza correlacionada dos dados deve ser considerada. DeLong *et al.* (1988) propuseram uma abordagem não paramétrica para analisar as áreas sob curvas ROC correlacionadas.

A razão de verossimilhança (RV) é outro indicador do desempenho de testes diagnósticos derivado da análise da curva ROC. Para um ponto de corte específico, é o quociente entre as probabilidades de determinado resultado entre indivíduos afetados e em não afetados pela condição testada. A RV positiva é calculada pela divisão da proporção de verdadeiros positivos pela proporção de falsos positivos (BECK & SCHULTZ, 1986; COSTA e NADANOVSKY, 2005b).

$$RV+ = \frac{S}{(1-E)} \quad (\text{EQUAÇÃO 12})$$

Pode ser interpretada como quantas vezes um resultado verdadeiro-positivo é mais provável que um falso-positivo. Em outras palavras, para um determinado ponto de corte ( $x$ ) cujo valor de RV positiva seja  $r$ , a probabilidade de um caso que tenha resultado do teste igual a  $x$  ser afetado pela condição testada é  $r$  vezes maior do que a de um caso com resultado menor que  $x$ . Valores altos de RV positiva são indicativos de testes de boa qualidade. A especificidade tem uma grande influência sobre o seu valor, pois quando elevada faz com que o denominador da expressão acima tenda a zero (COSTA e NADANOVSKY, 2005b).

Os valores ideais da área sob a curva ROC podem variar dependendo de sua aplicação. Swets (1988) sugere que se pode confiar que o teste apresenta boa acurácia, no caso de análise de mamografias, se área sob a sua curva ROC apresenta valores entre 0,80 a 0,90. Testes de previsão de chuva, quando avaliados pela curva ROC, podem ser considerados confiáveis quando sua área é de 0,75 a 0,90.

Cicchetti (2001) sugere que as estimativas de validade, no contexto do modelo de sensibilidade e especificidade, sejam avaliadas por referências que possam definir os valores aceitáveis ou suficientemente confiáveis ou válidos. Os níveis de acurácia diagnóstica geral, sensibilidade e especificidade, valores preditivos positivos e negativos, podem ser classificados em quatro categorias: 90 – 100% = excelente; 80 – 89% = bom; 70 – 79% razoável, e; < 70% = pobre. Essa mesma escala pode ser utilizada para classificar os níveis de concordância observada, utilizada no cálculo das estatísticas kappa, kappa ponderada e ICC. Conseqüentemente, os resultados desses coeficientes, corrigidos para a concordância obtida ao acaso, também podem ser classificados da mesma forma.



### **3 OBJETIVOS**

#### **3.1 OBJETIVO GERAL**

Verificar a validade dos índices multicomponentes DAI (índice de estética dental) e ICON (índice de complexidade, resultado e necessidade) para a necessidade de tratamento ortodôntico para a região de Belo Horizonte, Minas Gerais.

#### **3.2 OBJETIVOS ESPECÍFICOS**

São objetivos específicos desse trabalho:

- a) testar a validade dos índices DAI e ICON;
- b) verificar a sensibilidade, especificidade e a acurácia dos índices DAI e ICON;
- c) comparar a acurácia entre os índices DAI e ICON;
- d) estabelecer os pontos de corte adequados para a utilização dos índices DAI e ICON para necessidade de tratamento, em levantamentos epidemiológicos;
- e) verificar a confiabilidade da opinião do painel de especialistas;
- f) verificar a reprodutibilidade do pesquisador na utilização dos índices.

## 4 METODOLOGIA

O presente estudo foi desenvolvido em duas partes distintas:

- a) parte I – estudo de confiabilidade para calibração do examinador segundo os critérios dos índices DAI e ICON;
- b) parte II – validação dos índices DAI e ICON para a região de Belo Horizonte.

### 4.1 PARTE I: ESTUDO DE CONFIABILIDADE PARA CALIBRAÇÃO DO EXAMINADOR

O examinador pesquisador foi calibrado para os índices oclusais DAI e ICON, utilizados neste estudo, com base nos critérios publicados na literatura de referência (DANIELS & RICHMOND, 2000; JENNY & CONS, 1988; OMS, 1999). A calibração para os dois índices foi realizada em duas etapas principais. Inicialmente para obtenção do padrão de referência, foi feito um estudo dos critérios, um treinamento para sua aplicação e avaliação e, em seguida, o estabelecimento de um padrão de referência para os critérios de medida de cada um dos índices (FIG. 1). Concluindo esta etapa do estudo, foi realizada a calibração do examinador contra estes padrões (FIG. 2).

Como foram realizadas medições para dois índices distintos, tomou-se o cuidado de fazê-las em períodos separados de forma a evitar confusão na fixação dos critérios de cada índice por parte dos examinadores. O intervalo de tempo entre as sessões de aplicação de cada índice foi de, no mínimo 3 semanas, de forma que os examinadores pudessem não se lembrar da avaliação anterior. Entretanto, como os índices DAI e ICON são estruturalmente equivalentes (multicomponentes categóricos com resultado final numérico), a metodologia utilizada foi a mesma para ambos e será descrita de forma única.

Todos os exames e medições foram realizados por apenas um examinador, o autor do trabalho.

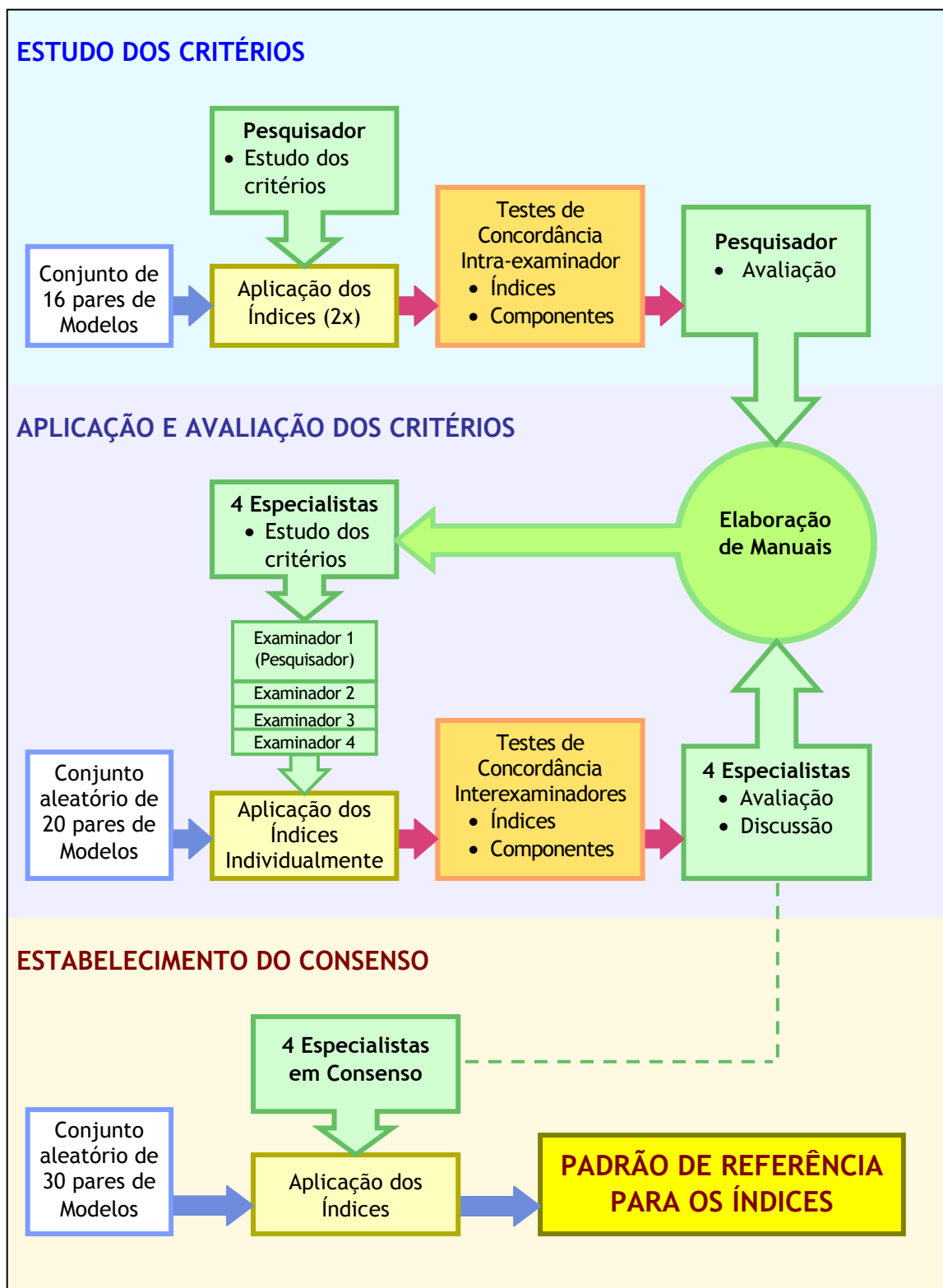


FIGURA 1 - Fluxograma da metodologia da obtenção do padrão de referência.

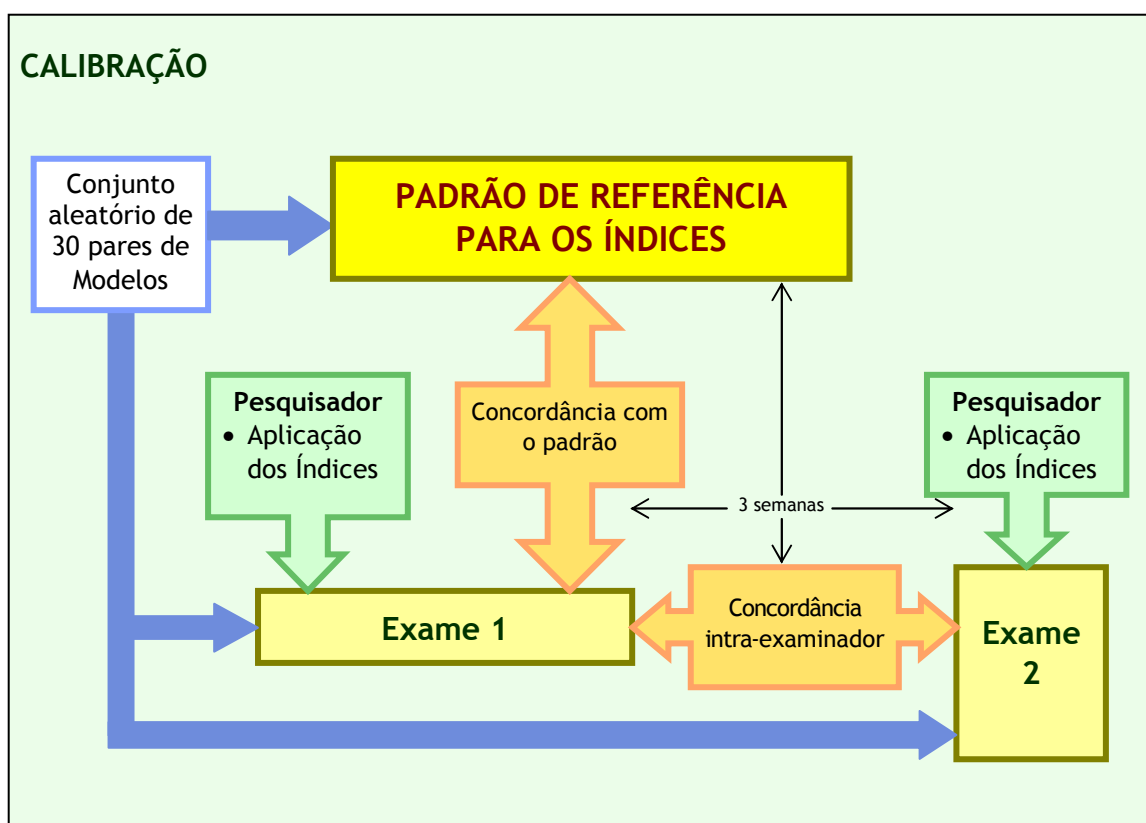


FIGURA 2 - Fluxograma da metodologia da calibração.

#### 4.1.1 ESTUDO DOS CRITÉRIOS

##### 4.1.1.1 MATERIAIS

Os seguintes materiais foram utilizados para o estudo dos critérios:

- conjunto de dezesseis pares de modelos em gesso representando três níveis de necessidade de tratamento ortodôntico;
- escala SCAN (APÊNDICE B);
- sonda periodontal tipo Universidade Carolina do Norte<sup>1</sup>;
- régua milimetrada de aço<sup>2</sup>;
- fichas para anotação dos valores dos índices (APÊNDICES C e D);
- etiquetas adesivas para numerar os modelos de estudo;
- canetas e lápis.

<sup>1</sup> Hu-Friedy® do Brasil. PCPUNC15BR. Catálogo de Produtos, 2ed. Rio de Janeiro. Disponível em <<http://www.hu-friedy.com.br>>

<sup>2</sup> marca Kawasa®, Tóquio, Japão.

Foi realizado um estudo dos critérios de cada um dos índices com a finalidade de verificar a operacionalidade da metodologia e, ao mesmo tempo, servir como treinamento inicial para a utilização dos índices nas próximas etapas do processo.

Um conjunto de 260 pares de modelos, do arquivo da Clínica do Curso de Especialização em Ortodontia da Faculdade de Odontologia da Universidade Federal de Minas Gerais (FO-UFMG) foi classificado segundo os critérios:

- a) escala SCAN (categorias de 1 a 10);
- b) problemas intra-arcos (normal, apinhamento, espaçamento, ausência de elemento dentário, outras anomalias do desenvolvimento dentário);
- c) problemas horizontais ântero-posteriores de forma genérica (classificação de Angle, relação de caninos e trespasse horizontal anterior);
- d) problemas verticais (normal, mordida profunda, mordida aberta);
- e) problemas transversais (normal, mordida cruzada, assimetrias nos arcos);
- f) fase de desenvolvimento da dentição (dentadura mista ou dentição permanente).

Foram excluídos 49 pares de modelos que se enquadraram nos seguintes critérios: modelos na fase de dentição decídua ou período inicial da dentadura mista, com presença de aparelho ortodôntico, síndromes e deformidades dento-faciais e aqueles cuja qualidade compromettesse a avaliação. Restaram 211 pares, cujas frequências das classificações da má oclusão, em relação à escala SCAN, são observadas na TAB. 32, APÊNDICE E.

Foram selecionados dezesseis pares com a seguinte distribuição de necessidade de tratamento ortodôntico: seis sem necessidade de tratamento, cinco no limite entre a não-necessidade e a necessidade e cinco com necessidade de tratamento.

O pesquisador estudou cuidadosamente os critérios publicados para os dois índices e redigiu manuais com os protocolos de aplicação dos mesmos (APÊNDICES A e B).

As medições foram realizadas nos dezesseis pares de modelos, em dias diferentes para cada índice. Após uma semana, os exames foram repetidos. Os valores atribuídos aos componentes dos índices foram anotados em fichas próprias para o cálculo dos valores finais para cada par de modelos.

Foi avaliada a concordância do pesquisador entre o primeiro e o segundo exames para cada um dos índices, através da abordagem gráfica de Bland-Altman. Para verificar quais os componentes apresentavam maior discordância entre os exames, foi feita uma avaliação com o teste kappa para cada um dos componentes de cada um dos índices. As observações e dúvidas do pesquisador foram anotadas para discussão na próxima etapa do estudo de confiabilidade.

O estudo dos critérios, cujo objetivo foi servir como treinamento e teste da metodologia, foi realizado com um número reduzido de casos. A análise dos resultados permitiu a familiarização com a metodologia, sendo inadequada uma análise estatística mais cuidadosa nesta fase.

Através da abordagem gráfica de Bland-Altman foi possível visualizar que o erro sistemático entre as duas avaliações com o DAI era pequeno (média das diferenças = 0,8) e que a diferença entre as duas avaliações realizadas apresentou limites de concordância (erro aleatório) de -11,4 até 13,0. No caso do DAI, os limites de concordância deveriam abranger, no máximo, o intervalo de -4,0 a 4,0.

Uma vantagem desta abordagem é a possibilidade de detectar quais os casos específicos onde o erro foi maior, ao contrário dos índices de concordância que apresentam uma avaliação global e inespecífica. Observou-se que há um caso cuja diferença das medidas entre os exames um e dois foi muito grande. Foi possível identificá-lo como sendo o caso número 4, cuja diferença foi igual a -19,6.

Quando observamos a concordância obtida pelo examinador entre os exames um e dois para os componentes do DAI, observou-se que esta concordância foi maior nos componentes apinhamento anterior ( $\kappa_w=0,805$ ) e diastema na linha média superior ( $\kappa_w=0,803$ ). Mas se o objetivo é melhorar a concordância, deve-se dar mais atenção aos componentes cuja concordância foi mais baixa: espaçamento anterior ( $\kappa_w=0,484$ ) e maior irregularidade anterior na mandíbula ( $\kappa_w=0,600$ ).

Na avaliação da concordância dos valores do ICON, o gráfico de Bland-Altman mostra um erro sistemático de 1,2 para o examinador entre os dois exames. Os limites de concordância calculados foram de -28,2 até 30,6, além do máximo de -18 a 18 recomendado pelos autores.

Os componentes do ICON que obtiveram uma maior concordância entre o primeiro e segundo exames foram a mordida cruzada ( $\kappa_s = 1,000$ ) e o componente estético

( $\kappa_w^2 = 0,857$ ). Os menores valores de concordância foram obtidos pelos componentes vertical ( $\kappa_w = 0,235$ ) e relação ântero-posterior bucal ( $\kappa_w = 0,407$ ).

#### 4.1.2 APLICAÇÃO E AVALIAÇÃO DOS CRITÉRIOS

##### 4.1.2.1 MATERIAIS

Os seguintes materiais foram utilizados para a aplicação e avaliação dos critérios:

- a) conjunto de vinte pares de modelos em gesso representando três níveis de necessidade de tratamento ortodôntico;
- b) escala SCAN (APÊNDICE B);
- c) sonda periodontal tipo Universidade Carolina do Norte®;
- d) régua milimetrada de aço;
- e) fichas para anotação dos valores dos índices (APÊNDICES C e D);
- f) manuais de referência para aplicação dos índices (APÊNDICES A e B);
- g) etiquetas adesivas para numerar os modelos de estudo;
- h) canetas e lápis.

##### 4.1.2.2 SELEÇÃO DO CONJUNTO DE VINTE PARES MODELOS

A partir do mesmo conjunto, agora com 195 pares de modelos de estudo do arquivo Curso de Especialização em Ortodontia da FO-UFMG, classificado pela pontuação da escala SCAN, vinte pares foram selecionados aleatoriamente de forma a abranger os três níveis de necessidade de tratamento ortodôntico (TAB. 33, APÊNDICE F). Esses modelos foram ordenados e nova numeração lhes foi atribuída a partir da geração de números aleatórios em planilha eletrônica<sup>3</sup>.

A aplicação dos critérios de cada um dos índices ao conjunto de vinte modelos de estudo e sua avaliação tiveram a finalidade de familiarizar e treinar quatro examinadores especialistas em ortodontia, incluindo-se o pesquisador. Esta etapa possibilitou identificar

---

<sup>3</sup> Microsoft® Office Excel 2003.

os pontos de discordância nas medidas realizadas, além levantar dúvidas e identificar os problemas de compreensão dos critérios entre os examinadores.

#### **4.1.2.3 ESTUDO DOS CRITÉRIOS PELO GRUPO DE EXAMINADORES**

Foram convidados a participar, como voluntários desta fase do estudo, três cirurgiões dentistas especialistas em ortodontia, vinculados a cursos de pós graduação ou extensão da FO-UFG, segundo o critério de disponibilidade.

Esse grupo de ortodontistas e o pesquisador estudaram, individualmente, as publicações de referência e os manuais elaborados pelo pesquisador para cada índice.

#### **4.1.2.4 APLICAÇÃO DOS ÍNDICES**

Em sessões de avaliação agendadas, em semanas separadas para cada índice, os examinadores aplicaram, individualmente, os critérios aos vinte pares de modelos numerados. Os valores dos componentes dos índices foram anotados em fichas específicas numeradas de acordo com os modelos examinados (APÊNDICES C e D). As sessões de avaliação duraram o tempo aproximado de quatro horas e cada examinador trabalhou no seu próprio ritmo, com intervalos para descanso. Ao final das sessões de avaliação foram discutidas e anotadas as dúvidas em relação à aplicação de determinados critérios. Em conjunto, foram consultadas as publicações de referência de forma a esclarecer dúvidas e aprimorar a redação dos manuais para aplicação dos índices.

Foi avaliada a concordância interexaminadores para os valores dos índices e de seus subcomponentes, para os seis pares de examinadores:  $1 \times 2$ ,  $1 \times 3$ ,  $1 \times 4$ ,  $2 \times 3$ ,  $2 \times 4$  e  $3 \times 4$ . Foram identificadas as principais discordâncias. Os pares de modelos que tiveram suas pontuações mais discordantes e os componentes de cada índice com valores mais baixos de concordância foram listados para discussão com o grupo de examinadores, para fixar os critérios de maneira uniforme e consensual e definir o conteúdo final dos manuais de referência.



### 4.1.3 ESTABELECIMENTO DO CONSENSO PARA OS CRITÉRIOS DE MEDIDA

#### 4.1.3.1 MATERIAIS

Os seguintes materiais foram utilizados para o estabelecimento do consenso para os critérios de medida:

- a) conjunto de trinta pares de modelos em gesso estratificados em dez níveis de necessidade de tratamento ortodôntico;
- b) escala SCAN (APÊNDICE B);
- c) sonda periodontal tipo Universidade Carolina do Norte®;
- d) régua milimetrada de aço;
- e) fichas para anotação dos valores dos índices (APÊNDICES C e D);
- f) manuais de referência para aplicação dos índices (APÊNDICES A e B);
- g) etiquetas adesivas para numerar os modelos de estudo;
- h) canetas e lápis.

#### 4.1.3.2 SELEÇÃO DO CONJUNTO DE TRINTA PARES DE MODELOS

A partir do conjunto remanescente de 175 pares de modelos, já classificados, do arquivo do Curso de Especialização em Ortodontia da FO-UFMG, trinta pares foram selecionados de forma a representar uma ampla variedade de tipos e gravidades de más oclusões na dentadura mista e dentição permanente, cuja frequência em relação à escala SCAN é apresentada na TAB. 34, APÊNDICE G.

O grupo de examinadores reuniu-se em duas sessões de avaliação, uma para cada índice, quando foram aplicados os critérios de cada um de seus componentes, de forma consensual, aos trinta pares de modelos. Os valores dos índices, assim atribuídos aos modelos, foram considerados como os padrões de referência contra os quais a calibração do examinador foi realizada.

#### 4.1.4 CALIBRAÇÃO

##### 4.1.4.1 MATERIAIS

Os seguintes materiais foram utilizados para a calibração:

- a) conjunto de trinta pares de modelos em gesso estratificados em dez níveis de gravidade da má oclusão;
- b) escala SCAN (APÊNDICE B);
- c) sonda periodontal tipo Universidade Carolina do Norte®;
- d) régua milimetrada de aço;
- e) fichas para anotação dos valores dos índices (APÊNDICES C e D);
- f) manuais de referência para aplicação dos índices (APÊNDICES A e B);
- g) etiquetas adesivas para numerar os modelos de estudo;
- h) canetas e lápis.

O pesquisador mediu os componentes (exame 1) conforme os critérios definidos anteriormente para cada índice nos mesmos trinta pares de modelos, três semanas após os padrões de referência terem sido estabelecidos. Os valores medidos foram registrados em fichas próprias (APÊNDICES C e D) e tabulados em planilha eletrônica para análise.

A concordância dos resultados obtidos pelo pesquisador, para o DAI e ICON, em relação ao padrão de referência foi analisada através do ICC e da abordagem gráfica de Bland-Altman.

As medições dos índices foram repetidas pelo pesquisador nos mesmos trinta pares de modelos, numa segunda sessão de calibração (exame 2), realizada três semanas após a primeira. Os procedimentos foram os mesmos empregados anteriormente com a finalidade de avaliar a concordância intraexaminador.

## 4.2 PARTE II: VALIDAÇÃO DOS ÍNDICES

A validação dos índices para necessidade de tratamento ortodôntico foi realizada em duas etapas principais: a obtenção do padrão-ouro numa escala nominal dicotômica e a medição dos índices, cujos resultados são escalas numéricas. Ambas foram

realizadas utilizando um mesmo conjunto de 108 pares de modelos de estudo. Os índices foram medidos em cada modelo pelo pesquisador calibrado e seus valores foram comparados ao padrão-ouro. Esse foi obtido a partir da opinião de vinte especialistas sobre a necessidade de tratamento ortodôntico (FIG. 3).

Da mesma forma como na calibração, a metodologia empregada para a validação de ambos os índices foi a mesma.

#### 4.2.1 MATERIAIS

Os seguintes materiais foram utilizados para a validação dos índices:

- a) conjunto de 108 pares de modelos em gesso numerados estratificados em dez níveis de gravidade da má oclusão;
- b) subconjunto de quarenta pares de modelos em gesso;
- c) escala SCAN (APÊNDICE B);
- d) sonda periodontal tipo Universidade Carolina do Norte®;
- e) régua milimetrada de aço;
- f) fichas para anotação dos valores dos índices (APÊNDICES C e D);
- g) orientação para avaliação de modelos por especialistas (APÊNDICE H);
- h) fichas numeradas para anotação dos valores referentes à opinião de necessidade de tratamento atribuídos aos modelos (APÊNDICE I);
- i) fichas para anotação dos valores atribuídos ao “ponto de tratamento indicado” (APÊNDICE J);
- j) etiquetas adesivas para numerar os modelos de estudo;
- k) canetas e lápis.



FIGURA 3 - Fluxograma da metodologia da validação.

## 4.2.2 SELEÇÃO DO CONJUNTO DE MODELOS DE ESTUDO

### 4.2.2.1 AMOSTRAGEM

Considerando a concordância de  $\kappa = 0,81$  ( $\sigma^2 = 0,15$ ) obtida pelo estudo de validação do ICON por Firestone *et al.* (2002a), uma estimativa do tamanho amostral ( $\alpha = 0,05$ ;  $\beta = 0,25$ ;  $\delta^2 = 0,01$ ) obteve um valor de 108 modelos de estudo utilizando a fórmula (MONTGOMERY, 2003):

$$n \cong \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2} \quad (\text{EQUAÇÃO 13}).$$

Do arquivo da Disciplina de Ortodontia da FO-UFGM, com 445 pares de modelos de estudo, foram selecionados 108 casos de acordo com os critérios de inclusão e exclusão, anteriormente citados

Os 445 pares modelos foram pontuados conforme a escala SCAN e categorizados pelos tipos de dentição e de má oclusão classificadas nos sentidos horizontal, vertical, transversal e em relação à discrepância de espaço nos arcos, da mesma forma como foi relatado anteriormente (TAB. 35, APÊNDICE L). De forma que o conjunto final fosse representativo de uma ampla variedade de tipos e gravidades de más oclusões, as categorias da escala SCAN deveriam conter variedades de tipo de dentição, bem como uma ampla variedade de tipos de má oclusão (TAB. 36, APÊNDICE M).

Os pares de modelos foram distribuídos nas categorias da escala SCAN, representando dez níveis de gravidade, conforme suas classificações de tipo dentição e de má oclusão. A quantidade de modelos variou entre as categorias de gravidade de forma a haver um maior número de modelos nas categorias intermediárias (GRÁFICO 21, APÊNDICE M). Isso porque os casos classificados nas extremidades da escala de necessidade de tratamento (necessidade máxima ou nenhuma necessidade) não testariam a capacidade discriminatória dos índices. As categorias de gravidade cinco e seis continham 13 pares de modelos cada, enquanto as categorias um e dez, seis pares cada. Foi gerada uma nova numeração aleatória em planilha eletrônica e todos os pares de modelos foram recodificados numericamente e etiquetados.

Com a finalidade de se avaliar a concordância intra-examinadores, um subconjunto de quarenta pares de modelos foi aleatoriamente selecionado para ser

reavaliado pelos especialistas e pelo pesquisador. Os exames foram repetidos respeitando um intervalo mínimo de três semanas entre as sessões para cada índice. Através da geração de números aleatórios, a numeração original do conjunto de 108 pares de modelos foi reordenada e os quarenta primeiros modelos nessa nova ordem foram escolhidos.

#### **4.2.3 OBTENÇÃO DO PADRÃO-OURO**

A opinião de um painel formado por ortodontistas foi a referência do “padrão ouro” sobre a indicação ou não para tratamento ortodôntico, para a validação dos índices DAI e ICON. O painel foi constituído por vinte mestres em ortodontia de Belo Horizonte que satisfizeram os seguintes critérios:

- a) ter título de mestre ou doutor registrado no Ministério da Educação;
- b) ser professor vinculado a uma Universidade sediada em Belo Horizonte;
- c) ter, no mínimo, cinco anos de experiência na clínica privada em ortodontia, em Belo Horizonte.

Os especialistas que satisfizeram critérios de inclusão, após contato oficial com as respectivas instituições de ensino superior, foram contatados por telefone e solicitados a participar voluntariamente do estudo. Foi feita uma apresentação do trabalho, seus objetivos e a explicação da forma de participação do colaborador e do Termo de Consentimento Livre e Esclarecido, que foi prontamente assinado por todos. Foi dada ênfase na necessidade de dedicação e exatidão no registro das opiniões solicitadas. Deveriam ter disponibilidade para comparecer às datas definidas para as avaliações inicial e final dos modelos de estudo.

Dos vinte especialistas participantes, seis eram professores na FO-UFMG e quatorze na Faculdade de Odontologia da Pontifícia Universidade Católica de Minas Gerais.

##### **4.2.3.1 OPINIÃO DO PAINEL DE ESPECIALISTAS**

Para coletar a opinião do painel de especialistas sobre a necessidade de tratamento ortodôntico foram avaliados todos os 108 pares de modelos do conjunto.

As avaliações foram agendadas por conveniência nos dias em que os professores estavam nas Universidades. Foram realizadas de uma até quatro sessões, dependendo do dia da semana e do tempo disponível no cronograma de atividades dos

professores. Os modelos foram dispostos de forma ordenada, segundo a numeração atribuída aleatoriamente em espaços cedidos pelas universidades, de forma a facilitar o acesso dos especialistas participantes.

As fichas de avaliação continham o número correspondente ao caso, a identificação em código do ortodontista, as orientações necessárias e uma escala numérica para ser marcada (APÊNDICE I).

Os especialistas foram instruídos, conforme os APÊNDICES H e I, a enquadrar a oclusão de cada par de modelos na escala tipo Likert com sete categorias, na qual o valor 1 significa “nenhuma ou mínima indicação para tratamento” e o valor 7 equivale à “indicação muito grande para tratamento”. Foram informados que deveriam trabalhar em seu próprio ritmo, sem limite de tempo (B EGLIN, 2001; FIRESTONE, 2002a; YOUNIS, 1997).

Cada modelo de estudo obteve uma pontuação única obtida pela média dos valores atribuídos por todos os especialistas conforme a escala de sete pontos.

Para obter o padrão-ouro, esta opinião média dos ortodontistas sobre a necessidade de tratamento ortodôntico para cada modelo foi dicotomizada nas categorias “sem necessidade” e “com necessidade”.

Com o objetivo de avaliar a concordância intra-examinadores, um subconjunto de quarenta modelos foi reavaliado pelo mesmo painel de especialistas, num intervalo de quatro semanas após a realização do primeiro exame, empregando a mesma metodologia já descrita.

Em seguida, foi solicitado que cada especialista utilizasse uma escala de sete pontos, idêntica à usada anteriormente (APÊNDICE J), para definir o valor no qual ou acima do qual ele acredita que o tratamento ortodôntico seria indicado. O “ponto de tratamento indicado” (PTI) foi obtido para cada um dos ortodontistas. A média dos vinte valores de PTI foi utilizada como ponto de corte único para dicotomizar a opinião média referente a cada um dos modelos, refletindo assim a opinião dos especialistas de forma dicotômica.

Os modelos cujas médias de opinião apresentaram valores menores que o PTI médio, foram classificados como “tratamento não indicado”, e os com médias iguais ou maiores que o PTI foram classificados como “tratamento indicado”.

#### 4.2.4 APLICAÇÃO DOS ÍNDICES PELO PESQUISADOR

O pesquisador calibrado aplicou os critérios estabelecidos para os índices ICON e DAI e fez a medição dos seus componentes, no conjunto de 108 pares de modelos, seis semanas após a última sessão de calibração, para cada índice. Os valores medidos em cada modelo foram anotados em fichas específicas (APÊNDICES C e D).

Após um intervalo de três semanas, o pesquisador repetiu a medição dos índices no subconjunto de 40 modelos, utilizando a mesma metodologia anteriormente descrita, com o objetivo de avaliar a concordância intra-examinador.

### 4.3 ANÁLISE DE DADOS

No estudo de confiabilidade, a concordância intra e interexaminadores foram avaliadas para os valores dos índices e para cada um dos seus subcomponentes separadamente, utilizando análises adequadas para cada tipo de variável.

As variáveis quantitativas ou de escala de intervalo tiveram a sua concordância avaliada através da abordagem gráfica de Bland-Altman e pelo ICC. Já as variáveis qualitativas categóricas ordinais foram avaliadas pela estatística kappa ponderada com ponderação quadrática, para variáveis com mais de cinco categorias, e ponderação linear para variáveis com cinco ou menos categorias. As variáveis categóricas dicotômicas foram avaliadas pela estatística kappa simples.

A abordagem de Bland-Altman tem a finalidade de elucidar os tipos de erros e as causas das discordâncias dentro de cada par de medidas realizadas num mesmo objeto. As diferenças dos valores atribuídos a cada par de modelos pelos examinadores foram traçadas num gráfico de dispersão contra as médias desses mesmos valores. As diferenças traçadas contra as médias de cada par de modelos permitem visualizar a distribuição do erro aleatório e localizar os casos específicos para cada valor do erro. As médias das diferenças representam o erro sistemático em relação à diferença zero (concordância absoluta). Foram calculados limites de confiança de 95% para avaliar a precisão do erro sistemático estimado e limites de concordância de 95% para avaliar a distribuição provável dos erros aleatórios em torno do valor médio. Os pressupostos da distribuição normal do erro aleatório e da independência do mesmo em relação às médias foram verificados e as correções para o cálculo dos limites de concordância implementadas, quando indicadas.



Os coeficientes de concordância para os valores dos índices e para os seus componentes, obtidos a partir dos testes ICC e kappa, respectivamente, tiveram seus resultados acompanhados por seus limites inferiores do intervalo de confiança de 95%. Esse limite inferior do intervalo de confiança é o menor resultado possível para o coeficiente, a partir do qual será avaliada a significância clínica.

Para avaliar os níveis de concordância, foi utilizada a escala proposta por Fleiss (1981) e modificada por Cicchetti (1992; 2001), dada a sua equivalência na avaliação dos níveis de concordância, tanto para o modelo de sensibilidade e especificidade quanto para os coeficientes de concordância.

Os níveis de concordância foram codificados em cores para facilitar a sua visualização (QUADRO 1).

#### QUADRO 1

Significância clínica para valores de kappa, ICC e concordância observada

| Valores de kappa ( $\kappa_s$ ),<br>kappa ponderado ( $\kappa_w$ )<br>ou ICC ( $r_i$ ) | Valores de<br>concordância<br>observada (%) | Níveis de Significância<br>Clínica ou Prática |
|--|---|---|
| < 0,40   | < 70  | Pobre   |
| 0,40 a 0,59  | 70 a 79                                     | Razoável                                      |
| 0,60 a 0,74  | 80 a 89                                     | Boa   |
| 0,75 a 1,00  | 90 a 100                                    | Excelente                                     |

FONTE - CICHETTI (2001) p. 697.

Os valores dos índices, atribuídos pelo pesquisador aos modelos, foram comparados com o padrão-ouro. Foram avaliadas a especificidade, a sensibilidade e a acurácia. Os valores preditivos positivos e negativos foram calculados a partir dos valores de sensibilidade e especificidade dos índices e da prevalência da necessidade de tratamento ortodôntico na população local.

Novos pontos de corte foram obtidos para cada índice, ajustados para os valores de sensibilidade e especificidade. A razão de verossimilhança positiva foi calculada para os dois índices, utilizando os pontos de corte originais recomendados pelos autores e os pontos de corte ajustados. Suas respectivas sensibilidade, especificidade, valores e preditivos positivos e negativos foram comparados de forma a estabelecer qual o ponto de corte mais adequado para utilização em levantamentos epidemiológicos. As comparações entre os índices foram realizadas utilizando-se os pontos de corte mais adequados.

A acurácia de cada um dos índices foi avaliada através da análise da curva ROC e os seus desempenhos foram comparados através da verificação da significância estatística da diferença entre as áreas das duas curvas ROC pelo método não paramétrico proposto por DeLong *et al.* (1988).

Foi testada a confiabilidade para os especialistas do painel e para o pesquisador. A concordância interexaminadores, por amostragem, em vinte combinações de pares de especialistas definidos aleatoriamente, e a concordância intra-examinadores para os vinte especialistas, foram verificadas por meio da estatística kappa ponderada. A confiabilidade intra-examinador do pesquisador para os valores dos índices foi testada através do ICC e da abordagem gráfica de Bland-Altman.

Todos os valores medidos foram tabulados em planilha eletrônica<sup>3</sup> para fins de análise estatística. Os dados digitados foram conferidos com as fichas originais com auxílio de um programa leitor de texto digital<sup>4</sup>. Os cálculos foram realizados na própria planilha ou em programa estatístico<sup>5</sup>.

Os seguintes cálculos e gráficos foram realizados em planilha eletrônica:

- a) valores de ICC e seus limites inferiores do intervalo de confiança de 95%;
- b) diferenças, médias, média das diferenças, intervalos de confiança de 95% da média das diferenças, limites de concordância de 95%;
- c) gráficos de Bland-Altman;
- d) razões de verossimilhança;
- e) valores preditivos positivos e negativos.

Os cálculos e gráficos abaixo foram realizados em programa estatístico:

- a) correlação de Spearman;
- b) regressão linear simples;
- c) valores de curtose e distorção;
- d) histogramas da distribuição dos erros;
- e) cálculos e gráficos da abordagem de Bland-Altman;
- f) valores de sensibilidade e especificidade;
- g) áreas sob a curva ROC e o gráfico da curva ROC;
- h) comparação da diferença entre duas curvas ROC.

---

<sup>4</sup> Free Natural Reader: software livre. Disponível em: <<http://www.naturalreaders.com/>>. Acesso em 26 nov. 2005.

<sup>5</sup> STATA®/SE for Windows®. Versão 8.0. College Station, Texas, 2003. Disponível em: <<http://www.stata.com>>.

#### 4.4 ASPECTOS ÉTICOS

O presente estudo, bem como o Termo de Consentimento Livre e Esclarecido foram aprovados pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, conforme Parecer nº ETIC 068/06 (APÊNDICE N e ANEXO A).

Os modelos de estudos utilizados neste estudo pertencem ao Departamento de Odontopediatria e Ortodontia e sua utilização foi autorizada (ANEXO B).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 ESTUDO DE CONFIABILIDADE PARA CALIBRAÇÃO DO EXAMINADOR

#### 5.1.1 DAI

##### 5.1.1.1 APLICAÇÃO E AVALIAÇÃO DOS CRITÉRIOS

Os resultados da concordância obtida entre os seis pares possíveis de examinadores para os valores do DAI em vinte pares de modelos foram avaliados através do ICC e são apresentados na TAB. 1.

TABELA 1

Resultados dos níveis de concordância dos pares de examinadores para os valores do DAI em vinte pares de modelos

| Pares de Examinadores | ICC (Limite Inferior do Intervalo de Confiança de 95%) |
|-----------------------|--|
| Ex. 1 x Ex. 2         | 0,865 (0,713)  |
| Ex. 1 x Ex. 3         | 0,972 (0,941)  |
| Ex. 1 x Ex. 4         | 0,977 (0,951)  |
| Ex. 2 x Ex. 3         | 0,851 (0,684)  |
| Ex. 2 x Ex. 4         | 0,871 (0,727)  |
| Ex. 3 x Ex. 4         | 0,982 (0,961)  |

Observou-se que o nível de concordância foi bom para os pares  $1 \times 2$ ,  $2 \times 3$  e  $2 \times 4$ ; e excelente para os pares  $1 \times 3$ ,  $1 \times 4$  e  $3 \times 4$ . Nota-se que os pares envolvendo o examinador 2 apresentaram menor concordância.

A TAB. 2 mostra os resultados do erro sistemático e seus limites de confiança de 95% e dos limites de concordância de 95%, utilizados nos gráficos de Bland & Altman (GRÁFICO 1).

O erro sistemático é relativo aos dois examinadores de cada par. Os maiores valores foram observados nos pares formados pelo examinador 2, cujos limites de confiança de 95% também foram maiores.

TABELA 2

Resultados do erro sistemático e seus limites de confiança e dos limites de concordância para os valores do DAI em vinte pares de modelos

| <b>Pares de Examinadores</b> | <b>Erro Sistemático</b> | <b>Intervalo de Confiança (95%) do Erro Sistemático</b> |     | <b>Limites de Concordância (95%)</b> |      |
|------------------------------|-------------------------|---|-----|--------------------------------------|------|
| Ex. 1 x Ex. 2                | 1,8                     | -0,8  | 4,4 | -9,7                                 | 13,3 |
| Ex. 1 x Ex. 3                | -0,5                    | -1,6  | 0,7 | -5,7                                 | 4,8  |
| Ex. 1 x Ex. 4                | 0,1                     | -1,0  | 1,1 | -4,7                                 | 4,8  |
| Ex. 2 x Ex. 3                | -2,3                    | -5,1  | 0,6 | -14,8                                | 10,3 |
| Ex. 2 x Ex. 4                | -1,8                    | -4,3  | 0,8 | -13,2                                | 9,7  |
| Ex. 3 x Ex. 4                | 0,5                     | -0,5  | 1,5 | -3,8                                 | 4,8  |

As concordâncias também foram menores nos pares formados pelo examinador 2, cujos limites de concordância atingiram os maiores valores absolutos.

Os casos de números 1, 3, 4, 14, 16, 18 e 20, que apresentaram maiores diferenças, foram identificados e separados para que os examinadores pudessem investigar os motivos das discordâncias.

Os pressupostos da distribuição normal do erro e da independência entre a diferença e a média dos valores para o cálculo dos limites de concordância não foram verificados nessa etapa do estudo, pois no momento não interessou avaliar os níveis específicos de concordância. O objetivo principal foi a comparação desses níveis para a identificação dos pontos de discordância e suas origens.

Foi possível identificar uma provável fonte de erro: a interpretação dos critérios pelo examinador 2. Porém isso não explica a totalidade dos erros. O erro sistemático, os erros aleatórios e os níveis de concordância dos outros pares de examinadores foram menores, mas ainda evidenciaram falta de concordância na medição do DAI.

Foram investigadas as concordâncias para as variáveis que compõem o valor final do índice, de forma isolada. Os resultados da concordância dos seis pares de examinadores para cada um dos dez componentes do DAI, avaliados pela estatística kappa, são apresentadas na TAB. 3.

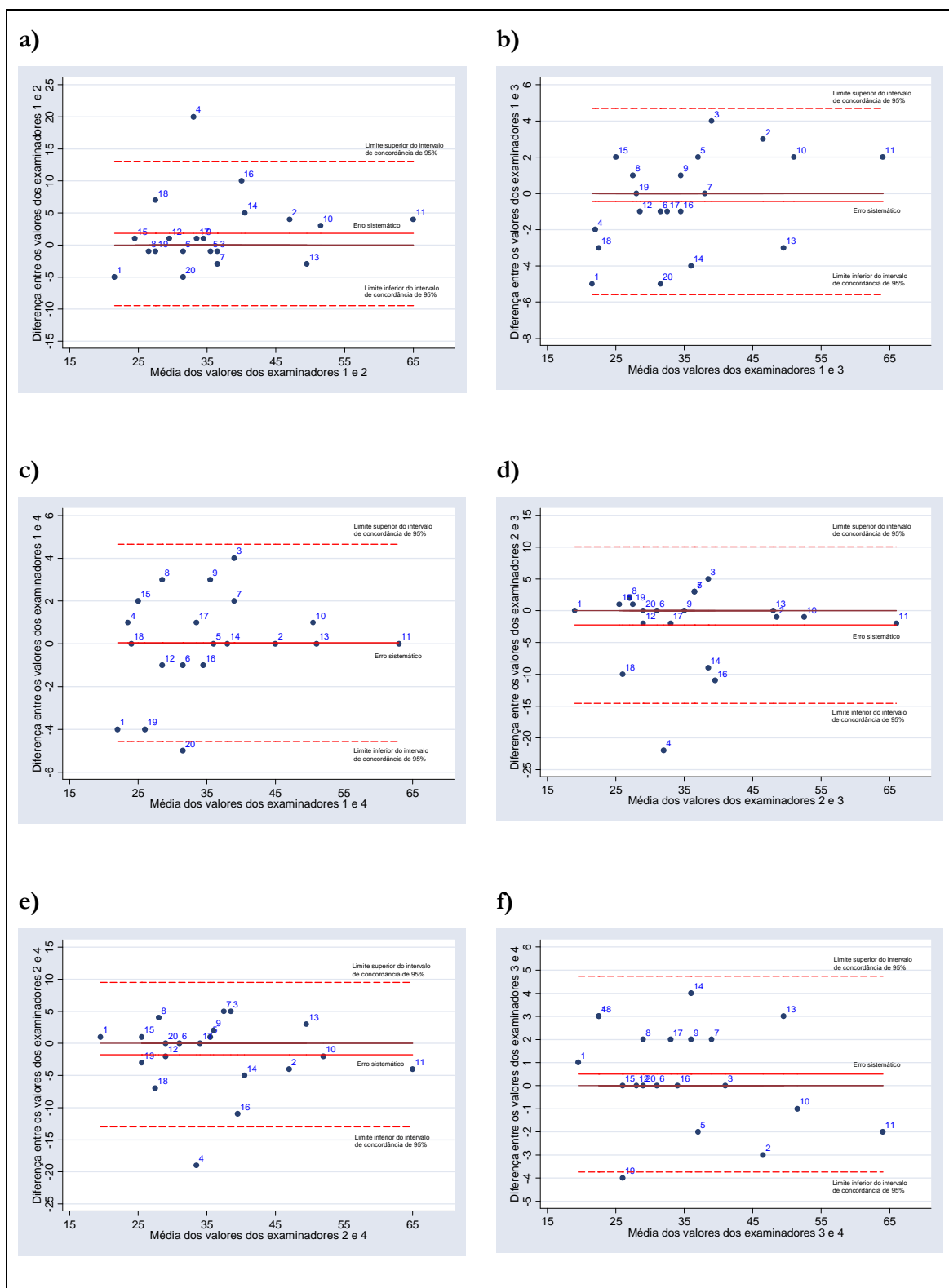


GRÁFICO 1 – Gráficos de Bland-Altman das diferenças versus as médias entre os pares de examinadores para o índice DAI em vinte pares de modelos: a) examinadores 1 e 2; b) examinadores 1 e 3; c) examinadores 1 e 4; d) examinadores 2 e 3; e) examinadores 2 e 4; f) examinadores 3 e 4

TABELA 3

Concordância dos pares de examinadores para os componentes do DAI em vinte pares de modelos

| Pares de Examinadores | Componentes do DAI e ponderação do coeficiente kappa |                          |             |                          |            |                          |            |                          |                        |                          |
|-----------------------|--|--------------------------|-------------|--------------------------|------------|--------------------------|------------|--------------------------|------------------------|--------------------------|
|                       | Ausentes   |                          | Apinhamento |                          | Espaços    |                          | Diastema   |                          | Irregularidade Maxilar |                          |
|                       | $\kappa_w$   | Limite inferior (IC 95%) | $\kappa_w$  | Limite inferior (IC 95%) | $\kappa_w$ | Limite inferior (IC 95%) | $\kappa_w$ | Limite inferior (IC 95%) | $\kappa_w^2$           | Limite inferior (IC 95%) |
| 1 x 2                 | 0,628  | 0,241                    | 0,792       | 0,596                    | 0,608      | 0,288                    | 0,716      | 0,591                    | 0,498                  | 0,206                    |
| 1 x 3                 | 0,931  | 0,842                    | 0,897       | 0,746                    | 0,675      | 0,375                    | 0,755      | 0,555                    | 0,457                  | 0,147                    |
| 1 x 4                 | 0,931  | 0,806                    | 0,845       | 0,688                    | 0,793      | 0,501                    | 1,000      | 1,000                    | 0,791                  | 0,681                    |
| 2 x 3                 | 0,674  | 0,266                    | 0,785       | 0,583                    | 0,938      | 0,802                    | 0,847      | 0,675                    | 0,976                  | 0,956                    |
| 2 x 4                 | 0,674  | 0,285                    | 0,734       | 0,522                    | 0,682      | 0,397                    | 0,716      | 0,589                    | 0,639                  | 0,451                    |
| 3 x 4                 | 1,000  | 1,000                    | 0,843       | 0,666                    | 0,744      | 0,455                    | 0,755      | 0,563                    | 0,562                  | 0,361                    |

| Pares de Examinadores | Componentes do DAI e ponderação do coeficiente kappa |                          |                   |                          |                      |                          |                |                          |                          |                          |
|-----------------------|--|--------------------------|-------------------|--------------------------|----------------------|--------------------------|----------------|--------------------------|--------------------------|--------------------------|
|                       | Irregularidade Mandibular                            |                          | Trespasse Maxilar |                          | Trespasse Mandibular |                          | Mordida Aberta |                          | Relação Ântero-posterior |                          |
|                       | $\kappa_w^2$   | Limite inferior (IC 95%) | $\kappa_w^2$      | Limite inferior (IC 95%) | $\kappa_w^2$         | Limite inferior (IC 95%) | $\kappa_s$     | Limite inferior (IC 95%) | $\kappa_w$               | Limite inferior (IC 95%) |
| 1 x 2                 | 0,765  | 0,661                    | 0,971             | 0,953                    | 0,966                | 0,942                    | 1,000          | 1,000                    | 0,459                    | 0,161                    |
| 1 x 3                 | 0,836  | 0,671                    | 0,953             | 0,927                    | 1,000                | 1,000                    | 1,000          | 1,000                    | 0,604                    | 0,219                    |
| 1 x 4                 | 0,946  | 0,893                    | 0,976             | 0,957                    | 1,000                | 1,000                    | 1,000          | 1,000                    | 0,625                    | 0,173                    |
| 2 x 3                 | 0,864  | 0,804                    | 0,979             | 0,964                    | 0,966                | 0,942                    | 1,000          | 1,000                    | 0,487                    | 0,166                    |
| 2 x 4                 | 0,798  | 0,697                    | 0,967             | 0,948                    | 0,966                | 0,941                    | 1,000          | 1,000                    | 0,375                    | 0,095                    |
| 3 x 4                 | 0,859  | 0,713                    | 0,970             | 0,958                    | 1,000                | 1,000                    | 1,000          | 1,000                    | 0,688                    | 0,317                    |

Observou-se que os níveis de concordância não são semelhantes entre os componentes do índice. Enquanto alguns componentes apresentaram níveis de concordância excelentes, para todos os pares de examinadores, outros tiveram valores no extremo mais baixo da escala. Alguns componentes apresentam heterogeneidade nos níveis de concordância, contrastando valores pobres, razoáveis, bons ou excelentes entre os pares.

Os resultados foram discutidos pelos examinadores e concluiu-se que os componentes cujos níveis de concordância foram mais baixos, de forma homogênea, foram os que apresentaram problemas comuns de interpretação dos critérios, ou dificuldade na execução da medida. Aqueles componentes cujos níveis de concordância apresentaram-se heterogêneos foram discutidos par a par, de forma a fixar os critérios uniformemente.

Foi constatado erro de interpretação no critério do componente “dentes visíveis ausentes” por parte do examinador 2. Na equação do DAI, esse componente é multiplicado por um coeficiente de 5,76, o que torna cada unidade de erro com maior potencial de alterar o valor final do índice. O componente “relação ântero-posterior”

apresentou os piores níveis de concordância para todos os pares de examinadores. Esse fato foi inesperado, pois o componente representa uma classificação comum e convencional do relacionamento entre os primeiros molares antagonistas. Foi constatado que, embora o critério fosse claro ao estabelecer os valores para cada situação de intercuspidação, o julgamento do significado clínico interferiu nas medições dos examinadores 2 e 4. Outra fonte de erro para as medidas desse componente foi a falta de padronização do ângulo de visão, causando um erro de paralaxe. Todos esses aspectos foram discutidos e o critério foi unificado.

A medição do componente “espaços” apresentou erro de interpretação em alguns modelos especificamente, o que foi devidamente exposto e discutido. Os componentes que foram medidos com a sonda periodontal, especialmente a “irregularidade maxilar” apresentaram problemas de padronização no posicionamento do instrumento e no arredondamento do valor obtido. Enquanto um examinador arredondou para o número inteiro mais próximo, conforme estabelecido pelos critérios, os outros três arredondaram, sistematicamente, para um número inteiro mais baixo.

### 5.1.1.2 CALIBRAÇÃO DO PESQUISADOR

O resultado do ICC ( $r_1 = 0,989$ ;  $r_1 \geq 0,980$ ) para concordância entre o padrão de referência do DAI e as medidas realizadas pelo pesquisador no conjunto de trinta pares de modelos é apresentado na TAB. 4. A concordância pode ser considerada excelente entre o valor atribuído pelo pesquisador, e o padrão de referência obtido por consenso.

TABELA 4

Concordância entre o pesquisador e o padrão de referência do DAI em trinta pares de modelos de estudo

| Pares de Examinadores | ICC (Limite Inferior do Intervalo de Confiança de 95%) |
|-----------------------|--|
| Padrão x pesquisador  | 0,989 (0,980)  |

O histograma da distribuição das diferenças entre as medidas realizadas pelo pesquisador e o padrão de referência não apresentou características de normalidade (GRÁFICO 2). Os valores de distorção (0,92;  $p = 0,028$ ) e de curtose (5,09;  $p = 0,023$ ) evidenciaram a falta de normalidade na distribuição dos erros ( $p = 0,0145$ ), possivelmente



causada por valores extremos, o que não é um problema grave, mas pode também significar que há associação entre as diferenças e as médias.

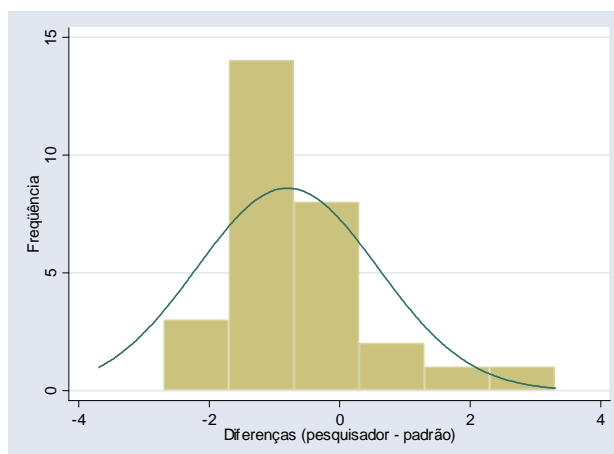


GRÁFICO 2 – Histograma das diferenças das medidas entre o pesquisador e o padrão de referência para o índice DAI

A possível associação entre as diferenças e as médias foi verificada. O coeficiente de correlação de Spearman ( $r_s = 0,0553$ ;  $p = 0,7718$ ) entre os módulos das diferenças e as médias confirma a baixa correlação entre as variáveis. O resultado da análise de regressão linear simples entre as diferenças e as médias sugere que não há dependência significativa entre essas variáveis ( $p = 0,283$ ).

Como o pressuposto da independência das variáveis foi atendido, os limites de concordância e seus intervalos de confiança foram calculados através da abordagem de Bland-Altman padrão, sem a necessidade de transformação dos dados ou correção dos limites pela análise de regressão dos módulos dos resíduos com a média.

O diagrama de dispersão entre as diferenças e as médias de cada observação permite uma verificação visual da relação entre as mesmas. O diagrama de dispersão entre os valores do pesquisador e do padrão mostra que os pontos não se afastam muito da linha de concordância e que há tendência a um desvio para um dos lados da linha, sugerindo um erro sistemático (GRÁFICO 3).

O erro sistemático observado foi de  $-0,7997$ , com intervalo de confiança de 95% entre  $-1,3191$  e  $-0,2803$ . O intervalo de concordância de 95% variou de  $-3,5259$  a

1,9266. Portanto, os limites de concordância apresentaram-se dentro dos valores preconizados pelos autores do índice (entre -4 e 4).

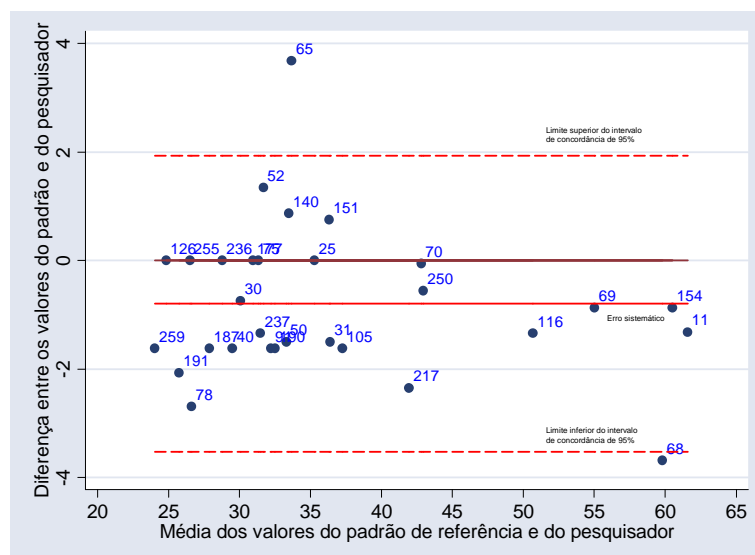


GRÁFICO 3 – Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas pelo pesquisador e o padrão de referência para o índice DAI, em trinta pares de modelos

### 5.1.1.3 CONCORDÂNCIA INTRA-EXAMINADOR

A concordância intra-examinador para os valores do DAI atribuídos pelo pesquisador, em dois exames repetidos, em trinta pares de modelos é apresentada na TAB. 5. Conforme o resultado do ICC ( $r_1 = 0,988$ ;  $r_1 \geq 0,978$ ), o nível de concordância foi considerado excelente.

TABELA 5

Concordância intra-examinador do pesquisador para o índice DAI em trinta pares de modelos

| Exames Repetidos pelo Pesquisador | ICC (Limite Inferior do Intervalo de Confiança (95%)) |
|-----------------------------------|---|
| 1 x 2                             | 0,988 (0,978)   |

O histograma da distribuição das diferenças entre as medidas realizadas pelo pesquisador em exames repetidos com intervalo de 3 semanas, apresentou características de

normalidade (GRÁFICO 4). O valor de distorção (-0,15;  $p = 0,699$ ) e de curtose (2,94;  $p = 0,665$ ) sugerem a normalidade na distribuição dos erros ( $p = 0,8452$ ).

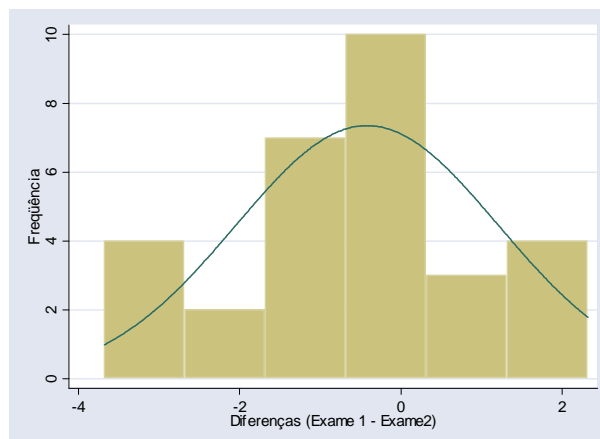


GRÁFICO 4 – Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice DAI, em trinta pares de modelos

O coeficiente de correlação de Spearman ( $r_s = -0,0362$ ;  $p = 0,8494$ ) entre os módulos das diferenças e as médias sugere a baixa correlação entre as variáveis. O resultado da análise de regressão linear simples entre as diferenças e as médias confirma que não há dependência significativa ( $p = 0,620$ ) entre essas variáveis. Dessa forma, os limites de concordância e seus intervalos de confiança foram calculados através da abordagem padrão.

O diagrama de dispersão das medidas repetidas do pesquisador mostra que os valores dos erros estão próximos da linha de concordância. Observou-se tendência a um desvio para um dos lados da linha, sugerindo erro sistemático (GRÁFICO 5).

O erro sistemático observado foi de -0,4199998 e seu intervalo de confiança de 95% foi de -1,0277401 a 0,18774044. O intervalo de concordância de 95% variou de -3,6100135 a 2,7700139. Os limites de concordância apresentaram-se dentro dos valores preconizados pelos autores do índice (entre -4 e 4).

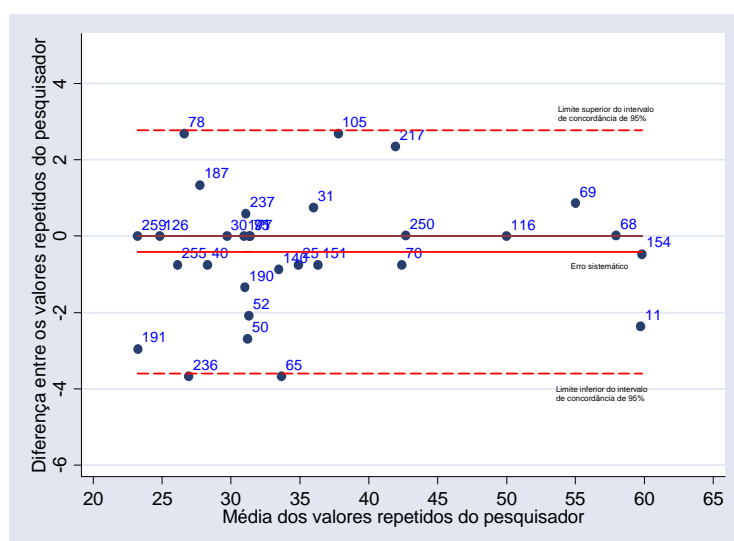


GRÁFICO 5 – Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice DAI, em trinta pares de modelos

Através do estudo de confiabilidade para calibração do pesquisador para o DAI, foi obtido um nível de concordância excelente tanto para o pesquisador e o padrão de referência, quanto para a concordância intra-examinador. O valor absoluto do erro sistemático da concordância entre o pesquisador e o padrão (IC 95% = -1,3191 a -0,2803) apresentou um valor de até 32,98% em relação ao limite aceitável 4 pontos, sendo que a amplitude do IC 95% representou 12,99% da amplitude desses limites. O erro sistemático da concordância intra-examinador (IC 95% = -1,02774 a 0,18774), em valor absoluto, representou até 25,69% desse limite, e sua amplitude atingiu 15,19% da amplitude dos limites. O erro aleatório encontrou-se dentro dos limites aceitáveis.

O DAI, conforme observado pelos examinadores, é um índice de fácil aprendizado e sua aplicação foi considerada objetiva. Algumas limitações do DAI foram observadas durante as sessões de calibração. O trespasse horizontal maxilar atribui pontuação maior para uma condição de normalidade (2 ou 3 mm) do que para a situação de mordida de borda-a-borda incisal de incisivos (0 mm). Ao determinar a necessidade de tratamento, o DAI desconsidera o impacto da má oclusão sobre a saúde bucal nos casos em que molares estão ausentes ou dentes posteriores apresentam inclinação axial incompatível com a função normal. Conforme já observado por Otuyemi e Noar (1996a), as mordidas cruzadas posteriores e as mordidas profundas excessivas não são medidas pelo DAI.

## 5.1.2 ICON

### 5.1.2.1 APLICAÇÃO E AVALIAÇÃO DOS CRITÉRIOS

Os resultados da concordância obtida para os valores do ICON entre os seis pares possíveis de examinadores em vinte pares de modelos foram avaliados através do ICC e são apresentados na TAB. 6.

TABELA 6

Resultados dos níveis de concordância dos pares de examinadores para os valores do ICON em vinte pares de modelos

| Pares de Examinadores | ICC (Limite Inferior do Intervalo de Confiança (95%)) |
|-----------------------|---|
| Ex. 1 x Ex. 2         | 0,855 (0,693)   |
| Ex. 1 x Ex. 3         | 0,767 (0,509)   |
| Ex. 1 x Ex. 4         | 0,939 (0,871)   |
| Ex. 2 x Ex. 3         | 0,732 (0,438)   |
| Ex. 2 x Ex. 4         | 0,875 (0,736)   |
| Ex. 3 x Ex. 4         | 0,706 (0,385)   |

Observou-se que o nível de concordância foi pobre para o par  $3 \times 4$ ; razoável para os pares  $1 \times 3$  e  $2 \times 3$ ; bom para os pares  $1 \times 2$  e  $2 \times 4$ , e; excelente para o par  $1 \times 4$ . Nota-se que os três pares constituídos pelo examinador 3 foram os que apresentaram menor concordância.

A TAB. 7 mostra os resultados do erro sistemático com seus limites de confiança de 95% e dos limites de concordância de 95%, utilizados nos gráficos de Bland-Altman (GRÁFICO 6).

TABELA 7

Resultados do erro sistemático e seus limites de confiança e dos limites de concordância para os valores do ICON em vinte pares de modelos

| Pares de Examinadores | Erro Sistemático | Intervalo de Confiança (95%) do Erro Sistemático |      | Limites de Concordância (95%) |      |
|-----------------------|------------------|--|------|-------------------------------|------|
| Ex. 1 x Ex. 2         | -1,3             | -8,0   | 5,5  | -31,5                         | 29,0 |
| Ex. 1 x Ex. 3         | -14,1            | -21,3  | -6,9 | -46,5                         | 18,3 |
| Ex. 1 x Ex. 4         | 4,7              | 0,8  | 8,6  | -12,7                         | 22,1 |
| Ex. 2 x Ex. 3         | -12,9            | -20,8  | -4,9 | -48,2                         | 22,5 |
| Ex. 2 x Ex. 4         | 6,0              | 0,5  | 11,4 | -18,3                         | 30,2 |
| Ex. 3 x Ex. 4         | 18,8             | 11,9   | 25,7 | -11,9                         | 49,5 |

Os maiores valores de erro sistemático foram observados nos pares formados pelo examinador 3, cujos limites de confiança de 95% também foram maiores, evidenciando a falta de precisão dessa estimativa.

Os casos de números 1, 7, 9 e 13, cujas diferenças foram maiores foram identificados e separados para que os examinadores pudessem investigar os motivos das discordâncias.

Os pressupostos para o cálculo dos limites de concordância não foram verificados nesta etapa do estudo. O objetivo principal foi a comparação dos valores dos erros para a identificação dos pontos de discordância e suas causas.

Foi possível identificar uma possível fonte de erro: a interpretação dos critérios pelo examinador 3. Porém isso não explica a totalidade dos erros. O erro sistemático, os erros aleatórios e os níveis de concordância dos outros pares de examinadores foram menores, mas ainda evidenciaram falta de concordância na medição do ICON.

Foi investigada a concordância para as variáveis que compõem o valor final do índice, de forma isolada. Os resultados da concordância dos seis pares de examinadores para cada um dos cinco componentes do ICON, avaliados pela estatística kappa, são apresentadas na TAB. 8.

TABELA 8

Concordância entre os pares de examinadores para os componentes do ICON

| Pares de Examinadores | Componentes do ICON e ponderação do coeficiente kappa |                          |              |                          |                 |                          |            |                          |                                |                          |
|-----------------------|---|--------------------------|--------------|--------------------------|-----------------|--------------------------|------------|--------------------------|--------------------------------|--------------------------|
|                       | Estético (Escala SCAN)                                |                          | Espaços      |                          | Mordida Cruzada |                          | Vertical   |                          | Relação Ântero-posterior bucal |                          |
|                       | $\kappa_w^2$  | Limite inferior (IC 95%) | $\kappa_w^2$ | Limite inferior (IC 95%) | $K_s$           | Limite inferior (IC 95%) | $\kappa_w$ | Limite inferior (IC 95%) | $\kappa_w$                     | Limite inferior (IC 95%) |
| 1 x 2                 | 0,813   | 0,598                    | 0,660        | 0,423                    | 0,634           | 0,279                    | 0,611      | 0,410                    | 0,284                          | 0,095                    |
| 1 x 3                 | 0,627   | 0,364                    | 0,936        | 0,769                    | 0,765           | 0,464                    | 0,500      | 0,237                    | 0,545                          | 0,250                    |
| 1 x 4                 | 0,883   | 0,723                    | 0,958        | 0,854                    | 0,886           | 0,671                    | 0,529      | 0,244                    | 0,582                          | 0,367                    |
| 2 x 3                 | 0,626   | 0,353                    | 0,721        | 0,568                    | 0,857           | 0,587                    | 0,437      | 0,182                    | 0,342                          | 0,048                    |
| 2 x 4                 | 0,872   | 0,677                    | 0,714        | 0,505                    | 0,737           | 0,403                    | 0,643      | 0,435                    | 0,335                          | 0,121                    |
| 3 x 4                 | 0,582   | 0,356                    | 0,932        | 0,784                    | 0,875           | 0,638                    | 0,363      | 0,090                    | 0,472                          | 0,294                    |

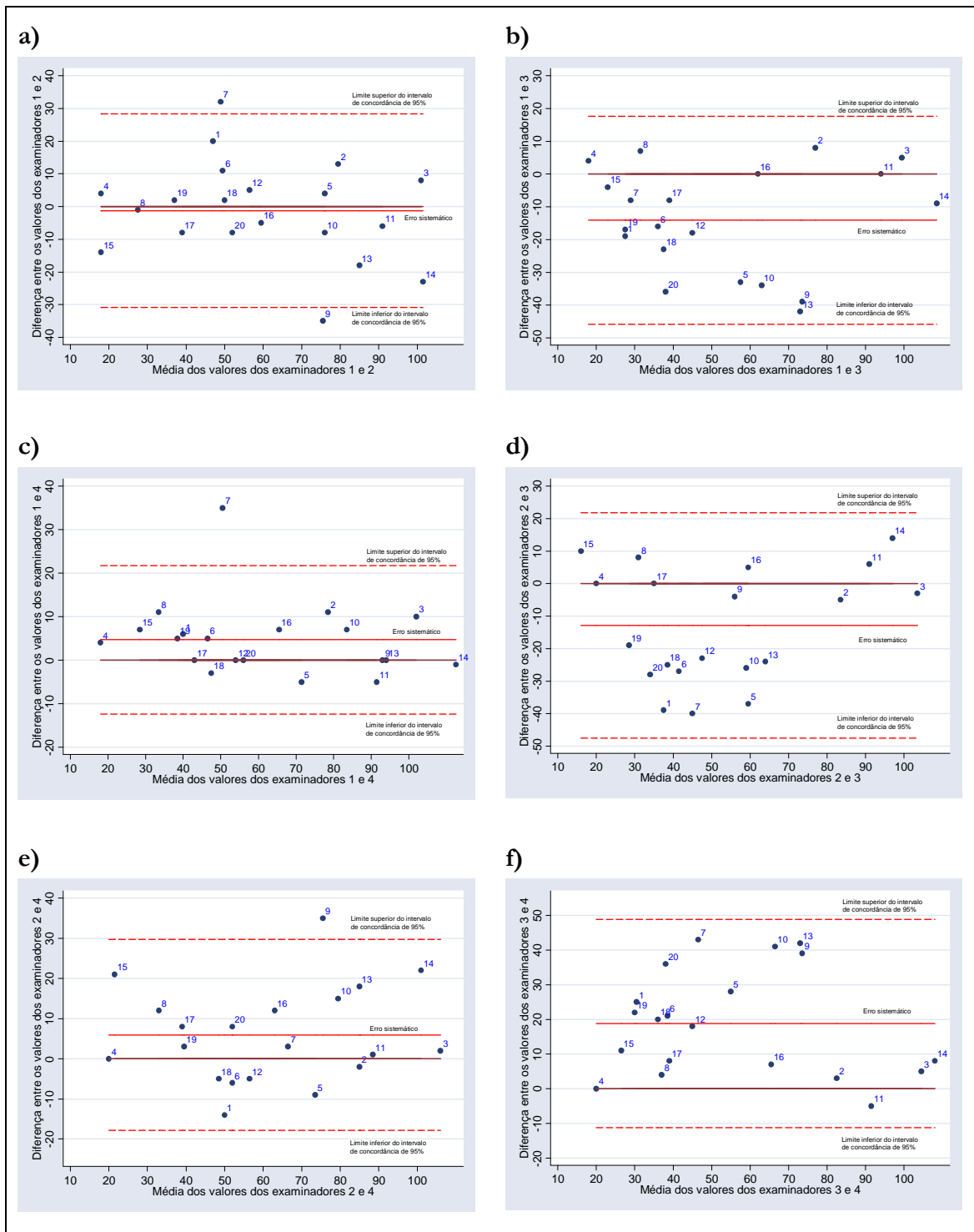


GRÁFICO 6 – Gráficos de Bland-Altman das diferenças versus as médias entre os pares de examinadores para o índice ICON em vinte pares de modelos: a) examinadores 1 e 2; b) examinadores 1 e 3; c) examinadores 1 e 4; d) examinadores 2 e 3; e) examinadores 2 e 4; f) examinadores 3 e 4

Observou-se que os níveis de concordância não são semelhantes entre os componentes do índice. Enquanto alguns componentes apresentaram níveis de concordância de razoável a excelente, outros tiveram valores classificados como pobre. Apenas o componente “relação ântero-posterior bucal” apresentou níveis de concordância pobres, para todos os pares de examinadores. O componente “vertical” obteve níveis de concordância entre pobre e razoável. Os componentes “estético (escala SCAN)”, “espaços” e “mordida cruzada” apresentaram níveis de concordância heterogêneos entre os pares de examinadores.

Os resultados foram discutidos pelos examinadores e concluiu-se que o componente “relação ântero-posterior bucal”, cujos níveis de concordância foram mais baixos, de forma homogênea, apresentou problemas de interpretação dos critérios ou dificuldade na execução da medida para todos os examinadores. Aqueles componentes cujos níveis de concordância apresentaram-se heterogêneos foram discutidos par a par, de forma a fixar os critérios uniformemente.

Foi constatado erro de interpretação no critério do componente “estético” por parte do examinador 3, que deliberadamente subestimou todas as suas medições. Na equação do ICON, esse componente é multiplicado por um coeficiente de 7, o que torna cada unidade de erro com maior potencial de alterar o valor final do índice. Esse componente foi considerado pelo grupo como sendo difícil de ser aprendido e reproduzido, dada sua subjetividade. O componente de “relação ântero-posterior bucal” apresentou os piores níveis de concordância para todos os pares de examinadores. Da mesma forma como ocorreu com o componente “relação ântero-posterior intermolares” do DAI, observou-se uma dificuldade em medir corretamente as relações entre as cúspides dos dentes posteriores. Foi constatado que, embora o critério fosse claro ao estabelecer os valores para cada situação de intercuspidação, o julgamento do seu significado clínico pode ter interferido nas medições dos quatro examinadores. Outra fonte de erro para as medidas desse componente foi a falta de padronização do ângulo de visão, causando um erro de paralaxe. Todos esses aspectos foram discutidos e o critério foi unificado.

A medição do componente “espaços” apresentou erro de interpretação em alguns modelos especificamente, o que foi devidamente exposto e discutido. Esse foi o componente considerado mais trabalhoso e demorado para ser medido. Além de exigir o cálculo da discrepância dento-alveolar da arcada superior, há algumas exceções em que o critério de medida é modificado, como por exemplo: se um dente incluso na arcada inferior



satisfizer as condições para que seja considerado impactado, a análise de espaço é desconsiderada, atribuindo-se a este componente a pontuação máxima.

### 5.1.2.2 CALIBRAÇÃO DO PESQUISADOR

O resultado do ICC ( $r_1 = 0,969$ ;  $r_1 \geq 0,943$ ) para concordância entre o padrão de referência do ICON e as medidas realizadas pelo pesquisador no conjunto de trinta pares de modelos é apresentado na TAB. 9. A concordância pode ser considerada excelente entre o valor atribuído pelo pesquisador e o padrão de referência obtido por consenso.

TABELA 9

Concordância entre o pesquisador e o padrão de referência do ICON

| Par de Medições      | ICC - Limite Inferior do Intervalo de Confiança (95%) |
|----------------------|---|
| Padrão x Pesquisador | 0,969 (0,943)   |

O histograma da distribuição das diferenças entre as medidas realizadas pelo pesquisador e o padrão de referência apresentou características de normalidade (GRÁFICO 7). O valor de distorção, próximo de zero ( $-0,09$ ;  $p = 0,804$ ), e o valor de curtose ( $2,70$ ;  $p = 0,949$ ) evidenciaram a normalidade ( $p = 0,9675$ ) na distribuição dos erros.

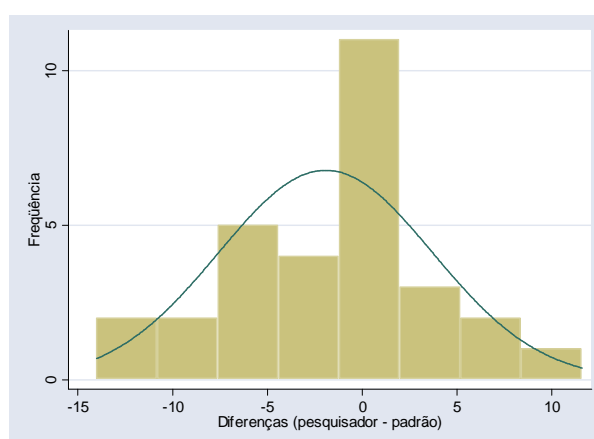


GRÁFICO 7 – Histograma das diferenças das medidas entre o pesquisador e o padrão de referência para o índice ICON, em trinta pares de modelos

Observou-se baixa correlação entre as diferenças e as médias ( $r_s = 0,1768$ ;  $p = 0,35$ ). A análise de regressão linear simples entre as diferenças e as médias também sugere que não há dependência significativa entre essas variáveis ( $p = 0,676$ ).

Como o pressuposto da independência das variáveis foi atendido, os limites de concordância e seus intervalos de confiança foram calculados através da abordagem padrão de Bland-Altman, sem a necessidade de transformação dos dados ou correção dos limites pela análise de regressão dos módulos dos resíduos com as médias.

O diagrama de dispersão entre as diferenças e as médias de cada observação, para os valores do pesquisador e do padrão, mostra que os pontos não se afastam muito da linha de concordância e que há tendência a um desvio para um dos lados da linha, sugerindo um erro sistemático (GRÁFICO 8).

O erro sistemático observado foi de  $-1,9333$  e seu intervalo de confiança de 95% foi de  $-4,0432$  a  $0,1765$ . O intervalo de concordância de 95% variou de  $-13,008$  a  $9,1413$ . Os limites de concordância apresentaram-se dentro dos valores preconizados pelos autores do índice (entre  $-18$  e  $18$ ).

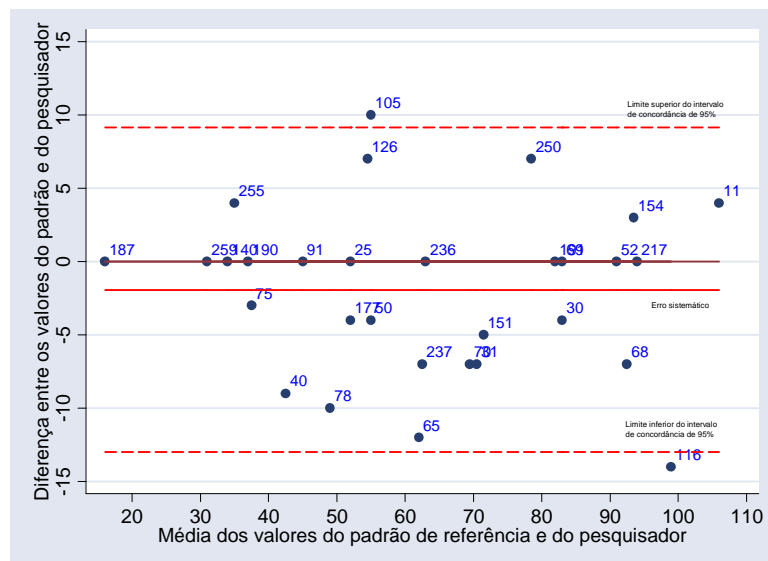


GRÁFICO 8 – Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas pelo pesquisador e o padrão de referência para o índice ICON, em trinta pares de modelos

### 5.1.2.3 CONCORDÂNCIA INTRA-EXAMINADOR

A concordância intra-examinador para os valores do ICON atribuídos pelo pesquisador, em dois exames repetidos, em trinta pares de modelos é apresentada na TAB. 10. Conforme o resultado do ICC ( $r_1 = 0,963$ ;  $r_1 \geq 0,931$ ), o nível de concordância foi considerado excelente.

TABELA 10

Concordância intra-examinador do pesquisador para o índice ICON, em trinta modelos

| Exames Repetidos pelo Pesquisador | ICC - Limite Inferior do Intervalo de Confiança (95%) |
|-----------------------------------|---|
| 1 x 2                             | 0,963 (0,931)   |

O histograma da distribuição das diferenças entre as medidas realizadas pelo pesquisador, em exames repetidos, apresentou características de normalidade (GRÁFICO 9). Os valores de distorção ( $-0,44$ ;  $p = 0,259$ ) e de curtose ( $3,59$ ;  $p = 0,233$ ) evidenciaram a normalidade na distribuição dos erros ( $p = 0,2286$ ).

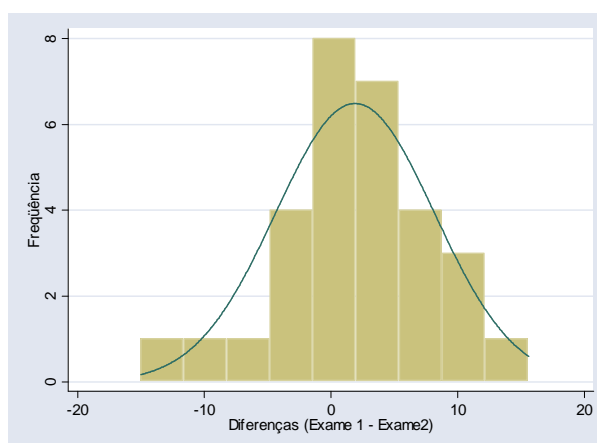


GRÁFICO 9 – Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice ICON, em trinta pares de modelos

O coeficiente de correlação de Spearman ( $r_s = -0,0735$ ;  $p = 0,6993$ ) entre os módulos das diferenças e as médias sugere a baixa correlação entre as variáveis. A análise de regressão linear simples entre as diferenças e as médias confirma que não há dependência significativa entre essas variáveis ( $p = 0,780$ ). Assim, os limites de concordância e seus intervalos de confiança foram calculados através da abordagem padrão de Bland-Altman.

O diagrama de dispersão entre as diferenças e as médias para cada observação, para os valores repetidos do pesquisador, mostra que os pontos não se afastam muito da linha de concordância e que há tendência a um desvio para um dos lados da linha, sugerindo um erro sistemático (GRÁFICO 10).

O erro sistemático observado foi de 1,9 com intervalo de confiança de 95% entre -0,4419 e 4,2419. O intervalo de concordância de 95% variou de -10,3926 a 14,1926. Portanto, os limites de concordância apresentaram-se dentro dos valores preconizados pelos autores do índice (entre -18 e 18).

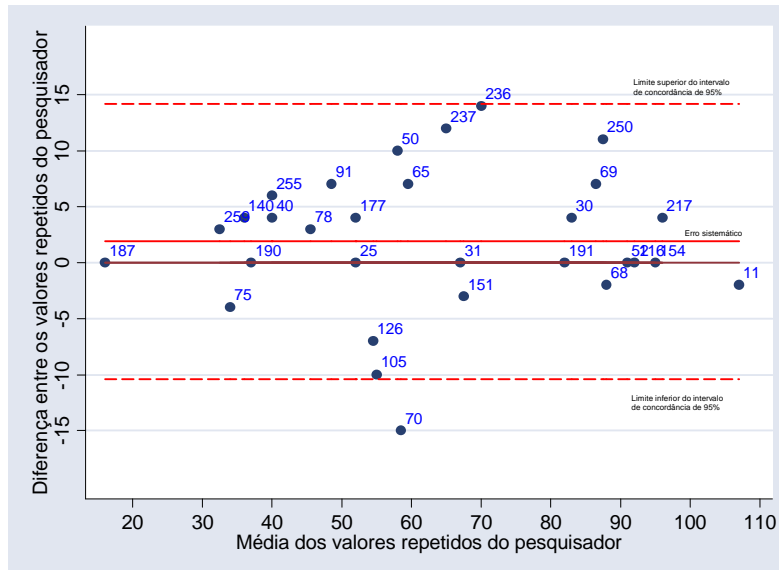


GRÁFICO 10 – Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice ICON, em trinta pares de modelos

Através do estudo de confiabilidade para calibração do pesquisador para o ICON, foi obtido um nível de concordância excelente tanto para o pesquisador e o padrão de referência, quanto para a concordância intra-examinador. O valor absoluto do erro sistemático da concordância entre o pesquisador e o padrão (IC 95% = -4,0432 a 0,1765) apresentou um valor de até 22,46% em relação ao limite aceitável 18 pontos, sendo que a amplitude do IC 95% representou 11,72% da amplitude desses limites. O erro sistemático da concordância intra-examinador (IC 95% = -0,4419 a 4,2419), em valor absoluto, representou até 23,57% desse limite, e sua amplitude atingiu 13,01% da amplitude dos limites. O erro aleatório encontrou-se dentro de limites aceitáveis.

Durante as sessões de calibração, as observações e dúvidas foram anotadas e discutidas. A primeira impressão ao estudar o ICON foi a dificuldade no aprendizado, principalmente de dois de seus componentes: o “estético” e o da “avaliação de espaços”. A dificuldade do primeiro reside na subjetividade de como deve ser feita a classificação baseada na aparência geral, e não em características anatômicas específicas. A escala é limitada em não mostrar alguns tipos de alterações morfológicas, como a mordida aberta, a mordida cruzada anterior, ou mesmo a ausência de dentes anteriores. Como o seu coeficiente é igual a sete, qualquer erro multiplicado por esse valor torna o componente “estético” o mais crítico em relação à concordância. Essas dificuldades também foram relatadas no estudo de validação do ICON realizado por Louwse *et al.* (2006).

A dificuldade da avaliação de espaços é dada pela complexidade das regras e suas exceções, que podem confundir um examinador pouco experiente.

Após o treinamento, a impressão de dificuldade diminuiu. O fato de ser constituído por apenas cinco componentes torna sua aplicação muito fácil de ser executada.

## 5.2 VALIDAÇÃO DOS ÍNDICES

### 5.2.1 PADRÃO-OURO: A OPINIÃO DOS ESPECIALISTAS

Observou-se alguma variabilidade na opinião, numa escala de 1 a 7, dos vinte especialistas em relação ao mesmo conjunto de 108 pares de modelos, sendo a média de 4,795 (D.P. =  $\pm 0,76$ ), o valor mínimo de 3,25 e o máximo de 6,44, atribuídos pelos especialistas 20 e 9, respectivamente (TAB. 11). O coeficiente de variação (CV) foi igual a 15,85%.

TABELA 11

Médias dos valores atribuídos ao mesmo conjunto de 108 pares de modelos por cada um dos especialistas

| <b>Examinadores</b> | <b>Médias</b> |
|---------------------|---------------|
| 1                   | 4,72          |
| 2                   | 4,24          |
| 3                   | 4,60          |
| 4                   | 5,47          |
| 5                   | 5,84          |
| 6                   | 4,39          |
| 7                   | 4,61          |
| 8                   | 3,54          |
| 9                   | 6,44          |
| 10                  | 5,00          |
| 11                  | 5,19          |
| 12                  | 4,85          |
| 13                  | 4,96          |
| 14                  | 5,09          |
| 15                  | 4,68          |
| 16                  | 5,76          |
| 17                  | 3,86          |
| 18                  | 4,93          |
| 19                  | 4,47          |
| 20                  | 3,25          |

Esse resultado é similar ao do estudo de validação do ICON, no qual Cooke (2006) também observou alguma variação na forma como 15 especialistas julgaram 153 modelos utilizando uma escala de 12 doze categorias. A amplitude das médias das opiniões baseadas nessa escala foi de 3,92. Entretanto, a comparação direta entre esses resultado e os do presente estudo é limitada devido à utilização de escalas diferentes.

A distribuição das médias das opiniões dos 20 especialistas para cada um dos 108 modelos, numa escala de 1 a 7 pontos, apresentou média igual a 4,795 ( $\pm 1,404$ ), mediana igual a 5,125 (TAB. 12, GRÁFICO 11).

TABELA 12

Distribuição das médias de opinião para os 108 modelos

| <b>Média</b> | <b>Desvio-padrão</b> | <b>Mediana</b> | <b>Mínimo</b> | <b>Máximo</b> |
|--------------|----------------------|----------------|---------------|---------------|
| 4,795        | 1,404                | 5,125          | 1,15          | 6,95          |

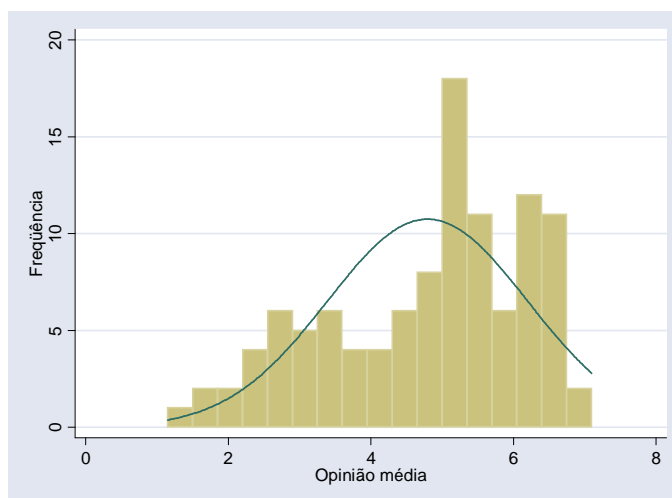


GRAFICO 11. Histograma da distribuição das médias das opiniões dos 20 especialistas para o conjunto de 108 modelos, numa escala de 1 a 7 pontos

O conjunto de modelos foi selecionado de forma que os casos fossem distribuídos através da escala de necessidade de tratamento, da seguinte forma: maior frequência de casos com necessidade intermediária, com proporção menor casos nos extremos inferior e superior da escala (TAB. 36 e GRÁFICO 21, APÊNDICE M). Essa distribuição possibilitou melhor avaliação do poder discriminatório do teste na faixa de valores onde há dificuldade na tomada de decisão. A média obtida pelo painel para o conjunto de modelos está acima do valor central da escala de opinião, igual a quatro. Isso pode ser decorrente do emprego de uma amostra que, para representar a grande variabilidade nos tipos de má oclusão, resultou num um grande número de modelos com necessidade de tratamento.

A variabilidade na opinião dos especialistas sobre a decisão da necessidade de tratamento ortodôntico pode ser considerada um reflexo da dificuldade em definir os sinais objetivos da má oclusão. Considerando que a necessidade de tratamento é uma medida do desvio de um ideal arbitrário, deveria haver um consenso a respeito desse ideal (RICHMOND *et al.*, 1994; 1992; SUMMERS, 1971; YOUNIS *et al.* 1997).

O ponto de corte utilizado para dicotomizar a opinião sobre necessidade de tratamento foi determinado pelos especialistas, que escolheram valores entre dois e cinco. O PTI, igual a 2,90, equivale à média dos valores escolhidos pelos vinte especialistas (TAB. 13).

O ponto de corte, obtido em estudos que utilizaram a mesma metodologia, foi igual a 3,25 (YOUNIS *et al.*, 1997) e 3,53 (BEGLIN *et al.*, 2001; FIRESTONE *et al.*, 2002).

TABELA 13

Ponto de tratamento indicado - PTI

| <b>Média dos PTI</b> | <b>Desvio-padrão</b> | <b>Mediana</b> | <b>Mínimo</b> | <b>Máximo</b> |
|----------------------|----------------------|----------------|---------------|---------------|
| 2,90                 | 0,79                 | 3,0            | 2             | 5             |

A determinação do ponto de corte classificou 86,11% dos casos como “com necessidade” de tratamento e 13,89% como “sem necessidade” (TAB. 14). De acordo com a estimativa da distribuição dos casos através da escala SCAN, cerca de 60% dos pares de modelos, teriam necessidade de tratamento, considerando apenas o aspecto estético.

Os estudos de Beglin *et al.* (2001) e Firestone *et al.* (2002a), que avaliaram respectivamente 156 e 170 modelos de estudo, apresentaram porcentagens de classificação mais equilibradas. Em ambos, o painel classificou 64% dos casos na categoria de necessidade de tratamento.

A diferença entre a taxa de indicação de necessidade de tratamento observada e a prevista na seleção da amostra pode ter sido consequência da média das opiniões ter apresentado um valor mais alto que o valor central da escala de opinião e do valor de PTI ter sido mais baixo do que os encontrados nos estudos citados acima.

TABELA 14

Distribuição da decisão dicotômica do painel

| <b>Decisão do Painel</b> | <b>n</b>   | <b>%</b>     |
|--------------------------|------------|--------------|
| <b>Com Necessidade</b>   | 93         | 86,11        |
| <b>Sem Necessidade</b>   | 15         | 13,89        |
| <b>Total</b>             | <b>108</b> | <b>100,0</b> |

### 5.2.1.1 CONFIABILIDADE DA OPINIÃO DOS ESPECIALISTAS

A concordância interexaminadores dos especialistas é apresentada na TAB. 15. Os níveis de concordância interexaminadores para os pares variaram de pobre a bom.

Os resultados de concordância intra-examinadores apresentaram níveis que variaram de pobre a excelente. O nível mais baixo de concordância intra-examinadores foi obtido por dois especialistas: os de número dois e dez. Apenas um deles atingiu o nível excelente, o de número onze (TAB. 16).



TABELA 15

Concordância interexaminadores por pares

| Pares de examinadores | $\kappa_w^2$ (Limite Inferior I.C. 95%) |
|-----------------------|---|
| 1 x 2                 | 0.631 (0.520)                           |
| 2 x 4                 | 0.565 (0.450)                           |
| 3 x 16                | 0.588 (0.463)                           |
| 4 x 14                | 0.502 (0.338)                           |
| 5 x 20                | 0.299 (0.212)                           |
| 6 x 12                | 0.651 (0.537)                           |
| 7 x 11                | 0.744 (0.662)                           |
| 8 x 18                | 0.499 (0.396)                           |
| 9 x 15                | 0.256 (0.144)                           |
| 10 x 17               | 0.644 (0.553)                           |
| 11 x 3                | 0.698 (0.602)                           |
| 12 x 1                | 0.577 (0.460)                           |
| 13 x 6                | 0.662 (0.555)                           |
| 14 x 19               | 0.528 (0.379)                           |
| 15 x 13               | 0.751 (0.663)                           |
| 16 x 9                | 0.423 (0.250)                           |
| 17 x 5                | 0.351 (0.235)                           |
| 18 x 10               | 0.692 (0.606)                           |
| 19 x 8                | 0.574 (0.462)                           |
| 20 x 7                | 0.548 (0.466)                           |

NOTA: medidas tomadas por amostragem

TABELA 16

Concordância intra-examinadores para os valores da opinião

| Examinador | $\kappa_w^2$ (Limite Inferior I.C. 95%) |
|------------|---|
| 1          | 0,826 (0,724)                           |
| 2          | 0,497 (0,217)                           |
| 3          | 0,808 (0,692)                           |
| 4          | 0,745 (0,529)                           |
| 5          | 0,711 (0,459)                           |
| 6          | 0,774 (0,640)                           |
| 7          | 0,846 (0,734)                           |
| 8          | 0,582 (0,406)                           |
| 9          | 0,645 (0,483)                           |
| 10         | 0,558 (0,370)                           |
| 11         | 0,859 (0,746)                           |
| 12         | 0,835 (0,710)                           |
| 13         | 0,822 (0,690)                           |
| 14         | 0,710 (0,537)                           |
| 15         | 0,697 (0,531)                           |
| 16         | 0,859 (0,744)                           |
| 17         | 0,758 (0,619)                           |
| 18         | 0,820 (0,678)                           |
| 19         | 0,604 (0,421)                           |
| 20         | 0,769 (0,639)                           |

Os estudos de Beglin *et al.* (2001) e Firestone *et al.* (2002a), para validação do DAI e do ICON respectivamente, obtiveram valores melhores de confiabilidade para o painel de especialistas. Beglin *et al.* (2001) observaram concordância interexaminadores de  $\kappa_w = 0,81$  (limite inferior do intervalo de confiança de 95%:  $\kappa_w \geq 0,81$ ) e a concordância intra-examinadores de  $\kappa_w = 0,92$  ( $\kappa_w \geq 0,90$ ), enquanto Firestone *et al.* (2002a) obtiveram valores para a concordância interexaminadores de  $\kappa_w = 0,81$  ( $\kappa_w \geq 0,81$ ) e intra-examinadores de  $\kappa_w = 0,92$  ( $\kappa_w \geq 0,90$ ).

A concordância para a opinião dos especialistas, observada no presente estudo, aproxima-se mais à relatada por Richmond *et al.* (1994), com valores de concordância intra-examinadores que variaram de 0,54 a 0,97 para necessidade estética e entre 0,12 e 0,89 para necessidade baseada na saúde dental. Níveis semelhantes de concordâncias foram obtidos por Richmond e Daniels (1998), cujos valores do limite inferior do IC de 95% do coeficiente kappa para a concordância intra-examinadores foram de 0,58 e 0,55, para necessidade baseada na estética e na saúde, respectivamente.

Younis *et al.* (1997) sugerem que a falta de concordância entre os especialistas cuja formação em ortodontia foi fora dos EUA, pode ser devida à falta de currículos comuns e de padrões de acreditação exigidos naquele país.

Para observar o efeito dos valores extremos sobre a distribuição da opinião dos especialistas, foram removidos os especialistas que atribuíram valores extremos e que apresentaram maior discordância (TAB. 15 e 16). Para a remoção, foram utilizados os seguintes critérios: concordância intra-examinador pobre, mais de 10% das observações fora do intervalo de confiança de 95% e concordância interexaminadores razoável ou pobre (TAB. 17).

TABELA 17

Valores obtidos pelos 108 modelos após remoção de sete especialistas discordantes e que apresentaram valores extremos

| Valores                             | Painel original | Painel após remoção dos especialistas discordantes |
|-------------------------------------|-----------------|--|
| <b>Média</b>                        | 4,795           | 4,792  |
| <b>Desvio-padrão</b>                | 0,76            | 0,65   |
| <b>Mínimo</b>                       | 3,25            | 3,25   |
| <b>Máximo</b>                       | 6,44            | 5,84   |
| <b>Mediana</b>                      | 4,79            | 4,79   |
| <b>Coeficiente de variação (CV)</b> | 15,85           | 13,56  |
| <b>Casos com necessidade</b>        | 93              | 93   |
| <b>PTI</b>                          | 2,90            | 2,94   |

Observa-se que, com a remoção dos sete especialistas mais discordantes e que apresentaram valores extremos, os valores de variabilidade mudaram enquanto os valores da média de opinião e do PTI não foram alterados de forma a modificar o número de casos classificados com necessidade de tratamento, que continuou igual a 93. Portanto, foi mantido o painel original com vinte especialistas, como o padrão-ouro da opinião da necessidade de tratamento ortodôntico.

## 5.2.2 VALORES DOS ÍNDICES

### 5.2.2.1 DAI

O examinador mediu os 108 pares de modelos de estudo conforme os critérios do DAI. Foi obtido um valor médio de 34,033 ( $\pm 8,422$ ). Esse valor encontra-se abaixo do valor central da escala do índice, que varia entre aproximadamente treze e oitenta (TAB. 18). Essa assimetria à esquerda corresponde à uma distribuição cujo valor central está mais próximo do ponto de corte igual a 31, sugerido pelos autores do índice (GRÁFICO 12). Uma maior frequência de valores um pouco acima do ponto de corte era esperada, pois o conjunto de modelos foi escolhido de forma a ter mais casos com necessidade de tratamento intermediária e menos casos com valores extremos na escala de necessidade. O conjunto de modelos apresentava uma estimativa de 60% dos casos com alguma necessidade de tratamento.

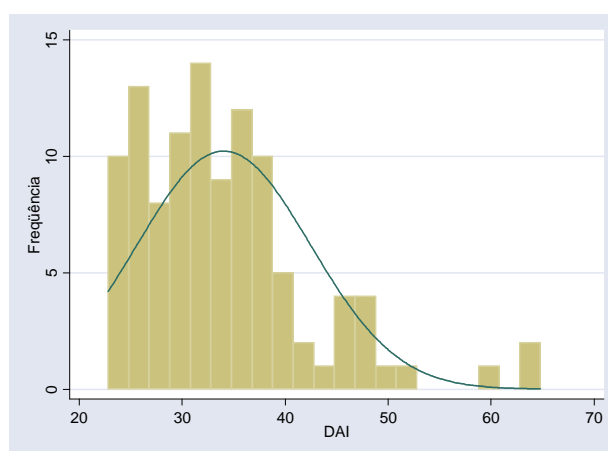


GRÁFICO 12. Distribuição dos valores do DAI medidos no conjunto de 108 pares de modelos

TABELA 18

Valores do DAI em 108 modelos de estudo

| <b>Média</b> | <b>Desvio-padrão</b> | <b>Mediana</b> | <b>Mínimo</b> | <b>Máximo</b> |
|--------------|----------------------|----------------|---------------|---------------|
| 34,033       | 8,422                | 32,565         | 22,8          | 64,92         |

Utilizando-se o ponto de corte igual a 31, originalmente preconizado pelos autores do DAI, observou-se que, enquanto os especialistas classificaram 93 (86,11%) pares de modelos com necessidade de tratamento, o DAI classificou 65 (60,19%) na mesma categoria. O DAI classificou corretamente 70,37% dos pares de modelos (TAB. 19).

TABELA 19

Distribuição dos resultados do DAI, com ponto de corte igual a 31, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos

| <b>Classificação do DAI</b> | <b>Decisão do Painel</b> |                        | <b>Total (%)</b> |
|-----------------------------|--------------------------|------------------------|------------------|
|                             | <b>Com necessidade</b>   | <b>Sem necessidade</b> |                  |
| <b>Com necessidade</b>      | 63                       | 2                      | 65 (60,19)       |
| <b>Sem necessidade</b>      | 30                       | 13                     | 43 (39,81)       |
| <b>Total (%)</b>            | <b>93 (86,11)</b>        | <b>15 (13,89)</b>      | <b>108 (100)</b> |

O DAI ter classificado menos casos com necessidade de tratamento do que o painel é uma questão relativa, pois é possível que o painel formado por ortodontistas de Belo Horizonte tenha uma opinião de tratar mais casos, evidenciando assim uma diferença geográfica na tomada de decisão. Entretanto esse é um problema de difícil avaliação e necessita de mais estudos para sua compreensão.

### 5.2.2.2 ICON

O examinador mediu os 108 pares de modelos de estudo conforme os critérios do ICON. Foi obtido um valor médio de 57,565 ( $\pm 23,509$ ), situado abaixo do valor central de uma escala que varia entre aproximadamente sete e 128 (TAB. 20). Observou-se que a distribuição dos valores medidos é deslocada para a esquerda, sendo que a média está mais próxima do ponto de corte igual a 43, sugerido pelos autores do índice (GRÁFICO 13). Uma frequência maior dos valores do ICON superior ao ponto de corte era esperada, pois o conjunto de modelos foi escolhido de forma a ter mais casos com necessidade de

tratamento intermediária e menos casos com valores extremos na escala de necessidade. O conjunto de modelos apresentava uma estimativa de 60% dos casos com alguma necessidade de tratamento.

TABELA 20

Valores do ICON em 108 modelos de estudo

| <b>Média</b> | <b>Desvio-padrão</b> | <b>Mediana</b> | <b>Mínimo</b> | <b>Máximo</b> |
|--------------|----------------------|----------------|---------------|---------------|
| 57,565       | 23,509               | 57             | 16            | 111           |

O ponto de corte preconizado pelos autores do ICON é igual a 43. Utilizando-se esse ponto de corte, observou-se que, enquanto os especialistas classificaram 93 (86,11%) pares de modelos com necessidade de tratamento, o ICON classificou 70 (64,81%) na mesma categoria. O ICON classificou corretamente 76,85% dos pares de modelos (TAB. 21)

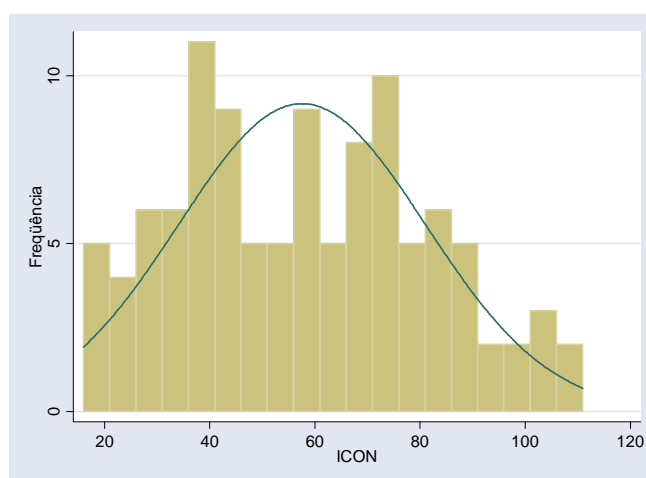


GRÁFICO 13. Distribuição dos valores do ICON medidos no conjunto de 108 pares de modelos

TABELA 21

Distribuição dos resultados do ICON, com ponto de corte igual a 43, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos

| Classificação do ICON | Decisão do Painel |                   | Total (%)        |
|-----------------------|-------------------|-------------------|------------------|
|                       | Com necessidade   | Sem necessidade   |                  |
| Com necessidade       | 69                | 1                 | 70 (64,81)       |
| Sem necessidade       | 24                | 14                | 38 (35,19)       |
| <b>Total (%)</b>      | <b>93 (86,11)</b> | <b>15 (13,89)</b> | <b>108 (100)</b> |

Conforme discutido anteriormente para o DAI, o fato do ICON ter classificado menos casos com necessidade de tratamento do que o painel pode refletir uma tendência dos ortodontistas de Belo Horizonte a sobretratamento.

### 5.2.2.3 REPRODUTIBILIDADE DO PESQUISADOR

#### 5.2.2.3.1 DAI

A concordância intra-examinador para os valores do DAI, atribuídos pelo pesquisador em dois exames repetidos, em quarenta pares de modelos, é apresentada na TAB. 22. Conforme o resultado do ICC ( $r_1 = 0,982$ ;  $r_1 \geq 0,969$ ), o nível de concordância foi considerado excelente. O estudo de Beglin *et al.* (2001) também obteve um valor excelente de concordância intra-examinador para o pesquisador calibrado, através estatística kappa, que com ponderação quadrática apresenta equivalência com o ICC (FLEISS & COHEN, 1973), de  $\kappa_w = 0,96$  (0,93).

TABELA 22

Concordância intra-examinador do pesquisador para o índice DAI, em quarenta pares de modelos

| Exames Repetidos pelo Pesquisador | ICC - Limite Inferior do Intervalo de Confiança (95%) |
|-----------------------------------|---|
| Ex. 1 (1 x 2)                     | 0,982 (0,969)   |

O histograma da distribuição das diferenças entre as medidas realizadas pelo pesquisador e o padrão de referência apresentou características de normalidade (GRÁFICO 14). Os valores de distorção ( $-0,55$ ;  $p = 0,121$ ) e de curtose ( $3,85$ ;  $p = 0,143$ ) sugerem a normalidade na distribuição dos erros ( $p = 0,1010$ ).

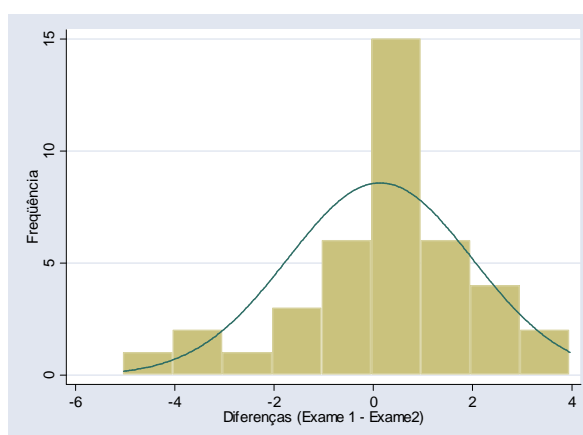


GRÁFICO 14 – Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice DAI, em quarenta pares de modelos

O coeficiente de correlação de Spearman ( $r_s = -0,1473$ ;  $p = 0,3643$ ) entre os módulos das diferenças e as médias confirma a baixa correlação entre as variáveis e os resultados da análise de regressão linear simples entre as diferenças e as médias sugerem que não há dependência significativa entre essas variáveis ( $p = 0,452$ ).

Como o pressuposto da independência das variáveis foi atendido, os limites de concordância e seus intervalos de confiança foram calculados através da abordagem padrão de Bland-Altman.

O diagrama de dispersão entre as diferenças e as médias de cada observação para os valores do pesquisador e do padrão mostra que os pontos não se afastam muito da linha de concordância e que a tendência do desvio para um dos lados da linha, sugere um baixo valor de erro sistemático (GRÁFICO 15).

O erro sistemático observado foi de 0,1365 e seu intervalo de confiança de 95% foi de -0,4577 a 0,7307. O intervalo de concordância de 95% variou de -3,5054 a 3,7783. Os limites de concordância apresentaram-se dentro dos valores preconizados pelos autores do índice (entre -4 e 4).

O valor absoluto do erro sistemático da concordância entre os exames repetidos pelo pesquisador (IC 95% = -0,4577 a 0,7307) apresentou um valor de até 18,27% em relação ao limite aceitável 4 pontos, sendo que a amplitude do IC 95% representou 14,86% da amplitude desses limites. O erro aleatório encontrou-se dentro dos limites aceitáveis. A verificação da concordância intra-examinador confirmou que o pesquisador manteve-se calibrado para o DAI, seis semanas após ter realizado o exercício de calibração.

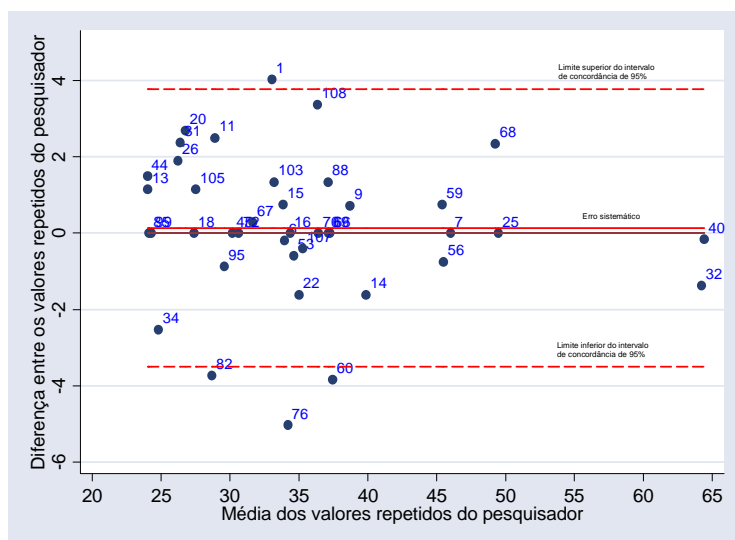


GRÁFICO 15 – Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice DAI, em quarenta pares de modelos

### 5.2.2.3.2 ICON

A concordância intra-examinador para os valores do ICON atribuídos pelo pesquisador, em dois exames repetidos, em quarenta pares de modelos é apresentada na TAB. 23. Conforme o resultado do ICC ( $r_1 = 0,972$ ;  $r_1 \geq 0,952$ ), o nível de concordância foi considerado excelente.

O estudo de Louwse *et al.* (2006) também obteve um valor médio excelente de concordância intra-examinador para o pesquisador calibrado ( $r_1 = 0,89$ ), embora não tenha sido relatado o valor do limite inferior do IC de 95% para o ICC. Já o trabalho de Firestone *et al.* (2002a) apresentou um bom nível de concordância intra-examinadores, avaliado através estatística kappa, que com ponderação quadrática apresenta equivalência com o ICC (FLEISS & COHEN, 1973), cujo valor foi de  $\kappa_w = 0,89$  (0,74).

TABELA 23

Concordância intra-examinador do pesquisador para o índice ICON, em quarenta modelos

| Exames Repetidos pelo Pesquisador | ICC - Limite Inferior do Intervalo de Confiança (95%) |
|-----------------------------------|---|
| 1 x 2                             | 0,972 (0,952)   |



O histograma da distribuição das diferenças entre as medidas realizadas pelo pesquisador e o padrão de referência apresentou características de normalidade (GRÁFICO 16). Os valores de distorção (0,13;  $p = 0,714$ ) e de curtose (3,82;  $p = 0,152$ ) evidenciaram a normalidade na distribuição dos erros ( $p = 0,3109$ ).

Observou-se baixa correlação entre os módulos das diferenças e as médias ( $r_s = 0,1144$ ;  $p = 0,4821$ ) e o resultado da análise de regressão linear simples entre as diferenças e as médias sugere que não há dependência significativa entre essas variáveis ( $p = 0,431$ ). Os limites de concordância e seus intervalos de confiança foram calculados através da abordagem padrão de Bland-Altman.

O diagrama de dispersão entre as diferenças e as médias de cada observação para os valores do pesquisador e do padrão mostra que os pontos estão próximos da linha de concordância e que a tendência ao desvio para um dos lados da linha sugere um baixo valor de erro sistemático (GRÁFICO 17).

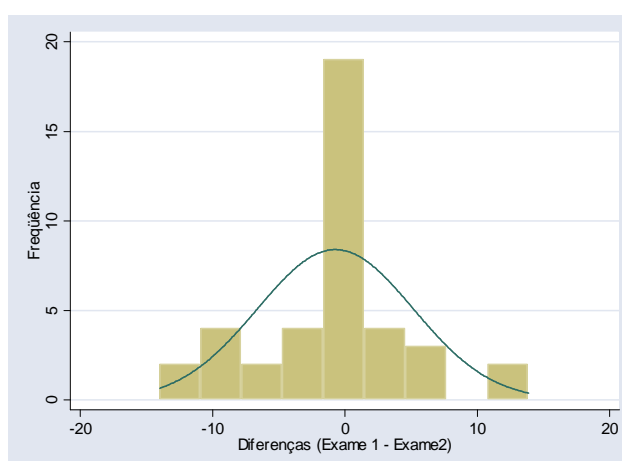


GRÁFICO 16 – Histograma das diferenças entre duas medidas repetidas pelo pesquisador para o índice ICON em quarenta pares de modelos.

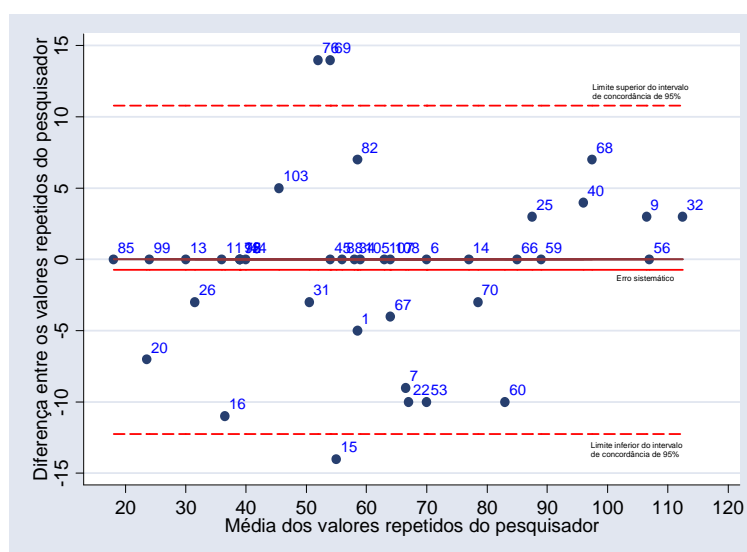


GRÁFICO 17 – Gráfico de Bland-Altman das diferenças versus as médias das medidas atribuídas em dois exames repetidos pelo pesquisador para o índice ICON, em quarenta pares de modelos

O erro sistemático observado foi de -0,725 e seu intervalo de confiança de 95% foi de -2,608 a 1,158. O intervalo de concordância de 95% variou entre -12,2651 e 10,8151. Os limites de concordância apresentaram-se dentro dos valores preconizados pelos autores do índice (entre -18 e 18).

O valor absoluto do erro sistemático da concordância entre os exames repetidos pelo pesquisador (IC 95% = -2,608 a 1,158) apresentou um valor de até 14,49% em relação ao limite aceitável 18 pontos, sendo que a amplitude do IC 95% representou 10,46% da amplitude desses limites. O erro aleatório encontrou-se dentro dos limites aceitáveis. A verificação da concordância intra-examinador confirmou que o pesquisador manteve-se calibrado para o ICON, seis semanas após ter realizado o exercício de calibração.

Os erros do examinador, até certo ponto, podem ser controlados através do treinamento de calibração. No entanto, a utilização de um único examinador impede a realização de testes de concordância interexaminadores para controlar os erros sistemáticos. A utilização do mesmo examinador em todo o processo, desde a seleção de modelos, passando pelo treinamento de calibração e, finalmente realizando as medidas e análises estatísticas pode ser uma fonte involuntária de erro. Mesmo com todos os cuidados no sentido de se evitar os erros conscientemente, as decisões e as medidas podem ser influenciadas por tendências pessoais, aptidão e experiência.

## 5.2.3 SENSIBILIDADE, ESPECIFICIDADE E ACURÁCIA

### 5.2.3.1 DAI

#### 5.2.3.1.1 Sensibilidade e especificidade

Os valores de sensibilidade e especificidade calculados para o DAI, utilizando o ponto de corte original, foram:  $S = 67,74\%$  e  $E = 86,67\%$  (TAB. 24). Enquanto a proporção de falsos-negativos encontrada foi igual a  $32,26\%$ , devido ao pobre nível de sensibilidade, um bom nível de especificidade fez com que a proporção de falsos-positivos fosse menor ( $13,33\%$ ).

Pressupondo a utilização do DAI em uma situação real de prevalência da necessidade de tratamento ortodôntico, os cálculos para valores preditivos positivo e negativo foram realizados com o dado de prevalência em Belo Horizonte ( $26,4\%$ ), obtido a partir do estudo de Marques *et al.* (2005). Foram encontrados os seguintes valores: VPP =  $62,38\%$ ; e VPN =  $89,17\%$ .

Esses valores podem ser interpretados como se, na prática, um dado caso classificado pelo índice como tendo necessidade, teria a chance de  $62,38\%$  de realmente necessitar de tratamento ortodôntico. Por outro lado, um caso classificado como sem necessidade teria a chance de  $89,17\%$  de não necessitar de tratamento.

A razão de verossimilhança para esse ponto de corte foi igual a 5,08. Um caso, cujo resultado do DAI fosse igual a 31, teria 5,08 mais chance de ser verdadeiro-positivo do que outro com um resultado menor do índice.

#### 5.2.3.1.2 Curva ROC

A área sob a curva ROC foi de  $81,83\%$  (IC de  $95\% = 71,21\%$  a  $92,44\%$ ), sugerindo que o modelo apresentou, no mínimo, um razoável nível de acurácia. O ponto de corte, igual a 31, pode ser localizado na curva através dos seus valores de sensibilidade e da proporção de falsos-positivos ( $1-E$ ) (GRÁFICO 18).

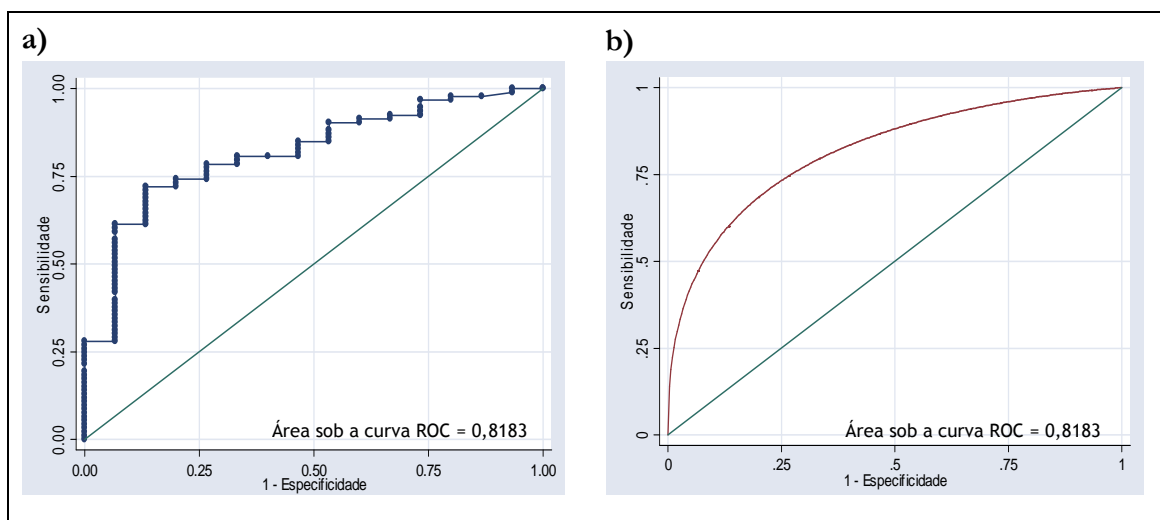


GRÁFICO 18 – Curva ROC para o DAI (n=108): a) representação dos 108 valores medidos; b) curva suavizada.

### 5.2.3.1.3 Ponto de corte ideal

Um ponto de corte igual a 30,4 fez com que o valor de sensibilidade aumentasse para um nível razoável (72,04%), enquanto a especificidade manteve-se boa, com o mesmo valor de 86,67% (GRÁFICO 18 e TAB. 24). O valor da sensibilidade aumentou sem que a especificidade diminuísse. Enquanto a proporção de falsos-positivos encontrada manteve-se com o valor de 13,33%, um valor aumentado para sensibilidade fez com que a proporção de falsos-negativos diminuísse para 27,96%.

Utilizando-se esse novo ponto de corte, igual a 30,4, ajustado para a melhor combinação dos valores de sensibilidade e especificidade, observou-se que, enquanto os especialistas classificaram 93 (86,11%) pares de modelos com necessidade de tratamento, o DAI classificou 69 (63,89%) na mesma categoria. Com esse novo ponto de corte, o DAI classificou corretamente 74,07% dos pares de modelos (TAB. 25).

TABELA 24

Sensibilidade e especificidade de pontos de corte do DAI em relação ao padrão-ouro

| Ponto de Corte | Sensibilidade (%) | Especificidade (%) |
|----------------|-------------------|--------------------|
| 30,04          | 75,27             | 73,33              |
| 30,11          | 74,19             | 73,33              |
| 30,2           | 74,19             | 80,00              |
| 30,27          | 73,12             | 80,00              |
| 30,29          | 72,04             | 80,00              |
| <b>30,4</b>    | <b>72,04</b>      | <b>86,67</b>       |
| 30,57          | 70,97             | 86,67              |
| 30,64          | 69,89             | 86,67              |
| 30,95          | 68,82             | 86,67              |
| <b>31</b>      | <b>67,74</b>      | <b>86,67</b>       |
| 31,06          | 66,67             | 86,67              |
| 31,22          | 65,59             | 86,67              |
| 31,41          | 64,52             | 86,67              |
| 31,54          | 63,44             | 86,67              |
| 31,73          | 62,37             | 86,67              |
| 31,97          | 61,29             | 86,67              |
| 32,02          | 61,29             | 93,33              |

TABELA 25

Distribuição dos resultados do DAI, com ponto de corte igual a 30,4, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos

| Classificação do DAI   | Decisão do Painel |                   | Total (%)        |
|------------------------|-------------------|-------------------|------------------|
|                        | Com necessidade   | Sem necessidade   |                  |
| <b>Com necessidade</b> | 67                | 2                 | 69 (63,89)       |
| <b>Sem necessidade</b> | 26                | 13                | 39 (36,11)       |
| <b>Total (%)</b>       | <b>93 (86,11)</b> | <b>15 (13,89)</b> | <b>108 (100)</b> |

A prevalência de necessidade de tratamento ortodôntico em Belo Horizonte, igual a 26,4%, foi utilizada para os cálculos dos valores preditivos positivo e negativo. Foram encontrados seguintes valores: VPP = 65,97%; e VPN = 89,63%.

Esses valores podem ser interpretados como se, na prática, um dado caso, classificado pelo índice como tendo necessidade, teria a chance de 65,97% de realmente necessitar de tratamento ortodôntico. Por outro lado, um caso classificado como sem necessidade teria a chance de 89,63% de não necessitar de tratamento.

A razão de verossimilhança para esse ponto de corte foi igual a 5,40. Um caso, cujo resultado do DAI fosse igual a 30,4, teria 5,40 mais chance de ser verdadeiro-positivo do que outro com um resultado menor do índice.

A TAB. 26 apresenta uma comparação entre os de valores de sensibilidade, especificidade, valores preditivos positivo e negativo e a porcentagem de classificação correta (acurácia geral) obtidos com os pontos de corte original e ajustado.

TABELA 26

Comparação das características de desempenho diagnóstico do DAI utilizando os pontos de corte original e ajustado na sua aplicação em 108 pares de modelos de estudo

| Ponto de corte do DAI | Sensibilidade | Especificidade | Valor preditivo positivo | Valor preditivo negativo | Classificações Corretas |
|-----------------------|---------------|----------------|--------------------------|--------------------------|-------------------------|
| 31 (original)         | 67,74         | 86,67          | 62,38                    | 89,17                    | 70,37                   |
| 30,4 (ajustado)       | 72,04         | 86,67          | 65,97                    | 89,63                    | 74,07                   |

Com o ponto de corte ajustado (30,4), observou-se uma melhora no nível de sensibilidade sem que a especificidade tenha diminuído. Os valores preditivos e a proporção de classificações corretas (acurácia geral) mantiveram seus níveis de significância clínica inalterados.

Considerando os resultados com os pontos de corte original e o ajustado no contexto da baixa prevalência da necessidade de tratamento ortodôntico, não há um ganho significativo em utilizar o ponto de corte ajustado. A diferença mais marcante ocorreu no valor de sensibilidade que passou do nível pobre de significância clínica para o nível razoável. Entretanto, os valores preditivos e a acurácia geral não tiveram seus níveis de significância clínica modificados com a alteração no ponto de corte. O aumento no valor da acurácia geral não deve ser considerado, pois é influenciado pela prevalência e foi calculado com base nos dados do conjunto de modelos escolhidos, cuja prevalência de necessidade de tratamento ortodôntico é alta (estimada em 60%), em relação à prevalência na população. Além disso, uma alteração de seis décimos no ponto de corte pode não ser prático na sua utilização. Assim, o ponto de corte original (31) pode ser considerado o ideal na utilização do DAI para a tomada de decisão da necessidade de tratamento ortodôntico, refletindo a opinião de um painel de 20 ortodontistas de Belo Horizonte.

## 5.2.3.2 ICON

### 5.2.3.2.1 Sensibilidade e especificidade

Os valores de sensibilidade e especificidade calculados para o ponto de corte original foram:  $S = 74,19\%$  e  $E = 93,33\%$  (TAB. 27). Enquanto a proporção de falsos-negativos encontrada foi igual a 25,81%, a proporção de falsos-positivos foi igual a 6,67%.

Pressupondo a utilização do ICON em uma situação real de prevalência da necessidade de tratamento ortodôntico, os cálculos para valores preditivos positivo e negativo foram realizados com o dado de prevalência em Belo Horizonte (26,4%), obtido a partir do estudo de Marques *et al.* (2005). Foram encontrados os seguintes valores: VPP = 79,96%; e VPN = 90,98%.

Os valores preditivos podem ser interpretados como se, na prática, um dado caso classificados pelo índice como tendo necessidade, teria a chance de 79,96% de realmente necessitar de tratamento ortodôntico. Por outro lado, um caso classificado como sem necessidade, teria a chance de 90,98% de não necessitar de tratamento.

A razão de verossimilhança para esse ponto de corte foi igual a 11,12. Um caso, cujo resultado do ICON fosse igual a 43, teria 11,12 mais chances de ser verdadeiro-positivo do que outro com um resultado menor do índice.

### 5.2.3.2.2 Curva ROC

A área sob a curva foi de 88,75% (IC de 95% = 78,57% a 98,92%), sugerindo que o modelo apresentou, no mínimo, um razoável nível de acurácia. O ponto de corte, igual a 43, pode ser localizado na curva através dos seus valores de sensibilidade e da proporção de falsos-positivos (1-E) (GRÁFICO 19).

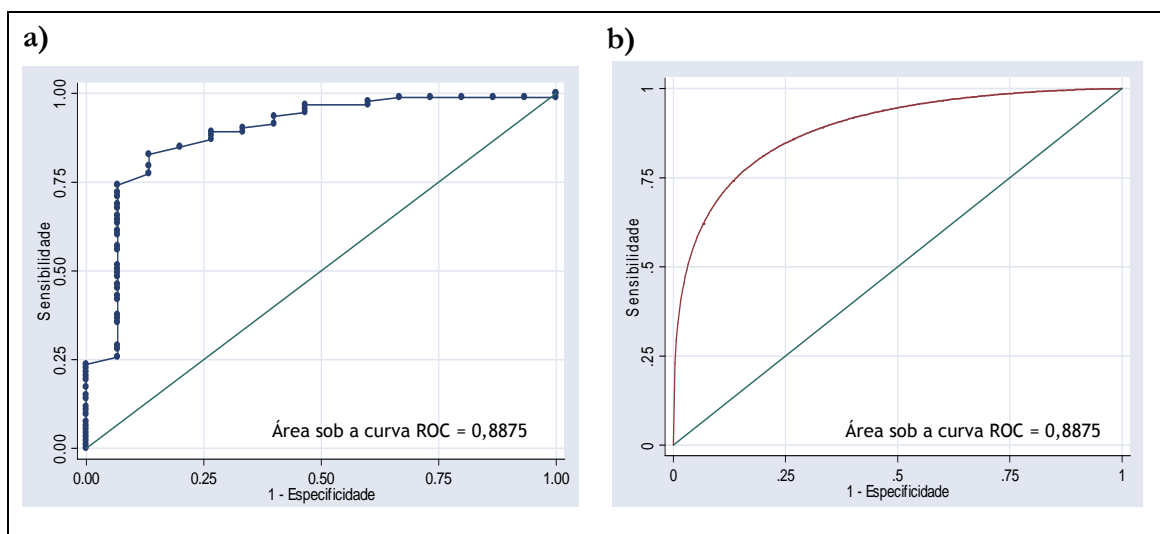


GRÁFICO 19 – Curva ROC para o ICON (n=108): a) representação dos 108 valores medidos; b) curva suavizada.

### 5.2.3.2.3 Ponto de corte ideal

Um ponto de corte igual a 40 fez com que o valor de sensibilidade aumentasse para 82,80%, embora a especificidade tenha diminuído para 86,67% (GRÁFICO 19 e TAB. 27). O valor da sensibilidade aumentou e a especificidade diminuiu. O valor mais alto para sensibilidade fez com que a proporção de falsos-negativos diminuísse para 17,20%, enquanto a proporção falsos-positivos aumentou para 13,33% em consequência da diminuição da especificidade.

Utilizando-se esse novo ponto de corte, igual a 40, ajustado para a melhor combinação dos valores de sensibilidade e especificidade, observou-se que, enquanto os especialistas classificaram 93 (86,11%) pares de modelos com necessidade de tratamento, o ICON classificou 79 (73,15%) na mesma categoria. Com esse novo ponto de corte, o ICON classificou corretamente 83,33% dos pares de modelos (TAB. 28).

Os valores de sensibilidade e especificidade do ICON, obtidos com o ponto de corte ajustado (40), são comparáveis aos resultados obtidos por Daniels e Richmond (2000) em seu estudo original ( $S = 0,85$ ;  $E = 0,84$ ; acurácia geral = 0,85).

Utilizando a prevalência de necessidade de tratamento ortodôntico em Belo Horizonte, igual a 26,4%, foram encontrados seguintes valores: VPP = 69,02%; e VPN = 93,35%.



TABELA 27

Sensibilidade e especificidade de pontos de corte do ICON em relação ao padrão-ouro

| Ponto de Corte | Sensibilidade (%) | Especificidade (%) |
|----------------|-------------------|--------------------|
| 30             | 94,62             | 53,33              |
| 32             | 93,55             | 60,00              |
| 33             | 91,40             | 60,00              |
| 34             | 90,32             | 66,67              |
| 35             | 89,25             | 66,67              |
| 36             | 89,25             | 73,33              |
| 37             | 88,17             | 73,33              |
| 38             | 87,10             | 73,33              |
| 39             | 84,95             | 80,00              |
| <b>40</b>      | <b>82,80</b>      | <b>86,67</b>       |
| 41             | 79,57             | 86,67              |
| 42             | 77,42             | 86,67              |
| <b>43</b>      | <b>74,19</b>      | <b>93,33</b>       |
| 45             | 72,04             | 93,33              |
| 47             | 70,97             | 93,33              |
| 48             | 68,82             | 93,33              |
| 49             | 67,74             | 93,33              |
| 52             | 65,59             | 93,33              |

TABELA 28

Distribuição dos resultados do ICON, com ponto de corte igual a 40, e da decisão dicotômica do painel de especialistas (padrão-ouro) em 108 pares de modelos

| Classificação do ICON  | Decisão do Painel |                   | Total (%)        |
|------------------------|-------------------|-------------------|------------------|
|                        | Com necessidade   | Sem necessidade   |                  |
| <b>Com necessidade</b> | 77                | 2                 | 79 (73,15)       |
| <b>Sem necessidade</b> | 16                | 13                | 29 (26,85)       |
| <b>Total (%)</b>       | <b>93 (86,11)</b> | <b>15 (13,89)</b> | <b>108 (100)</b> |

Esses valores podem ser interpretados como se, na prática, um determinado caso classificado como tendo necessidade pelo índice teria a chance de 69,02% de realmente necessitar de tratamento ortodôntico. Por outro lado, um caso classificado como sem necessidade, teria a chance de 93,35% de não necessitar de tratamento.

A razão de verossimilhança para esse ponto de corte foi igual a 6,21. Um caso, cujo resultado do ICON fosse igual a 40, teria 6,21 mais chances de ser verdadeiro-positivo do que outro com um resultado menor do índice.

A TAB. 29 apresenta uma comparação entre os de valores de sensibilidade, especificidade, valores preditivos positivo e negativo e a porcentagem de classificação correta (acurácia geral) obtidos com os pontos de corte original e ajustado.

TABELA 29

Comparação das características de desempenho diagnóstico do ICON utilizando os pontos de corte original e ajustado na sua aplicação em 108 pares de modelos de estudo

| Ponto de corte do ICON | Sensibilidade | Especificidade | Valor preditivo positivo | Valor preditivo negativo | Classificações Corretas |
|------------------------|---------------|----------------|--------------------------|--------------------------|-------------------------|
| 43 (original)          | 74,19         | 93,33          | 79,96                    | 90,98                    | 76,85                   |
| 40 (ajustado)          | 82,80         | 86,67          | 69,02                    | 93,35                    | 83,33                   |

Com o ponto de corte ajustado (40), observou-se uma melhora no nível de sensibilidade e uma diminuição no nível de especificidade. O valor preditivo positivo teve seu nível de significância clínica reduzido de razoável para pobre, enquanto o valor preditivo negativo manteve-se no nível excelente. A proporção de classificações corretas (acurácia geral) obteve melhora em seu nível de significância clínica, passando de razoável para bom.

Considerando os resultados com os pontos de corte original e o ajustado no contexto da baixa prevalência da necessidade de tratamento ortodôntico, não há um ganho significativo em utilizar o ponto de corte ajustado.

A diferença mais marcante com o novo ponto de corte ocorreu no valor de sensibilidade que passou do nível razoável de significância clínica para o nível bom. Em consequência disso, a proporção de classificações corretas aumentou consideravelmente. Entretanto deve ser observado que a acurácia geral é influenciada pela prevalência e esse valor foi calculado com base nos dados do conjunto de modelos escolhidos, cuja prevalência de necessidade de tratamento ortodôntico é alta (60%), em relação à prevalência na população.

Os valores preditivos, também influenciados pela prevalência, foram calculados a partir do valor da prevalência na população (26,4%) e oferecem uma melhor perspectiva sobre a performance do índice, caso fosse utilizado numa situação de prevalência real. O valor preditivo negativo não teve seu nível de significância clínica alterado com a alteração no ponto de corte. Como a especificidade apresenta grande influência nos valores preditivos, especialmente em situações de baixa prevalência, a utilização do ponto de corte ajustado, com a redução no valor da especificidade, causou uma importante diminuição no

valor preditivo positivo. O valor preditivo positivo, que representa a proporção dos verdadeiros positivos entre os classificados positivos pelo índice, atinge níveis melhores como reflexo de valores mais altos de especificidade. Um índice utilizado como teste de rastreamento em situações de baixa prevalência deve priorizar a especificidade.

Além disso, uma alteração de três unidades mantém o ponto de corte próximo ao valor originalmente preconizado e pode não ser prático na sua utilização. Assim, o ponto de corte original (43) pode ser considerado o ideal na utilização do ICON para a tomada de decisão da necessidade de tratamento ortodôntico, refletindo a opinião de um painel de 20 ortodontistas de Belo Horizonte.

Alterar o valor do ponto de corte pode ser uma questão política. Um ponto de corte mais baixo inclui mais casos na categoria “tratamento indicado” e pode aumentar a concordância do índice com o padrão-ouro dos especialistas (validade). Por outro lado, a má oclusão é um problema comum e autolimitante, sendo que apenas os casos mais graves geram conseqüências à saúde. Além disso, o tratamento ortodôntico é geralmente longo e apresenta custos consideráveis. Portanto, quando há limitação de recursos, é razoável que os pontos de corte usados sejam posicionados num valor mais alto e rígido na seleção dos casos com “tratamento indicado” para limitar a utilização destes recursos escassos.

### **5.2.3.3 COMPARAÇÃO ENTRE OS ÍNDICES**

Considerando a prevalência de 26,4% para Belo Horizonte, um bom índice de rastreamento de necessidade de tratamento ortodôntico seria aquele com um maior VPP, portando com maior valor de especificidade e uma menor parcela de falsos positivos. A parcela de falsos positivos é indesejável por acarretar custos emocional e financeiro. Por outro lado, o resultado falso negativo não vem a ser um grande problema, pois má oclusão não apresenta riscos graves à saúde. Assim o valor mais alto de especificidade deve ser priorizado e o valor de sensibilidade deve ser equilibrado de forma a não aumentar demasiadamente o número de falsos negativos.

Comparando-se os valores relacionados à validade do DAI e do ICON (TAB. 30), utilizando seus respectivos pontos de corte originais, observou-se que os valores obtidos pelo ICON refletem melhor a opinião do painel de especialistas.

TABELA 30

Desempenho diagnóstico do DAI e do ICON utilizando os pontos de corte originais na sua aplicação em 108 pares de modelos de estudo

|   | DAI                 | ICON                |
|---|---------------------|---------------------|
| Sensibilidade (%)                           | 67,74               | 74,19               |
| Especificidade (%)                          | 86,67               | 93,33               |
| Valor preditivo positivo (%)*               | 62,38               | 79,96               |
| Valor preditivo negativo (%)*               | 89,17               | 90,98               |
| Razão de verossimilhança positiva           | 5,08                | 11,12               |
| Acurácia: área sob a curva ROC (%) (IC 95%) | 81,83 (71,21-92,44) | 88,75 (78,57-98,92) |

NOTA: os valores preditivos positivo e negativo foram calculados a partir de dados secundários sobre a prevalência de necessidade de tratamento ortodôntico em Belo Horizonte (MARQUES *et al.* 2005)

A comparação entre as medidas de acurácia (área sob a curva ROC) de cada um dos índices é apresentada na TAB. 31. O ICON parece ser um melhor discriminador para necessidade de tratamento do que o DAI, porque a sua área sob a curva ROC foi significativamente maior ( $p=0,0435$ ) (GRÁFICO 20).

TABELA 31

Comparação entre porcentagens das áreas sob as curvas ROC do DAI e do ICON

|      | Area ROC (%) | Erro-padrão (%) | Intervalo de confiança de 95% |        |
|------|--------------|-----------------|-------------------------------|--------|
| DAI  | 81,83        | 5,42            | 71,213                        | 92,443 |
| ICON | 88,75        | 5,19            | 78,571                        | 98,920 |

( $p=0,0435$ )

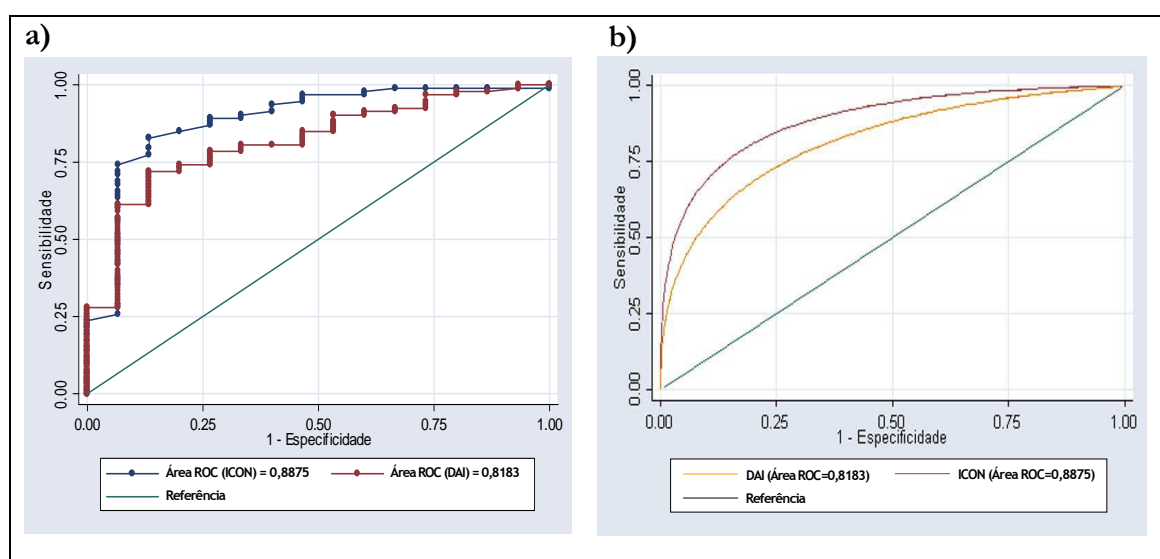


GRÁFICO 20 – Comparação entre as áreas sob as curvas ROC para o ICON e para o DAI: a) representação dos 108 valores medidos; b) curvas suavizadas.

Uma possível causa para os menores valores de validade obtidos pelo DAI, em relação ao ICON, é a desconsideração, em sua estrutura, das características relacionadas ao impacto da má oclusão sobre a saúde bucal (DANYLUK, 1999; OTUYEMI & NOAR 1996a). Assim, os casos considerados pelo painel com necessidade de tratamento devido a fatores não só relacionados à estética, mas ao conceito de oclusão ideal funcional, podem não ter sido classificados pelo DAI. Por outro lado, O ICON, ao considerar tanto a estética como as características funcionais, teria maior chance de refletir esses aspectos da tomada de decisão do painel.

Outro aspecto em relação à construção do DAI, que pode limitar a generalização do seu uso, é a forma como foram obtidos seus pontos de corte e a recomendação de sua validade tanto para países desenvolvidos e em desenvolvimento. Os pontos de corte foram obtidos através da correlação entre a distribuição de freqüências do valor numérico do DAI com a distribuição de categorias de gravidade de má oclusão segundo o relatório do Centro Nacional de Estatísticas de Saúde dos Estados Unidos da América (*National Center for Health Statistics – NCHS*). Os níveis de necessidade de tratamento foram deduzidos a partir da comparação do percentual acumulado entre as amostras (BEGLIN *et al.*, 2001; JENNY *et al.*, 1993, 1996a). A recomendação da generalização da sua validade foi baseada numa correlação entre opiniões de estudantes de quatro países a respeito do impacto estético da oclusão, avaliada em 25 fotografias (CONS *et al.*, 1989).

A utilização de um painel de ortodontistas de uma região geográfica específica impede a generalização de seus resultados. Há evidência de que o país onde o especialista trabalha tem efeito na sua avaliação da necessidade de tratamento ortodôntico (RICHMOND, 1998). Dessa forma, a validade de um índice depende do painel de especialistas que serve como o padrão-ouro. A qualidade dessa validação depende da minimização de erros de diferentes origens.

Após busca nas bases de dados *Medical Literature Analysis and Retrieval System Online* (MEDLINE), *Literatura Latino Americana e do Caribe em Ciências da Saúde* (LILACS) e *Scientific Electronic Library Online* (SciELO), até a data de 11 de novembro de 2007, utilizando os termos de pesquisa ICON, DAI, ortodontia, maloclusão, sensibilidade e especificidade, validade, necessidades e demanda de serviços de saúde, reprodutibilidade dos testes e necessidade de tratamento, não foram encontrados relatos de validação do DAI ou do ICON no Brasil, ou mesmo de estudos sobre a confiabilidade da opinião de

especialistas brasileiros sobre a necessidade de tratamento ortodôntico. Sugere-se que a confiabilidade dos valores do padrão-ouro utilizado seja melhor investigada.

Um nível baixo de confiabilidade do painel validador torna o próprio padrão-ouro muito subjetivo. Enquanto não houver padrões mais objetivos e baseados em evidência para a decisão de tratamento ortodôntico a opinião dos especialistas continua sendo a melhor referência. Em seu estudo, Lowerse *et al.* (2006), eliminaram os especialistas que obtiveram menores valores de confiabilidade. Conforme pode ser observado na TAB. 17, no presente estudo, a remoção dos especialistas com a confiabilidade mais baixa não alterou os valores da opinião do painel utilizada como padrão-ouro. Considerando a variância na opinião subjetiva do painel, índices poderiam ser ferramentas educacionais úteis para melhorar a concordância dos profissionais em relação à decisão da necessidade de tratamento ortodôntico.

Mesmo que o ICON tenha apresentado acurácia significativamente maior que o DAI, ambos podem ser considerados válidos para utilização como medida de necessidade normativa de tratamento ortodôntico em testes de rastreamento em levantamentos epidemiológicos. Numa situação de emprego dos índices em uma população real, algumas questões poderiam ser levantadas sobre a sua validade: se o valor medido pelo índice é alterado durante o período de desenvolvimento da oclusão; se a validade das medidas verificadas em modelos de estudo corresponde às das medidas em humanos; e se a decisão sobre a necessidade de tratamento corresponde à necessidade normativa que pode ou não estar relacionada à percepção do paciente ou de seus pais.

Conforme foi observado por Summers (1971), o desenvolvimento da oclusão pode alterar os valores medidos por um índice. Tarvitt e Freer (1998) observaram que ocorreu uma grande alteração na oclusão no período crítico entre os 12 e os 16 anos de idade, a qual reduziu, em média, 5,1 pontos no valor medido pelo DAI dentro desse período. Essas alterações dependem de quais componentes são medidos pelo índice. No caso do DAI, por exemplo, os componentes que mais contribuíram para essa alteração temporal foram o trespassamento horizontal maxilar, apinhamento e espaçamentos anteriores, diastemas na linha média e relacionamento intermolares. Esses componentes representam características muito mutáveis no desenvolvimento da oclusão. Como cada índice utiliza medidas de diferentes características da oclusão em seu cálculo, o impacto das alterações do desenvolvimento da oclusão no seu valor final pode não ser o mesmo para índices distintos. No presente estudo o DAI foi medido conforme adaptação para a dentadura

mista, proposta pela OMS, (1999), que considera algumas variações decorrentes do desenvolvimento da oclusão nos componentes dentes visíveis ausentes, espaçamento nos segmentos anteriores e relação ântero-posterior de molares. Com o objetivo de minimizar o efeito do desenvolvimento da oclusão na redução da pontuação do DAI, algumas características fisiológicas da dentadura mista, especialmente na “fase do patinho feio”, deveriam ser desconsideradas ou modificadas na avaliação: diastema na linha média, maior irregularidade anterior na maxila e na mandíbula e mordida aberta anterior. Mais estudos seriam necessários para determinar a magnitude dessas alterações temporais específicas para cada índice e estabelecer qual a correção mais adequada para cada caso: adaptações na estrutura do índice, nos pesos de categorias específicas ou até mesmo uma correção no valor final obtido, de acordo com a fase da dentição medida. Nesse estudo, foram utilizados dezesseis casos (14,8%) com dentadura mista. Se for considerado o efeito temporal, a pontuação para esses casos seria reduzida, o que potencialmente diminuiria o número de casos classificados com necessidade de tratamento pelo DAI.

A confiabilidade e a validade de avaliações realizadas em modelos de estudo comparadas à situação clínica podem ser questionadas. Esses aspectos necessitam de estudos adicionais, dada a importância da comparação de dados retrospectivos, geralmente obtidos de modelos de estudo, e de informações atuais obtidas clinicamente. Acredita-se que as medidas objetivas, provenientes de modelos de estudo, apresentam alta confiabilidade se comparadas às obtidas em exames clínicos. Buchanan *et al.* (1994) obtiveram concordância substancial quando compararam as medidas do componente de saúde dental do IOTN entre modelos e avaliação clínica. Ao comparar medidas subjetivas, como o componente estético do IOTN (escala SCAN), entre modelos de estudo e exame clínico, também foi obtida concordância substancial. Embora estudos dessa natureza não tenham sido realizados para o DAI e o ICON, Beglin *et al.* (2001) afirmam que a aplicação da maioria dos índices oclusais pode ser feita em modelos de estudo e não requer a presença do paciente, pois em geral testes de diagnóstico não exigem a presença do paciente, já que a fonte dos dados (sangue, tecido ou imagem) foi adquirida.

Num estudo realizado em escolares brasileiros, Peres *et al.* (2002) relataram que, entre as características biométricas da oclusão, a presença de apinhamento incisal e trespasse horizontal estavam associados à insatisfação com a aparência. Foi verificado também que a necessidade de tratamento ortodôntico foi melhor percebida pelos adolescentes que apresentaram irregularidade anterior na mandíbula, trespasse horizontal e

diastema anterior. Porém, os resultados não são conclusivos, visto que as características medidas foram somente alguns componentes do DAI que obtiveram melhores valores de concordância pelo examinador. Porém, mesmo se fosse utilizado em sua forma integral, o DAI não registra todas as características oclusais que poderiam estar relacionados à percepção de necessidade e satisfação com aparência.

De Oliveira e Sheiham (2004) avaliaram médias realizadas com o IOTN e o impacto na qualidade de vida de adolescentes brasileiros e relataram que, para melhorar as avaliações normativas de necessidade de tratamento ortodôntico, há necessidade incorporar medidas de fatores subjetivos relacionadas à qualidade de vida. Entretanto, esses autores utilizaram apenas o componente dental do IOTN, desconsiderando o componente estético do índice que poderia incluir esses fatores subjetivos. É possível que a escala SCAN incorpore parte dessa medida subjetiva da percepção da necessidade pelo paciente.

De forma geral, a utilização de índices oclusais para determinar a necessidade de tratamento ortodôntico apresenta limitações, inerentes ao procedimento, desde a determinação do padrão-ouro até a incorporação da percepção do paciente sobre a necessidade do tratamento. O DAI e o ICON, apesar de parecerem semelhantes estruturalmente, apresentam diferenças significativas em suas validades. Por apresentar maior acurácia e contemplar importantes componentes relacionados à saúde, além dos aspectos estéticos, o ICON poderia substituir o DAI como índice de necessidade de tratamento ortodôntico.



## 6 CONCLUSÃO

De acordo com a metodologia empregada no presente estudo, pode-se concluir que:

- a) a utilização dos índices como testes de rastreamento em levantamentos epidemiológicos é recomendada, tendo sido verificada boa acurácia para o DAI (81,83%) e para o ICON (88,75%), sendo este último superior;
- b) o ICON apresentou razoável sensibilidade (74,19%) e excelente especificidade (93,33%), empregando o ponto de corte preconizado por seus idealizadores;
- c) o DAI apresentou pobre sensibilidade (67,74%) e boa especificidade (86,67%), empregando o ponto de corte preconizado por seus idealizadores;
- d) o painel de especialistas apresentou concordância intra-examinadores que variou dos níveis pobre a excelente, e a concordância interexaminadores variou de pobre a boa; o que evidencia a necessidade da especialidade desenvolver critérios mais objetivos de diagnóstico da necessidade de tratamento ortodôntico;
- e) o pesquisador apresentou excelente reprodutibilidade na aplicação dos índices DAI ( $r_1 \geq 0,969$ ) e ICON ( $r_1 \geq 0,952$ ).

## REFERÊNCIAS<sup>6</sup>

1. ANTUNES, J. L. F.; PERES, M. A. *Fundamentos de odontologia: epidemiologia da saúde bucal*. Rio de Janeiro: Guanabara Koogan, 2006. 472 p.
2. BECK, J. R.; SHULTZ, E. K. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch Pathol Lab Med*. Chicago, v. 110, n. 1, p. 13-20, Jan. 1986.
3. BEGLIN, F. M.; FIRESTONE, A. R.; VIG, K. W. L.; BECK, F. M.; KUTHY, R. A.; WADE, D. A comparison of the reliability and validity of 3 occlusal indexes of orthodontic treatment need. *Am. J. Orthod. Dentofacial Orthop*. Saint Louis, v. 120, n. 3, p. 240-246, Sep. 2001.
4. BENTELE, M. J.; VIG, K. W. L.; SHANKER, S.; BECK, F. M. Efficacy of training dental students in the index of orthodontic treatment need. *Am. J. Orthod. Dentofacial Orthop*. Saint Louis, v. 122, n. 5, p. 456-462, Nov. 2002.
5. BLAND, J. M.; ALTMAN, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* Thousand Oaks, v. 8, n. 4, p. 135-160, 1999.
6. BLAND, J. M.; ALTMAN, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. London, v. 1, n. 8476, p. 307-310, Feb. 1986.
7. BRASIL. Ministério da Saúde - Divisão Nacional de Saúde Bucal. *Projeto SB 2003: Condições de Saúde Bucal da População Brasileira 2002-2003: Resultados Principais*. Brasília: Ministério da Saúde, 2004.
8. BRESOLIN, D. Controle e prevenção da maloclusão. In: PINTO, V. G. *Saúde Bucal Coletiva*. 4ed. São Paulo: Livraria Editora Santos, 2000. Cap. 18, p. 473-479.
9. BROOK, P. H.; SHAW, W. C. The development of an index of orthodontic treatment priority. *Eur. J. Orthod*. Oxford, v. 11, n. 3; p. 309-320, Aug. 1989.
10. BROWN, R.; RICHMOND, S. An update on the analysis of agreement for orthodontic indices. *Eur. J. Orthod*. Oxford, v. 27, n. 3, p. 286-291, Jun. 2005.
11. BUCHANAN, I. B.; DOWNING, A.; STIRRUPS, D. R. A comparison of the index of orthodontic treatment need applied clinically and to diagnostic records. *Br. J. Orthod*. Oxford, v. 21, n. 2, p. 185-188, May 1994.
12. CARLOS, J. P. Evaluation of indices of malocclusion. *Int. Dent. J.* Den Haag, v. 20, n. 4, p. 606-617, Dec. 1970.
13. CICCETTI, D. V. Assessing inter-rater reliability for rating scales: resolving some basic issues. *Br. J. Psychiatry*. London, v. 129, p. 452-456, 1976.
14. CICCETTI, D. V. Methodological commentary: the precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. *J. Clin. Exp. Neuropsychol*. Lisse, v. 23, n. 5, p. 695-700, Oct. 2001.

---

<sup>6</sup> Foram seguidas as seguintes normas da ABNT: NBR-6023, NBR-6024, NBR-6032, NBR-14724.

15. CICCETTI, D. V.; VOLKMAR, F.; SPARROW, S. S.; COHEN, D.; FERMANIAN, J.; ROURKE, B. P. Assessing the reliability of clinical scales when the data have both nominal and ordinal features: proposed guidelines for neuropsychological assessments. *J. Clin. Exp. Neuropsychol.* Lisse, v. 14, n. 5, p. 673-686, Sep. 1992.
16. COHEN, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* Durham, v. 20, n. 1, p. 37-46, 1960.
17. COHEN, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* Washington, v. 70, n. 4, p. 213-220, Oct. 1968.
18. COHEN, L. K.; JAGO, J. D. Toward the formulation of sociodental indicators. *Int. J. Health Serv.* Westport, v. 6, n. 4, p. 681-698, 1976.
19. CONS, N. C.; JENNY, J.; KOHOUT, F. J.; SONGPAISAN, Y.; JOTIKASTIRA, D. Utility of the dental aesthetic index in industrialized and developing countries. *J. Pub. Health Dent.* Raleigh, v. 49, n. 3, p. 163-166, Summer, 1989.
20. COOKE, M. Assessment of the validity of the index of complexity, outcome and need (ICON), and handicapping labio-lingual deviations with California modifications (HLD CalMod). *Am. J. Orthod. Dentofacial Orthoped.* Saint Louis, v. 130, n. 6, p. 806. Dec. 2006. Thesis abstract.
21. COSTA, A. J. L.; NADANOVSKY, P. Validade na pesquisa odontológica In: LUIZ, R. R.; COSTA, A. J. L.; NADANOVSKY, P. *Epidemiologia e bioestatística na pesquisa odontológica.* Rio de Janeiro: Atheneu, 2005a. Cap. 6, p. 125-141.
22. COSTA, A. J. L.; NADANOVSKY, P. Teste diagnóstico e rastreamento (screening) em saúde bucal. In: LUIZ, R. R.; COSTA, A. J. L.; NADANOVSKY, P. *Epidemiologia e bioestatística na pesquisa odontológica.* Rio de Janeiro: Atheneu, 2005b. Cap. 9, p. 181-198.
23. DANIELS, C.; RICHMOND, S. The development of the index of complexity, outcome and need (ICON). *Br. J. Orthod.* Oxford, v. 27, n. 2, p. 149-162, Jun. 2000.
24. DANYLUK, K.; LAVELLE, C.; HASSARD, T. Potential application of the dental aesthetic index to prioritize the orthodontic service needs in a public funded dental program. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 116, n. 3, p. 279-286, Sep. 1999.
25. DEGUZMAN, L.; BAHIRAEI, D.; VIG, K. W.; VIG, P. S.; WEYANT, R. J.; O'BRIEN, K. The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 107, n.2, p. 172-176, Feb. 1995.
26. DELONG, E. R.; DELONG, D. M.; CLARKE-PEARSON, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* Washington, v. 44, n.3, p. 837-845, Sep. 1988.
27. DE OLIVEIRA, C. M.; SHEIHAM, A. Orthodontic treatment and its impact on oral health-related quality of life in Brazilian adolescents. *Br. J. Orthod.* Oxford, v. 31, n. 1, p. 20-27, Jan. 2004.
28. DOHOO, I.; MARTIN, W.; STRYHN, H. *Veterinary epidemiologic research.* Charlottetown: AVC, 2003. 706 p.
29. DRAKER, H. L. Handicapping labio-lingual deviations: a proposed index for public health purposes. *Am. J. Orthod.* Saint Louis, v. 46, n. 4, p. 295-305, Apr. 1960.

30. ESPELAND, L. V.; IVARSSON, K.; STENVIK, A.; ALSTAD, T. A. Perception of malocclusion in 11-year-old children: a comparison between personal and parental awareness. *Eur. J. Orthod.* Oxford, v 14, n. 5, p.350-358, Oct. 1992.
31. EVANS, R.; SHAW, W. Preliminary evaluation of an illustrated scale for rating dental attractiveness. *Eur. J. Orthod.* Oxford, v.9, n. 4, p. 314-318, Nov. 1987.
32. FIRESTONE, A. R.; BECK, F. M.; BEGLIN, F. M.; VIG, K. W. L. Validity of the index of complexity, outcome and need (ICON) in determining orthodontic treatment need. *Angle Orthod.* Appleton, v. 72, n. 1, p. 15-20, Feb. 2002a.
33. FIRESTONE, A. R.; BECK, F. M.; BEGLIN, F. M.; VIG, K. W. L. Evaluation of the peer assessment rating (PAR) index as an index of orthodontic treatment need. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 122, n. 5, p. 463-469, Nov. 2002b.
34. FLEISS, J. L. *Statistical methods for rates and proportions*. 2 ed. New York: John Wiley & Sons, 1981. 321 p.
35. FLEISS, J. L. *The Design and analysis of clinical experiments*. New York: John Wiley & Sons, 1986, 432 p.
36. FLEISS, J. L.; COHEN, J. The equivalence of weighted kappa and the intra-class correlation coefficient as measures of reliability. *Educ. Psychol. Meas.* Durham, v. 33, n. 3, p. 613-619, Autumn, 1973.
37. FLEISS, J. L.; COHEN, J.; EVERITT, B. S. Large sample standard errors of kappa and weighted kappa. *Psychol. Bull.* Washington, v. 72, n. 5, p. 323-327, 1969.
38. FOX, N. A.; DANIELS, C.; GILGRASS, T. A comparison of the Index of Complexity Outcome and Need (ICON) with the Peer Assessment Rating (PAR) and the Index of Orthodontic Treatment Need (IOTN). *Br. Dent. J.* London, v. 193, n. 4, p. 225-230, Aug. 2002.
39. GOCHMAN, D. S.; The measurement and development of dentally relevant motives. *J. Pub. Health Dent.* Raleigh, v. 35, n. 3, p. 160-164, Summer, 1975.
40. GORDIS, L. *Epidemiologia*. 2 ed. Rio de Janeiro: Revinter, 2004. 302 p.
41. GRABER, L. W.; LUCKER, G. W. Dental esthetic self-evaluation and satisfaction. *Am. J. Orthod.* Saint Louis, v. 77, n. 2, p. 163-173, Feb. 1980.
42. HANLEY, J. A.; McNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* Detroit, v. 143, n. 1, p. 29-36, Apr. 1982.
43. HOLMES, A. The prevalence of orthodontic treatment need. *Br. J. Orthod.* Oxford, v. 19, n. 3, p. 177-182, Oct. 1992a.
44. HOLMES, A. The subjective need and demand for orthodontic treatment. *Br. J. Orthod.* Oxford, v. 19, n. 4, p. 287-297, Nov. 1992b.
45. JÄRVINEN, S. Indexes for orthodontic treatment need. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 120, n. 3, p. 237-239, Sep. 2001.
46. JENNY, J.; CONS, N. C. Establishing malocclusion severity levels on the dental aesthetic index (DAI) scale. *Aust. Dent. J.* North Sydney, v. 41 n. 1, p. 43-46. Feb. 1996a.
47. JENNY, J.; CONS, N. C. *Guidelines for using the DAI: a supplement to DAI – the dental aesthetic index*. Iowa City: College of Dentistry, The University of Iowa, 1988. 56 p.

48. JENNY, J.; CONS, N. C.; Comparing and contrasting two orthodontic indices, the index of orthodontic treatment need and the dental aesthetic index. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 110, n. 4, p. 410-416, Oct. 1996b.
49. JENNY, J.; CONS, N. C.; KOHOUT, F. J.; JAKOBSEN, J. Predicting handicapping malocclusions using the dental aesthetic index (DAI). *Int. Dent. J.* Den Haag, v. 43 n. 2, p. 128-132. Apr. 1993.
50. KATTNER, P. F.; SCHNEIDER, B. J. Comparison of Roth appliance and standard edgewise appliance treatment results. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 103, n. 1, p. 24-32, Jan. 1993.
51. KEAY, P. A.; FREER, T. J.; BASFORD, K. E. Orthodontic treatment need and the dental aesthetic index. *Aus. Orthod. J.* Brisbane, v. 13, n. 1, p. 4-7, Oct. 1993.
52. LANDIS, J. R.; KOCH, C. G. The measurement of observer agreement for categorical data. *Biometrics.* Washington, v. 33, n. 1, p. 159-174, Mar. 1977.
53. LEWITT, D. W.; VIROLAINEN, K. Conformity and independence in adolescents' motivation for orthodontic treatment. *Child Dev.* Chicago, v. 39, n. 4, p. 1189-1200, Dec. 1968.
54. LIEPA, A.; URTANE, I.; RICHMOND, S.; DUNSTAN, F. Orthodontic treatment need in Latvia. *Eur. J. Orthod.* Oxford, v. 25, n. 3, p. 279-284, Jun. 2003.
55. LINDER-ARONSON, S. Orthodontics in the Swedish Public Dental Health Service. *Trans. Eur. Orthod. Soc.* London, p. 233-240. 1974.
56. LINDSAY, S. J. E.; HODGKINS, J. F. W. Children's perception of their own malocclusions. *Br. J. Orthod.* Oxford, v. 10, n. 1, p. 13-20, Jan. 1983.
57. LOWERSE, T. J.; AARTMAN, I. H. A.; KRAMER, G. J. C.; PRAHL-ANDERSEN, B. The reliability and validity of the index of complexity, outcome and need for determining treatment need in Dutch orthodontic practice. *Eur. J. Orthod.* Oxford, v. 28, n. 2, p. 58-64, Feb. 2006.
58. LUIZ, R. R. Erros de Mensuração. In: *Epidemiologia e bioestatística na pesquisa odontológica.* Rio de Janeiro: Atheneu, 2005. Cap. 5, p. 91-124.
59. LUTHER, F. Orthodontics and the temporomandibular joint: where are we now? Part 2. Functional occlusion, malocclusion and TMD. *Angle Orthod.* Appleton, v. 68, n. 4, p. 305-318, Aug. 1998.
60. MARQUES, L. S.; BARBOSA, C. C.; RAMOS-JORGE, M. L.; PORDEUS, I. A.; PAIVA, S. M. Prevalência da maloclusão e necessidade de tratamento ortodôntico em escolares de 10 a 14 anos de idade em Belo Horizonte, Minas Gerais, Brasil: enfoque psicossocial. *Cad. Saúde Pública.* Rio de Janeiro, v. 21, n. 4, p. 1099-1106, jul-ago, 2005.
61. MARQUES, L. S.; RAMOS-JORGE, M. L.; PAIVA, S. M.; PORDEUS, I. A. Malocclusion: esthetic impact and quality of life among Brazilian schoolchildren. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 129, n. 3, p. 424-427, Mar. 2006.
62. McKIERNAN, E. X. F.; McKIERNAN, F.; JONES, M. L. Psychological profiles and motives of adults seeking orthodontic treatment. *Int. J. Adult. Orthod. Orthognath. Surg.* Lombard, v. 7, n. 3, p. 187-198, 1992.
63. McLAIN, J. B.; PROFFITT, W. R. Oral health status in the United States: prevalence of malocclusion. *J. Dent. Educ.* Washington, v. 49, n. 6, p. 386-396, Jun. 1985.
64. METZ, C. E. Basic principles of ROC analysis. *Semin. Nucl. Med.* Orlando, v.8, n.4, p. 283-298, Oct. 1978.

65. MONTGOMERY, D. C.; RUNGER, G. C. *Estatística aplicada e probabilidade para engenheiros*. Rio de Janeiro: LTC, 2003. 463 p.
66. NELSON, S. Epidemiology for the practicing orthodontist. *Semin. Orthod.* Orlando, v. 5, n. 2, p. 77-84, Jun. 1999.
67. NIELSEN, I. L. Vertical malocclusions: etiology, development, diagnosis and some aspects of treatment. *Angle Orthod.* Appleton, v. 61, n. 4, 247-260, Dec. 1991.
68. NIELSEN, I. L. Vertical malocclusions: etiology, development, diagnosis and some aspects of treatment – Author's response. *Angle Orthod.* Appleton, v. 62, n. 2, p. 87, Jun. 1992.
69. ORGANIZAÇÃO MUNDIAL DE SAÚDE. *Levantamentos básicos em saúde bucal*. 4ed. São Paulo: Santos, 1999. 66 p.
70. OTUYEMI, O. D.; NOAR, J. H. Variability in recording and grading the need for orthodontic treatment using the handicapping malocclusion assessment record, occlusal index and dental aesthetic index. *Community Dent. Oral. Epidemiol.* Copenhagen, v. 24, n. 3, p. 222-224, Jun. 1996a.
71. OTUYEMI, O. D.; NOAR, J. H. A comparison between DAI and SCAN in estimating orthodontic treatment need. *Int. Dent. J.* Den Haag, v. 46, n. 1, p. 35-40, Feb. 1996b.
72. PEREIRA, M. G. Aferição de eventos. In: PEREIRA, M. G. *Epidemiologia: teoria e prática*. Rio de Janeiro: Guanabara Koogan, 1995. Cap. 17, p. 354-376.
73. PERES, K. G.; TRAEBERT, E. S. A.; MARCENES, W. Diferenças entre autopercepção e critérios normativos na identificação de oclusopatias. *Rev Saúde Pública* São Paulo, v. 36, n. 2, p. 230-236, 2002.
74. PETERSEN, B.; DAHLSTRÖM, L. Perception of treatment need among orthodontic patients compared with professionals. *Acta Odontol. Scand.* Oslo, v. 56, n. 5, p.299-302, Oct. 1998.
75. POLEDNAK, A. P. Geographic variation in postmastectomy breast reconstruction rates. *Plast. Reconstr. Surg.* Philadelphia, v. 106, n. 2, p. 298-301, Aug. 2000.
76. PRAHL-ANDERSEN, B. The need for orthodontic treatment. *Angle Orthod.* Appleton, v. 48, n. 1, p. 1-9, Jan. 1978.
77. RICHMOND, S.; BUCHANAN, I. B.; BURDEN, J. B.; O'BRIEN, K. D.; ANDREWS, M.; ROBERTS, C. T.; TURBILL, E. A. Calibration of dentists in the use of occlusal indices. *Community Dent. Oral Epidemiol.* Copenhagen, v. 23, n. 3, p. 173-176, Jun. 1995.
78. RICHMOND, S.; DANIELS, C. P. International comparisons of professional assessments in orthodontics: Part 1--Treatment need. *Am. J. Orthod Dentofacial Orthop.* Saint Louis, v. 113, n. 2, p. 180-185, Feb. 1998.
79. RICHMOND, S.; O'BRIEN, K. D.; ROBERTS, C. T.; ANDREWS, M. Dentists variation in the determination of orthodontic treatment need. *Br. J. Orthod.* Oxford, v. 21, n. 1, p. 65-68, Feb. 1994.
80. RICHMOND, S.; SHAW, W. C.; O'BRIEN, K. D.; BUCHANAN, I. B.; JONES, R.; STEPHENS, S. D.; ROBERTS, C. T.; ANDREWS, M. The development of the PAR index (peer assessment rating):reliability and validity. *Eur. J. Orthod.* Oxford, v. 14, n. 2, p. 125-139, Apr. 1992.
81. ROBERTS, C. T.; RICHMOND, S. The design and analysis of reliability studies for the use of epidemiological and audit indices in orthodontics. *Br. J. Orthod.* Oxford, v. 24, n. 2, p. 139-147, May. 1997.

82. SALZMANN, J. A. Handicapping malocclusion assessment to establish treatment priority. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 54, n. 10, p. 749-765, Oct. 1968.
83. SHAW, W. C. Factors influencing the desire for orthodontic treatment. *Eur. J. Orthod.* Oxford, v. 3, n. 3, p. 151-162, 1981.
84. SHAW, W. C.; LEWIS, H. G.; ROBERTSON, N. R. E. Perception of malocclusion. *Br. Dent. J.* London, v. 138, n. 6, p. 211-216, Mar. 1975
85. SHAW, W. C.; RICHMOND, S.; O'BRIEN, K. D. Quality control in orthodontics: risk/benefit considerations. *Br. Dent. J.* London, v. 170, n. 5, p. 33-37, Jan. 1991a.
86. SHAW, W. C.; RICHMOND, S.; O'BRIEN, K. D. The use of occlusal indices: a European perspective. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 107, n. 1, p. 1-10, Jan. 1995.
87. SHAW, W. C.; RICHMOND, S.; O'BRIEN, K. D.; BROOK, P. Quality control in orthodontics: indices of treatment need and treatment standards. *Br. Dent. J.* London, v. 170, n. 3, p. 107-112, Feb. 1991b.
88. SHEIHAM, A. A. A determinação de necessidade de tratamento odontológico: um abordagem social. In: PINTO, V. G. *Saúde Bucal Coletiva*. 4ed. São Paulo: Santos, 2000. Cap. 6, p. 223-250.
89. SOLOW, B. Guest editorial: orthodontic screening and thirty party financing. *Eur. J. Orthod.* Oxford, v. 17, n. 1, p. 79-83, Feb. 1995.
90. STATISTICS.COM. Disponível em: <http://www.statistics.com/resources/glossary/v/validity.php>. Acessado em: 26 de novembro de 2005.
91. STEDMAN, T. L. *Stedman's medical dictionary*. 26ed. Baltimore: Williams & Wilkins, 1995. 2029 p.
92. SUMMERS, C. J. The occlusal index: a system for identifying and scoring occlusal disorders. *Am. J. Orthod. Dentofacial Orthop.* Saint Louis, v. 59, n. 6, p. 552-567, Jun. 1971.
93. SWETS, J. A. Measuring the accuracy of diagnostic systems. *Science*. Washington, v. 240, n. 3, p. 1285-1293, Jun. 1988.
94. TARVIT, D. J.; FREER, T. J. Assessing malocclusion: the time factor. *Br. J. Orthod.* Oxford, v. 25, n. 1, p.31-34, Feb. 1998.
95. TULLOCH, J. F. C.; SHAW, W. C.; UNDERHILL, C.; SMITH, A.; JONES, G.; JONES, M. A comparison of attitudes toward orthodontic treatment in British and American communities. *Am. J. Orthod.* Saint Louis, v. 85, n. 3, p. 253-259, Mar. 1984.
96. WILT, T. J.; COWPER, D. C.; GAMMACK, J. K.; GOING, D. R.; NUGENT, S.; BOROWSKY, S. J. An evaluation of radical prostatectomy at Veterans Affairs Medical Centers: time trends and geographic variation in utilization and outcomes. *Med. Care*. Philadelphia, v. 37, n. 10, p. 1046-1056, Oct. 1999.
97. WINTER, A. A. Measurement of the millimeter markings of periodontal probes. *J. Periodontal Res.* Copenhagen, v. 50, n. 9, p.483-485, Sep. 1979.
98. YOUNIS, J. W.; VIG, K. W. L.; RINCHUSE, D. J.; WEYANT, R. J. A validation study of three indexes of orthodontic treatment need in the United States. *Community Dent. Oral Epidemiol.* Copenhagen, v. 25, n. 5, p. 358-362, Oct. 1997.

# APÊNDICES

## APÊNDICE A

### MANUAL DE REFERÊNCIA: Critérios para Aplicação do DAI para Necessidade de Tratamento Ortodôntico

O índice de estética dental (DAI) é um índice quantitativo, proposto em 1986 por CONS e JENNY (JENNY & CONS, 1996a; 1988) para avaliar o grau de impacto estético da dentição. São avaliadas dez características oclusais conforme padrões socialmente definidos como aceitáveis para estética dental. Cada uma das características é pontuada, multiplicada pelos seu respectivo coeficiente (peso) e os resultados são totalizados com a adição de uma constante (QUADRO 2). O resultado final é um valor único que combina os aspectos estéticos e físicos da oclusão (JENNY & CONS, 1996a). O DAI é um valor que varia de 13 a 80 ou mais, que pode ser categorizado através de pontos de corte.

**QUADRO 2:** Componentes da equação de regressão do DAI com seus respectivos coeficientes (pesos)

| Componentes do DAI  | Valor | Coeficientes | subtotal     |
|---|-------|--------------|--------------|
| Número de dentes visíveis ausentes  |       | 5,76         |              |
| Apinhamento anterior (0=nenhum, 1=um segmento, 2=2segmentos)              |       | 1,15         |              |
| Espaçamento anterior (0=nenhum, 1=um segmento, 2=2segmentos)              |       | 1,31         |              |
| Diastema na linha média (mm)  |       | 3,13         |              |
| Maior irregularidade anterior na maxila (mm)                              |       | 1,34         |              |
| Maior irregularidade anterior na mandíbula (mm)                           |       | 0,75         |              |
| Trespasse horizontal maxilar (mm)   |       | 1,62         |              |
| Trespasse horizontal mandibular (mm)                                      |       | 3,68         |              |
| Mordida aberta anterior (mm)  |       | 3,69         |              |
| Relação ântero-posterior de molares (0=normal, 1=1/2 cúspide, 2=1cúspide) |       | 2,69         |              |
| <b>Constante (adicionar aos subtotais)</b>                                |       |              | <b>13,36</b> |
| <b>Valor total do índice DAI</b>  |       |              |              |

FONTE: JENNY & CONS, 1988, adaptada.

Para categorização dicotômica em “com necessidade” (valores iguais ou maiores que 31) e “sem necessidade” (valores menores que 31) de tratamento ortodôntico, o valor do ponto de corte recomendado é igual a 31 (QUADRO. 3) (B EGLIN *et al.*, 1991; JENNY & CONS, 1996a).

**QUADRO 3:** Categorização do DAI a partir dos pontos de corte

| Pontos de corte | Categorias   |                           |
|-----------------|--|---------------------------|
|                 | Ordinais   | Dicotômicas               |
| 13 a 25         | Oclusão normal ou mínima má oclusão                              | Tratamento não necessário |
| 26 a 30         | Má oclusão definida (tratamento eletivo)                         | Tratamento necessário     |
| 31 a 35         | Má oclusão grave (tratamento altamente desejável)                | Tratamento necessário     |
| 36 a mais       | Má oclusão muito grave ou incapacitante (tratamento obrigatório) | Tratamento necessário     |

FONTE: JENNY & CONS, 1996a; 1996b; B EGLIN *et al.* 1991; adaptado.



## CÓDIGOS E CRITÉRIOS DE AVALIAÇÃO

É recomendável que a avaliação seja realizada em indivíduos que já não possuam dentes decíduos, ou seja, em idades a partir dos 12 anos, embora o índice possa ser adaptado para a dentadura mista com algumas alterações. Em vez de contar o número de dentes anteriores ausentes, o espaço resultante da recente esfoliação de dente decíduo não deve ser considerado se o sucessor permanente parecer estar em vias de irromper logo (JENNY & CONS, 1988; 1996a; OMS, 1999). As medições não são realizadas em dentes não irrompidos. Considerando as características fisiológicas do desenvolvimento da oclusão, outras adaptações para medições na dentadura mista foram feitas neste estudo para os componentes: dentes visíveis ausentes, diastema na linha média, maior irregularidade anterior na maxila e na mandíbula, mordida aberta anterior e relação ântero-posterior de molares. Estas modificações estão descritas para cada componente separadamente.

Cada componente é avaliado como presente ou ausente ou medido em milímetros inteiros. Apenas para a relação ântero-posterior de molares o desvio mais grave do normal é registrado. Jenny e Cons (1988) recomendam que as medidas sejam realizadas com um paquímetro de Boley ou com um abaixador de língua de madeira e um lápis bem apontado. Como instrumento de medida, a Organização Mundial de Saúde (OMS) recomenda a utilização de uma sonda CPI (OMS, 1999). Será utilizado método de medida recomendado pela OMS e utilizado no levantamento epidemiológico nacional SB-2003 (BRASIL, 2001). Entretanto, para maior precisão de medida, será utilizada a sonda milimetrada modelo Universidade Carolina do Norte # 15 (WINTER, 1979) (FIG.4).

O formulário para aplicação do DAI é semelhante ao mostrado no QUADRO 1. O tempo médio para a aplicação é de aproximadamente 3 minutos.



**FIGURA 4.:** Sondas periodontais CPI-OMS (a) e Universidade Carolina do Norte (b).  
 FONTE: ANTUNES & PERES, 2006.(a); Hu-Friedy do Brasil, 2006 (b).

### Dentes Visíveis Ausentes

São contados os dentes visíveis ausentes. O número de incisivos, caninos e pré-molares que faltam pode variar de 0 a 10 para cada arcada. A contagem deve iniciar-se no segundo pré-molar direito, continuando para a região anterior até o segundo pré-molar esquerdo (FIG.5).

O total de dentes ausentes para cada arcada, maxilar e mandibular, deve ser anotado em seus respectivos campos (JENNY & CONS, 1988; OMS, 1999).

#### Observações:

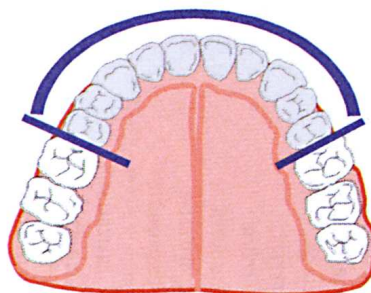
Há situações em que dentes ausentes não serão contados como espaço, caso:

- os espaços resultantes de ausências dentárias estejam fechados;
- haja algum dentes decíduo ainda no arco e o sucessor não irrompido;

- o dente ausente tenha sido substituído por prótese fixa.

#### **Modificações para dentadura mista:**

- O espaço deixado pela esfoliação de dente decíduo não será contado caso esteja na época ou seqüência de troca corretas e não haja indícios da ausência do permanente sucessor.



**FIGURA 5.:** Região para avaliação de dentes ausentes.

FONTE: ANTUNES & PERES, 2006

### **Apinhamento Anterior**

A presença de apinhamento nos segmentos anteriores, compreendendo quatro incisivos, deve ser observada nas arcadas superior e inferior. Considera-se apinhamento quando o **espaço entre os caninos direito e esquerdo** é insuficiente para acomodar os quatro incisivos alinhados. O apinhamento na região anterior é pontuado da seguinte maneira (JENNY & CONS, 1988; OMS, 1999):

- 0 → sem apinhamento nos segmentos superior ou inferior;
- 1 → apinhamento em um segmento (superior ou inferior);
- 2 → apinhamento em dois segmentos (superior e inferior);

#### **Observações:**

Caso haja alguma **dúvida**, a pontuação mais baixa deverá ser registrada.

Se um ou ambos os **caninos estiverem deslocados** e os incisivos alinhados, não deverá ser considerado apinhamento.

Se há **espaço suficiente** para o alinhamento dos incisivos mas estes encontram-se deslocados da linha do arco ou girovertidos, **não deverá ser registrado como apinhamento.**

### **Espaçamento anterior**

Espaços, ou falta de contato interproximal, nos segmentos anteriores superior e inferior deverão ser considerados espaçamento quando o **espaço entre os caninos direito e esquerdo** for maior que o necessário para o alinhamento normal dos incisivos. Nos casos de falta de espaço no segmento anterior, mesmo que haja falta de contato interproximal ou espaço entre dois ou mais incisivos, não é considerado espaçamento (JENNY & CONS, 1988; OMS, 1999).

- 0 → sem espaçamento
- 1 → espaçamento em um segmento (superior ou inferior)
- 2 → espaçamento em dois segmentos (superior e inferior)

**Observações:**

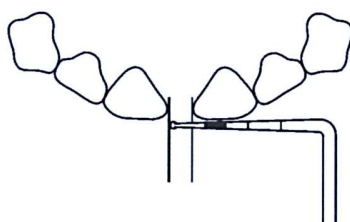
Caso haja alguma dúvida, a pontuação mais baixa deverá ser registrada.

O espaço resultante da perda de um dente decíduo não deverá ser registrado como espaçamento se houver indicação de que o sucessor permanente irromperá logo.

### Diastema na linha média

O diastema na linha média é o espaço entre os incisivos centrais permanentes superiores e é registrado conforme sua medida em milímetros (FIG.6). Esta medida poderá ser realizada em qualquer nível das superfícies mesiais dos incisivos centrais e o valor, em milímetros, pode ser arredondado para um número inteiro mais próximo.

Para padronizar a medição e reduzir a possibilidade de erro, esta deverá ser realizada na altura do equador dos dentes adjacentes ou no ponto de maior convexidade da superfície proximal, onde deveria ser o ponto de contato (JENNY & CONS, 1988; OMS, 1999).



**FIGURA 6.:** Diastema na linha mediana.  
FONTE: OMS, 1999.

**Modificações para dentadura mista:**

O diastema na linha média superior é fisiológico até o final da dentadura mista. Um diastema de até um milímetro não será marcado caso os caninos permanentes ainda não tenham irrompido.

### Maior irregularidade anterior na maxila

São consideradas irregularidades anteriores na maxila as giroversões e deslocamentos em relação ao alinhamento normal dos quatro incisivos superiores. Podem ocorrer com ou sem apinhamento. A maior irregularidade observada entre os incisivos adjacentes ou entre os incisivos laterais e caninos é medida com a sonda periodontal. A ponta da sonda é apoiada na superfície vestibular do incisivo mais linguvertido ou girovertido, perpendicularmente à linha do arco e paralelamente ao plano oclusal (FIG.7). A irregularidade é estimada, em milímetros, através da referência marcada na sonda. O valor, em milímetros, pode ser arredondado para um número inteiro mais próximo (JENNY & CONS, 1988; OMS, 1999).

**Observações:**

Podem ocorrer irregularidades com ou sem apinhamento. Se há espaço suficiente para o alinhamento dos quatro incisivos, a irregularidade deverá ser registrada, enquanto que o apinhamento não.

**Modificações para dentadura mista:**

As irregularidades entre incisivos laterais e caninos não serão marcadas se os caninos permanenetes ainda não irromperam.



**FIGURA 7.:** Maior irregularidade anterior na maxila.

FONTE: OMS, 1999. Modificada.

**Maior irregularidade anterior na mandíbula**

As irregularidades anteriores na mandíbula são as giroversões e deslocamentos em relação ao alinhamento normal dos quatro incisivos inferiores. As medidas são realizadas e registradas da mesma forma como descritas acima para a maxila (JENNY & CONS, 1988; OMS, 1999).

**Modificações para dentadura mista:**

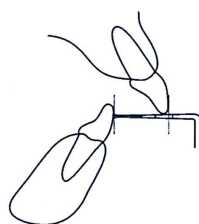
As irregularidades entre incisivos laterais e caninos não serão marcadas se os caninos permanenetes ainda não irromperam.

**Trespasse horizontal positivo excessivo (sobressaliência maxilar anterior)**

É a distância entre a borda incisal vestibular do incisivo superior mais protruído e a superfície vestibular do incisivo inferior correspondente, observada em oclusão cêntrica, e medida com a sonda periodontal colocada em contato com a superfície vestibular do incisivo inferior, paralelamente ao plano oclusal (FIG.8) e perpendicular à linha do arco. Se o incisivo superior estiver girado, utilizar o ponto mais protruído como referência de medida. O valor, em milímetros, é arredondado para um número inteiro mais próximo (JENNY & CONS, 1988; OMS, 1999).

**Observações:**

Esta característica **não** deverá ser medida caso os incisivos superiores estejam **ausentes**, ou haja **mordida cruzada anterior**. Nestes casos a medida será zero.



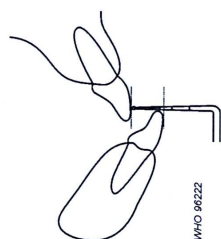
**FIGURA 8.:** Trespasse horizontal positivo. FONTE: OMS, 1999.

### Sobressaliência mandibular anterior

Esta característica é registrada caso haja mordida cruzada anterior, ou haja qualquer incisivo inferior protruído (ou labiovertido) em relação ao seu antagonista. A medida é realizada com a sonda periodontal sobre o incisivo inferior mais protruído em relação ao seu antagonista (FIG.9). É registrado o maior trespasse horizontal mandibular encontrado (protrusão mandibular) ou mordida cruzada, da mesma forma como descrito acima para o trespasse horizontal excessivo, arredondando-se os milímetros para o número inteiro mais próximo (JENNY & CONS, 1988; OMS, 1999).

#### Observações:

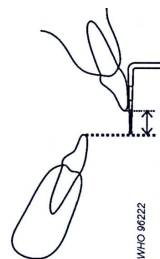
Não deverá ser registrada a sobressaliência mandibular anterior caso apenas parte da borda incisal de um incisivo inferior girado esteja protruída em relação ao seu antagonista.



**FIGURA 9.:** Trespasse horizontal negativo. FONTE: OMS, 1999.

### Mordida aberta anterior

A falta de trespasse vertical entre incisivos antagonistas é definida como mordida aberta anterior. A medida da mordida aberta é estimada com a utilização da sonda periodontal, colocada de forma perpendicular ao plano oclusal nos dentes onde a mordida aberta apresenta a maior medida (FIG.10). Com o objetivo de padronizar a mensuração e reduzir erros, a medida deve ser realizada preferencialmente a partir do meio da borda incisal dos dentes envolvidos. O valor, em milímetros, é arredondado para o número inteiro mais próximo (JENNY & CONS, 1988; OMS, 1999).



**FIGURA 10.:** Medição da mordida aberta anterior.  
FONTE: OMS, 1999.

#### Modificações para dentadura mista:

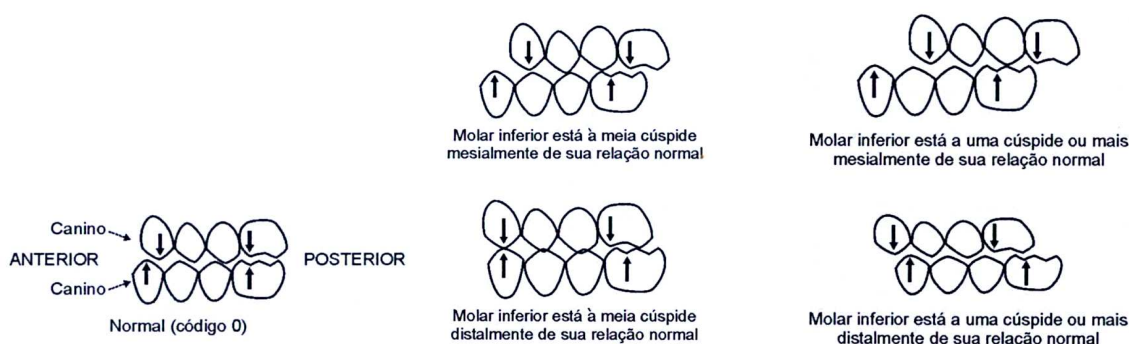
Se a mordida aberta é fisiológica e transicional durante a irrupção dos incisivos, a mordida aberta não será marcada.

## Relação intermolar ântero-posterior

Normalmente esta avaliação é realizada baseada na intercuspidação dos primeiros molares permanentes superiores e inferiores. Se os primeiros molares não puderem ser utilizados devido à sua ausência, irrupção incompleta, alterações na anatomia por cáries extensas ou restaurações, avalia-se então o relacinamento de caninos e pré-molares permanentes. Ambos os lados são avaliados em oclusão e somente o lado com maior desvio é registrado (FIG.11). Em caso de dúvida, a pontuação mais baixa deverá ser marcada (JENNY & CONS, 1988; OMS, 1999).

Os seguintes valores são utilizados:

- 0 → normal (relação de classe I de Angle);
- 1 → meia cúspide: méso- ou distoclusão por meia cúspide, até uma cúspide (exclusive);
- 2 → uma cúspide: méso- ou distoclusão por uma cúspide ou mais.



**FIGURA 11.:** Intercuspidação nos segmentos posteriores. FONTE: OMS, 1999.

### Modificações para dentadura mista:

Se os molares antagonistas estiverem em relação de topo-a-topo (molar inferior meia cúspide distalmente à sua relação normal) e a relação dos planos terminais dos molares decíduos for de plano reto, este componente será marcado com o código zero (0). Esta relação de molares é fisiológica e é a mais comum, pois será corrigida naturalmente com o deslocamento do molar inferior para o espaço livre de Nance.

## APÊNDICE B

### MANUAL DE REFERÊNCIA: Critérios para Aplicação do ICON para Necessidade de Tratamento Ortodôntico

O índice de complexidade, resultado e necessidade (ICON) de tratamento ortodôntico é um índice quantitativo multifuncional que foi desenvolvido em 2000 por RICHMOND e DANIELS para medir a complexidade, a aceitabilidade do resultado e a necessidade do tratamento ortodôntico. Possui cinco componentes: estético, avaliação de espaço, mordida cruzada, relação vertical anterior e relação ântero-posterior dos segmentos bucais. Devem ser atribuídos valores a todos os componentes e cada um é categorizado conforme seu respectivo critério. Os valores obtidos para cada componente são multiplicados por seus pesos específicos e então somados para obtenção do índice final. Para cada componente são utilizados os pesos listados no QUADRO 4.

**QUADRO 4 : Componentes do ICON**

| Componentes                    | Valor | Pesos | subtotal |
|--------------------------------|-------|-------|----------|
| Estético                       |       | × 7   |          |
| Avaliação de espaço            |       | × 5   |          |
| Mordida cruzada                |       | × 5   |          |
| Relação vertical anterior      |       | × 4   |          |
| Relação ântero-posterior bucal |       | × 3   |          |
| <b>Valor do índice (total)</b> |       |       |          |

FONTE: DANIELS & RICHMOND, 2000.

Para necessidade de tratamento ortodôntico, o ponto de corte recomendado é igual a 43. Aqueles casos cujos valores finais obtidos pelo índice forem maiores que 43 serão considerados com necessidade de tratamento ortodôntico.

#### Pressupostos:

- Exceto para o componente estético, dentes decíduos não são considerados na pontuação, exceto no caso de elemento decíduo que deva ser preservado no arco para manter espaço protético em caso de agenesia do substituto permanente.

#### APLICAÇÃO DO ÍNDICE

O tempo médio para a sua aplicação é de aproximadamente 1 a 3 minutos.

#### Componente: Estético (Escala SCAN)

É utilizada a escala ilustrada desenvolvida por EVANS e SHAW (1987).

A dentição é comparada à escala ilustrada que corresponde à gradação do comprometimento estético que vai desde o valor um (mínimo comprometimento estético) até dez (máximo comprometimento estético), conforme a FIG. 12. A classificação do grau de comprometimento estético deve ser obtida sem a necessidade de coincidir exatamente as características oclusais anatômicas do caso em análise com uma ilustração específica da escala. A escala serve como referência de um *continuum* subjacente desde o menor até o maior grau de comprometimento estético.

Algumas características oclusais que têm influência no grau impacto estético são bem representadas pela escala: apinhamento ou deslocamento de pontos de contato, nivelamento,

mordida profunda, presença de diastemas, dentes anteriores com alteração de forma ou ausentes, assimetrias definidas, atresia da maxila, trespasse horizontal excessivo.

Algumas características, embora serem notoriamente de grande influência no impacto estético e psico-social, não são representadas pela escala de fotografias, como: fenda palatina ou labial, mordida aberta anterior, espaços resultantes da perda de elementos dentários anteriores, mordida cruzada anterior (*overjet* negativo).



**FIG. 12:** Escala SCAN (Componente Estético)

#### Componente: Avaliação de Espaços

Esta variável tenta quantificar as discrepâncias de espaço na arcada superior ou a presença de dentes impactados em ambas as arcadas. As discrepâncias negativas (apinhamento) e positivas (espaços) são categorizadas conforme o QUADRO 5.



**QUADRO 5 : Componente avaliação de espaços**

| <b>Pontuação</b>             | <b>0</b>       | <b>1</b>   | <b>2</b>   | <b>3</b>    | <b>4</b>     | <b>5</b>                   |
|------------------------------|----------------|------------|------------|-------------|--------------|----------------------------|
| Apinhamento no arco superior | Menos que 2 mm | 2,1 a 5 mm | 5,1 a 9 mm | 9,1 a 13 mm | 13,1 a 17 mm | > 17 mm ou dente impactado |
| Espaçamento no arco superior | Até 2 mm       | 2,1 a 5 mm | 5,1 a 9 mm | > 9 mm      |              |                            |

A soma do diâmetro méso-distal dos dentes permanentes presentes ou prestes a irromper do arco superior é comparada ao espaço disponível, mesial aos últimos elementos irrompidos. Pode ser necessário a utilização de uma régua milimetrada. No caso de dentes permanentes perdidos, inclusos ou não irrompidos na fase de dentadura mista, são sugeridos diâmetros méso-distais médios de 7mm para pré-molares e caninos inferiores e de 8mm para caninos superiores. No caso de dentes homólogos irrompidos, estes podem servir como referência mais precisa da largura méso-distal do elemento ausente.

A profundidade da curva de Spee e a protrusão não são consideradas no cálculo da discrepância de espaço.

À presença de dentes impactados, em qualquer uma das arcadas, é atribuída pontuação máxima (categoria 5). Para ser considerado impactado, o elemento deve estar necessariamente incluso.

Um dente incluso é considerado impactado nas seguintes condições:

- Localização ectópica ou impactado contra um dente adjacente (excluindo terceiros molares, mas incluindo dentes supranumerários).
- Quando menos que 4 mm de espaço está disponível entre os dois dentes permanentes adjacentes.

Dentes decíduos retidos (agenesia do permanente sucessor) e dentes supra numerários irrompidos devem ser considerados como espaço ao menos que devam ser preservados no arco para manter espaço protético. Espaços resultantes de perda dentária por trauma ou exodontia também são considerados no cálculo.

Pode ser necessária a utilização de radiografias para confirmar casos de agenesia e supranumerários.

### **Componente: Mordida Cruzada**

Numa relação transversal normal, a cúspide palatina de um dente posterior superior oclui preferencialmente na fossa oclusal do seu antagonista, ou pelo menos, entre as suas pontas de cúspides lingual e vestibular. A mordida cruzada é considerada presente no segmento bucal quando há uma relação transversal de ponta de cúspide a ponta de cúspide, ou pior, tanto no sentido lingual quanto vestibular. Podem estar envolvidos um ou mais dentes com ou sem deslocamento mandibular (mordida cruzada funcional).

No segmento anterior, a mordida cruzada é definida como relacionamento de incisivos ou caninos superiores de borda a borda ou por lingual dos oponentes inferiores. Uma mordida cruzada no segmento posterior ou anterior, ou em ambos, é categorizada com o valor 1. A ausência de mordida cruzada recebe o valor 0.

### **Componente: Relação Vertical Anterior**

Este componente inclui tanto a mordida aberta (exeto as condições do desenvolvimento da oclusão) quanto a mordida profunda na região anterior. Se ambas as alterações estão presentes,

deve ser marcada somente a situação com pontuação mais alta conforme protocolo de pontuação do QUADRO 6. A mordida profunda é medida na parte mais profunda da sobremordida nos incisivos. A mordida aberta pode ser medida com uma régua milimetrada a partir do meio da borda incisal do dente superior com maior desvio. O toque incisal, ou trespasse vertical igual a zero, é considerado mordida completa (categorizado como 0).

Para o componente relação vertical anterior, as categorias de 0 a 4 são atribuídas conforme o QUADRO 6.

**QUADRO 6 : Componente vertical anterior**

| Pontuação                   | 0  | 1  | 2   | 3                                 | 4             |
|-----------------------------|--|--|---|-----------------------------------|---------------|
| Mordida aberta em incisivos | Mordida completa                                 | Menos que 1 mm   | 1,1 a 2 mm  | 2,1 a 4 mm                        | Mais que 4 mm |
| Sobremordida em incisivos   | Cobertura de até $\frac{1}{3}$ do dente inferior | Entre $\frac{1}{3}$ a $\frac{2}{3}$ de cobertura do dente inferior | Dente inferior entre $\frac{2}{3}$ a totalmente coberto | Dente inferior totalmente coberto |               |

### Componente: Relação Ântero-Posterior no Segmento Bucal

A avaliação deste componente inclui caninos, pré-molares e molares. O relacionamento ântero-posterior das cúspides é avaliado e cada lado é classificado independentemente numa escala de 3 categorias. Os valores obtidos em cada lado são somados para obter o valor consolidado deste componente. Relacionamento exclusivamente cúspide-embradura (Classe I, Classe II ou Classe III de Angle) recebe o valor 0. Relacionamento cúspide-cúspide recebe o valor 2. Qualquer relacionamento entre cúspide-embradura e cúspide-cúspide, excluindo estas condições limites, recebe o valor 1.

**QUADRO 7: Resumo do protocolo da avaliação dos componentes do ICON.**

| Componentes                        | Critérios (valores válidos)  | Pontuação / Categorias                             |  |                                    |                    |              |                            |
|------------------------------------|--|--|--|------------------------------------|--------------------|--------------|----------------------------|
|                                    |  | 0  | 1  | 2                                  | 3                  | 4            | 5                          |
| Estético                           | Escala AC-IOTN (1-10)  | Utilizar escala AC-IOTN                            |  |                                    |                    |              |                            |
| Apinhamento no arco superior       | Marcar só o valor mais alto para apinhamento (0-5) ou espaço (0-3)           | Menos que 2 mm                                     | 2,1 a 5 mm                                   | 5,1 a 9 mm                         | 9,1 a 13 mm        | 13,1 a 17 mm | > 17 mm ou dente impactado |
| Espaçamento no arco superior       |  | Até 2 mm   | 2,1 a 5 mm                                   | 5,1 a 9 mm                         | > 9 mm             |              |                            |
| Mordida Cruzada                    | Relação transversal cúspide-cúspide ou pior (0-1)                            | Ausência de mordida cruzada                        | Presença de mordida cruzada                  |                                    |                    |              |                            |
| Mordida aberta em incisivos        | Marcar só o valor mais alto para mordida aberta (0-4) ou sobremordida (0-3). | Mordida completa (trespasse = 0)                   | Menos que 1 mm                               | 1,1 a 2 mm                         | 2,1 a 4 mm         | > 4 mm       |                            |
| Sobremordida em incisivos          | Sobremordida: cobertura dos incisivos inferiores                             | Até $\frac{1}{3}$ do dente                         | $\frac{1}{3}$ a $\frac{2}{3}$ de cobertura   | $\frac{2}{3}$ a totalmente coberto | Totalmente coberto |              |                            |
| Ântero-posterior no segmento bucal | Lados direito e esquerdo somados (0-4)                                       | Só relação cúspide-embradura (Classe I, II ou III) | Outra relação até cúspide-cúspide, exclusive | Relação cúspide-cúspide            |                    |              |                            |

## APÊNDICE C

FICHA PARA O DAI DATA: \_\_\_ / \_\_\_ / \_\_\_\_

 Conj. \_\_\_ modelos: \_\_\_\_\_ ( \_\_\_º ex.) Modelo Núm.: \_\_\_\_\_

| Componentes do DAI (valores possíveis)                                   | Valores   |
|--|-----------|
| Dentes ausentes (pré-mol. a pré-mol. – 0 a 10 por arcada)                | maxila    |
|  | mandíbula |
| Apinham anterior (0, 1=1segm., 2=2segm.)                                 |           |
| Espaçam. anterior (0, 1=1segm., 2=2segm.)                                |           |
| Diastema na linha média sup. (mm arred.)                                 |           |
| Maior irreg. anterior na maxila (mm arred.)                              |           |
| Maior irreg. anterior na mandíbula (mm arred.)                           |           |
| Tresp. horiz. max. – sobressaliência (mm arred. – topo-a-topo = 0)       |           |
| Tresp. horiz. mand. – mord. cruz. anterior (mm arred.)                   |           |
| Mord. aberta anterior (mm arred.)  |           |
| Rel. ânt.-post. de molares (lado c/ maior desvio: 0, 1=½cúsp., 2=1cúsp.) |           |

## APÊNDICE D

FICHA PARA O ICON DATA: \_\_\_ / \_\_\_ / \_\_\_\_

 Conj. \_\_\_ modelos: \_\_\_\_\_ ( \_\_\_º ex.) Modelo Núm.: \_\_\_\_\_

| Componente (valores possíveis)            | Valor                       |
|---|-----------------------------|
| Estético (CE-IOTN); (1-10)                |                             |
| Espaços                                   | Apinham. superior (0-5)     |
|   | Espaçam. superior (0-3)     |
| Mordida cruzada (0-1)                     |                             |
| Vertical                                  | Mord. aberta anterior (0-4) |
|   | Sobremordida (0-3)          |
| Rel. ântero-posterior bucal (d + e) (0-4) |                             |

## APÊNDICE E

TABELA 32

Frequência de classificações de má oclusão em relação à classificação pela escala SCAN no conjunto de 211 modelos da FO-UFMG

| Classificação da Má Oclusão |                   | Classificação pela Escala SCAN |           |           |           |          |           |           |           |           |           | Subtotal (%)       |
|-----------------------------|-------------------|--------------------------------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-----------|-----------|--------------------|
|                             |                   | 1                              | 2         | 3         | 4         | 5        | 6         | 7         | 8         | 9         | 10        |                    |
| Intra-arcos                 | Normal            | 5                              | 9         | 7         | 4         | 0        | 2         | 0         | 0         | 2         | 0         | 29 (13,7)          |
|                             | Apinhamento       | 1                              | 9         | 19        | 11        | 1        | 17        | 9         | 27        | 14        | 8         | 116 (55,0)         |
|                             | Espaços           | 1                              | 4         | 0         | 10        | 4        | 0         | 0         | 1         | 1         | 2         | 23 (10,9)          |
|                             | Ausências, outros | 1                              | 6         | 1         | 4         | 4        | 6         | 2         | 4         | 5         | 10        | 43 (20,4)          |
|                             | <b>Total</b>      | <b>8</b>                       | <b>28</b> | <b>27</b> | <b>29</b> | <b>9</b> | <b>25</b> | <b>11</b> | <b>32</b> | <b>22</b> | <b>20</b> | <b>211 (100,0)</b> |
| Horizontal                  | Classe I          | 5                              | 21        | 20        | 15        | 5        | 20        | 2         | 17        | 1         | 6         | 112 (53,1)         |
|                             | Classe II         | 3                              | 6         | 5         | 13        | 3        | 4         | 9         | 10        | 21        | 10        | 84 (39,8)          |
|                             | Classe III        | 0                              | 1         | 2         | 1         | 1        | 1         | 0         | 5         | 0         | 4         | 15 (7,1)           |
|                             | <b>Total</b>      | <b>8</b>                       | <b>28</b> | <b>27</b> | <b>29</b> | <b>9</b> | <b>25</b> | <b>11</b> | <b>32</b> | <b>22</b> | <b>20</b> | <b>211 (100,0)</b> |
| Vertical                    | Normal            | 5                              | 19        | 16        | 19        | 2        | 21        | 6         | 12        | 5         | 7         | 112 (53,1)         |
|                             | Mordida Profunda  | 3                              | 6         | 2         | 6         | 6        | 2         | 5         | 8         | 12        | 9         | 59 (28,0)          |
|                             | Mordida Aberta    | 0                              | 3         | 9         | 4         | 1        | 2         | 0         | 12        | 5         | 4         | 40 (19,0)          |
|                             | <b>Total</b>      | <b>8</b>                       | <b>28</b> | <b>27</b> | <b>29</b> | <b>9</b> | <b>25</b> | <b>11</b> | <b>32</b> | <b>22</b> | <b>20</b> | <b>211 (100,0)</b> |
| Transversal                 | Normal            | 8                              | 19        | 14        | 22        | 8        | 12        | 9         | 14        | 13        | 11        | 130 (61,6)         |
|                             | Mordida Cruzada   | 0                              | 6         | 13        | 6         | 0        | 12        | 2         | 18        | 9         | 7         | 73 (34,6)          |
|                             | Assimetrias       | 0                              | 3         | 0         | 1         | 1        | 1         | 0         | 0         | 0         | 2         | 8 (3,8)            |
|                             | <b>Total</b>      | <b>8</b>                       | <b>28</b> | <b>27</b> | <b>29</b> | <b>9</b> | <b>25</b> | <b>11</b> | <b>32</b> | <b>22</b> | <b>20</b> | <b>211 (100,0)</b> |

## APÊNDICE F

TABELA 33

Conjunto de vinte pares de modelos utilizados na aplicação e avaliação dos critérios

| Escala<br>SCAN | Níveis de<br>Necessidade | Frequência | Dentição |            | Total (%)         |
|----------------|--------------------------|------------|----------|------------|-------------------|
|                |                          |            | Mista    | Permanente |                   |
| 1              |                          | 2          |          |            |                   |
| 2              | sem                      | 4          | 1        | 9          | 10 (50,0)         |
| 3              | necessidade              | 1          |          |            |                   |
| 4              |                          | 3          |          |            |                   |
| 5              | Limite entre             | 1          |          |            |                   |
| 6              | necessidade e            | 1          | 2        | 3          | 5 (25,0)          |
| 7              | não necessidade          | 3          |          |            |                   |
| 8              |                          | 3          |          |            |                   |
| 9              | com                      | 1          | 1        | 4          | 5 (25,0)          |
| 10             | necessidade              | 1          |          |            |                   |
| <b>Total</b>   |                          | <b>20</b>  | <b>4</b> | <b>16</b>  | <b>20 (100,0)</b> |



## APÊNDICE H

### AVALIAÇÃO DE MODELOS POR ESPECIALISTAS

#### **ATENÇÃO:**

É importante que a avaliação seja feita com critério e que sua opinião registrada seja a mais sincera possível.

Solicitamos sua avaliação destes modelos de estudo quanto à **necessidade de tratamento ortodôntico**. Considere apenas as características representadas pelos modelos. Não considere fatores que poderiam influenciar sua opinião subjetiva, tais como: idade, sexo, classe sócio-econômica.

## APÊNDICE I

### FICHA PARA AVALIAÇÃO DE MODELOS POR ESPECIALISTAS

Examinador: \_\_\_\_\_

Data: \_\_\_\_/\_\_\_\_/\_\_\_\_

Conjunto de modelos: 108 (\_\_\_\_ª avaliação)

Modelo Núm.: \_\_\_\_\_

Na sua opinião, em que extensão esta oclusão necessita de tratamento ortodôntico? Por favor, marque o número correspondente na escala abaixo:

|                    |   |   |   |   |   |   |                 |
|--------------------|---|---|---|---|---|---|-----------------|
| NENHUMA/<br>MÍNIMA |   |   |   |   |   |   | MUITO<br>GRANDE |
| 1                  | 2 | 3 | 4 | 5 | 6 | 7 |                 |

## APÊNDICE J

## FICHA PARA REGISTRAR O “PONTO DE TRATAMENTO INDICADO”

Examinador: \_\_\_\_\_ Data: \_\_\_\_/\_\_\_\_/\_\_\_\_

Na escala de 7 pontos utilizada nesta sessão para a avaliação dos modelos, indique o valor no qual ou acima dele você acha que o tratamento ortodôntico é indicado.

|                    |   |   |   |   |   |                 |
|--------------------|---|---|---|---|---|-----------------|
| NENHUMA/<br>MÍNIMA |   |   |   |   |   | MUITO<br>GRANDE |
| 1                  | 2 | 3 | 4 | 5 | 6 | 7               |



## APÊNDICE L

TABELA 35

Frequência de classificações de má oclusão em relação à classificação pela escala SCAN no conjunto de 445 modelos da FO-UFGM

| Classificação das<br>Más Oclusões |                   | Classificação pela Escala SCAN |           |           |           |           |           |           |           |           |           | Totais (%)         |
|-----------------------------------|-------------------|--------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|--------------------|
|                                   |                   | 1                              | 2         | 3         | 4         | 5         | 6         | 7         | 8         | 9         | 10        |                    |
| Intra-<br>arcos                   | Normal            | 11                             | 22        | 11        | 8         | 3         | 2         | 3         | 0         | 4         | 0         | 64 (14,4)          |
|                                   | Apinhamento       | 2                              | 21        | 43        | 25        | 6         | 33        | 22        | 37        | 19        | 10        | 218 (49,0)         |
|                                   | Espaços           | 1                              | 9         | 6         | 24        | 11        | 1         | 2         | 2         | 7         | 4         | 67 (15,1)          |
|                                   | Ausencias, outros | 2                              | 8         | 12        | 12        | 6         | 13        | 9         | 12        | 10        | 12        | 96 (21,6)          |
|                                   | <b>Total</b>      | <b>16</b>                      | <b>60</b> | <b>72</b> | <b>69</b> | <b>26</b> | <b>49</b> | <b>36</b> | <b>51</b> | <b>40</b> | <b>26</b> | <b>445 (100,0)</b> |
| Hori-<br>zontal                   | Classe I          | 11                             | 41        | 47        | 35        | 10        | 33        | 9         | 23        | 5         | 6         | 220 (49,4)         |
|                                   | Classe II         | 5                              | 17        | 22        | 31        | 13        | 14        | 27        | 22        | 31        | 13        | 195 (43,8)         |
|                                   | Classe III        | 0                              | 2         | 3         | 3         | 3         | 2         | 0         | 6         | 4         | 7         | 30 (6,7)           |
|                                   | <b>Total</b>      | <b>16</b>                      | <b>60</b> | <b>72</b> | <b>69</b> | <b>26</b> | <b>49</b> | <b>36</b> | <b>51</b> | <b>40</b> | <b>26</b> | <b>445 (100,0)</b> |
| Verti-<br>cal                     | Normal            | 12                             | 44        | 46        | 40        | 9         | 36        | 16        | 23        | 11        | 8         | 245 (55,1)         |
|                                   | Mordida Profunda  | 4                              | 12        | 14        | 19        | 14        | 7         | 17        | 12        | 21        | 12        | 132 (29,7)         |
|                                   | Mordida Aberta    | 0                              | 4         | 12        | 10        | 3         | 6         | 3         | 16        | 8         | 6         | 68 (15,3)          |
|                                   | <b>Total</b>      | <b>16</b>                      | <b>60</b> | <b>72</b> | <b>69</b> | <b>26</b> | <b>49</b> | <b>36</b> | <b>51</b> | <b>40</b> | <b>26</b> | <b>445 (100,0)</b> |
| Trans-<br>versal                  | Normal            | 14                             | 43        | 45        | 50        | 22        | 21        | 26        | 23        | 21        | 11        | 276 (62,0)         |
|                                   | Mordida Cruzada   | 2                              | 14        | 25        | 17        | 2         | 24        | 10        | 27        | 17        | 9         | 147 (33,0)         |
|                                   | Assimetrias       | 0                              | 3         | 2         | 2         | 2         | 4         | 0         | 1         | 2         | 6         | 22 (4,9)           |
|                                   | <b>Total</b>      | <b>16</b>                      | <b>60</b> | <b>72</b> | <b>69</b> | <b>26</b> | <b>49</b> | <b>36</b> | <b>51</b> | <b>40</b> | <b>26</b> | <b>445 (100,0)</b> |

## APÊNDICE M

TABELA 36

Frequência de classificações de má oclusão em relação à classificação pela escala SCAN no conjunto de 108 modelos

| Classificação das<br>Más Oclusões |                   | Classificação pela Escala SCAN |           |           |           |           |           |           |           |           |          | Totais (%)         |
|-----------------------------------|-------------------|--------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|--------------------|
|                                   |                   | 1                              | 2         | 3         | 4         | 5         | 6         | 7         | 8         | 9         | 10       |                    |
| Den-<br>tição                     | Mista             | 0                              | 0         | 3         | 1         | 2         | 0         | 5         | 3         | 1         | 1        | 16 (14,8)          |
|                                   | Permanente        | 6                              | 11        | 9         | 11        | 11        | 13        | 7         | 9         | 10        | 5        | 92 (85,2)          |
|                                   | <b>Total</b>      | <b>6</b>                       | <b>11</b> | <b>12</b> | <b>12</b> | <b>13</b> | <b>13</b> | <b>12</b> | <b>12</b> | <b>11</b> | <b>6</b> | <b>108 (100,0)</b> |
| Intra-<br>arcos                   | Normal            | 4                              | 5         | 1         | 2         | 1         | 0         | 3         | 0         | 2         | 0        | 18 (16,7)          |
|                                   | Apinhamento       | 1                              | 4         | 5         | 4         | 4         | 7         | 5         | 5         | 5         | 2        | 42 (38,9)          |
|                                   | Espaços           | 0                              | 2         | 3         | 3         | 6         | 1         | 0         | 1         | 1         | 2        | 19 (17,6)          |
|                                   | Ausências, outros | 1                              | 0         | 3         | 3         | 2         | 5         | 4         | 6         | 3         | 2        | 29 (26,9)          |
|                                   | <b>Total</b>      | <b>6</b>                       | <b>11</b> | <b>12</b> | <b>12</b> | <b>13</b> | <b>13</b> | <b>12</b> | <b>12</b> | <b>11</b> | <b>6</b> | <b>108 (100,0)</b> |
| Hori-<br>zontal                   | Classe I          | 4                              | 5         | 6         | 5         | 5         | 7         | 5         | 5         | 3         | 0        | 45 (41,7)          |
|                                   | Classe II         | 2                              | 5         | 5         | 5         | 6         | 5         | 7         | 6         | 4         | 3        | 48 (44,4)          |
|                                   | Classe III        | 0                              | 1         | 1         | 2         | 2         | 1         | 0         | 1         | 4         | 3        | 15 (13,9)          |
|                                   | <b>Total</b>      | <b>6</b>                       | <b>11</b> | <b>12</b> | <b>12</b> | <b>13</b> | <b>13</b> | <b>12</b> | <b>12</b> | <b>11</b> | <b>6</b> | <b>108 (100,0)</b> |
| Verti-<br>cal                     | Normal            | 5                              | 10        | 6         | 7         | 6         | 7         | 5         | 6         | 5         | 1        | 58 (53,7)          |
|                                   | Mordida Profunda  | 1                              | 1         | 5         | 2         | 5         | 4         | 6         | 3         | 3         | 3        | 33 (30,6)          |
|                                   | Mordida Aberta    | 0                              | 0         | 1         | 3         | 2         | 2         | 1         | 3         | 3         | 2        | 17 (15,7)          |
|                                   | <b>Total</b>      | <b>6</b>                       | <b>11</b> | <b>12</b> | <b>12</b> | <b>13</b> | <b>13</b> | <b>12</b> | <b>12</b> | <b>11</b> | <b>6</b> | <b>108 (100,0)</b> |
| Trans-<br>versal                  | Normal            | 4                              | 9         | 9         | 8         | 10        | 7         | 8         | 4         | 4         | 0        | 63 (58,3)          |
|                                   | Mordida Cruzada   | 2                              | 2         | 2         | 3         | 2         | 5         | 4         | 7         | 5         | 2        | 34 (31,5)          |
|                                   | Assimetrias       | 0                              | 0         | 1         | 1         | 1         | 1         | 0         | 1         | 2         | 4        | 11 (10,2)          |
|                                   | <b>Total</b>      | <b>6</b>                       | <b>11</b> | <b>12</b> | <b>12</b> | <b>13</b> | <b>13</b> | <b>12</b> | <b>12</b> | <b>11</b> | <b>6</b> | <b>108 (100,0)</b> |

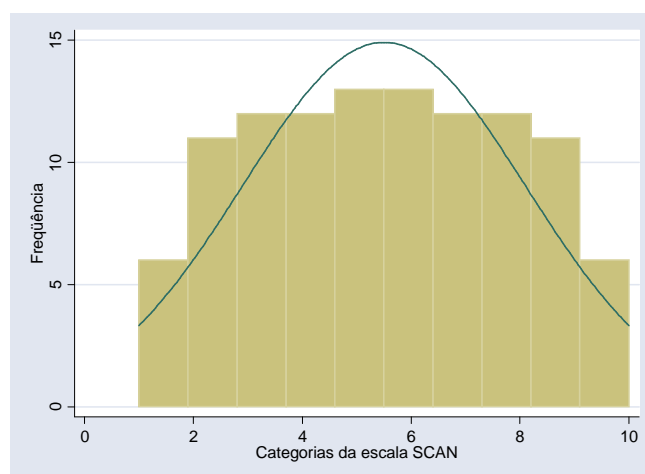


GRÁFICO 21 – Histograma da distribuição das categorias da escala SCAN no conjunto de 108 modelos.

## APÊNDICE N

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DO ESPECIALISTA COLABORADOR

Prezado(a) ortodontista,

Você está sendo convidado(a) a participar como voluntário de um estudo de minha autoria, sob a orientação dos Profs. Dr. Allyson Nogueira Moreira e Dra. Cláudia Silami de Magalhães, desenvolvido na Faculdade de Odontologia da UFMG com o título: “Necessidade de tratamento ortodôntico: validação do ICON (índice de complexidade, resultado e necessidade) e do DAI (índice de estética dental) para a cidade de Belo Horizonte”.

O objetivo deste trabalho é a validação de índices oclusais para a avaliação da necessidade de tratamento ortodôntico frente à opinião de ortodontistas de Belo Horizonte que possuam notório saber na especialidade.

Sua participação é fundamental para o desenvolvimento do trabalho. Consiste em avaliar 108 modelos de estudo e registrar sua opinião quanto à necessidade de tratamento ortodôntico em duas sessões de avaliação em datas diferentes. O tempo estimado é de 70 minutos para a primeira sessão e de 20 minutos para a segunda.

O anonimato e o sigilo das informações serão garantidos. Em caso de dúvida, coloco-me à disposição para solucioná-las pelos telefones: 3296-3767, 9974-0577 e pelo email: rodrigo@norremose.odo.br.

Qualquer dúvida quanto ao compromisso ético desta pesquisa, o (a) senhor (a) poderá consultar o COMITÊ DE ÉTICA EM PESQUISA DA UFMG localizado no Campus da Pampulha, na Av. Antônio Carlos, 6627. Prédio da Reitoria - Sala 701, Pampulha. CEP: 31270-901 – Belo Horizonte-MG.

Caso deseje participar do estudo, solicito que assine o termo de consentimento livre e esclarecido.

---

Rodrigo Norremose Costa - Mestrando –UFMG

#### CONSENTIMENTO

Considerando a exposição acima, como participante voluntário do estudo de validação, afirmo que fui devidamente esclarecido sobre a finalidade e objetivo da pesquisa, quanto ao caráter confidencial de minhas respostas e também sobre a utilização das informações emitidas por mim. Minhas dúvidas foram esclarecidas e declaro ter recebido cópia desta autorização.

Por esse instrumento, eu \_\_\_\_\_ autorizo o pesquisador Rodrigo Norremose Costa a utilizar as minhas respostas e outros dados obtidos com a minha participação no trabalho de pesquisa. E que em qualquer momento da pesquisa posso retirar este consentimento.

Belo Horizonte, \_\_\_\_\_ de \_\_\_\_\_ de 2006

---

Assinatura do CD Ortodontista

## ANEXOS

### ANEXO A

Universidade Federal de Minas Gerais  
Comitê de Ética em Pesquisa da UFMG - COEP

**Parecer nº. ETIC 068/06**

**Interessado: Prof. Dr. Allyson Nogueira Moreira**  
**Departamento de Odontologia Restauradora**  
**Faculdade de Odontologia - UFMG**

### DECISÃO

O Comitê de Ética em Pesquisa da UFMG – COEP aprovou no dia 12 de abril de 2006, o projeto de pesquisa intitulado **“Necessidade de Tratamento Ortodôntico: Validação do ICON (índice de Complexidade, Resultado e Necessidade) e do DAÍ (Índice de Estética Dental) para a Cidade de Belo Horizonte”** bem como o Termo de Consentimento Livre e Esclarecido do referido projeto.

O relatório final ou parcial deverá ser encaminhado ao COEP um ano após o início do projeto.

  
**Profa. Dra. Maria Elena de Lima Perez Garcia**  
**Presidente do COEP/UFMG**

## ANEXO B

## AUTORIZAÇÃO DO DEPARTAMENTO DE ODONTOPEDIATRIA E ORTODONTIA

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FACULDADE DE ODONTOLOGIA  
DEPARTAMENTO DE ODONTOPEDIATRIA E ORTODONTIA

Belo Horizonte, 23 de Maio de 2005.

Ilma. Sra.  
Profª. Maria de Lourdes Andrade Massara  
Chefe do Departamento de Odontopediatria e Ortodontia

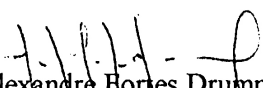
Prezada Professora,

Em resposta ao ofício encaminhado pelo aluno do curso de mestrado, Rodrigo Norremose Costa, autorizo a utilização dos modelos de estudo da clínica de ortodontia, dentro das dependências do departamento.

Quanto a solicitação de vagas na clínica de ortodontia para tratamento dos pacientes participantes do estudo proposto pelo aluno, não podemos garantir estes tratamentos em prazos determinados, só nos disponibilizando a colocá-los em uma lista de espera como todos os pacientes que nos procuram.

Colocando-me à sua disposição.

Atenciosamente,

  
Alexandre Fortes Drummond  
Coordenador do Curso de Especialização em Ortodontia