

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM FARMÁCIA**

**AVALIAÇÃO DA QUALIDADE DA REDAÇÃO DE ARTIGOS CIENTÍFICOS DE  
ACURÁCIA DIAGNÓSTICA PUBLICADOS EM PERIÓDICOS NACIONAIS E  
INTERNACIONAL UTILIZANDO O *STARD* E O *QUADAS***

**FERNANDO RODRIGUES SAGEBIN**

**FLORIANÓPOLIS**

**2009**

**FERNANDO RODRIGUES SAGEBIN**

**AVALIAÇÃO DA QUALIDADE DA REDAÇÃO DE ARTIGOS CIENTÍFICOS DE  
ACURÁCIA DIAGNÓSTICA PUBLICADOS EM PERIÓDICOS NACIONAIS E  
INTERNACIONAL UTILIZANDO O *STARD* E O *QUADAS***

**Dissertação apresentada ao Programa de  
Pós-Graduação em Farmácia como  
requisito parcial à obtenção do grau de  
Mestre em Farmácia**

**Orientador: Prof. Dr. Arício Treitinger**

**FLORIANÓPOLIS**

**2009**

Ficha catalográfica elaborada na DECTI da BU/UFSC

- S129 Sagebin, Fernando Rodrigues  
Avaliação da qualidade da redação de artigos científicos de acurácia diagnóstica publicados em periódicos nacionais e internacional utilizando o stard e o quadas [dissertação]/ Fernando Rodrigues Sagebin ; orientador, Arício Treitinger. -Florianopolis, SC : 2009.  
106 f.: tabs., mapas
- Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências de Saúde. Programa de Pós-Graduação em Farmácia
- Inclui bibliografia
1. Farmacia. I.Treitinger, Aricio. II.Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Farmácia. III. Título.

CDU 615.12

**Dedico este trabalho à minha família.**

## AGRADECIMENTOS

Muitas pessoas fizeram parte desta jornada, e todas merecem meu agradecimento, porém algumas, de modo muito especial.

Agradeço primeiramente ao meu orientador, Prof. Dr. Arício Treitinger, pela oportunidade de trabalhar ao seu lado e por ter acreditado em mim.

Agradeço de forma especial ao Prof. Dr. Marcos José Machado pelos ensinamentos, paciência e amizade.

A todos os professores do Programa de Pós-Graduação em Farmácia da Universidade Federal de Santa Catarina pela contribuição em minha formação. Agradeço de forma especial ao Prof. Dr. Paulo Ogliari do Departamento de Informática e Estatística CTC-INE.

Aos meus colegas de mestrado, de maneira especial à Pâmela Soares, Júnior André da Rosa, Natália Scaraveli e Andréa do Livramento pela amizade, apoio e momentos de descontração.

Aos amigos, alguns de longa data outros recentes, mas que marcaram profundamente essa trajetória: Carolina Bianchi, Kelly González, Fernanda Ramos, Thales Preissler, Moacir Amorim, Carlos Alberto Amaral, Patrícia Pozzatti e Silvia Borgmann.

Agradeço imensamente às minhas tias Lygia Armando e Marlene Henriqson pelo carinho e apoio incondicional sem os quais essa jornada não seria possível.

Agradeço à minha família, meu pai Paulo, minha mãe Ana Lúcia, minha irmã Ana Paula e cunhado Cesar Corrêa pelo incentivo, carinho e amor.

Meu sincero: muito obrigado!

**“Nada é permanente, salvo a mudança.”  
Heráclito de Éfeso (540 a.C.-470 a.C.)**

## RESUMO

O registro completo e preciso da pesquisa diagnóstica é essencial para aprimorar a exatidão e integridade da redação dos estudos de acurácia diagnóstica, a fim de permitir aos leitores analisar o potencial para a incorporação de vieses aos estudos e para avaliar sua aplicabilidade. Para avaliação da qualidade de redação de artigos de acurácia diagnóstica foram aplicados os instrumentos STARD (*Standards for Reporting of Diagnostic Accuracy*) e QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*) aos artigos publicados em seis periódicos de medicina laboratorial: *Clinical Biochemistry*, *Jornal Brasileiro de Patologia e Medicina Laboratorial (JBPML)*, *LAES&HAES*, *NewsLab*, *Revista Brasileira de Análises Clínicas (RBAC)* e *Revista Brasileira de Ciências Farmacêuticas (RBCF)* entre os anos de 2002 e 2006. Um total de 101 artigos preencheu os critérios de inclusão (*Clinical Biochemistry*=53, *JBPML*=11, *LAES&HAES*=15, *NewsLab*=09, *RBAC*=13, *RBCF*=0). O número médio de itens do STARD e QUADAS atendidos nos artigos avaliados foi  $15,9 \pm 2,2$  ( $63,6\% \pm 8,7\%$ ) e  $9,0 \pm 1,8$  ( $64,3\% \pm 12,6\%$ ) para o periódico *Clinical Biochemistry*,  $13,4 \pm 2,9$  ( $53,6\% \pm 11,5\%$ ) e  $8,1 \pm 1,4$  ( $57,8\% \pm 10,3\%$ ) para o *JBPML*,  $10,6 \pm 3,6$  ( $42,2\% \pm 14,5\%$ ) e  $6,3 \pm 1,8$  ( $45,2\% \pm 12,8\%$ ) para o *LAES&HAES*,  $11,1 \pm 3,4$  ( $44,4\% \pm 13,8\%$ ) e  $7,3 \pm 2,0$  ( $52,4\% \pm 14,3\%$ ) para o *NewsLab* e  $12,0 \pm 3,2$  ( $48,1\% \pm 12,6\%$ ) e  $7,1 \pm 2,7$  ( $50,5\% \pm 19,2\%$ ) para o *RBAC*. Foram avaliadas as mudanças nas proporções que os itens do STARD e QUADAS foram atendidos através dos cinco anos analisados através do teste de  $\chi^2$  Armitage para tendências lineares. Ficou demonstrada diferença estatística significativa para a *Revista Brasileira de Análises Clínicas* na aplicação do STARD ( $p=0,0021$ ) e QUADAS ( $p=0,0002$ ) e para a revista *NewsLab* apenas na aplicação do STARD ( $p=0,0324$ ). Estes resultados demonstram que a qualidade da redação dos artigos de acurácia diagnóstica nos periódicos selecionados está abaixo da objetivada por estes instrumentos, ou seja, que o delineamento, condução, análise e resultados destes estudos possam ser compreendidos pelos leitores através de uma redação completa e transparente e que a qualidade de redação continua similar à de antes da publicação destes instrumentos no ano de 2003 na maioria dos periódicos analisados.

Palavras-chave: qualidade de redação, acurácia diagnóstica, STARD, QUADAS.

## ABSTRACT

Complete and accurate reporting of diagnostic research are essential to improve the accuracy and completeness of reporting of studies of diagnostic accuracy in order to allow readers to assess the potential for bias in the study and to evaluate its generalisability. To evaluate quality of reporting in diagnostic accuracy articles, the Standards for Reporting of Diagnostic Accuracy (STARD) and Quality Assessment of Diagnostic Accuracy Studies (QUADAS) were used to evaluate the quality of each article published between 2002 and 2006 in six laboratory medicine journals, the Clinical Biochemistry, *Jornal Brasileiro de Patologia e Medicina Laboratorial* (JBPML), LAES&HAES, NewsLab, *Revista Brasileira de Análises Clínicas* (RBAC) and *Revista Brasileira de Ciências Farmacêuticas* (RBCF). A total of 101 reports met the inclusion criteria (Clinical Biochemistry=53, JBPML=11, LAES&HAES=15, NewsLab=09, RBAC=13, RBCF=0). The mean number of STARD and QUADAS reported items in evaluated studies was  $15,9 \pm 2,2$  ( $63,6\% \pm 8,7\%$ ) and  $9,0 \pm 1,8$  ( $64,3\% \pm 12,6\%$ ) for Clinical Biochemistry,  $13,4 \pm 2,9$  ( $53,6\% \pm 11,5\%$ ) and  $8,1 \pm 1,4$  ( $57,8\% \pm 10,3\%$ ) for JBPML,  $10,6 \pm 3,6$  ( $42,2\% \pm 14,5\%$ ) and  $6,3 \pm 1,8$  ( $45,2\% \pm 12,8\%$ ) for LAES&HAES,  $11,1 \pm 3,4$  ( $44,4\% \pm 13,8\%$ ) and  $7,3 \pm 2,0$  ( $52,4\% \pm 14,3\%$ ) for NewsLab and  $12,0 \pm 3,2$  ( $48,1\% \pm 12,6\%$ ) and  $7,1 \pm 2,7$  ( $50,5\% \pm 19,2\%$ ) for RBAC. The change in proportion of times that STARD and QUADAS items were reported across the five years of publication were evaluated by using  $\chi^2$  test for linear trend. Only RBAC ( $p=0,0021$  for STARD and  $p=0,0002$  for QUADAS) and NewsLab ( $p=0,0324$  for STARD) showed significant statistical difference. This results show which the quality of reporting in diagnostic accuracy articles in the selected journals is suboptimal. After publication of the STARD and QUADAS tools in 2003, the quality of reporting of diagnostic accuracy studies remained similar to pre-STARD and pre-QUADAS publication in most journals reviewed.

Keywords: quality of reporting, diagnostic accuracy, STARD, QUADAS.



## SUMÁRIO

1. INTRODUÇÃO.....	15
1.1 Considerações gerais .....	15
1.2 Conhecimento, informação e qualidade da informação .....	16
1.3 Medicina baseada em evidências e Diagnóstico Laboratorial.....	18
1.4 Estudos de acurácia diagnóstica .....	20
1.5 Avaliação da qualidade dos artigos de acurácia diagnóstica.....	24
2. JUSTIFICATIVA.....	30
3. OBJETIVOS.....	31
3.1 Objetivo Geral .....	31
3.2 Objetivos Específicos .....	31
4. MATERIAIS E MÉTODOS.....	32
4.1 Critérios de escolha e inclusão dos periódicos científicos a serem avaliados.....	32
4.2 Tipo de estudos incluídos .....	32
4.3 Amostragem .....	32
4.3.1 Localização e Fonte dos estudos .....	32
4.3.2 Critérios de inclusão de artigos científicos.....	33
4.4 Extração dos dados .....	33
4.5 Análise e síntese dos dados coletados .....	34
4.6 Análise estatística .....	34
5. RESULTADOS .....	36
5.1 Localização e recuperação dos estudos de acurácia diagnóstica.....	36
5.2 Características das revistas e dos artigos de acurácia diagnóstica.....	37
5.2.1 <i>Clinical Biochemistry</i> .....	41
5.2.2 Jornal Brasileiro de Patologia e Medicina Laboratorial .....	41
5.2.3 LAES&HAES.....	42
5.2.4 NewsLab.....	42
5.2.5 Revista Brasileira de Análises Clínicas.....	43
5.2.6 Revista Brasileira de Ciências Farmacêuticas .....	43
5.3 Resultados verificados com o roteiro STARD .....	45

5.3.1 <i>Clinical Biochemistry</i> .....	45
5.3.2 Jornal Brasileiro de Patologia e Medicina Laboratorial .....	50
5.3.3 LAES&HAES.....	54
5.3.4 NewsLab.....	58
5.3.5 Revista Brasileira de Análises Clínicas .....	62
5.4 Resultados verificados com o instrumento QUADAS .....	66
5.4.1 <i>Clinical Biochemistry</i> .....	68
5.4.2 Jornal Brasileiro de Patologia e Medicina Laboratorial .....	70
5.4.3 LAES&HAES.....	73
5.4.4 NewsLab.....	75
5.4.5 Revista Brasileira de Análises Clínicas .....	78
5.5 Análise da variação dos itens atendidos do STARD e QUADAS no período analisado ...	80
5.6 Avaliação dos itens do STARD e QUADAS concernentes à validade interna e externa dos artigos avaliados .....	83
5.6.1 <i>Clinical Biochemistry</i> .....	83
5.6.1.1 STARD .....	83
5.6.1.2 QUADAS .....	84
5.6.2 Jornal Brasileiro de Patologia e Medicina Laboratorial .....	84
5.6.2.1 STARD .....	84
5.6.2.2 QUADAS .....	85
5.6.3 LAES&HAES.....	86
5.6.3.1 STARD .....	86
5.6.3.2 QUADAS .....	86
5.6.4 NewsLab.....	87
5.6.4.1 STARD .....	87
5.6.4.2 QUADAS .....	88
5.6.5 Revista Brasileira de Análises Clínicas .....	88
5.6.5.1 STARD .....	88
5.6.5.2 QUADAS .....	89
5.7 Análise da variação dos itens concernentes à validade interna e externa atendidos do STARD e QUADAS no período analisado .....	90
5.8 Correlação entre STARD e QUADAS .....	91
6. DISCUSSÃO .....	92
7. CONCLUSÕES .....	106

8. REFERÊNCIAS BIBLIOGRÁFICAS .....	107
ANEXO A – <i>STARD (Standards for Reporting of Diagnostic Accuracy)</i> .....	127
ANEXO B – <i>QUADAS (Quality Assessment of Diagnostic Accuracy Studies)</i> .....	129

## LISTA DE TABELAS

<b>Tabela 1:</b> Tabela de contingência 2 x 2.....	22
<b>Tabela 2:</b> Principais fontes de viés e variabilidade.....	28
<b>Tabela 3:</b> Número de artigos de acurácia diagnóstica recuperados nos periódicos, totais por ano analisado e número total de artigos publicados por periódico .....	38
<b>Tabela 4:</b> Características e índices bibliográficos dos periódicos científicos analisados.....	39
<b>Tabela 5:</b> Distribuição da presença de resumo e resumo estruturado entre os artigos recuperados.....	40
<b>Tabela 6:</b> Número médio de itens do STARD atendidos nos diferentes periódicos em números absolutos e percentuais e valores mínimos e máximos encontrados nos artigos avaliados.....	45
<b>Tabela 7:</b> Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico <i>Clinical Biochemistry</i> .....	47
<b>Tabela 8:</b> Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.....	52
<b>Tabela 9.</b> Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.....	56
<b>Tabela 10.</b> Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.....	60
<b>Tabela 11.</b> Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC .....	64
<b>Tabela 12.</b> Número médio de questões com resposta “sim” na aplicação do QUADAS nos diferentes periódicos analisados .....	67
<b>Tabela 13.</b> Número médio de questões com resposta “não claro” na aplicação do QUADAS nos diferentes periódicos analisados .....	67
<b>Tabela 14.</b> Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico <i>Clinical Biochemistry</i> .....	69
<b>Tabela 15.</b> Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.....	71

<b>Tabela 16.</b> Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.....	74
<b>Tabela 17.</b> Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.....	76
<b>Tabela 18.</b> Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC.....	79

## LISTA DE FIGURAS

<b>Figura 1:</b> Diagrama do processo de procura e seleção dos artigos de acurácia diagnóstica nos periódicos selecionado.....	36
<b>Figura 2:</b> Distribuição das áreas das análises clínicas dos artigos de acurácia diagnóstica recuperados nos (a) periódicos nacionais, (b) na totalidade dos artigos analisados, (c) <i>Clinical Biochemistry</i> , (d) Jornal Brasileiro de Patologia e Medicina Laboratorial, (e) LAES&HAES, (f) NewsLab e (g) Revista Brasileira de Análises Clínicas .....	44
<b>Figura 3:</b> Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico <i>Clinical Biochemistry</i> .....	46
<b>Figura 4:</b> Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico JBPML.....	51
<b>Figura 5:</b> Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico LAES&HAES .....	55
<b>Figura 6:</b> Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico NewsLab.....	59
<b>Figura 7:</b> Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação na RBAC .....	63
<b>Figura 8:</b> Percentual de itens com resposta “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico <i>Clinical Biochemistry</i> .....	68
<b>Figura 9:</b> Número de itens com resposta “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico JBPML.....	71
<b>Figura 10:</b> Número de itens com resposta “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico LAES&HAES.....	73
<b>Figura 11:</b> Número de itens com respostas “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação na revista NewsLab.....	76
<b>Figura 12:</b> Número de itens do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação na RBAC .....	78

<b>Figura 13:</b> Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro STARD nos diferentes anos analisados nos periódicos <i>Clinical Biochemistry</i> , RBAC e JBPML.....	81
<b>Figura 14:</b> Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro STARD nos diferentes anos analisados nos periódicos NewsLab e LAES&HAES.....	81
<b>Figura 15:</b> Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro QUADAS nos diferentes anos analisados nos periódicos <i>Clinical Biochemistry</i> , RBAC e JBPML.....	82
<b>Figura 16:</b> Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro QUADAS nos diferentes anos analisados na revista NewsLab e LAES&HAES.....	82
<b>Figura 17:</b> Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro STARD e QUADAS concernentes às questões relacionadas à validade interna na RBAC.....	90
<b>Figura 18:</b> Diagrama de dispersão das variáveis respostas afirmativas STARD e QUADAS na totalidade de artigos avaliados (n=101).....	91

**LISTA DE ABREVIATURAS**

BIOSIS - *Biological Abstracts-Current Contents*

CONSORT - *Consolidated Standards of Reporting Trials*

EMBASE - *Excerpta Medica Database*

FN - Falso Negativo

FP - Falso Positivo

JAMA - *Journal of the American Medical Association*

JBPML – *Jornal Brasileiro de Patologia e Medicina Laboratorial*

LILACS - *Literatura Latino-Americana e do Caribe em Ciências da Saúde*

MBE - *Medicina Baseada em Evidências*

MEDLINE - *Medical Literature Analysis and Retrieval System Online*

MeSH - *Medical Subject Headings*

MLBE - *Medicina Laboratorial Baseada em Evidências*

QUADAS - *Quality Assessment of Diagnostic Accuracy Studies*

QUOROM - *Quality of Reporting of Meta-analyses*

RBAC - *Revista Brasileira de Análises Clínicas*

RBCF - *Revista Brasileira de Ciências Farmacêuticas*

ROC - *Receiver Operator Characteristic*

SciELO - *Scientific Electronic Library Online*

SJR - *SCImago Journal & Country Rank*

STROBE - *Strengthening the Reporting of Observational Studies in Epidemiology*

STARD - *Standards for Reporting Diagnostic Accuracy*

VN – Verdadeiro Negativo

VP – Verdadeiro Positivo



## 1. INTRODUÇÃO

### 1.1 Considerações gerais

A qualidade das informações divulgadas pelas publicações científicas passa, necessariamente, pela clareza da redação dos estudos por seus autores. Portanto, a baixa qualidade de redação de um artigo científico dificulta ou impossibilita ao leitor julgar adequadamente acerca da validade do mesmo (SMIDT *et al.*, 2005).

Os testes diagnósticos, como parte essencial da prática médica, devem contar com especial atenção de sua avaliação, pois apesar dos notáveis avanços técnicos e da melhoria contínua de sua precisão analítica, a avaliação da acurácia diagnóstica de novos testes tem recebido, freqüentemente, apenas uma modesta atenção. Além disso, estudos demonstram que muitos não seguem mínimos padrões metodológicos para sua correta avaliação, fato este que pode superestimar seu valor diagnóstico (REID; LACHS; FEINSTEIN, 1995; LIJMER *et al.*, 1999, LUMBRERAS-LACARRA; RAMOS-RINCON; HERNANDES-AGUADO, 2004).

Iniciativas que visam a avaliação sistemática da redação de tipos específicos de estudos têm sido adotadas por diversas publicações. No caso dos testes diagnósticos, o STARD (*Standards for Reporting Diagnostic Accuracy*) foi elaborado e publicado em 2003 objetivando a adição de todas as informações relevantes à redação de um manuscrito e que muitas vezes são omitidas nos estudos de acurácia diagnóstica (BOSSUYT *et al.*, 2003a). Ao adotar um roteiro ou *checklist* para a redação correta de artigos científicos, a questão advinda é se este fator realmente contribuirá para melhora significativa da redação. A iniciativa CONSORT (*Consolidated Standards of Reporting Trials*) para a redação de ensaios clínicos controlados e randomizados demonstrou melhora significativa na redação deste tipo específico de artigo apenas 2 anos após sua publicação (MOHER; JONES; LEPAGE, 2001).

No presente estudo, a avaliação da qualidade de redação dos estudos de acurácia diagnóstica, foi realizada com a utilização do *checklist* STARD e do instrumento QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*). Ambos complementam-se, pois enquanto o STARD visa a qualidade da redação, o QUADAS, desenvolvido e validado de forma sistemática por Whiting *et al.* (2003), objetiva a análise da qualidade de estudos de acurácia diagnóstica incluídos em revisões sistemáticas. Sendo a qualidade de um estudo intimamente ligada à sua redação, a aplicação de ambos possibilitaria constatar se há associação entre a qualidade metodológica de um estudo e sua qualidade de redação.

Com a seleção dos periódicos a serem incluídos neste estudo, procurou-se heterogeneidade quanto à indexação e público-alvo entre os periódicos de circulação nacional, para que as publicações na área das Análises Clínicas no Brasil pudessem ser parcialmente representadas e nos possibilitassem um panorama da qualidade de redação dos artigos de acurácia diagnóstica nas publicações nacionais. Quanto ao periódico internacional eleito para o presente estudo, o mesmo promoveu, em seu ano de publicação, a iniciativa STARD como forma de melhorar a redação dos estudos de acurácia diagnóstica e por isso estaria com maior probabilidade de demonstrar alguma melhora na qualidade de redação dos artigos de acurácia diagnóstica. O período de avaliação remete ao fato de ambos (STARD e QUADAS) terem sido publicados em 2003 e possibilitar desta forma uma avaliação sobre o impacto destes na qualidade de redação dos artigos de acurácia diagnóstica nas publicações nacionais e na publicação internacional avaliadas.

## **1.2 Conhecimento, informação e qualidade da informação**

O conhecimento é objeto de estudo da filosofia desde a Grécia antiga, passando pelos mais importantes pensadores do século XX até a atualidade. Recentemente, tem-se testemunhado uma rápida expansão do interesse de outras áreas pelo conhecimento (STENMARK, 2001). Com o advento de evoluções nos campos da ciência e tecnologia, mudanças significativas na produção do conhecimento marcaram o final da segunda metade do século passado (NEVES, 2006), tornando o conhecimento humano atualmente rico, abundante e bastante abrangente (MOURA, 1996).

Importante no entendimento da criação do conhecimento é a diferenciação entre dados, informações e conhecimento. Vários autores buscam destacar essas diferenças, porém não há um consenso entre suas definições. Estes conceitos são tratados, normalmente, em um sentido hierárquico, em que dados são simples fatos que se tornam informações quando combinados de forma compreensível. Uma informação é transformada em conhecimento quando é possível ligá-la a outras informações, avaliando-a e entendendo seu significado em um contexto específico (SILVA, 2004). Outros autores conceituam informação como o conhecimento que pode ser transmitido sem a perda de sua integridade, sugerindo que informação é uma forma de conhecimento (STENMARK, 2001).

As diversas definições dadas para o conhecimento normalmente consideram essa hierarquia e convergem para a idéia de que o conhecimento é formado por informação, que pode ser expressa, verbalizada, e é relativamente estática e estável (formato explícito), em completo relacionamento com uma característica mais subjetiva e não palpável, que está na mente das pessoas e é relativamente instável ou dinâmica, e que envolve experiência, contexto, interpretação e reflexão (formato tácito) (SILVA, 2004; STENMARK, 2001).

A informação em seu formato explícito, entendida como meio de organização deste conhecimento, está dispersa em um enorme conjunto de fontes, sejam centros de pesquisas, bases de dados, empresas, bibliotecas entre outras (MOURA, 1996). As discussões advindas da filosofia da ciência e das ciências sociais, aliadas aos desafios das novas formas de comunicação através das redes de computadores, exigem que se enfrentem novas questões na abordagem da informação, como por exemplo, sua qualidade (NEHMY; PAIM, 1998).

A análise da qualidade da informação esbarra, inicialmente, na dificuldade de se conceituar qualidade (CASTRO, 2008; PAIM; NEHMY; GUIMARÃES, 1996). Vários termos são utilizados para a abordagem da avaliação da informação de forma similar à da qualidade, e entre eles o mais freqüente é o uso de *valor* como equivalente à qualidade. Apesar de valor possuir conotações desde filosóficas até econômicas, o uso neste contexto atribui a valor a significação mais próxima de *valor de uso* e não *valor de troca*, utilizado na economia (NEHMY; PAIM, 1998). De um modo geral, qualidade é entendida como a categoria mais abrangente sob a qual são agrupados fatores ou indicadores de diferentes significados, desta forma, a qualidade da informação é considerada como uma categoria multidimensional englobando a dimensão transcendente (filosófica ou metafísica), a intrínseca (características técnicas) e a contingencial (aspectos práticos) (PAIM; NEHMY; GUIMARÃES, 1996).

Este modelo multidimensional demonstra a complexidade própria do processo de avaliação da informação. Assim, focando os atributos intrínsecos, estes se referem aos valores inerentes ao dado, ou enfim, à informação. Os valores incluem dimensões como validade, confiabilidade, precisão, completeza, novidade, pertinência, atualidade, significado através do tempo, abrangência, as quais mantêm entre si uma estreita inter-relação. Validade pressupõe integridade da fonte de informação e forma de registro fiel ao fato que representa. A confiabilidade significa credibilidade no conteúdo e na fonte da informação. Relaciona-se com a idéia de autoridade cognitiva como prestígio, respeito, reputação da fonte, autor ou instituição. A precisão tem o sentido aproximado de exatidão, correção ou à forma de registro fiel ao fato representado. A completeza implica a inclusão de todos os dados necessários,

relativos a um determinado problema. E, finalmente, a novidade representa o novo, o recente. A atualidade implica consonância com o ritmo de produção da informação. Deve-se notar que a relação entre os diferentes atributos intrínsecos da qualidade da informação é extremamente forte, dificultando o estabelecimento de fronteiras entre um e outro, como por exemplo, entre os atributos precisão e a validade, que têm significados muito próximos. Na verdade, a integridade da noção de qualidade da informação pressupõe necessariamente a presença do conjunto dos atributos intrínsecos (PAIM; NEHMY; GUIMARÃES, 1996).

Quanto à informação na área da saúde, a avaliação da qualidade de informação encontrada nas publicações científicas envolve ao menos dois aspectos: sua validade interna (validade) e sua validade externa (aplicabilidade). Com o advento da medicina baseada em evidências (MBE) como novo paradigma da prática médica a qualidade das informações presentes nas publicações científicas tornou-se de fundamental importância.

### **1.3 Medicina baseada em evidências e Diagnóstico Laboratorial**

Há uma tendência atual de a prática clínica ser norteada pela corrente chamada medicina baseada em evidências. A MBE é definida como o uso consciente, explícito e criterioso das melhores evidências disponíveis na literatura médica para tomadas de decisões em relação ao manejo de pacientes (SACKETT *et al.*, 1996) e se apóia no tripé formado pela epidemiologia clínica, pela bioestatística e pela informática médica (VASCONCELOS-SILVA; CASTIEL, 2005). São consideradas evidências estudos clínicos publicados em periódicos científicos ou bancos de dados eletrônicos, sob forma de artigos originais, resumos estruturados de artigos originais, revisões sistemáticas e diretrizes (BERWANGER; AVEZUM; GUIMARÃES, 2006).

A qualidade da informação torna-se imprescindível no contexto da MBE, pois esta adota ser a evidência o principal fator a substanciar a tomada de decisão clínica e esta só será adequada se contiver boa qualidade (MCQUEEN, 2001; TRENTI, 2003). A qualidade almejada engloba, desde o correto delineamento metodológico para que as suas conclusões não sejam comprometidas e sejam transformadas em conhecimento facilitador da prática médica, até a adequada redação dos estudos (MCQUEEN, 2001).

A MBE transforma o paradigma da prática médica tradicional, assumindo que a intuição, a experiência clínica não sistematizada e razões fisiopatológicas não são suficientes para a tomada de decisão médica e promove a análise crítica das evidências de pesquisas clínicas (GUYATT; COOK; HAYNES, 2004). Ela está baseada na necessidade do uso do

conhecimento em seus dois formatos; o tácito resultante das observações clínicas individuais não sistemáticas e o explícito presente nos resultados de experimentos devidamente registrados (GUYATT; COOK; HAYNES, 2004; MCQUEEN, 2001; PRICE, 2000).

Diversos fatores estão relacionados para induzir o estabelecimento de uma cultura baseada em evidências em todas as áreas da medicina e outras áreas como a odontologia (CHIAPPELLI *et al.*, 2003) e a enfermagem (CRUZ; PIMENTA, 2005). Entre estes, estão o número crescente de publicações na área biomédica. Estima-se que sejam publicados mais de 30 mil volumes de periódicos biomédicos e mais de 17 mil novos títulos de livros a cada ano impossibilitando a adequada atualização dos profissionais. Aliados a este fator, encontra-se o fenômeno global do aumento dos custos dos sistemas de saúde, deficiências no processo educacional e a presença de um usuário dos sistemas de saúde mais crítico, ao menos nos países desenvolvidos, que procura o melhor diagnóstico e terapia em um menor espaço de tempo e que tem acesso à informação em meio eletrônico (MCQUEEN, 2001; TRENTI, 2003). Contudo, apesar das aclamações como novo paradigma da prática médica e de enfrentar críticas por sua grande ênfase estatística e abordagem administrativa, ela tem se mostrado útil em outras áreas da medicina, como a medicina laboratorial (PRICE, 2000; TRENTI, 2003).

A Medicina Laboratorial Baseada em Evidências (MLBE) faz uso da epidemiologia clínica, estatística e ciências sociais combinadas à patologia bioquímica e molecular para avaliar a efetividade de testes diagnósticos na tomada de decisões clínicas e desfecho dos pacientes (MCQUEEN, 2001; TRENTI, 2003). Os testes diagnósticos são uma ferramenta valiosa que o clínico lança mão para, aliado ao seu juízo crítico e conhecimento prévio, estabelecer a etiologia das queixas e anormalidades dos pacientes (FLORES, 2005; TATSIONI, 2005). Para a prática da medicina sob o novo paradigma da MBE, tanto o clínico quanto o analista clínico devem estar familiarizados com as questões que dizem respeito à precisão, exatidão e acurácia de um determinado teste, pois é baseada nestes conceitos básicos que estará centrada a discussão da aplicabilidade e da validade do resultado para o paciente em questão (FLORES, 2005).

A abordagem baseada em evidências é, portanto, altamente recomendável na medicina laboratorial, pois diferenças significativas em estratégias diagnósticas entre diferentes hospitais para a mesma apresentação clínica assim como consideráveis dados demonstrando uso inapropriado ou ineficaz de testes laboratoriais vem se tornando uma constante na área (TRENTI, 2003). Destacam-se, nesse processo, a demora do reconhecimento de intervenções clinicamente efetivas e sua subsequente aplicação na prática e também exemplos de

intervenções implementadas e depois consideradas sem utilidade (TRENTI, 2003). As revisões sistemáticas são consideradas, atualmente, fontes das melhores evidências disponíveis e demonstram que, na literatura científica, um expressivo número de estudos diagnósticos tem baixa qualidade devido a falhas fundamentais em seus delineamentos (MCQUEEN, 2001; MCQUEEN, 2003; TRENTI, 2003).

Apesar do grande desafio da medicina laboratorial em demonstrar seu papel no desfecho clínico dos pacientes assim como seu papel em fornecer valor adicional aos tratamentos médicos (TRENTI, 2003), é inegável a relevância que o laboratório clínico possui nas questões relativas a diagnóstico, terapêutica e prognóstico, portanto, o uso adequado dos testes laboratoriais é fundamental para a aplicação e o êxito da MLBE (FLORES, 2005).

A introdução constante de novos testes e o avanço da tecnologia médica em geral impossibilita a manutenção da atualização dos profissionais em relação ao arsenal de testes diagnósticos disponíveis, bem como de sua real aplicabilidade (FLORES, 2005). Por este motivo, a avaliação acurada de testes diagnósticos é um pré-requisito para a introdução destes nos sistemas de saúde, na tentativa de reduzir o uso de testes laboratoriais inapropriados e o uso dos recursos finitos na atual e permanente análise de custos e benefícios (FLORES, 2005; TRENTI, 2003).

#### **1.4 Estudos de acurácia diagnóstica**

Os testes diagnósticos são utilizados com o propósito de aumentar a certeza sobre a presença ou ausência de uma doença, monitorar uma progressão clínica, dar suporte a tratamentos ou analisar um prognóstico. Além disso, os testes diagnósticos devem apresentar benefícios clínicos, como influenciar a tomada de decisões clínicas, o manejo dos pacientes e o bem-estar dos mesmos. Portanto, testes que não possuam estas características são considerados obsoletos. Além disso, devem ser suficientemente seguros para que não induzam decisões por tratamentos inapropriados e preocupações desnecessárias (VAN DEN BRUEL et. al. 2007).

A partir do trabalho de Ledley e Lusted (1959), classificando os testes diagnósticos de modo hierárquico, várias foram as tentativas de classificar os estudos sobre testes diagnósticos (BOSSUYT, 2006; SACKETT; HAYNES, 2002; VAN DEN BRUEL et. al. 2007). A classificação dos testes diagnósticos de maneira hierárquica pode ser dividida em estudos que visam: avaliação de desempenho técnico (precisão, exatidão, intervalo analítico, interferentes, variação biológica e estabilidade da amostra); avaliação de desempenho

diagnóstico (sensibilidade, especificidade, razões de probabilidade positivas e negativas, curva ROC); avaliação de benefícios clínicos (estratégias diagnósticas e terapêuticas, melhorias na aderência medicamentosa, redução de toxicidade e efeitos adversos), avaliação de impacto organizacional (redução do tempo de internação, redução do tempo de utilização do quadro de funcionários, utilização reduzida de outros recursos e redução de riscos) e avaliação de decisões sobre custos e efetividades (PRICE, 2000; VAN DEN BRUEL et. al. 2007).

Entre os estudos que avaliam o desempenho diagnóstico estão os estudos de acurácia diagnóstica, onde esta se refere à habilidade de um teste em identificar uma condição de interesse (BRUNS, 2003; VAN DEN BRUEL et. al. 2007). O termo *acurácia* não deve ser confundido com acurácia analítica, avaliada nos estudos de desempenho técnico. Em medicina, testes diagnósticos são usados no diagnóstico, prognóstico, monitoramento e triagem. Portanto, acurácia analítica deve, logicamente, preceder estudos de acurácia diagnóstica (BRUNS *et al.* 2000).

Em estudos de acurácia diagnóstica, os resultados de um ou mais testes sob avaliação (*index test*) são comparados com os resultados do padrão de referência (*reference standard*), ambos realizados em pacientes suspeitos de possuírem uma condição de interesse. O termo *teste* se refere a qualquer método utilizado para obter informação adicional sobre o estado de saúde do paciente. Inclui exames clínicos, testes laboratoriais, exames de imagem, testes funcionais e histopatológicos. A condição de interesse ou condição alvo pode se referir a uma doença em particular, um estágio da doença, uma condição de saúde ou qualquer outra condição identificável que induza intervenção clínica, como testes diagnósticos ou iniciação, modificação e término do tratamento. O padrão de referência é considerado o melhor método disponível para estabelecer a presença ou ausência da condição de interesse e pode ser um único método ou uma combinação de métodos para estabelecer a presença da condição alvo (BOSSUYT *et al.*, 2003c).

Nos estudos clássicos de acurácia diagnóstica, pacientes suspeitos de possuírem a condição alvo se submetem ao *index test*. Todos os pacientes são, então, verificados pelo padrão de referência e ambos os testes devem ser interpretados utilizando mascaramento tipo duplo-cego (WHITTING *et al.*, 2004). O delineamento mais adequado é um estudo de corte transversal em que o teste avaliado é comparado ao teste padrão de referência em uma coorte de pacientes que são selecionados de uma população clinicamente relevante, isto é, pacientes em que o teste será aplicado na prática clínica (KNOTTNERUS; MURIS, 2003; VAN DEN BRUEL et. al. 2007).

Apesar da maioria dos testes diagnósticos possuírem resultados contínuos ou múltiplos, o agrupamento em categorias ou a aplicação do valor de corte possibilita a classificação dos resultados de um teste em duas categorias: positivos ou negativos. A partir desta dicotomização é possível representar a comparação entre o teste avaliado e o padrão de referência através de uma tabela de contingência 2 x 2 (Tabela 1). Os indicadores do desempenho do teste que derivam desta tabela são: a sensibilidade, especificidade, valores preditivos positivos e negativos, razões de verossimilhança positivos e negativos, área sob curvas ROC e razões de chance diagnóstica (GLAS *et al.*, 2003, LANGLOTZ, 2003).

**Tabela 1.** Tabela de contingência 2 x 2.

		Teste de referência	
		Condição-alvo presente	Condição-alvo ausente
Teste	Positivo	VP	FP
	Negativo	FN	VN

VP=verdadeiros positivos; FP=falsos positivos; FN=falsos negativos e VN=verdadeiros negativos

Os indicadores mais utilizados são sensibilidade e especificidade. Define-se sensibilidade como a proporção de pessoas com uma doença que tem um teste positivo, ou seja, a capacidade que o teste apresenta de detectar os indivíduos verdadeiramente positivos. A especificidade é a proporção de indivíduos sem a doença que apresentam um teste negativo ou a capacidade do teste em detectar os indivíduos verdadeiramente negativos (AGRESTI, 2002; FLETCHER; FLETCHER, 2006; LANGLOTZ, 2003; PINHO; MATTOS, 2002). Sensibilidade e especificidade são propriedades inerentes a cada teste e não variam substancialmente a não ser por mudanças na técnica ou por erros na sua aplicação (PEREIRA, 2006). Desta forma, um teste útil para descartar a doença deve ter alta sensibilidade e um teste útil para confirmar a doença deve ter alta especificidade (TRENTI, 2003).

Os valores preditivos são a probabilidade da doença, dados os resultados de um teste e são dependentes da prevalência do evento estudado. O valor preditivo positivo é a probabilidade da doença em um paciente com um resultado positivo (anormal) do teste. O valor preditivo negativo é a probabilidade de não ter a doença quando o resultado do teste for negativo (normal). O valor preditivo é também chamado de probabilidade posterior ou pós-teste, isto é, a probabilidade de ter a doença após o conhecimento do resultado do teste



(FLETCHER; FLETCHER, 2006, GLAS *et al.*, 2003; HONEST; KHAN, 2002; LANGLOTZ, 2003).

As razões de verossimilhança são uma forma alternativa de descrever o desempenho de um teste diagnóstico e é definida como a probabilidade de um determinado resultado em pessoas com a doença dividida pela probabilidade do resultado em pessoas sem a doença. Ela expressa quantas vezes mais (ou menos) é provável encontrar um resultado positivo em pessoas doentes, em comparação com as não-doentes (FLETCHER; FLETCHER, 2006, GLAS *et al.*, 2003; HONEST; KHAN, 2002; LANGLOTZ, 2003).

Um modo alternativo de expressar a relação entre a sensibilidade e a especificidade de um teste diagnóstico é construir uma curva, chamada de curva ROC (*receiver operator characteristic*). Ela é construída por meio da representação gráfica da taxa de verdadeiros-positivos (sensibilidade) contra a taxa de falso-positivos (1-especificidade). Os valores nos eixos vão de uma probabilidade de 0 a 1,0 ou, em outras palavras, de 0 a 100%. Por ser construída a partir de pares de valores de sensibilidade e especificidade ela demonstra a gravidade do contrabalanço entre a sensibilidade e a especificidade de um teste e pode ser utilizada para decisão do limiar de decisão clínica (ponto de corte) a partir do qual os resultados são considerados anormais, sendo especialmente valiosas para comparar testes alternativos para um mesmo diagnóstico (FLETCHER; FLETCHER, 2006; SHAPIRO, 1999; ZWEIG; CAMPBELL, 1993). Testes com bom poder discriminatório concentram-se no canto esquerdo superior da curva ROC, enquanto testes com menor poder discriminatório apresentam curvas que ficam mais próximas da diagonal que vai da esquerda inferior para a direita superior (TRENTI, 2003). A acurácia geral de um teste pode também ser descrita como a área sob a curva ROC, segundo a qual quanto maior for a área, melhor será o teste. (FLETCHER; FLETCHER, 2006; SHAPIRO, 1999; ZWEIG; CAMPBELL, 1993)

Outra estimativa empregada para expressar a acurácia são as razões de chance diagnósticas (*diagnostic odds ratio*). Estas não dependem da prevalência e é a razão entre a chance do teste avaliado ser positivo quando há doença comparada com a chance do mesmo resultado do teste em pacientes sem a doença. Alternativamente pode ser interpretado como a razão de chances de doença em testes positivos em relação às chances de doença em testes negativos (GLAS *et al.*, 2003; HONEST; KHAN, 2002; PANAGEAS, 2003).

### 1.5 Avaliação da qualidade dos artigos de acurácia diagnóstica

A avaliação da qualidade das publicações científicas encontra-se, atualmente, associada ao seu corpo editorial e ao processo conhecido como revisão por pares (*peer review*). A revisão por pares é definida como a avaliação dos achados de uma pesquisa por sua qualidade, significância e originalidade por especialistas qualificados que atuam como sentinelas entre as descobertas científicas e sua publicação. Suas revisões são uma tentativa de assegurar a qualidade da informação científica e atuam na redução da desinformação e confusão (BENOS *et al.*, 2006; LANDKROON *et al.*, 2006).

O sistema de revisão por pares teve seu início no século XVIII, evoluindo até seu formato atual que possui consideráveis variações em sua aplicação. Em geral incorpora um processo que compreende a distribuição aos revisores, a sua avaliação e a chegada a um consenso a respeito do mérito dos manuscritos submetidos, mérito esse, evidenciado pela aceitação ou rejeição por parte da publicação (BENOS *et al.*, 2006; GOODMAN *et al.*, 1994). Apesar de suas deficiências, caracterizadas por atrasos desnecessários até a publicação de manuscritos e até mesmo ineficácia geral do processo, ela fornece uma oportunidade formal aos autores de avaliar a reação ao seu trabalho bem como permite a possível detecção e subsequente correção de erros ou falhas antes que este venha a se tornar de domínio público. Desta forma, a revisão por pares introduz um valor adicional na publicação (BENOS *et al.*, 2006).

Nas últimas décadas, muitas publicações têm reconhecido a importância da avaliação da qualidade da informação na área da saúde por parte de seus leitores e algumas iniciativas têm sido desenvolvidas neste sentido (BOSSUYT, 2004). Em 1981 foi publicada pelo *Canadian Medical Association Journal* uma série de cinco artigos que explicavam como avaliar criticamente a literatura médica sob o título *How to read clinical journals* (SACKETT, 1981). A aceitação desta série e os novos conhecimentos acumulados geraram outra série de artigos publicada pelo *Journal of the American Medical Association* (JAMA) intitulada *User's guide to the medical literature* (OXMAN; SACKETT; GUYATT, 1993). Também com esta mesma perspectiva foram introduzidos o modelo de artigo e resumo estruturado e os requisitos uniformes de Vancouver para manuscritos submetidos a jornais biomédicos (INTERNATIONAL COMMITTEE OF MEDICAL JOURNAL EDITORS, 1997).

Apesar da definição da palavra qualidade não ser fácil, entende-se que a avaliação da qualidade de uma publicação envolve ao menos três aspectos: sua validade interna, sua validade externa ou aplicabilidade e a análise estatística ou importância (CASTRO, 2008). Os critérios de validade interna para a pesquisa diagnóstica se referem às características do estudo que o protegem contra erros ou vieses, como a utilização de um padrão de referência válido, a definição adequada do limiar de positividade do padrão de referência e o próprio planejamento do estudo. Critérios de validade externa fornecem informações sobre a aplicabilidade, generalização dos resultados do estudo e julgamento se o teste foi avaliado de acordo com padrões metodológicos aceitáveis. Enquadram-se neste critério a descrição adequada das características demográficas da população, a execução e a reprodutibilidade do teste (DEVILLÉ *et al.*, 2002).

Estudos de acurácia diagnóstica são necessários para validar novos testes antes de estes serem incorporados na prática clínica e os índices de acurácia gerados permitem ao clínico avaliar a utilidade potencial do novo teste. Entretanto, estudos conduzidos inadequadamente e com baixa qualidade de redação são propensos a vieses que, por sua vez, podem levar a avaliações excessivamente otimistas do valor diagnóstico do novo teste (LIJMER *et al.*, 1999). Desta forma, resultados exagerados podem levar à adoção prematura do uso destes, podendo levar a tomada de decisão incorreta pelos clínicos. Uma avaliação rigorosa dos testes diagnósticos antes da sua introdução na prática clínica pode, não somente, reduzir o número de consequências clínicas indesejadas relacionadas a estimativas equivocadas de acurácia diagnóstica, mas também reduzir os custos dos sistemas de saúde evitando testes desnecessários (BOSSUYT *et al.*, 2003c). Diversas publicações têm demonstrado que a qualidade de redação dos estudos de acurácia diagnóstica é frequentemente pobre, tornando difícil o julgamento acerca da validade, incorporação de vieses e aplicabilidade a pacientes em outros cenários clínicos (BOSSUYT, 2004).

Vieses estão presentes no estudo quando distorções são introduzidas como consequência de defeitos no delineamento e condução do estudo. Por esta razão um estudo de acurácia diagnóstica com vieses pode produzir avaliações do teste que diferem do seu verdadeiro desempenho. Em contrapartida, variabilidade surge das diferenças entre os estudos, como por exemplo, em termos da população avaliada, protocolo do teste ou definição da condição de interesse. Embora a variabilidade não leve a avaliações com viés do desempenho do teste, ela pode limitar a aplicabilidade dos resultados e, portanto é importante considerá-la na avaliação do estudo (WHITING *et al.*, 2004). As principais fontes de viés e variação são sumarizadas na Tabela 2.

Diversos autores têm avaliado a qualidade dos estudos diagnósticos. Reid, Lachs e Feinstein (1995) apontaram diversas limitações metodológicas das pesquisas em testes diagnósticos publicados em periódicos científicos internacionais. Eles examinaram a aderência a sete padrões metodológicos de epidemiologia clínica que freqüentemente não eram apresentados na redação dos artigos científicos. Estes incluíam características demográficas (idade e sexo) e clínicas (severidade, duração da doença e comorbidades) da população estudada, análise dos subgrupos de pacientes, prevenção contra vieses, precisão dos resultados para a acurácia do teste, apresentação de resultados indeterminados e da reprodutibilidade do teste. Em 112 estudos de testes diagnósticos publicados entre os anos 1978 e 1993, estes autores demonstraram que a aplicação dos sete critérios, apesar de apresentar melhora no decorrer do período avaliado, continuavam não sendo utilizados com freqüência satisfatória. Por exemplo, apenas 8% dos artigos incluíam cálculo da acurácia diagnóstica para subgrupos relevantes de pacientes e viés de verificação foi evitado por não mais que 46% dos artigos avaliados.

Lijmer *et al.* (1999) demonstraram que muitos estudos de testes diagnósticos continham deficiências metodológicas e estas podiam resultar em uma superestimação da acurácia diagnóstica dos testes. Demonstraram que estudos de caso-controle que utilizavam pacientes saudáveis como controles levavam também à superestimação da acurácia diagnóstica, comparados aos estudos de coorte. Além disso, o conhecimento dos resultados do teste em avaliação e o uso de informações clínicas na interpretação do teste padrão de referência também tinham a mesma conseqüência. Os critérios avaliados por estes pesquisadores, diziam respeito ao delineamento do estudo, incluindo seleção dos pacientes, à coleta dos dados e à própria redação do artigo. Na avaliação de 218 artigos científicos, apenas 6,8% satisfizeram todos os 8 critérios utilizados e 30% satisfizeram 6 ou mais critérios.

O primeiro roteiro para a redação de estudos de acurácia diagnóstica foi proposto por Bruns *et al* (2000), baseados no conhecimento de que a tecnologia dos testes diagnósticos demonstrava constante melhoria de acurácia analítica e precisão, mas a avaliação da acurácia diagnóstica recebia apenas uma modesta atenção. Incluíram em seu roteiro, itens que objetivavam a boa informação, com ênfase nos critérios de inclusão dos pacientes nos estudos e na clareza de redação dos resultados.

Em 2004, Lumbreras-Lacarra, Ramos-Rincon e Hernández-Aguado avaliaram 79 estudos de testes diagnósticos laboratoriais publicados nos periódicos *Clinical Chemistry* e *Clinical Chemistry and Laboratory Medicine* nos anos de 1996, 2001 e 2002, utilizando os sete padrões metodológicos utilizados por Reid, Lachs e Feinstein (1995). O número médio

dos critérios satisfeitos pelos artigos científicos analisados foi de 2,3 em 1996, 2,7 em 2001 e 3,4 em 2002 demonstrando, ainda, uma grande subutilização dos mesmos.

Ao longo das últimas décadas, várias publicações têm reconhecido a importância da qualidade de redação dos artigos publicados e vários roteiros ou *checklists* para avaliar a qualidade de redação de tipos específicos de estudos foram desenvolvidos (BOSSUYT, 2004; MCQUEEN, 2001; SMIDT *et al.*, 2005; SMIDT *et al.*, 2006a). Entre estes se encontram os roteiros de como devem ser redigidos artigos científicos que relatam resultados dos ensaios clínicos controlados e randomizados como o *CONSORT*, dos estudos observacionais como o *STROBE* (*Strengthening the Reporting of Observational studies in Epidemiology*), das revisões sistemáticas de ensaios clínicos randomizados como o *QUOROM* (*Quality of Reporting of Meta-analyses*) e dos estudos de acurácia diagnóstica como o *STARD* (BEGG *et al.*, 1996; BOSSUYT *et al.*, 2003c; MOHER *et al.*, 1999; VANDENBROUCKE *et al.*, 2007).

Em 1996, um grupo de pesquisadores desenvolveu uma relação de itens a serem incluídos em um checklist para estudos de acurácia diagnóstica. Esta relação foi publicada primeiramente no periódico *Clinical Chemistry* para comentários, e uma versão corrigida foi publicada neste mesmo periódico em 2000 e incorporada às informações aos autores. Este *checklist* foi então utilizado pelos editores na revisão dos estudos de acurácia diagnóstica publicados na *Clinical Chemistry* durante os anos de 2001 e 2002. Em janeiro de 2003 o *STARD* (*Standards for Reporting Diagnostic Accuracy*) foi publicado simultaneamente em oito jornais médicos internacionais (*Radiology*, *American Journal of Clinical Pathology*, *Annals of Internal Medicine*, *British Medical Journal*, *Clinical Biochemistry*, *Clinical Chemistry*, *Clinical Chemistry of Laboratory Medicine* e *Lancet*). (BOSSUYT, 2004). O roteiro consiste em uma lista de 25 itens e na recomendação do uso de um diagrama delineando o estudo e o fluxo dos pacientes através deste. Os itens são arranjados sob os títulos: título, resumo e palavras-chave (1 item), introdução (1 item), métodos (11 itens), resultados (11 itens) e discussão (1 item). Este pode ser empregado pelo autor de um artigo científico avaliando testes diagnósticos de forma a assegurar que todas as informações relevantes estejam apresentadas em sua redação (BOSSUYT *et al.*, 2003a; BRUNS, 2003; SMIDT *et al.*, 2006a; SMIDT *et al.*, 2006b).

**Tabela 2.** Principais fontes de viés e variabilidade

Fonte	Viés ou Variação	Descrição
<b>População</b>		
Características demográficas	Variação	Teste pode ter desempenho diferente em amostras diferentes. Portanto, podem levar à variações na estimativa do desempenho do teste.
Severidade da doença	Variação	Diferença na severidade da doença entre os estudos pode levar à diferenças na estimativa do desempenho do teste
Prevalência da doença	Variação	A prevalência da condição alvo varia de acordo com o local e pode afetar a estimativa de desempenho do teste. Viés de contexto faz com que os responsáveis pela interpretação dos testes considerem estes mais frequentemente positivos em locais de alta prevalência da doença, podendo afetar a estimativa de desempenho do teste.
Seleção distorcida dos pac.	Variação	O processo de seleção determina a composição da amostra do estudo. Se o processo de seleção não objetiva incluir um espectro de pacientes similar ao da população em que o teste vai ser usado na prática, os resultados do estudo podem ter aplicabilidade limitada.
<b>Protocolo do teste: materiais e métodos</b>		
Execução do teste	Variação	Uma descrição suficiente da execução do index teste e do padrão de referência é importante porque variações na medida da acurácia diagnóstica podem ser resultado de diferenças na execução do teste.
Tecnologia do teste	Variação	Quando as características do teste mudam ao longo do tempo como resultado de melhorias tecnológicas ou a experiência do operador do teste, estimativas do desempenho do teste podem ser afetadas.
Paradoxo tratamento ou viés de progressão da doença	Viés	Viés de progressão da doença ocorre quando o index test é realizado após um tempo extraordinariamente longo após o padrão de referência, então a doença está em um estágio mais avançado quando o padrão de referência é realizado. Paradoxo de tratamento ocorre quando o tratamento é iniciado com base no resultado do index test e o padrão de referência é aplicado após o início do tratamento.
<b>Padrão de referência e procedimentos de verificação</b>		
Padrão de referência inapropriado	Viés	Padrões de referência imperfeitos podem gerar viés na medida da acurácia diagnóstica do index test.
Viés de verificação diferencial	Viés	Parte dos resultados do index test são verificados por um padrão de referência diferente.
Viés de verificação parcial	Viés	Apenas uma amostra selecionada dos pacientes que se submeteram ao index test são verificados pelo padrão de referência
<b>Interpretação</b>		
Viés de revisão	Viés	Interpretação dos index e padrão de referência são influenciados pelo conhecimento dos resultados do outro teste. Viés de revisão diagnóstica ocorre quando os resultados do index test são conhecidos quando o padrão de referência é interpretado. Viés de revisão do teste ocorre quando os resultados do padrão de referência são conhecidos enquanto o index test é interpretado.
Viés de revisão clínica	Viés	A disponibilidade de informações de dados clínicos, como idade, sexo e sintomas durante a interpretação dos resultados do teste pode afetar as estimativas de desempenho do teste.
Viés de incorporação	Viés	O resultado do index tes é usado para estabelecer o diagnóstico final.
Variabilidade do observador	Variação	A reprodutibilidade do resultado do teste é um dos determinantes da acurácia diagnóstica de um index test. Devido a variação nos procedimentos laboratoriais ou observadores, pode não ser consistente em gerar os mesmos resultados quando repetido. Em duas ou mais observações do mesmo estudo diagnóstico, variabilidade intraobservador ocorre quando a mesma pessoa obtém diferentes resultados, e interobservador quando 2 ou mais pessoas discordam.
<b>Análise</b>		
Manejo de resultados indeterminados	Viés	Um teste diagnóstico pode produzir resultados não interpretáveis que variam em frequência dependendo do teste. Estes problemas, frequentemente não são relatados nos estudos de eficácia do teste. E estes resultados são simplesmente removidos da análise. Isto pode levar a um viés da análise das características do teste.
Escolha arbitrária do ponto de corte	Variação	A seleção do ponto de corte para o index test que maximize a sensibilidade e especificidade do teste pode levar a medidas muito otimistas do desempenho do teste. O desempenho deste cutoff em outra população pode não ser a mesma como no estudo original.

Fonte: adaptado de Whiting et al. (2004)

Vários itens do roteiro *STARD* foram incorporados ao instrumento *QUADAS* (*Quality Assessment of Diagnostic Accuracy Studies*), desenvolvido com a finalidade de verificar a qualidade de artigos científicos de acurácia diagnóstica, incluídos em revisões sistemáticas, (HATJIMIHAİL, 2006; TATSIONI *et al.*, 2005). A qualidade objetivada pelo instrumento está relacionada tanto aos critérios de validade interna quanto externa desses estudos. Ele foi desenvolvido de maneira sistemática por Whiting *et al.* (2003) baseando-se em três revisões sistemáticas sobre o assunto e também em um grupo de discussão envolvendo pesquisadores altamente qualificados na área. O instrumento *QUADAS* foi desenvolvido e disponibilizado em 2003 e consiste em um checklist estruturado compreendendo 14 questões. Os itens fornecem uma abordagem padronizada da análise da qualidade e cobre o espectro de pacientes, a escolha do padrão de referência, viés de progressão de doença, viés de verificação, viés de revisão, viés de revisão clínica, execução do teste, saída de pacientes no decorrer do estudo e resultados indeterminados (WHITING *et al.*, 2003). Adaptações do instrumento *QUADAS* têm sido propostas, baseadas em tipos particulares de metodologia de testes diagnósticos como, por exemplo, nas ciências “ômicas” (genômica, proteômica e metabolômica) (LUMBRERAS *et al.*, 2008).

O instrumento *QUADAS* tem sido amplamente utilizado por pesquisadores constando seu uso em inúmeras revisões sistemáticas em condições como dengue, dermatite atópica, fraturas vertebrais, lesão labral, osteoporose e dispnéia (BLACKSELL *et al.*, 2006; BRENNINKMEIJER *et al.*, 2007; HENSCHKE; MAHER; REFSHAUGE, 2008; MUNRO; HEALY, 2008; RUD *et al.*, 2008; WORSTER *et al.*, 2008). O *QUADAS*, avaliado por Whiting *et al.* (2006) quanto à sua validade e utilidade, demonstrou concordância entre três revisores em 30 estudos incluídos em revisões sistemáticas com concordância sobre todos os itens acima de 80%.

## 2. Justificativa

O desenvolvimento contínuo de testes diagnósticos gera um grande número de artigos científicos que analisam sua eficácia. Uma avaliação rigorosa destes, antes de sua introdução na prática clínica pode, não somente reduzir o número de conseqüências clínicas indesejadas, mas também reduzir os custos dos sistemas de saúde ao evitar a incorporação de testes sem utilidade diagnóstica. Importante ressaltar que a avaliação dos testes diagnósticos passa, necessariamente, pela redação dos estudos por seus autores e que esta deve ser precisa e transparente permitindo aos leitores avaliar a probabilidade de incorporação de vieses durante a condução do estudo.

Apesar dos visíveis esforços para a melhoria na qualidade da metodologia e da redação de artigos científicos, como a publicação de diretrizes para a redação de tipos específicos de estudos, evidências apontam para uma baixa qualidade de redação dos artigos de acurácia diagnóstica publicados nos principais periódicos médicos internacionais e mesmo após a publicação do STARD não foi demonstrada tendência pronunciada em direção a uma melhora desta qualidade.

Estes fatos aliados a uma carência de informações específicas a respeito da qualidade de redação dos artigos de acurácia diagnóstica visando a avaliação de testes laboratoriais em periódicos nacionais e internacionais torna este trabalho uma tentativa de apresentar um panorama desta qualidade em periódicos científicos nacionais e em um periódico internacional na área das Análises Clínicas.

Alguns periódicos nacionais analisados, apesar de não possuírem indexação, têm sua importância ligada à sua ampla utilização por profissionais e pesquisadores, um grande número de leitores estimados por edição, além de distribuição gratuita. Quanto ao período proposto (2002 a 2006) para a análise das publicações, este pretende avaliar se houve melhora significativa desde a publicação do roteiro STARD e do instrumento QUADAS, visto que ambos foram inicialmente publicados em 2003.



### 3. Objetivos

#### 3.1 Objetivo Geral

Avaliar por meio dos instrumentos *STARD* e *QUADAS* a qualidade da redação de artigos científicos publicados em periódicos brasileiros e estrangeiro de circulação nacional e internacional, de importância para as análises clínicas publicados no período de 2002 a 2006.

#### 3.2 Objetivos Específicos

- 1) Localizar e recuperar artigos de acurácia diagnóstica de testes laboratoriais nas revistas nacionais Revista Brasileira de Ciências Farmacêuticas, Jornal Brasileiro de Patologia e Medicina Laboratorial, Revista Brasileira de Análises Clínicas, LAES&HAES, NewsLab;
- 2) Localizar e recuperar artigos de acurácia diagnóstica de testes laboratoriais na revista internacional *Clinical Biochemistry*;
- 3) Avaliar nos estudos recuperados a qual área das análises clínicas pertencem;
- 4) Avaliar a qualidade da redação dos artigos recuperados, utilizando-se o roteiro *STARD* e o instrumento *QUADAS*;
- 5) Comparar a qualidade da redação dos periódicos entre si e entre periódicos nacionais e o internacional;
- 6) Verificar a influência da publicação do *STARD* e do *QUADAS* na qualidade da redação dos artigos a partir da análise da quantidade de itens atendidos nos diferentes periódicos;
- 6) Inferir considerações sobre a qualidade da redação dos artigos a partir da análise da validade interna e validade externa.

## **4. Materiais e métodos**

### **4.1 Critérios de escolha e inclusão dos periódicos científicos a serem avaliados:**

Os critérios de escolha e inclusão dos periódicos científicos basearam-se em sua ampla utilização por profissionais e pesquisadores, classificação pelo programa Qualis CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) para a área de farmácia no triênio 2004-2006, avaliação pelo JCR (*Journal Citation Reports*) e disponibilidade de acesso.

A partir da aplicação destes critérios de escolha foram incluídos no presente estudo: um periódico científico de circulação internacional, com maior fator de impacto e com acesso completo no portal CAPES, três periódicos científicos de circulação nacional, com acesso completo no portal CAPES, independente de suas avaliações pelo JCR; e dois periódicos científicos de circulação nacional, independentes de suas avaliações pelo JCR e de sua disponibilidade de acesso completo no portal CAPES.

### **4.2 Tipo de estudos incluídos:**

Estudos primários (na forma de artigos científicos publicados em periódicos nacionais e internacionais especializados na área de Análises Clínicas) avaliando a acurácia diagnóstica de testes laboratoriais.

### **4.3 Amostragem:**

#### **4.3.1 Localização e Fonte dos estudos**

Como fontes do estudo foram utilizadas todas as edições dos periódicos científicos impressos nos anos 2002, 2003, 2004, 2005 e 2006. Também foram utilizadas todas as edições dos periódicos científicos disponibilizados por meio eletrônico no portal CAPES e nos sítios (*sites*) dos próprios periódicos nos anos 2002, 2003, 2004, 2005 e 2006.

Foram utilizados os periódicos nacionais, Revista Brasileira de Análises Clínicas (Rio de Janeiro), Jornal Brasileiro de Patologia e Medicina Laboratorial (Rio de Janeiro), Revista Brasileira de Ciências Farmacêuticas (São Paulo), NewsLab (São Paulo) e LAES-HAES (São Paulo). *Clinical Biochemistry* (Canadá) foi o periódico internacional eleito para o presente estudo.

#### **4.3.2 Critérios de inclusão de artigos científicos:**

Os artigos foram incluídos se a data de publicação estivesse entre os anos 2002 e 2006, o assunto tratasse de pesquisa em testes diagnósticos, se enquadrasse como estudos primários de acurácia diagnóstica e a população investigada fosse constituída de seres humanos.

O mestrando e seu orientador avaliaram independentemente os títulos, resumos e palavras-chave de todos os artigos para determinar se os mesmos atendiam aos critérios de inclusão. Não havendo dúvidas, o texto completo foi então recuperado e lido por ambos os revisores. No caso de discordância entre a inclusão ou não de algum artigo um terceiro revisor, também professor ligado ao programa de pós-graduação da UFSC, foi consultado.

#### **4.4 Extração dos dados**

Os dois revisores, independentemente, aplicaram a cada artigo o roteiro *STARD* (Anexo A) e o instrumento *QUADAS* (Anexo B). Neste procedimento, os revisores determinaram se cada item dos formulários havia sido adequadamente descrito no texto. Foram coletados os dados descrevendo os detalhes bibliográficos do artigo (periódico e ano) e a que área das análises clínicas cada artigo científico pertencia. Divergências entre os dados coletados foram resolvidas através de consenso entre os revisores, e a opinião do terceiro revisor foi ouvida caso estas persistissem.

#### 4.5 Análise e síntese dos dados coletados

Para cada item do formulário *STARD*, foi calculado o número total de artigos que o atenderam. Para cada artigo foi contado o número total de itens do *STARD* (variação de 0 a 25) como indicação da qualidade da redação dos artigos científicos. Os itens 8, 9, 10, 11, 13 e 24 são relacionados ao teste avaliado e ao padrão de referência e foi contado ½ item para o teste avaliado e ½ item para o padrão de referência.

Para cada item do formulário *QUADAS*, que compreende 14 questões, os artigos são avaliados através de três tipos de respostas possíveis: resposta “sim”, “não” ou “não claras” ao item. Para cada item do formulário *QUADAS*, foi calculado o número total de artigos que o atenderam. Para cada artigo foi contado o número total de itens do *QUADAS* para cada uma das respostas possíveis como indicação da qualidade dos artigos científicos. A avaliação da qualidade do estudo é realizada em conformidade com a quantidade de respostas “sim” obtidas, portanto, artigos com maior número de respostas “sim” foram considerados de maior qualidade.

As características principais dos estudos primários foram sumarizadas em tabelas contendo detalhes bibliográficos, tipo de estudo e o resultado da avaliação de qualidade do estudo.

#### 4.6 Análise estatística

Para avaliar a concordância entre os revisores foram realizados os testes estatísticos de Kappa (LANDIS; KOCH, 1977). Para a interpretação dos valores estimados de Kappa foi utilizada a classificação sugerida por Landis e Koch (1977), indicando a força de concordância: pobre (kappa <0,00), fraco (kappa 0,00-0,20), regular (kappa 0,21-0,40), moderado (kappa 0,41-0,60), forte (kappa 0,61-0,80) e quase perfeito (kappa 0,81-1,00). O coeficiente de correlação de *Pearson* foi calculado para comparação entre os escores do *STARD* e *QUADAS*. Para a avaliação da presença de tendência linear pela aplicação do roteiro *STARD* e instrumento *QUADAS* foi utilizado o teste de  $\chi^2$  Armitage. Valores de  $P < 0,05$  foram considerados significativos estatisticamente.

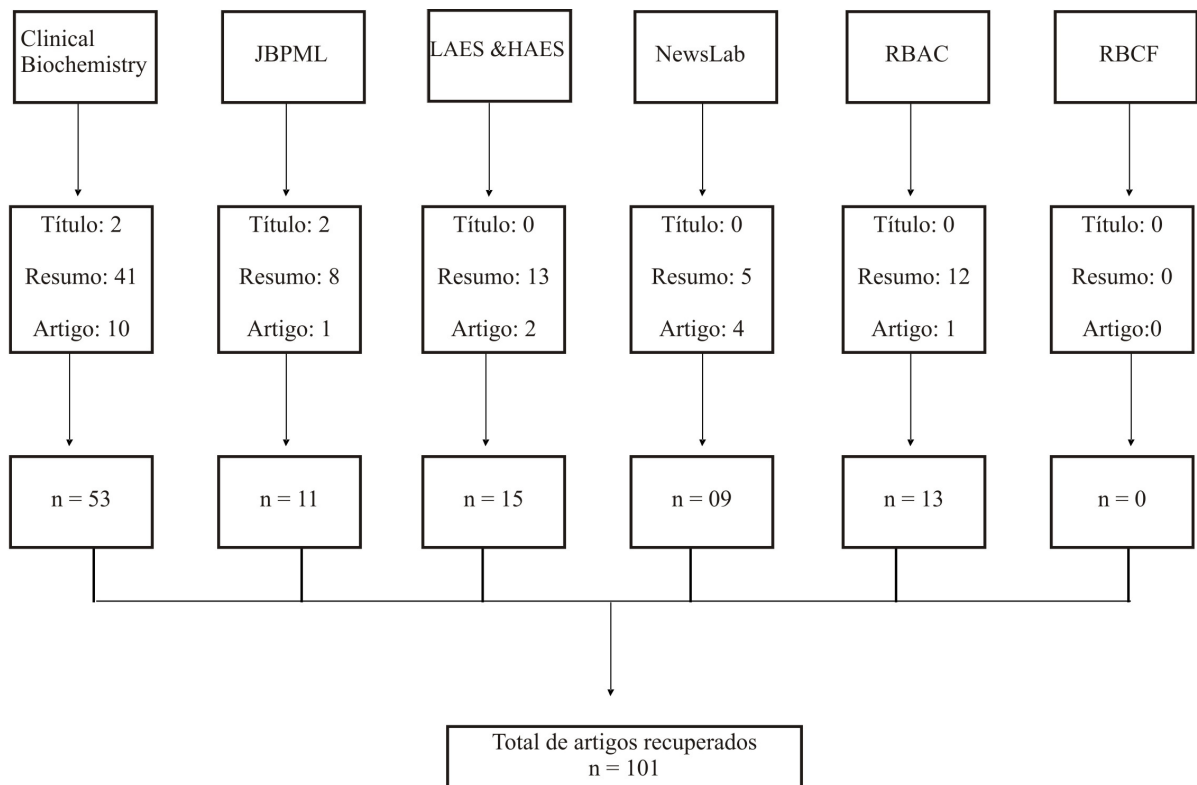
Os dados serão analisados utilizando-se os programas computacionais MedCalc® for Windows versão 9.3.0.0 (Copyright© 1993-2007 Frank Schoonjans; MedCalc Software; Broekstraat 52, B-9030 Mariakerke, Belgium) e STATISTICA for Windows (STATSOFT, Inc 1998).

## 5. Resultados

### 5.1 Localização e recuperação dos estudos de acurácia diagnóstica

Os periódicos científicos incluídos neste estudo foram *Clinical Biochemistry*, *Jornal Brasileiro de Patologia e Medicina Laboratorial*, *LAES&HAES*, *NewsLab*, *Revista Brasileira de Análises Clínicas* e *Revista Brasileira de Ciências Farmacêuticas*. Foram localizados e recuperados 101 artigos de acurácia diagnóstica. A *Revista Brasileira de Ciências Farmacêuticas* não apresentou artigos de acurácia diagnóstica no período analisado.

O diagrama apresentado na Figura 1 mostra de forma esquemática o processo de seleção e recuperação dos artigos nos periódicos selecionados. O processo de busca dos artigos de acurácia diagnóstica foi realizado por meio da leitura dos títulos e resumos de todos os artigos publicados no período definido do estudo. Quando houve indicação pelo título e resumo de tratar-se de um artigo de interesse, mas faltavam evidências claras dos critérios de inclusão definidos, a leitura integral do artigo foi então realizada. Título, resumo e artigo, indicados no diagrama, representam a ordem e em qual etapa da busca tornou-se claro tratar-se de artigo de acurácia diagnóstica.



**Figura 1.** Diagrama do processo de procura e seleção dos artigos de acurácia diagnóstica nos periódicos selecionados.

O número total de artigos de acurácia diagnóstica nos periódicos, sua distribuição nos anos analisados e o número total de artigos publicados são apresentados na Tabela 3. O número de artigos de acurácia diagnóstica recuperados por periódico, variou entre 9 e 53, correspondendo a 3,57 e 7,55% do total de artigos publicados. *Clinical Biochemistry*, JBPML, LAES&HAES, NewsLab e RBAC, publicaram no período de 2002 a 2006 um total de 702, 275, 241, 252 e 261 artigos, respectivamente.

Foram comparadas as proporções dos artigos de acurácia diagnóstica em relação aos totais publicados nos periódicos nacionais (48/1029) e no periódico internacional (53/702). Os artigos de acurácia, nos periódicos nacionais, correspondiam a 4,66% do total publicado e no periódico internacional a 7,55%. Ficou demonstrada diferença estatística significativa entre as proporções nos periódicos nacionais e internacional ( $P=0,011$ ) pelo teste de  $\chi^2$  (Tabela 3).

## 5.2 Características das revistas e dos artigos de acurácia diagnóstica

As características dos periódicos analisados como classificação no Qualis CAPES na área de Farmácia, indexação, fator de impacto, índice H, periodicidade e disponibilidade foram coletadas. Dentre as seis revistas analisadas apenas a *Clinical Biochemistry* e o JBPML são avaliados pelo JCR possuindo um fator de impacto de 2,072 e 0,0593, respectivamente. O portal SJR (*SCImago Journal & Country Rank*) apresenta a avaliação da revista *Clinical Biochemistry*, JBPML e RBCF com índices de 0,303, 0,037 e 0,046 no ano de 2007. Quanto à indexação, a revista *Clinical Biochemistry* é indexada pelo MEDLINE (*Medical Literature Analysis and Retrieval System Online*); JBPML, RBAC e RBCF são indexados pelo SciELO (*Scientific Electronic Library Online*) e/ou LILACS (Literatura Latino-Americana e do Caribe em Ciências da Saúde), enquanto as revistas NewsLab e LAES&HAES não possuem indexação (Tabela 4).

Os artigos de acurácia diagnóstica foram também classificados de acordo com as áreas de conhecimento específico da grande área de Análises Clínicas as quais pertenciam. As áreas encontradas foram: a bioquímica, a imunologia, a hematologia, a citologia, a parasitologia, a microbiologia, a urinálise, a genética e a histoquímica. Os gráficos de setor demonstrando a distribuição entre as diferentes áreas são apresentados na Figura 2.

Na avaliação dos periódicos nacionais, incluindo todos os artigos de acurácia diagnóstica localizados e recuperados nos periódicos nacionais (n=48), a área predominante dos procedimentos diagnósticos avaliados foi a imunologia (36%), seguida pela microbiologia

**Tabela 3.** Número de artigos de acurácia diagnóstica recuperados nos periódicos, totais por ano analisado e número total de artigos publicados por periódico.

Periódicos	Artigos publicados por ano					Artigos recuperados n (% do total)	Nº de artigos publicados
	2002	2003	2004	2005	2006		
Jornal Brasileiro de Patologia e Medicina Laboratorial	3	2	3	3	-	11 (3,98)	275
LAES & HAES	2	1	2	4	6	15 (6,64)	241
NewsLab	4	1	1	2	1	09 (3,57)	252
Revista Brasileira de Análises Clínicas	2	2	2	3	4	13 (4,98)	261
Subtotal	11	6	8	12	11	48 (4,66)*	1029
<i>Clinical Biochemistry</i>	7	9	11	14	12	53 (7,55)*	702
Totais	18	15	19	26	23	101(5,83)	1731

\*Diferença estatística significativa  $\chi^2$  ( $P=0,011$ )



**Tabela 4.** Características e índices bibliográficos dos periódicos científicos analisados.

Periódico (ano de início da publicação)	Qualis CAPES <sup>a</sup> Farmácia	Indexação	Fator de Impacto <sup>b</sup> (2007)	SJR <sup>c</sup> (2007)	Índice H <sup>d</sup>	Periodicidade	Disponibilidade	
							Acesso	Período disponível
<i>Clinical Biochemistry</i> (1967)	Internacional A	MEDLINE	2,072	0,303	43	Mensal	Impresso e Online	1995 ao presente <sup>†</sup>
Jornal Brasileiro de Patologia e Medicina Laboratorial Laes Haes (1979)	Nacional A	SciELO, LILACS	0,0593	0,037	1	Trimestral	Impresso e Online	2001 ao presente <sup>†</sup>
NewsLab (1993)	Nacional C	-	-	-	-	Bimestral	Impresso	1979 ao presente
Revista Brasileira de Análises Clínicas (1969)	Nacional C	-	-	-	-	Bimestral	Impresso e Online	2003 ao presente <sup>†</sup>
Revista Brasileira de Ciências Farmacêuticas (1939)	Nacional B	LILACS	-	-	-	Trimestral	Impresso e Online	2004 ao presente <sup>†</sup>
Revista Brasileira de Ciências Farmacêuticas (1939)	Internacional C	SciELO	-	0,046	5	Trimestral	Impresso e Online	2004 ao presente <sup>†</sup>

<sup>a</sup>Lista de revistas utilizadas para a divulgação da produção intelectual dos programas de pós-graduação *stricto sensu*, classificados quanto ao âmbito de circulação (Local, Nacional, Internacional) e à qualidade (A, B, C), por área de avaliação (CAPES, 2008).

<sup>b</sup>Fator de Impacto da Thomson Scientific, que calcula para cada uma das revistas indexadas no *Journal Citation Reports* (JCR), a média de citações recebidas no ano de referência pelos artigos publicados nos dois anos anteriores (UNIVERSIDADE DE SÃO PAULO, 2008).

<sup>c</sup>O portal SJR (*SCImago Journal & Country Rank*) utiliza o índice bibliográfico Scopus, permitindo a publicação seletiva de relatórios dinâmicos com indicadores baseados nas citações entre os trabalhos científicos indexados no Scopus, de 1996 a 2007, possibilitando classificar o desempenho e impacto de revistas científicas e de países (BIBLIOTECA VIRTUAL EM SAÚDE, 2008).

<sup>d</sup>O índice H, proposto por Hirsch (2005), quantifica a produtividade individual e o impacto de cientistas baseando-se nos seus artigos (*papers*) mais citados, sendo ele o número de artigos com citações maiores ou iguais a esse número.

<sup>†</sup> Período referente ao acesso online.

(15%), bioquímica (15%), citologia (8%), parasitologia, hematologia e histoquímica com 6% e genética e urinálise com 4% (Fig. 2a). Na avaliação geral, incluindo todos os artigos de acurácia diagnóstica (n=101), a área predominante dos procedimentos diagnósticos avaliados foi a bioquímica, seguidos pela imunologia que, juntos, perfizeram 71% do total de artigos avaliados, seguidos por microbiologia (7%), genética (6%), citologia (4%), histoquímica (3%), urinálise (3%), hematologia (3%) e parasitologia (3%) (Fig 2b).

A presença de resumo e resumo estruturado nos artigos recuperados foi analisada. A presença de resumo variou, entre os periódicos, de 67 e 100% dos artigos. Quanto à estruturação do resumo, o periódico *Clinical Biochemistry* apresentou esta característica em 96% dos artigos, enquanto os periódicos LAES&HAES e NewsLab não apresentaram resumos estruturados. O JBPML apresentou 54% de artigos com resumo estruturados e a RBAC 23% (Tabela 5).

**Tabela 5.** Distribuição da presença de resumo e resumo estruturado entre os artigos recuperados

	Resumo		
	Ausente (%)	Presente	
		Estruturado (%)	Não estruturado (%)
Periódicos			
<i>Clinical Biochemistry</i>	0(0)	51(96)	2(4)
Jornal Brasileiro de Patologia e Medicina Laboratorial	0(0)	6(54)	5(46)
LAES&HAES	1(6)	0(0)	14(94)
NewsLab	3(33)	0(0)	6(67)
Revista Brasileira de Análises Clínicas	0(0)	3(23)	10(77)

Quanto à originalidade dos artigos de acurácia diagnóstica recuperados, apenas os periódicos LAES&HAES e NewsLab apresentaram artigos não-originais, representando 60 e 56% de artigos respectivamente.

### 5.2.1 *Clinical Biochemistry*

O periódico *Clinical Biochemistry* é a publicação oficial da Sociedade Canadense de Químicos Clínicos e teve seu início em 1967. Tem como público-alvo, químicos, bioquímicos, imunologistas e biólogos, publicando artigos relacionados à biologia molecular, bioquímica, química e imunologia para investigação clínica e diagnóstica, terapia e monitoramento de doenças humanas (ELSEVIER, 2008).

Embora o Qualis CAPES classifique o periódico como Internacional A, ele não se encontra incluído na relação dos periódicos para a área de Farmácia, sendo o único periódico analisado indexado pelo MEDLINE. Possui periodicidade mensal, acesso completo online de 1995 ao presente e seu índice H foi 43.

Quanto às áreas abrangidas pelos artigos de acurácia diagnóstica selecionados, o periódico *Clinical Biochemistry* apresentou um maior número de artigos avaliando procedimentos diagnósticos na área de bioquímica (75%), seguido pela área de imunologia (15%), genética (8%) e urinálise (2%) (Fig 2c).

### 5.2.2 **Jornal Brasileiro de Patologia e Medicina Laboratorial**

O JBPML, a continuação do Jornal Brasileiro de Patologia, é a publicação oficial da Sociedade Brasileira de Patologia Clínica, da Sociedade Brasileira de Patologia e da Sociedade Brasileira de Citopatologia. O periódico publica artigos originais, artigos de revisão, relatos de caso, comunicações rápidas e cartas aos editores. Publica artigos na área de medicina laboratorial, patologia clínica, patologia e citopatologia (SCIELO, 2008b).

É classificado pelo Qualis CAPES como Nacional A na área de Farmácia, sendo indexado pelo SciELO e LILACS. Sua periodicidade é trimestral, pode-se acessar online seus artigos completos a partir de 2001 e seu índice H foi 1.

O JBPML apresentou um maior número de artigos de acurácia diagnóstica na área de bioquímica (28%), seguido por histoquímica (27%), imunologia e citologia, ambas com 18% e microbiologia (9%) (Fig. 2d).

### 5.2.3 LAES&HAES

A revista LAES&HAES divulga trabalhos científicos relacionados com atividades em laboratórios de análises clínicas, patologia clínica, hematologia, anatomia patológica, microbiologia, bancos de sangue, hemocentros, biologia molecular, medicina nuclear e administração hospitalar. Os dados obtidos no sítio eletrônico da revista informam que a tiragem é de 16.117 exemplares por edição, com média de 6 leitores por exemplar da revista, totalizando, aproximadamente, 96.702 leitores por edição. O conteúdo editorial inclui artigos científicos avaliados por um corpo editorial, com ênfase na prática e pesquisa do setor diagnóstico nacional e internacional (LAES&HAES, 2008).

É classificada pelo Qualis CAPES como Nacional C na área de farmácia, não possuindo indexação. Sua periodicidade é bimestral e os artigos não estão disponíveis online no site da revista, sendo disponibilizados apenas na forma impressa.

A revista LAES&HAES apresentou um predomínio de artigos de acurácia diagnóstica avaliando procedimentos diagnósticos na área de imunologia (61%) seguidos por microbiologia, genética e hematologia, todos com 13% do total (Fig.2e).

### 5.2.4 NewsLab

A revista NewsLab publica bimestralmente editoriais, relatos de casos, revisões, casos educacionais, resumos de teses entre outros. Os artigos científicos abordam conteúdos da área de biologia, citologia e anatomo-patologia, infectologia, biologia molecular, genética, virologia, microbiologia e hematologia. No sítio eletrônico da revista consta que a mesma é enviada a 11.000 leitores com média de 6 leitores por exemplar, totalizando, aproximadamente, 66.000 leitores por edição (NEWSLAB, 2008).

É classificada pelo Qualis CAPES como Nacional C na área de Farmácia, não possuindo indexação. Os artigos estão disponíveis de forma impressa a partir de 1993 e online no site da revista de 2003 até o presente.

A Revista NewsLab apresentou maior homogeneidade na distribuição entre as áreas nos artigos de acurácia diagnóstica avaliados, sendo imunologia a área mais frequente (23%) seguida por bioquímica (22%), microbiologia (22%) e hematologia, parasitologia e citologia, todas com 11% do total (Fig. 2f).

### **5.2.5 Revista Brasileira de Análises Clínicas**

A RBAC é o periódico oficial da Sociedade Brasileira de Análises Clínicas e publica trimestralmente editoriais, artigos originais, relatos de casos e artigos de revisão. Os artigos científicos abordam conteúdos da área de bioquímica, biofísica, anatomia patológica e patologia clínica, medicina laboratorial e farmácia (SOCIEDADE BRASILEIRA DE ANÁLISES CLÍNICAS, 2008).

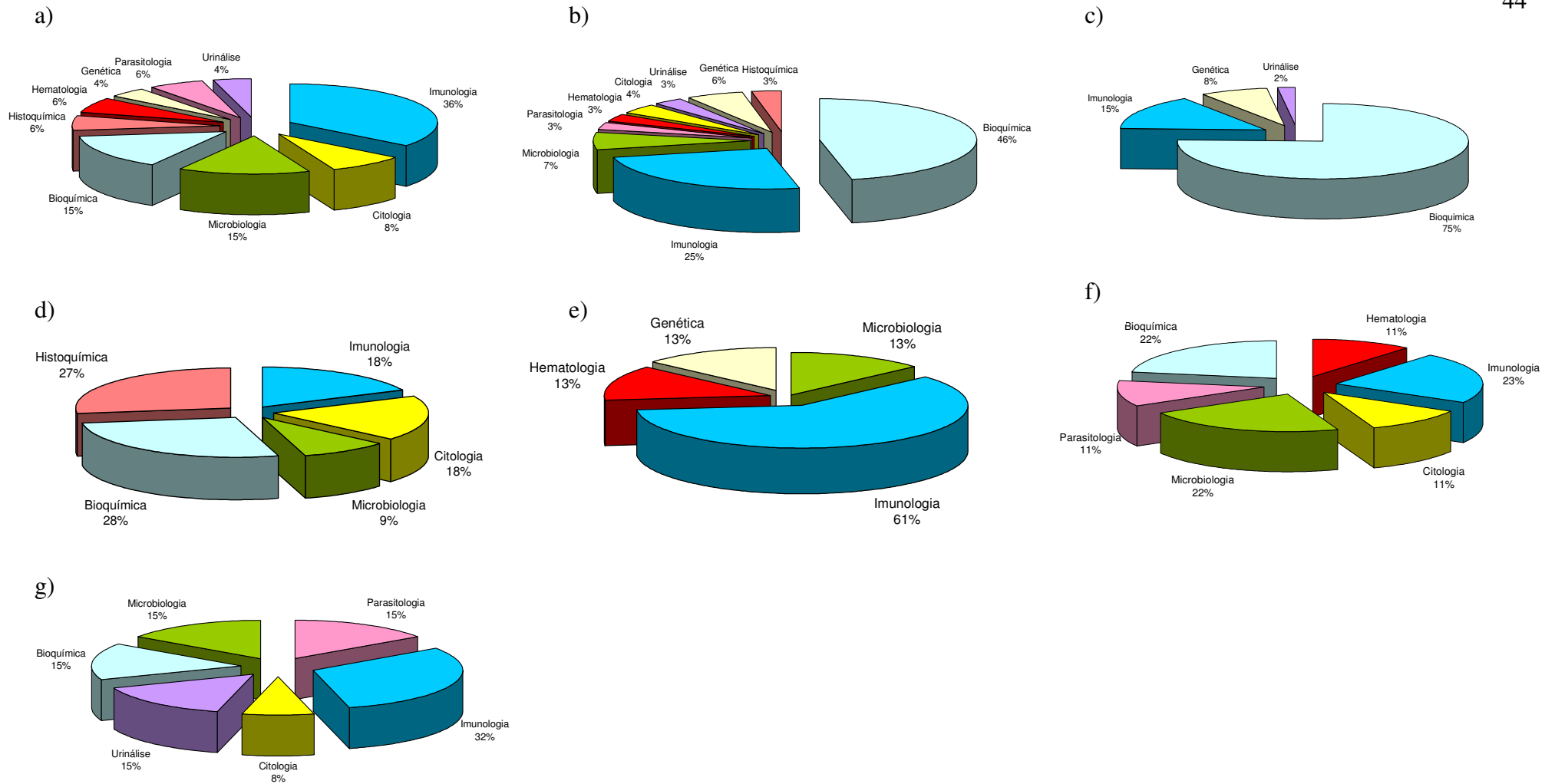
É disponibilizada na forma de artigos com texto completo na forma impressa a partir de 1969 e de livre acesso pelo sítio eletrônico da revista a partir de 2004. É classificada pelo Qualis CAPES como Nacional B na área de Farmácia, sendo indexada pelo LILACS.

A RBAC demonstrou homogeneidade entre as áreas das Análises Clínicas nos artigos de acurácia diagnóstica analisados. A imunologia foi a área mais freqüente (32%) seguida por parasitologia (15%), microbiologia (15%), bioquímica (15%), urinálise (15%) e citologia (8%) (Fig. 2g).

### **5.2.6 Revista Brasileira de Ciências Farmacêuticas**

A RBCF é editada pela Faculdade de Ciências Farmacêuticas da Universidade de São Paulo (USP), com periodicidade trimestral. Publica artigos originais relacionados com as áreas de conhecimento das Ciências Farmacêuticas e sua origem remonta aos Anais de Farmácia e Odontologia da USP, periódico iniciado em 1939. Em 1963, teve início a Revista da Faculdade de Farmácia e Bioquímica da USP, que, em 1970, recebeu o título de Revista de Farmácia e Bioquímica da USP. Em 1999, foi totalmente reformulada, recebendo a denominação de Revista Brasileira de Ciências Farmacêuticas (SCIELO, 2008a).

É classificada pelo Qualis CAPES como Internacional C na área de Farmácia, possuindo indexação pelo SciELO. Os artigos estão disponíveis na íntegra online a partir de 2004. Não foi avaliada pelo JCR em 2007, pelo SJR seu índice é de 0,046 e o índice H de 5.



**Figura 2.** Distribuição das áreas das análises clínicas dos artigos de acurácia diagnóstica recuperados nos (a) periódicos nacionais, (b) na totalidade dos artigos analisados, (c) *Clinical Biochemistry*, (d) *Jornal Brasileiro de Patologia e Medicina Laboratorial*, (e) *LAES&HAES*, (f) *NewsLab* e (g) *Revista Brasileira de Análises Clínicas*.

### 5.3 Resultados verificados com o roteiro STARD

A partir da aplicação das 25 questões do STARD foi obtido o número médio de itens atendidos nos cinco diferentes periódicos. Na análise dos artigos de acurácia diagnóstica de todos os periódicos analisados, a maioria dos artigos apresentava qualidade abaixo da objetivada pela iniciativa STARD. A concordância interobservador foi de 98% dos itens ( $k=0,96$ ). O tempo médio necessário para aplicação do checklist foi de aproximadamente 1 hora por artigo. Na avaliação da qualidade de redação dos artigos de acurácia diagnóstica, o número médio de itens atendidos variou de 10,6 ( $\pm 3,6$ ) para a revista LAES&HAES a 15,9 ( $\pm 2,2$ ) para o periódico *Clinical Biochemistry* e o percentual de itens atendidos variou de 20 a 82% nos diferentes artigos avaliados. Nenhum artigo alcançou mais de 90% dos itens do roteiro STARD (Tabela 6). Quanto à apresentação de um diagrama, recomendado pelo *checklist*, o periódico *Clinical Biochemistry* apresentou apenas um artigo contendo o diagrama. Os periódicos NewsLab e LAES&HAES apresentaram um artigo contendo o diagrama cada, mas tratava-se de um mesmo artigo não-original publicado em ambas revistas.

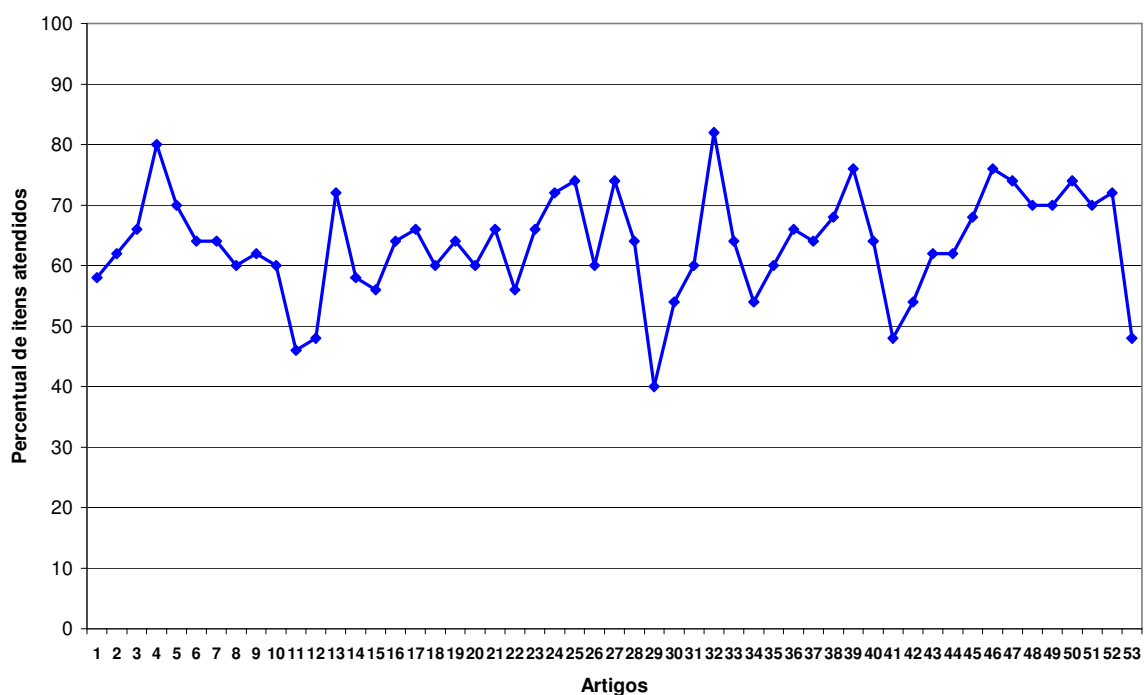
**Tabela 6.** Número médio de itens do STARD atendidos nos diferentes periódicos em números absolutos e percentuais e valores mínimos e máximos encontrados nos artigos avaliados.

Periódico	STARD (média $\pm$ desvio padrão)	STARD % (média $\pm$ desvio padrão)	Min-max (%)
<i>Clinical Biochemistry</i>	15,9 $\pm$ 2,2	63,6 $\pm$ 8,7	40-82
Jornal Brasileiro de Patologia e Medicina Laboratorial	13,4 $\pm$ 2,9	53,6 $\pm$ 11,5	32-68
LAES&HAES	10,6 $\pm$ 3,6	42,2 $\pm$ 14,5	20-78
NewsLab	11,1 $\pm$ 3,4	44,4 $\pm$ 13,8	20-60
Revista Brasileira de Análises Clínicas	12,0 $\pm$ 3,2	48,1 $\pm$ 12,6	24-62

#### 5.3.1 *Clinical Biochemistry*

Na análise dos artigos de acurácia diagnóstica do periódico *Clinical Biochemistry* (n=53) se verificou um percentual médio de 63,6% ( $\pm 8,7$ ) de itens atendidos para os artigos analisados (AGEWALL, 2003; ATALAY *et al.*, 2005; BIANCHI *et al.*, 2004; BOLANDPARVAZ *et al.*, 2004; BUHLING *et al.*, 2003; CAKIR *et al.*, 2005; CHANG *et al.*, 2006; CHEN *et al.*, 2005; COLLINSON *et al.*, 2006; DORIZZI *et al.*, 2006; EISSA *et al.*,

2004; EISSA *et al.*, 2005; EL-MASRY *et al.*, 2006; EL-SALAHY, 2002a; EL-SALAHY, 2002b; ERBAGCI *et al.*, 2002; ESKICORAPCI *et al.*, 2006; FABRIS *et al.*, 2006; GONZÁLEZ *et al.*, 2002; GONZÁLEZ *et al.*, 2005; GÜLTEPE *et al.*, 2003; HUTCHINSON *et al.*, 2005; KADOKAWA *et al.*, 2003; KASSIM *et al.*, 2004; KIR *et al.*, 2003; KOÇAK *et al.*, 2004; LYON *et al.*, 2003; MABROUK; ALI-LABIB, 2003; MAHMOUD; EL-GENDI; AHMED, 2005; MARSILLACH *et al.*, 2005; MOHAMMADZADEH *et al.*, 2003; MOLINARI *et al.*, 2004; PAL *et al.*, 2004; PENTTILÄ *et al.*, 2002; PETERSEN *et al.*, 2004; PORCEL *et al.*, 2005; QIU; FANG; HE, 2006; REY *et al.*, 2004; SAPIN *et al.*, 2003; SEKEROGLU *et al.*, 2002; SÈNE *et al.*, 2006; SUTTON-RILEY *et al.*, 2006; SUZUKI *et al.*, 2002; UZUN *et al.*, 2005; VANDERHEYDEN *et al.*, 2006; WEBER *et al.*, 2005; WOO *et al.*, 2005; YAGMUR *et al.*, 2006; YAMAN *et al.*, 2005; ZAMBON *et al.*, 2004; ZAMBON *et al.*, 2006; ZANINOTTO *et al.*, 2005; ZHANG *et al.*, 2004). O artigo com maior qualidade de redação apresentou 82% de itens atendidos e aquele com menor, 40% (Tabela 6). A Figura 3 demonstra o percentual de itens do STARD alcançado por cada artigo individualmente.



**Figura 3.** Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico *Clinical Biochemistry*.

O atendimento dos itens individualmente variou entre 3,8% (questão 20) e 100% (questão 25). Apenas a questão 25 obteve resposta “sim” em todos os 53 artigos analisados. Verificou-se que os itens atendidos em mais de 90% dos artigos analisados foram: item 2, 3, 4, 6, 7, 8a, 9a, 15, 16, 21 e 25. Estes estão relacionados à definição do objetivo do estudo



como avaliação da acurácia diagnóstica de um teste laboratorial ou comparação da acurácia entre testes (questão 2), à descrição da população estudada, critérios de exclusão e local de coleta de dados (questão 3), à forma de recrutamento dos participantes (questão 4), à descrição da coleta de dados (questão 6), à descrição do padrão de referência (questão 7), à descrição da execução completa do teste em avaliação (questão 8a), à definição das unidades e limites superiores de positividade do teste em avaliação (9a), à descrição das características clínicas e demográficas da população estudada (questão 15), à descrição do número de participantes nas diferentes etapas do estudo (questão 16), à avaliação da acurácia diagnóstica (questão 21) e à descrição da aplicabilidade clínica dos resultados do estudo (questão 25) (Tabela 7).

Os itens atendidos por menos de 10% dos artigos foram a 10a, 13b e 20. Estes estão relacionados à descrição do número, treinamento e especialidade das pessoas que executaram e interpretaram o teste em avaliação (questão 10a), à descrição de métodos para cálculo da reprodutibilidade do teste padrão de referência (questão 13b) e à descrição de eventos adversos pela execução do teste avaliado ou do teste padrão de referência (questão 20) (Tabela 7).

**Tabela 7.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico *Clinical Biochemistry*.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
<b>Título</b>							
1.	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	3(42,8)	3(33,3)	6(54,5)	7(50)	6(50)	25(47,2)
<b>Introdução</b>							
2.	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	6(85,7)	8(88,9)	11(100)	13(92,8)	10(83,3)	48(90,5)
<b>Métodos</b>							
3.	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	7(100)	8(88,9)	10(90,9)	13(92,8)	11(91,6)	49(92,4)
4.	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	7(100)	7(77,8)	11(100)	14(100)	12(100)	51(96,2)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 7.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico *Clinical Biochemistry*.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
5.	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	3(42,8)	3(33,3)	4(36,3)	10(71,4)	9(75)	29(54,7)
6.	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	7(100)	9(100)	11(100)	14(100)	11(91,6)	52(98,1)
7.	The reference standard and its rationale.	7(100)	7(77,7)	10(90,9)	13(92,8)	11(91,6)	48(90,5)
8.	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for						
	a) index test	6(85,7)	9(100)	11(100)	14(100)	12(100)	52(98,1)
	b) reference standard	5(71,4)	6(66,6)	7(63,6)	6(42,8)	8(66,6)	32(60,4)
9.	Definition of and rationale for the units, cut-offs and/or categories of the results of the						
	a) index test	7(100)	9(100)	11(100)	13(92,8)	12(100)	52(98,1)
	b) reference standard	6(85,7)	6(66,6)	8(72,7)	11(78,6)	10(83,3)	41(77,3)
10.	The number, training and expertise of the persons executing and reading the						
	a) index test	0(0)	1(11,1)	0(0)	1(7,1)	1(8,3)	3(5,7)
	b) reference standard	0(0)	1(11,1)	0(0)	2(14,2)	4(33,3)	7(13,2)
11.	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.						
	a) index test	1(14,3)	0(0)	2(18,2)	2(14,2)	1(8,3)	6(11,3)
	b) reference standard	1(14,3)	2(22,2)	3(27,3)	1(7,1)	5(41,6)	12(22,6)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 7.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico *Clinical Biochemistry*.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
12.	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	6(85,7)	6(66,6)	11(100)	12(85,7)	11(91,6)	46(86,8)
13.	Methods for calculating test reproducibility, if done.						
	a) for the index test	1(14,3)	1(11,1)	2(18,2)	2(14,2)	1(8,3)	7(13,2)
	b) for the reference standard	1(14,3)	0(0)	0(0)	1(7,1)	0(0)	2(3,8)
<b>Resultados</b>							
14.	When study was performed, including beginning and end dates of recruitment.	3(42,8)	1(11,1)	6(54,5)	4(28,5)	3(25)	17(32)
15.	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	7(100)	8(88,9)	11(100)	13(92,8)	11(91,6)	50(94,3)
16.	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	5(71,4)	9(100)	10(90,9)	14(100)	12(100)	50(94,3)
17.	Time-interval between the index tests and the reference standard, and any treatment administered in between.	4(57,1)	4(44,4)	5(45,4)	6(42,8)	6(50)	25(47,2)
18.	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	6(85,7)	6(66,6)	7(63,6)	11(78,6)	11(91,6)	41(77,3)
19.	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	4(57,1)	2(22,2)	5(45,4)	3(21,4)	5(41,6)	19(35,8)
20.	Any adverse events from performing the index tests or the reference standard.	0(0)	0(0)	1(9,1)	0(0)	1(8,3)	2(3,8)
21.	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	7(100)	9(100)	10(90,9)	14(100)	12(100)	52(98,1)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 7.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico *Clinical Biochemistry*.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
22.	How indeterminate results, missing data and outliers of the index tests were handled.	6(85,7)	9(100)	9(81,8)	10(71,4)	9(75)	43(81,1)
23.	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	4(57,1)	2(22,2)	4(36,3)	3(21,4)	8(66,6)	21(39,6)
24.	Estimates of test reproducibility, if done.						
	a) for the index test	4(57,1)	6(66,6)	7(63,6)	5(35,7)	3(25)	25(47,2)
	b) for the reference standard	2(28,6)	2(22,2)	2(18,2)	0(0)	1(8,3)	7(13,2)
<b>Discussão</b>							
25.	Discuss the clinical applicability of the study findings.	7(100)	9(100)	11(100)	14(100)	12(100)	53(100)

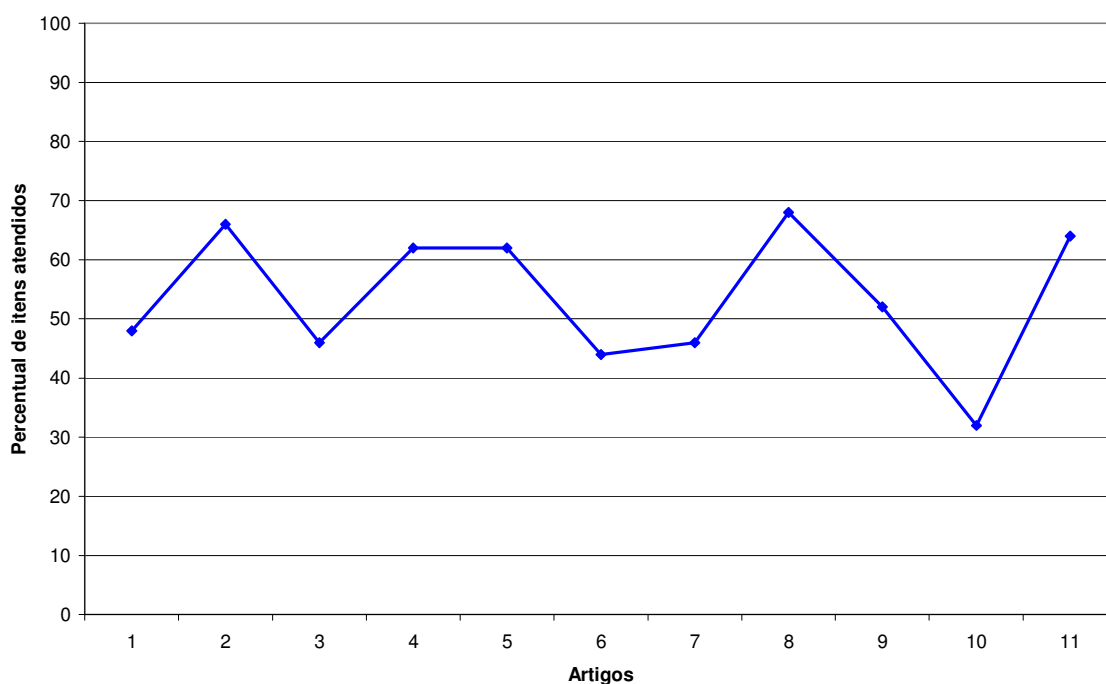
\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

### 5.3.2 Jornal Brasileiro de Patologia e Medicina Laboratorial

Na análise dos artigos de acurácia diagnóstica do JBPML (n=11) se verificou um percentual médio de 53,6% ( $\pm 11,5$ ) de itens atendidos para os artigos analisados (AMARAL *et al.*, 2003; BOTTINI *et al.*, 2005; BRUDER *et al.*, 2004; CASTRO *et al.*, 2004; GUGELMIN; TORRES, 2005; LIMA; GROTO, 2002; MACEDO *et al.*, 2002; MARINHO *et al.*, 2004; PEREIRA *et al.*, 2003; PINHO; MATTOS, 2002; SEGURA; MAGALHÃES, 2005). O artigo com melhor qualidade de redação apresentou 68% de respostas “sim” e aquele com menor, 32% (Tabela 6). A Figura 4 apresenta o número de itens do STARD alcançado por cada artigo individualmente.

O atendimento dos itens individualmente variou entre 0% e 100%. Os itens atendidos por mais de 90% dos artigos analisados foram os itens 2, 6, 9<sup>a</sup>, 21 e 25. Estes estão relacionados à definição do objetivo do estudo como avaliação da acurácia diagnóstica de um teste laboratorial ou comparação da acurácia diagnóstica entre testes (questão 2), à descrição da coleta de dados (questão 6), à descrição das unidades e limites superiores de positividade do teste em avaliação (questão 9<sup>a</sup>), à apresentação da avaliação da acurácia diagnóstica

(questão 21) e à discussão da aplicabilidade clínica dos resultados do estudo (questão 25) (Tabela 8).



**Figura 4.** Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico JBPML.

As questões respondidas afirmativamente por menos de 10% dos artigos foram a 10b, 11<sup>a</sup>, 11b, 13<sup>a</sup>, 13b, 20, 23, 24<sup>a</sup>, 24b. Estas questões estão relacionadas à descrição do número, treinamento e especialidade das pessoas que executaram e interpretaram o teste em avaliação (questão 10<sup>a</sup>), à avaliação cegada dos leitores dos testes para os resultados do teste em avaliação e padrão de referência (11<sup>a</sup> e 11b respectivamente), à descrição de métodos para cálculo da reprodutibilidade do teste em avaliação e do padrão de referência (questão 13<sup>a</sup> e 3b respectivamente), à descrição de eventos adversos pela execução do teste (questão 20), à descrição da variabilidade da acurácia diagnóstica entre subgrupos, revisores ou centros (questão 23) e à avaliação da reprodutibilidade do teste em avaliação e do padrão de referência (24<sup>a</sup> e 24b). Os itens não atendidos por nenhum dos artigos analisados foram a 10b, 11<sup>a</sup>, 11b e 23 (Tabela 8).

**Tabela 8.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*				
		2002	2003	2004	2005	Total
<b>Título</b>						
1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	3(100)	1(50)	0(0)	2(66,6)	6(54,5)
<b>Introdução</b>						
2.	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	3(100)	2(100)	3(100)	3(100)	11(100)
<b>Métodos</b>						
3.	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	1(33,3)	1(50)	1(100)	1(33,3)	6(54,5)
4.	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	3(100)	1(50)	3(100)	2(66,6)	9(81,8)
5.	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? <input type="checkbox"/> F not, specify how participants were further selected.	0(0)	2(100)	0(0)	1(33,3)	3(27,3)
6.	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	3(100)	2(100)	3(100)	3(100)	11(100)
7.	The reference standard and its rationale.	2(66,6)	2(100)	2(66,6)	2(66,6)	8(72,7)
8.	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for .					
	a) index test	2(66,6)	2(100)	3(100)	2(66,6)	9(81,8)
	b) reference standard	0(0)	1(50)	2(66,6)	1(33,3)	4(36,3)
9.	Definition of and rationale for the units, cut-offs and/or categories of the.					
	a) index test	3(100)	2(100)	3(100)	3(100)	11(100)
	b) reference standard	3(33,3)	2(100)	1(33,3)	2(66,6)	6(54,5)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 8.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*				
		2002	2003	2004	2005	Total
10.	The number, training and expertise of the persons executing and reading the.					
	a) index test	0(0)	2(100)	0(0)	0(0)	2(18,2)
	b) reference standard	0(0)	1(50)	0(0)	0(0)	1(9,1)
11.	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.					
	a) index test	0(0)	1(50)	0(0)	0(0)	1(9,1)
	b) reference standard	0(0)	1(50)	0(0)	0(0)	1(9,1)
12.	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	2(66,6)	0(0)	2(66,6)	2(66,6)	6(54,5)
13.	Methods for calculating test reproducibility, if done.					
	a) for the index test	0(0)	0(0)	0(0)	0(0)	0(0)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Resultados</b>						
14.	When study was performed, including beginning and end dates of recruitment.	1(33,3)	2(100)	1(33,3)	2(66,6)	6(54,5)
15.	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	0(0)	0(0)	3(100)	1(33,3)	4(36,3)
16.	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	3(100)	2(100)	2(66,6)	2(66,6)	9(81,8)
17.	Time-interval between the index tests and the reference standard, and any treatment administered in between.	1(33,3)	2(100)	2(66,6)	1(33,3)	6(54,5)
18.	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	2(66,6)	1(50)	2(66,6)	0(0)	5(45,4)
19.	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	3(100)	1(50)	1(33,3)	3(100)	8(72,7)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 8.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.

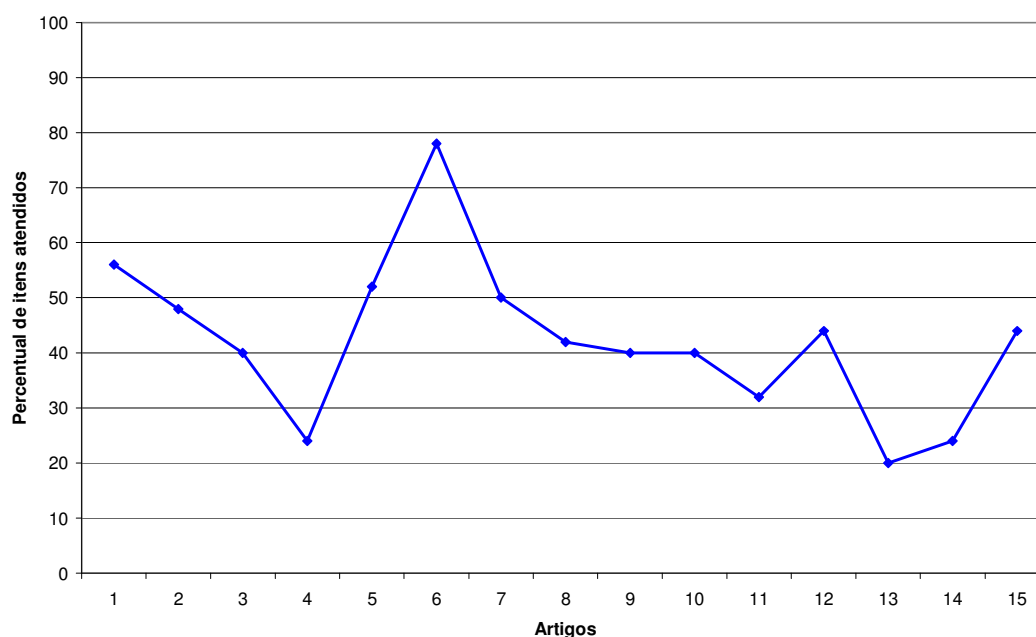
Item	Descrição dos itens (conforme original)	Número de artigos por ano*				
		2002	2003	2004	2005	Total
20.	Any adverse events from performing the index tests or the reference standard.	0(0)	0(0)	0(0)	0(0)	0(0)
21.	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	3(100)	2(100)	3(100)	3(100)	11(100)
22.	How indeterminate results, missing data and outliers of the index tests were handled.	3(100)	2(100)	2(66,6)	2(66,6)	9(81,8)
23.	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	1(33,3)	0(0)	0(0)	0(0)	1(9,1)
24.	Estimates of test reproducibility, if done.					
	a) for the index test	0(0)	0(0)	0(0)	0(0)	0(0)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Discussão</b>						
25.	Discuss the clinical applicability of the study findings.	3(100)	2(100)	3(100)	3(100)	11(100)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

### 5.3.3 LAES&HAES

Na análise dos artigos de acurácia diagnóstica da revista LAES&HAES (n=15) verificou-se um percentual médio de 42,2% ( $\pm 14,5$ ) de itens do STARD atendidos para os artigos analisados (EGGERDING; TRISCOTT, 2003; FISCHER *et al.*, 2006; GUILLOUX; HAMBERGER, 2005; KRAJDEN *et al.*, 2004; LY *et al.*, 2006; OLIVEIRA *et al.*, 2006; OLIVEIRA; CASTRO; LANGHI, 2004; SÁEZ-ALQUÉZAR *et al.*, 2005b; SÁEZ-ALQUÉZAR *et al.*, 2006; SANTOS *et al.*, 2005; SANTOS *et al.*, 2002; SAUERBREI; WUTZLER, 2006; SHIRIT *et al.*, 2004; TOZZOLI *et al.*, 2006; YOUNG *et al.*, 2002). O artigo com melhor qualidade de redação apresentou 78% de itens atendidos e aquele com menor qualidade, 20% (Tabela 6). A Figura 5 ilustra o número de itens do STARD alcançado por cada artigo individualmente.





**Figura 5.** Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico LAES&HAES.

O atendimento dos itens individualmente variou entre 0% e 86,7%. Não foram verificados itens atendidos por mais de 90% dos artigos analisados, portanto as questões respondidas afirmativamente em mais de 80% dos artigos foram as questões 2, 6, 21 e 25. Estes itens estão relacionados à definição do objetivo do estudo como avaliação da acurácia diagnóstica de um teste laboratorial ou comparação da acurácia entre testes (questão 2), à descrição da coleta de dados (questão 6), à apresentação da avaliação da acurácia diagnóstica (questão 21) e à discussão da aplicabilidade clínica dos resultados do estudo (questão 25). Nenhum dos itens foi atendido pela totalidade dos artigos analisados (Tabela 9).

Os itens atendidos por menos de 10% dos artigos foram a 10a, 10b, 13b, 15, 20, 24b. Estes estão relacionados à descrição do número, treinamento e especialidade das pessoas que executaram e interpretaram o teste em avaliação e o padrão de referência (questão 10a e 10b), à descrição de métodos para cálculo da reprodutibilidade do teste padrão de referência (questão 13b), à descrição das características clínicas e demográficas da população estudada (questão 15), à descrição de eventos adversos pela execução do teste (questão 20) e à avaliação da reprodutibilidade do teste padrão de referência (24b). As questões que não obtiveram respostas afirmativas em nenhum artigo analisado foram a 10a, 20 e 24b (Tabela 9).

**Tabela 9.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
<b>Título</b>							
1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	2(100)	1(100)	0(0)	2(50)	3(50)	8(53,3)
<b>Introdução</b>							
2.	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	1(50)	1(100)	1(50)	4(100)	6(100)	13(86,6)
<b>Métodos</b>							
3.	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	0(0)	0(0)	1(50)	2(50)	2(33,3)	5(33,3)
4.	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	1(50)	1(100)	1(50)	2(50)	3(50)	8(53,3)
5.	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? <input type="checkbox"/> F not, specify how participants were further selected.	1(50)	0(0)	0(0)	2(50)	0(0)	3(20)
6.	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	2(100)	1(100)	2(100)	4(100)	4(66,6)	13(86,6)
7.	The reference standard and its rationale.	1(50)	1(100)	1(50)	2(50)	3(50)	8(53,3)
8.	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for.						
	a) index test	2(100)	0(0)	2(100)	3(75)	3(50)	10(66,6)
	b) reference standard	2(100)	0(0)	0(0)	1(25)	2(33,3)	5(33,3)
9.	Definition of and rationale for the units, cut-offs and/or categories of the results of the.						
	a) index test	1(50)	1(100)	2(100)	3(75)	2(33,3)	9(60)
	b) reference standard	1(50)	0(0)	1(50)	2(50)	1(16,6)	5(33,3)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 9.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
10.	The number, training and expertise of the persons executing and reading the.						
	a) index test	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	b) reference standard	0(0)	0(0)	0(0)	1(25)	0(0)	1(6,6)
11.	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.						
	a) index test	0(0)	0(0)	1(50)	0(0)	1(16,6)	2(13,3)
	b) reference standard	0(0)	0(0)	0(0)	1(25)	1(16,6)	2(13,3)
12.	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	0(0)	0(0)	0(0)	2(50)	1(16,6)	3(20)
13.	Methods for calculating test reproducibility, if done.						
	a) for the index test	0(0)	0(0)	1(50)	1(25)	0(0)	2(13,3)
	b) for the reference standard	0(0)	0(0)	0(0)	1(25)	0(0)	1(6,6)
<b>Resultados</b>							
14.	When study was performed, including beginning and end dates of recruitment.	2(100)	0(0)	1(50)	2(50)	1(16,6)	6(40)
15.	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	0(0)	0(0)	0(0)	1(25)	0(0)	1(6,6)
16.	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	2(100)	0(0)	2(100)	4(100)	5(83,3)	13(86,6)
17.	Time-interval between the index tests and the reference standard, and any treatment administered in between.	1(50)	1(100)	1(50)	3(75)	1(16,6)	7(46,6)
18.	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	1(50)	0(0)	0(0)	1(25)	0(0)	2(13,3)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

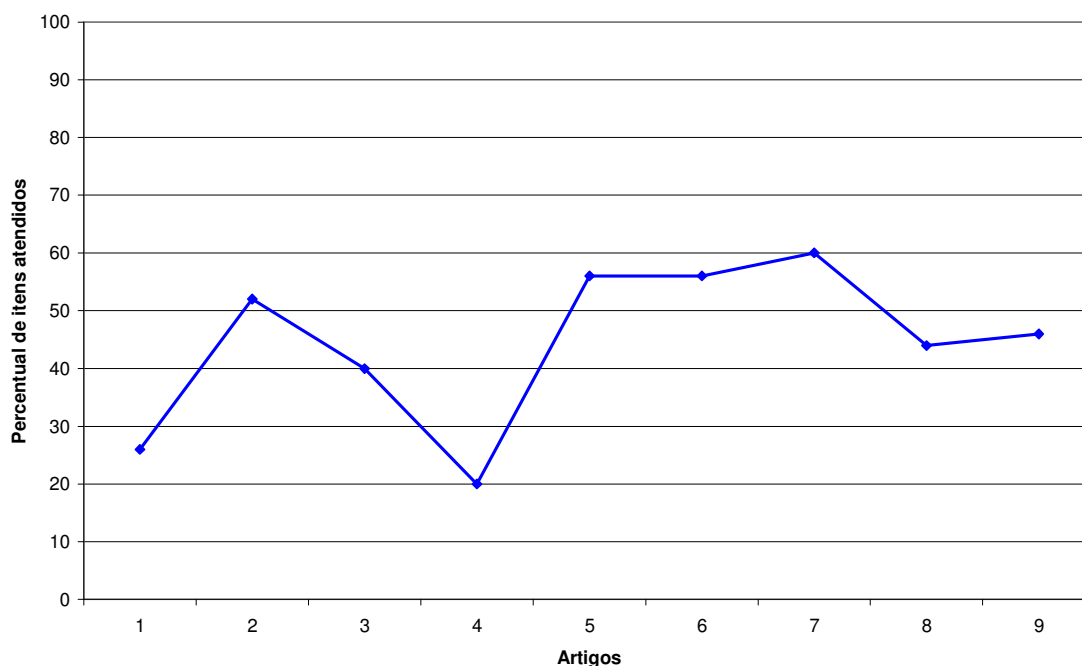
**Tabela 9.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
19.	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	2(100)	0(0)	1(50)	4(100)	4(66,6)	11(73,3)
20.	Any adverse events from performing the index tests or the reference standard.	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
21.	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	2(100)	1(100)	1(50)	3(75)	5(83,3)	12(80)
22.	How indeterminate results, missing data and outliers of the index tests were handled.	2(100)	1(100)	1(50)	2(50)	4(66,6)	10(66,6)
23.	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	1(50)	0(0)	0(0)	1(25)	0(0)	2(13,3)
24.	Estimates of test reproducibility, if done.						
	a) for the index test	0(0)	1(100)	1(50)	2(50)	0(0)	4(26,6)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Discussão</b>							
25.	Discuss the clinical applicability of the study findings.	2(100)	1(100)	2(100)	4(100)	4(66,6)	13(86,6)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

### 5.3.4 NewsLab

Na análise dos artigos de acurácia diagnóstica do periódico NewsLab (n=09) se verificou um percentual médio de 44,4% ( $\pm 13,8$ ) de respostas “sim” para os artigos analisados (CONCEIÇÃO *et al.*, 2005; EGGERDING; TRISCOTT, 2003; HOSHINO-SHIMIZÙ *et al.*, 2002; KUOKKANEN *et al.*, 2006; LOBO; FEIJÓ, SIMÕES-BARBOSA, 2002; RAO, 2002; ROSSATO; SILVA, 2004; SADOYAMA; GONTIJO FILHO, 2002; SÁEZ-ALQUÉZAR *et al.*, 2005a). O artigo com melhor qualidade de redação apresentou 60% de respostas “sim” e aquele com menor, 20% (Tabela 6). A Figura 6 apresenta o número de itens do STARD alcançado por cada artigo individualmente.



**Figura 6.** Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico NewsLab.

O atendimento dos itens individualmente variou entre 0% e 100%. Não se verificou itens atendidos em mais de 90% dos artigos analisados, portanto itens atendidos em mais de 80% dos artigos foram: o item 6, 9a, 16, 17, 21 e 22. Estes estão relacionados à descrição da coleta de dados (questão 6), à definição das unidades e limites superiores de positividade do teste em avaliação (questão 9a), à descrição do número de participantes nas diferentes etapas do estudo (questão 16), à descrição do intervalo de tempo entre o teste avaliado e o padrão de referência (questão 17), à apresentação da avaliação da acurácia diagnóstica (questão 21) e à descrição de como foi o manejo dos resultados indeterminados, perdidos ou *outliers* (questão 22). Apenas as questões 16 e 21 obtiveram respostas “sim” em todos os artigos avaliados na revista NewsLab (Tabela 10).

Os itens não atendidos por nenhum estudo foram a 5, 10a, 10b, 13b, 20, 24b. Estas questões estão relacionadas à descrição da amostragem dos participantes (questão 5), à descrição do número, treinamento e especialidade das pessoas que executaram e interpretaram o teste em avaliação e o padrão de referência (questão 10a e 10b), à descrição de métodos para cálculo da reprodutibilidade do teste padrão de referência (questão 13b), à descrição de eventos adversos pela execução do teste (questão 20) e à avaliação da reprodutibilidade do teste padrão de referência (questão 24b) (Tabela 10).

**Tabela 10.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
<b>Título</b>							
1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1(25)	1(100)	1(100)	0(0)	1(100)	4(44,4)
<b>Introdução</b>							
2.	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	2(50)	1(100)	1(100)	2(100)	1(100)	7(77,7)
<b>Métodos</b>							
3.	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	0(0)	0(0)	1(100)	0(0)	0(0)	1(11,1)
4.	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	2(50)	1(100)	1(100)	2(100)	0(0)	6(66,6)
5.	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? <input type="checkbox"/> F not, specify how participants were further selected.	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
6.	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	3(75)	1(100)	1(100)	2(100)	1(100)	8(88,8)
7.	The reference standard and its rationale.	1(25)	1(100)	1(100)	2(100)	1(100)	6(66,6)
8.	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for.						
	a) index test	1(25)	1(100)	0(0)	1(50)	1(100)	4(44,4)
	b) reference standard	2(50)	0(0)	0(0)	1(50)	0(0)	3(33,3)
9.	Definition of and rationale for the units, cut-offs and/or categories of the results of the.						
	a) index test	3(75)	1(100)	1(100)	2(100)	1(100)	8(88,8)
	b) reference standard	1(25)	1(100)	1(100)	2(100)	1(100)	6(66,6)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 10.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
10	The number, training and expertise of the persons executing and reading the.						
	a) index test	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	b) reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.						
	a) index test	1(25)	0(0)	0(0)	1(50)	0(0)	2(22,2)
	b) reference standard	1(25)	0(0)	0(0)	1(50)	0(0)	2(22,2)
12.	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	0(0)	1(100)	1(100)	1(50)	0(0)	3(33,3)
13.	Methods for calculating test reproducibility, if done.						
	a) for the index test	0(0)	0(0)	0(0)	1(50)	0(0)	1(11,1)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Resultados</b>							
14.	When study was performed, including beginning and end dates of recruitment.	2(50)	0(0)	1(100)	1(50)	0(0)	4(44,4)
15.	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	0(0)	0(0)	1(100)	0(0)	0(0)	1(11,1)
16.	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	4(100)	1(100)	1(100)	2(100)	1(100)	9(100)
17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	4(100)	1(100)	0(0)	2(100)	1(100)	8(88,8)
18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	0(0)	0(0)	0(0)	0(0)	1(100)	1(11,1)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 10.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.

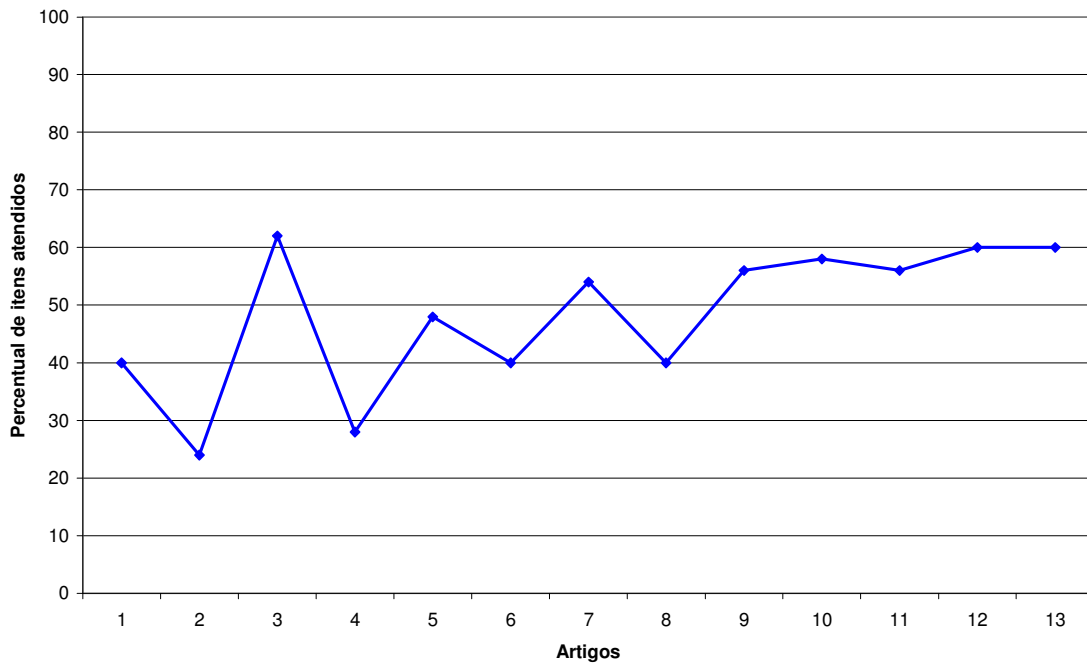
Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	2(50)	1(100)	0(0)	0(0)	1(100)	4(44,4)
20	Any adverse events from performing the index tests or the reference standard.	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	4(100)	1(100)	1(100)	2(100)	1(100)	9(100)
22	How indeterminate results, missing data and outliers of the index tests were handled.	4(100)	1(100)	1(100)	2(100)	0(0)	8(88,8)
23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	0(0)	0(0)	0(0)	1(50)	0(0)	1(11,1)
24	Estimates of test reproducibility, if done.						
	a) for the index test	0(0)	1(100)	0(0)	1(50)	0(0)	2(22,2)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Discussão</b>							
25	Discuss the clinical applicability of the study findings.	1(25)	1(100)	1(100)	2(100)	1(100)	6(66,6)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

### 5.3.5 Revista Brasileira de Análises Clínicas

Na análise dos artigos de acurácia diagnóstica da RBAC (n=13) se verificou um percentual médio de 48,1% ( $\pm 12,6$ ) de itens atendidos para os artigos analisados (CARVALHO *et al.*, 2002; CARVALHO *et al.*, 2005; DE BEM *et al.*, 2005; LIMA *et al.*, 2006; MARTINS *et al.*, 2006; MAUÉS *et al.*, 2003; MOTTA *et al.*, 2003; RAMOS; PIZZOLITTO; PIZZOLITTO, 2006; SANTOS; SOARES, 2005; SILVA *et al.*, 2004; STINGHEN; NASCIMENTO; LEONART, 2004; UTIYAMA *et al.*, 2002; YOSHIDA *et al.*, 2006). O artigo com melhor qualidade de redação apresentou 62% de respostas “sim” e aquele com menor, 24% (Tabela 6). A Figura 7 mostra o número de itens do STARD alcançado por cada artigo individualmente.





**Figura 7.** Número de itens do STARD atendidos por cada artigo individualmente em ordem cronológica de publicação na RBAC.

O atendimento dos itens individualmente variou entre 0% e 100%. Verificou-se que os itens atendidos em mais de 90% dos artigos analisados foram os itens 8a, 21 e 25. Estes estão relacionados à descrição da execução completa do teste em avaliação (questão 8a), à apresentação da avaliação da acurácia diagnóstica (questão 21) e à discussão da aplicabilidade clínica dos resultados do estudo (questão 25). As questões 21 e 25 coincidiam com as questões que obtiveram respostas “sim” em todos os artigos avaliados (Tabela 11).

Os itens atendidos em menos de 10% dos artigos foram 5, 10a, 10b, 11a, 11b, 13a, 13b, 20, 24a, 24b. Estas questões estão relacionadas à descrição da amostragem dos participantes (questão 5), à descrição do número, treinamento e especialidade das pessoas que executaram e interpretaram o teste em avaliação e o padrão de referência (questão 10a e 10b), à avaliação cegada dos leitores dos testes para os resultados do teste em avaliação e padrão de referência (11a e 11b respectivamente), à descrição de métodos para cálculo da reprodutibilidade do teste em avaliação e do padrão de referência (questão 13a e 3b respectivamente), à descrição de eventos adversos pela execução do teste (questão 20) e à avaliação da reprodutibilidade do teste em avaliação e do padrão de referência (24a e 24b). Os itens que não obtiveram respostas afirmativas em nenhum estudo foram 10a, 10b, 11a, 11b, 13a, 13b, 24a e 24b (Tabela 11).

**Tabela 11.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
<b>Título</b>							
1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1(50)	2(100)	0(0)	3(100)	4(100)	10(77)
<b>Introdução</b>							
2.	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	1(50)	1(50)	1(50)	3(100)	4(100)	10(77)
<b>Métodos</b>							
3.	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	0(0)	1(50)	2(100)	2(66,6)	2(50)	7(54)
4.	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	1(50)	1(50)	2(100)	2(66,6)	3(75)	9(69)
5.	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	0(0)	0(0)	0(0)	1(33,3)	0(0)	1(7,7)
6.	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	1(50)	1(50)	1(50)	3(100)	4(100)	10(77)
7.	The reference standard and its rationale.	0(0)	1(50)	1(50)	3(100)	4(100)	9(69)
8.	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for.						
	a) index test	2(100)	2(100)	2(100)	2(66,6)	4(100)	12(92)
	b) reference standard	0(0)	1(50)	1(50)	1(33,3)	3(75)	6(46)
9.	Definition of and rationale for the units, cut-offs and/or categories of the results of the.						
	a) index test	2(100)	1(50)	2(100)	1(33,3)	4(100)	10(77)
	b) reference standard	0(0)	1(50)	1(50)	1(33,3)	4(100)	7(54)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 11.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
10	The number, training and expertise of the persons executing and reading the.						
	a) index test	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	b) reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.						
	a) index test	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	b) reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
12.	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	1(50)	1(50)	1(50)	1(33,3)	2(50)	6(46)
13.	Methods for calculating test reproducibility, if done.						
	a) for the index test	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Resultados</b>							
14.	When study was performed, including beginning and end dates of recruitment.	0(0)	1(50)	1(50)	2(66,6)	2(50)	6(46)
15.	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	1(50)	0(0)	2(100)	1(33,3)	1(25)	5(38)
16.	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	1(50)	1(50)	1(50)	1(33,3)	4(100)	8(61)
17.	Time-interval between the index tests and the reference standard, and any treatment administered in between.	0(0)	0(0)	1(50)	2(66,6)	4(100)	7(54)
18.	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	1(50)	1(50)	0(0)	1(33,3)	0(0)	3(23)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 11.** Atendimento individual dos itens do roteiro STARD para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
19.	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	1(50)	1(50)	1(50)	2(66,6)	4(100)	9(69)
20.	Any adverse events from performing the index tests or the reference standard.	0(0)	1(50)	0(0)	0(0)	0(0)	1(7,7)
21.	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	2(100)	2(100)	2(100)	3(100)	4(100)	13(100)
22.	How indeterminate results, missing data and outliers of the index tests were handled.	1(50)	2(100)	1(50)	2(66,6)	4(100)	10(77)
23.	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	0(0)	1(50)	0(0)	0(0)	1(25)	2(15)
24.	Estimates of test reproducibility, if done.						
	a) for the index test	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
	b) for the reference standard	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
<b>Discussão</b>							
25.	Discuss the clinical applicability of the study findings.	2(100)	2(100)	2(100)	3(100)	4(100)	13(100)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

#### 5.4 Resultados verificados com o instrumento QUADAS

A análise e avaliação da qualidade de redação dos artigos foram realizadas também através do instrumento QUADAS. A partir da aplicação das 14 questões do QUADAS foi obtido o número médio de questões com resposta “sim” nos diferentes periódicos analisados. A concordância interobservador foi de 96% dos itens ( $k=0,92$ ). O tempo médio necessário para aplicação do checklist foi de aproximadamente 45 minutos por artigo. A maioria dos artigos apresentava qualidade mediana de redação considerando o alcance de suas respostas “sim”. Na avaliação da qualidade de redação dos artigos de acurácia diagnóstica, o número médio de itens atendidos variou de 6,3 ( $\pm 1,8$ ) para a revista LAES&HAES a 9,0 ( $\pm 1,8$ ) para o periódico *Clinical Biochemistry* e o percentual de itens atendidos variou de 14,3 a 85,7% nos diferentes artigos avaliados. Nenhum artigo alcançou mais de 90% dos itens do

QUADAS. O número médio de itens atendidos e seus desvios-padrão são apresentados na Tabela 12. O número médio de itens classificados como “não claro” e seus desvios-padrão são apresentados na Tabela 13.

Entre os periódicos nacionais, apenas a revista NewsLab apresentou artigos com mais de 80% dos itens do QUADAS atendidos enquanto a revista LAES&HAES não apresentou nenhum artigo atendendo mais de 70% dos itens. A revista *Clinical Biochemistry* apresentou 3,8% de artigos que atendendo mais de 80% dos itens do QUADAS.

**Tabela 12.** Número médio de questões com resposta “sim” na aplicação do QUADAS nos diferentes periódicos analisados.

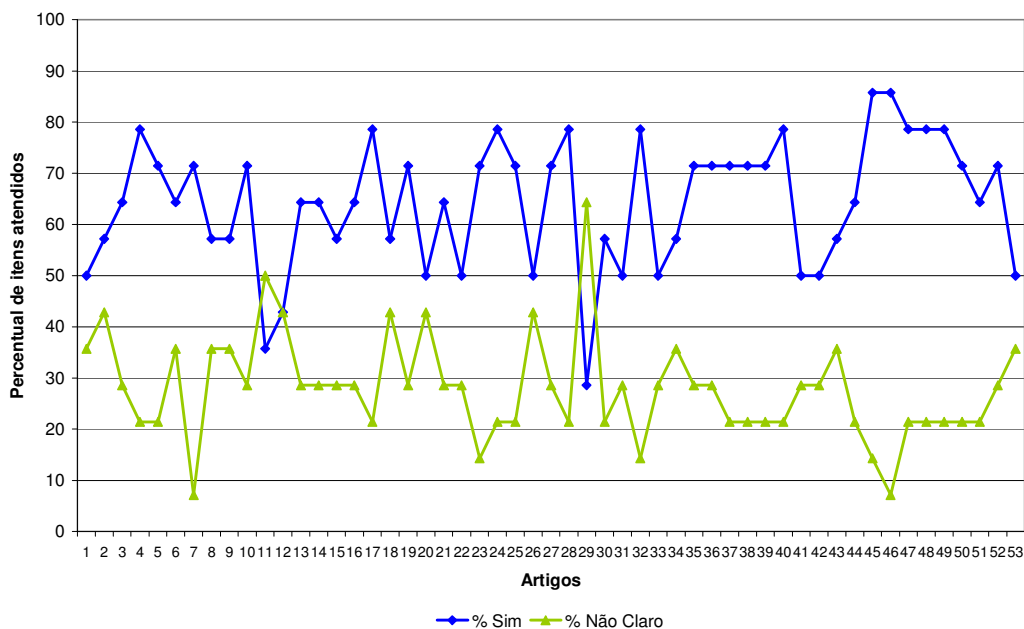
Periódico	QUADAS SIM (média ± desvio padrão)	QUADAS % SIM (média ± desvio- padrão)	Min-max (%)
<i>Clinical Biochemistry</i>	9,0 ± 1,8	64,3±12,6	28,6-85,7
Jornal Brasileiro de Patologia e Medicina Laboratorial	8,1 ± 1,4	57,8±10,3	42,8-71,4
LAES&HAES	6,3 ± 1,8	45,2±12,8	21,4-64,3
NewsLab	7,3 ± 2,0	52,4±14,3	35,7-85,7
Revista Brasileira de Análises Clínicas	7,1 ± 2,7	50,5±19,2	14,3-78,6

**Tabela 13.** Número médio de questões com resposta “não claro” na aplicação do QUADAS nos diferentes periódicos analisados.

Periódico	QUADAS NÃO CLARO (média ± desvio padrão)	QUADAS % NÃO CLARO (média ± desvio- padrão)	Min-max (%)
<i>Clinical Biochemistry</i>	3,9 ± 1,4	28,0 ± 10,3	7,1 - 64,3
Jornal Brasileiro de Patologia e Medicina Laboratorial	3,4 ± 1,8	24,0 ± 12,9	7,1 - 42,8
LAES&HAES	5,3 ± 1,3	38,1 ± 9,2	21,4 - 50
NewsLab	4,9 ± 1,4	34,9 ± 10,4	14,3 - 50
Revista Brasileira de Análises Clínicas	5,1 ± 1,9	36,8 ±14,0	14,3 - 64,3

### 5.4.1 Clinical Biochemistry

Na análise dos artigos de acurácia diagnóstica dos periódicos *Clinical Biochemistry* (n=53) se verificou um percentual médio de 64,3% ( $\pm 12,6$ ) de respostas “sim” para os artigos analisados. O artigo com melhor qualidade de redação apresentou 85,7% de respostas “sim” e aquele com menor, 28,6% (Tabela 12). Para as respostas “não claro” o percentual médio foi de 28% ( $\pm 10,3$ ) (Tabela 13) e para as respostas “não” foi de 7,3 ( $\pm 7,2$ ). A Figura 8 ilustra o percentual de respostas “sim” e “não claro” do QUADAS alcançado por cada artigo individualmente.



**Figura 8.** Percentual de itens com resposta “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico *Clinical Biochemistry*.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 4, 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Elas estão associadas, respectivamente, ao período adequado de realização entre o teste avaliado e o padrão de referência para que não houvesse alteração na condição de saúde do paciente (questão 4), à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11, respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 14).

**Tabela 14.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico *Clinical Biochemistry*.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?	7(100)	8(88,9)	11(100)	13(92,8)	11(91,6)	50(94,3)
2.	Were selection criteria clearly described?	6(85,7)	5(55,5)	7(63,6)	11(78,5)	10(83,3)	39(73,6)
3.	Is the reference standard likely to correctly classify the target condition?	7(100)	5(55,5)	8(72,7)	12(85,7)	11(91,6)	43(81,1)
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	3(42,8)	4(44,4)	5(45,4)	6(42,8)	7(58,3)	25(47,2)
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	6(85,7)	6(66,6)	11(100)	10(71,4)	12(100)	45(84,9)
6.	Did patients receive the same reference standard regardless of the index test result?	6(85,7)	7(77,7)	11(100)	10(71,4)	12(100)	46(86,8)
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	6(85,7)	6(66,6)	9(81,8)	13(92,8)	11(91,6)	45(84,9)
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	6(85,7)	8(88,9)	10(90,9)	13(92,8)	11(91,6)	48(90,5)
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	3(42,8)	3(33,3)	3(27,3)	5(35,7)	6(50)	20(37,7)
10	Were the index test results interpreted without knowledge of the results of the reference standard?	1(14,3)	0(0)	2(18,2)	2(14,3)	1(8,3)	6(11,3)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 14.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico *Clinical Biochemistry*.

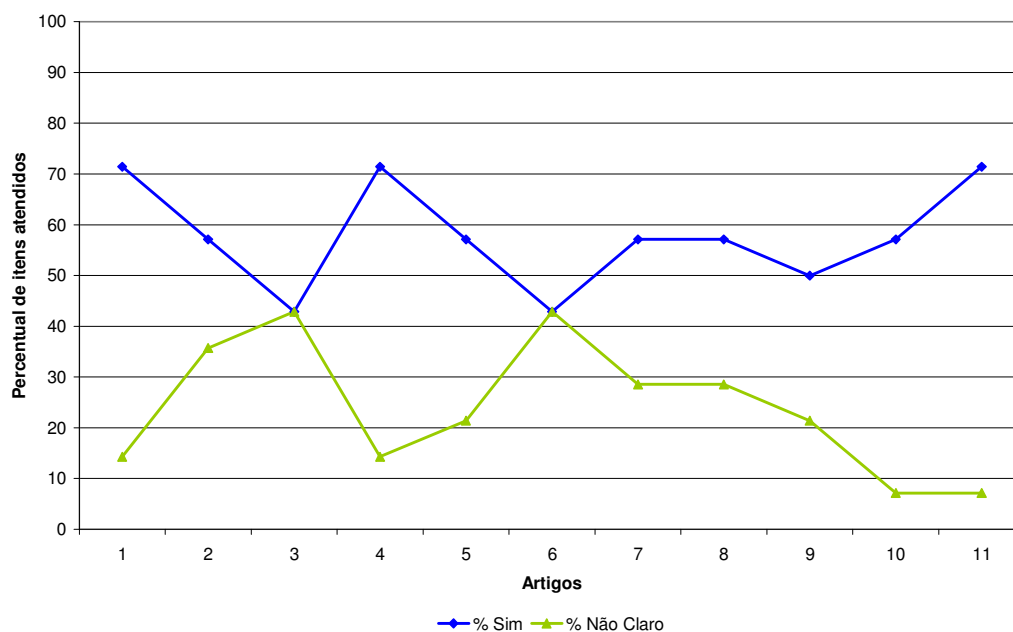
Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	1(14,3)	2(22,2)	3(27,3)	2(14,3)	5(41,6)	13(24,5)
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	0(0)	0(0)	0(0)	1(7,1)	0(0)	1(1,9)
13.	Were uninterpretable/intermediate test results reported?	6(85,7)	9(100)	10(90,9)	13(92,8)	11(91,6)	49(92,4)
14.	Were withdrawals from the study explained?	6(85,7)	9(100)	10(90,9)	13(92,8)	9(75)	47(88,7)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

#### 5.4.2 Jornal Brasileiro de Patologia e Medicina Laboratorial

Na análise dos artigos de acurácia diagnóstica do JBPML (n=11) verificou-se um percentual médio de 57,8% ( $\pm 10,3$ ) de respostas “sim” para os artigos analisados. O artigo com melhor qualidade de redação apresentou 71,4% de respostas “sim” e aquele com menor, 42,8% (Tabela 12). Para as respostas “não claro” o percentual médio foi de 24% ( $\pm 12,9$ ) e para as respostas “não” foi de 18,2 ( $\pm 8,0$ ) (Tabela 13). A figura 9 apresenta o número respostas “sim” e “não claro” do QUADAS alcançado por cada artigo individualmente.





**Figura 9.** Número de itens com resposta “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico JBPML.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 4, 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Elas estão associadas, respectivamente, ao período adequado de realização entre o teste avaliado e o padrão de referência para que não houvesse alteração na condição de saúde do paciente (questão 4), à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 15).

**Tabela 15.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005		
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	3(100)	2(100)	3(100)	2(66,6)	10(90,9)	
2	Were selection criteria clearly described?	0(0)	1(50)	1(33,3)	1(33,3)	3(27,3)	
3	Is the reference standard likely to correctly classify the target condition?	3(100)	2(100)	1(33,3)	2(66,6)	8(72,7)	

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

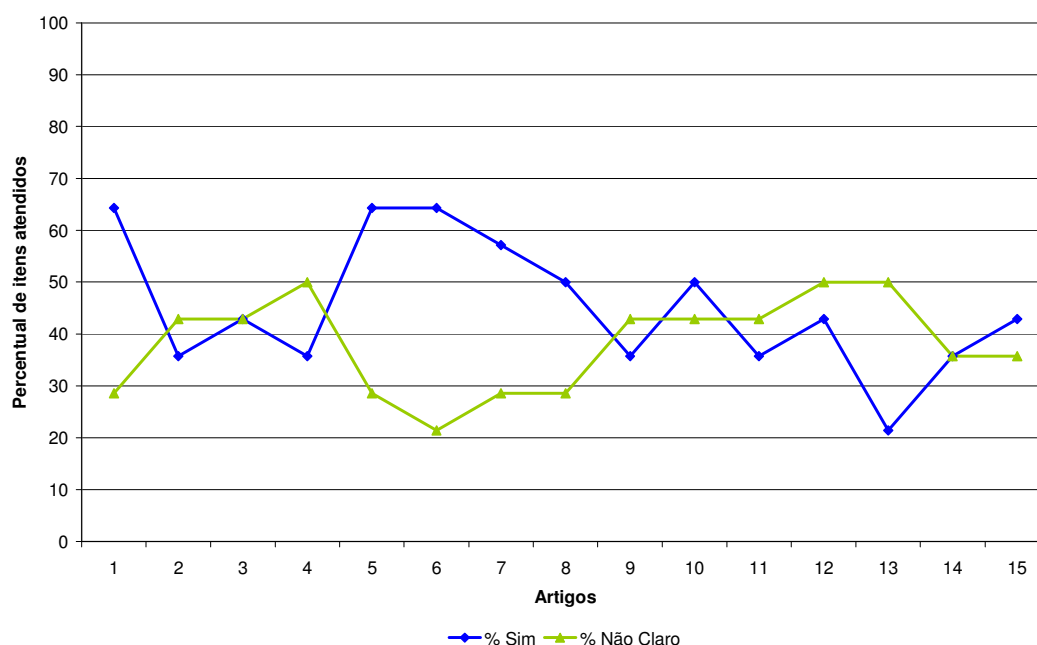
**Tabela 15.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico JBPML.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*				
		2002	2003	2004	2005	Total
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	1(33,3)	2(100)	2(66,6)	1(33,3)	6(54,5)
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	3(100)	2(100)	2(66,6)	3(100)	10(90,9)
6.	Did patients receive the same reference standard regardless of the index test result?	3(100)	1(50)	2(66,6)	3(100)	9(81,8)
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	3(100)	0(0)	2(66,6)	1(33,3)	6(54,5)
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	1(33,3)	2(100)	3(100)	2(66,6)	8(72,7)
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	0(0)	0(0)	1(33,3)	1(33,3)	2(18,2)
10.	Were the index test results interpreted without knowledge of the results of the reference standard?	0(0)	1(50)	0(0)	1(33,3)	2(18,2)
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	1(33,3)	1(50)	1(33,3)	1(33,3)	4(36,3)
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	1(33,3)	0(0)	0(0)	2(66,6)	3(27,3)
13.	Were uninterpretable/ intermediate test results reported?	3(100)	2(100)	2(66,6)	3(100)	10(90,9)
14.	Were withdrawals from the study explained?	2(66,6)	2(100)	2(66,6)	2(66,6)	8(72,7)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

### 5.4.3 LAES&HAES

Na análise dos artigos de acurácia diagnóstica do periódico LAES&HAES (n=15) verificou-se um percentual médio de 45,2% ( $\pm 12,8$ ) de respostas “sim” para os artigos analisados. O artigo com melhor qualidade de redação apresentou 64,3% de respostas “sim” e aquele com menor qualidade, 21,4% (Tabela 12). Para as respostas “não claro” o percentual médio foi de 38,1% ( $\pm 9,2$ ) (Tabela 13) e para as respostas “não” foi de 16,7% ( $\pm 7,5$ ). O número respostas “sim” e “não claro” do QUADAS alcançado por cada artigo individualmente é apresentado na Figura 10.



**Figura 10.** Número de itens com resposta “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação no periódico LAES&HAES.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 1, 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Elas estão associadas, respectivamente, à população selecionada para o estudo (questão 1), à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11, respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 16).

**Tabela 16.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	0(0)	1(100)	0(0)	1(25)	1(16,6)	3(20)
2.	Were selection criteria clearly described?	0(0)	0(0)	0(0)	1(25)	0(0)	1(6,6)
3.	Is the reference standard likely to correctly classify the target condition?	1(50)	1(100)	1(50)	3(75)	2(33,3)	8(53,3)
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	1(50)	1(100)	1(50)	3(75)	1(16,6)	7(46,6)
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	2(100)	0(0)	2(100)	4(100)	5(83,3)	13(86,6)
6.	Did patients receive the same reference standard regardless of the index test result?	2(100)	0(0)	2(100)	4(100)	5(83,3)	13(86,6)
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	2(100)	1(100)	2(100)	4(100)	5(83,3)	14(93,3)
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	1(50)	0(0)	1(50)	1(25)	1(16,6)	4(26,6)
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	1(50)	0(0)	0(0)	1(25)	0(0)	2(13,3)
10	Were the index test results interpreted without knowledge of the results of the reference standard?	0(0)	0(0)	1(50)	0(0)	1(16,6)	2(13,3)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

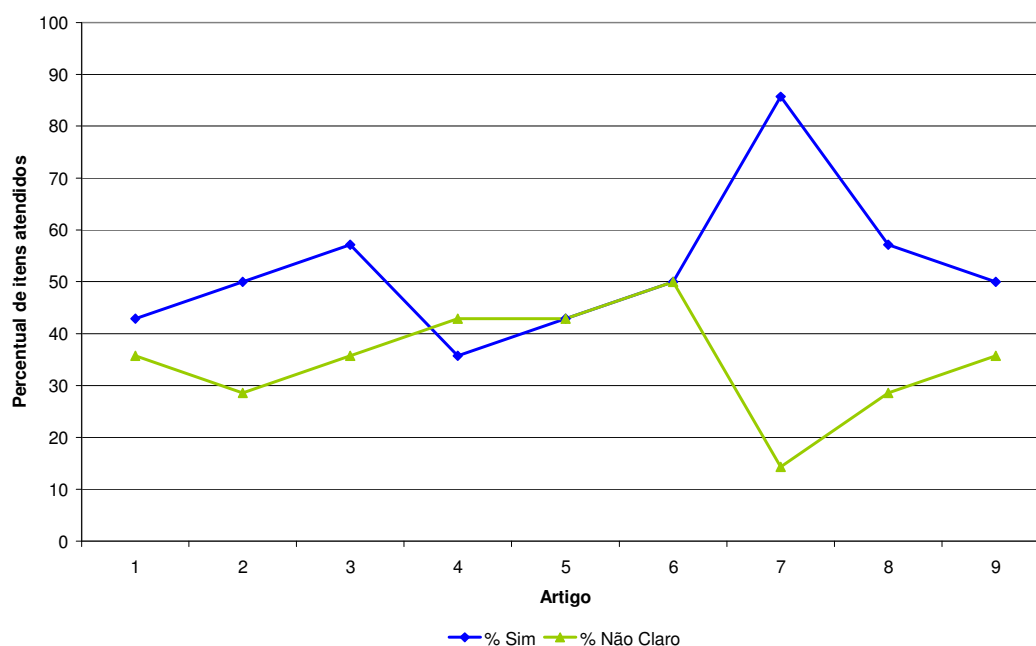
**Tabela 16.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico LAES&HAES.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	0(0)	0(0)	0(0)	1(25)	1(16,6)	2(13,3)
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
13.	Were uninterpretable/ intermediate test results reported?	2(100)	1(100)	2(100)	3(75)	6(100)	14(93,3)
14.	Were withdrawals from the study explained?	2(100)	1(100)	2(100)	3(75)	6(100)	14(93,3)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

#### 5.4.4 Revista NewsLab

Na análise dos artigos de acurácia diagnóstica da Revista NewsLab (n=09) verificou-se um percentual médio de 52,4% ( $\pm 14,3$ ) de respostas “sim” para os artigos analisados. O artigo com melhor qualidade de redação apresentou 85,7% de respostas “sim” e aquele com menor qualidade, 35,7% (Tabela 12). Para as respostas “não claro” o percentual médio foi de 34,9% ( $\pm 10,4$ ) (Tabela 13) e para as respostas “não” foi de 12,7 ( $\pm 8,6$ ). A figura 11 mostra o número respostas “sim” e “não claro” do QUADAS alcançado por cada artigo individualmente.



**Figura 11.** Número de itens com respostas “sim” e “não claro” do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação na revista NewsLab.

Apenas a questão 13 obteve respostas “sim” em todos os artigos e está associada à apresentação de resultados duvidosos ou não interpretáveis. As questões 1, 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à população selecionada para o estudo, questão 1, independência na interpretação do teste avaliado e do teste padrão, questões 10 e 11 e conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação, questão 12 (Tabela 17).

**Tabela 17.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	0(0)	1(100)	1(100)	2(100)	0(0)	4(44,4)
2	Were selection criteria clearly described?	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
3	Is the reference standard likely to correctly classify the target condition?	1(25)	1(100)	1(100)	2(100)	1(100)	6(66,6)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

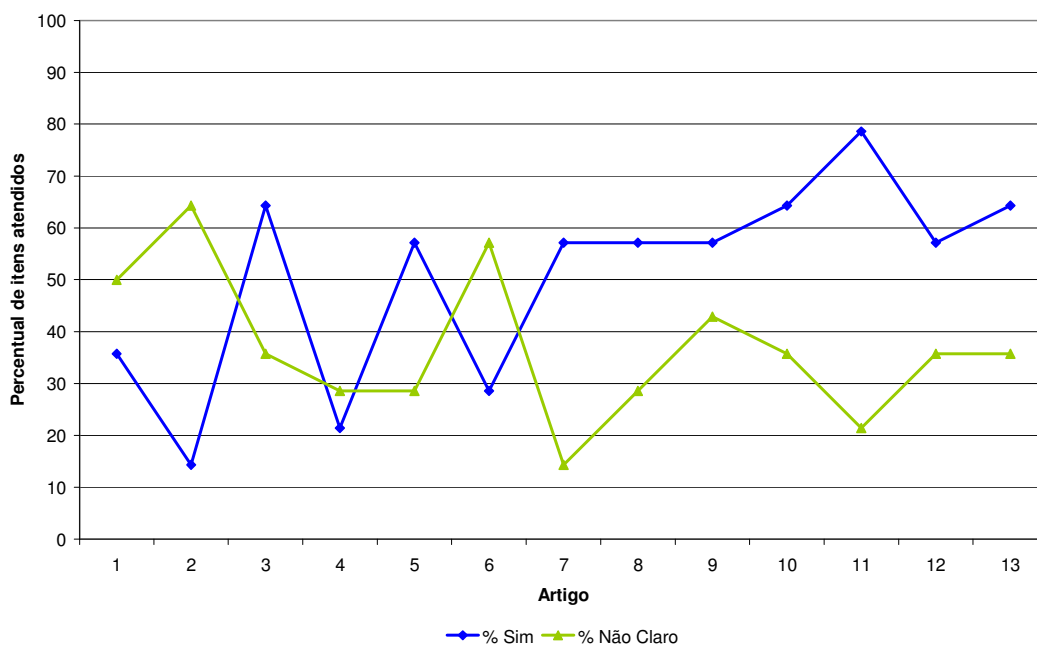
**Tabela 17.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico NewsLab.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	4(100)	1(100)	0(0)	2(100)	1(100)	8(88,8)
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	4(100)	0(0)	1(100)	2(100)	1(100)	8(88,8)
6.	Did patients receive the same reference standard regardless of the index test result?	2(50)	0(0)	1(100)	2(100)	1(100)	6(66,6)
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	3(75)	1(100)	1(100)	2(100)	1(100)	8(88,8)
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	1(25)	0(0)	0(0)	1(50)	1(100)	3(33,3)
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	1(25)	0(0)	0(0)	1(50)	0(0)	2(22,2)
10.	Were the index test results interpreted without knowledge of the results of the reference standard?	1(25)	0(0)	0(0)	1(50)	0(0)	2(22,2)
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	1(25)	0(0)	0(0)	1(50)	0(0)	2(22,2)
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
13.	Were uninterpretable/ intermediate test results reported?	4(100)	1(100)	1(100)	2(100)	1(100)	9(100)
14.	Were withdrawals from the study explained?	4(100)	1(100)	1(100)	2(100)	0(0)	8(88,8)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

### 5.4.5 Revista Brasileira de Análises Clínicas

Na análise dos artigos de acurácia diagnóstica da RBAC (n=13) se verificou um percentual médio de 50,5% ( $\pm 19,2$ ) de respostas “sim” para os artigos analisados. O artigo com melhor qualidade de redação apresentou 78,6% de respostas “sim” e aquele com menor qualidade, 14,3% (Tabela 12). Para as respostas “não claro” o percentual médio foi de 36,8% ( $\pm 14,0$ ) e para as respostas “não” foi de 12,6% ( $\pm 14,6$ ) (Tabela 13). A Figura 12 apresenta o número respostas “sim” do QUADAS alcançado por cada artigo individualmente.



**Figura 12.** Número de itens do instrumento QUADAS atendidos por cada artigo individualmente em ordem cronológica de publicação na RBAC.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 18).



**Tabela 18.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC.

Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?	2(100)	1(50)	2(100)	2(66,6)	2(50)	9(69,2)
2.	Were selection criteria clearly described?	0(0)	0(0)	1(50)	1(33,3)	1(25)	3(23)
3.	Is the reference standard likely to correctly classify the target condition?	0(0)	2(100)	0(0)	2(66,6)	4(100)	8(61,5)
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	0(0)	0(0)	1(50)	2(66,6)	4(100)	7(58,3)
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	0(0)	1(50)	1(50)	2(66,6)	4(100)	8(61,5)
6.	Did patients receive the same reference standard regardless of the index test result?	0(0)	1(50)	1(50)	2(66,6)	4(100)	8(61,5)
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	1(50)	1(50)	2(100)	2(66,6)	4(100)	10(77)
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	2(100)	2(100)	1(50)	2(66,6)	4(100)	11(84,6)
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	0(0)	1(50)	1(50)	1(33,3)	2(50)	5(38,5)
10	Were the index test results interpreted without knowledge of the results of the reference standard?	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

**Tabela 18.** Atendimento individual dos itens do instrumento QUADAS para avaliação dos artigos de acurácia diagnóstica de acordo com o ano de publicação no periódico RBAC.

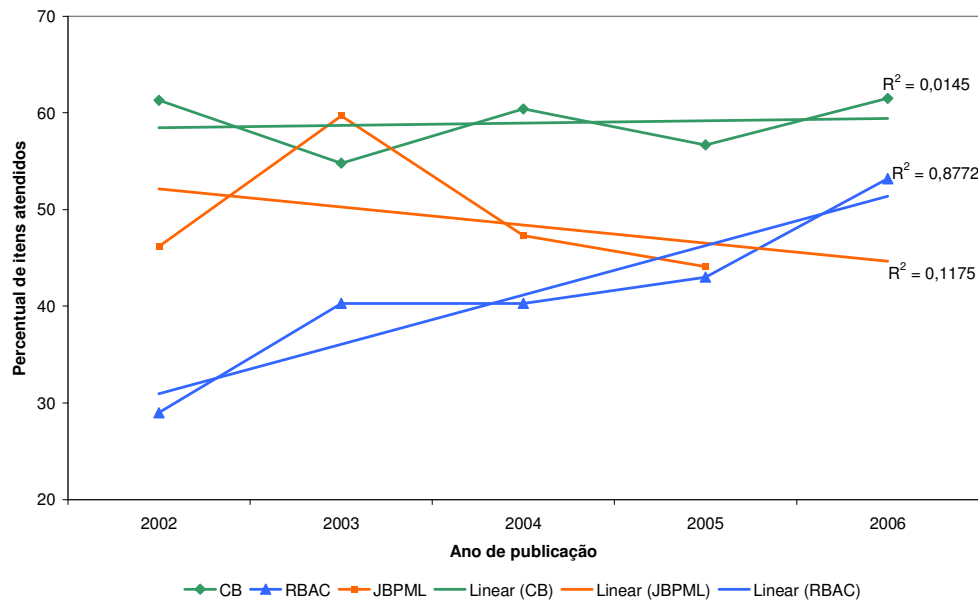
Item	Descrição dos itens (conforme original)	Número de artigos por ano*					Total
		2002	2003	2004	2005	2006	
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	0(0)	0(0)	0(0)	1(33,3)	0(0)	1(7,7)
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	0(0)	0(0)	0(0)	1(33,3)	0(0)	1(7,7)
13.	Were uninterpretable/ intermediate test results reported?	1(50)	2(100)	1(50)	3(100)	4(100)	11(84,6)
14.	Were withdrawals from the study explained?	1(50)	1(50)	1(50)	3(100)	4(100)	10(77)

\*Números entre parênteses são percentuais, calculados com base no número total de artigos de acurácia diagnóstica recuperados por ano.

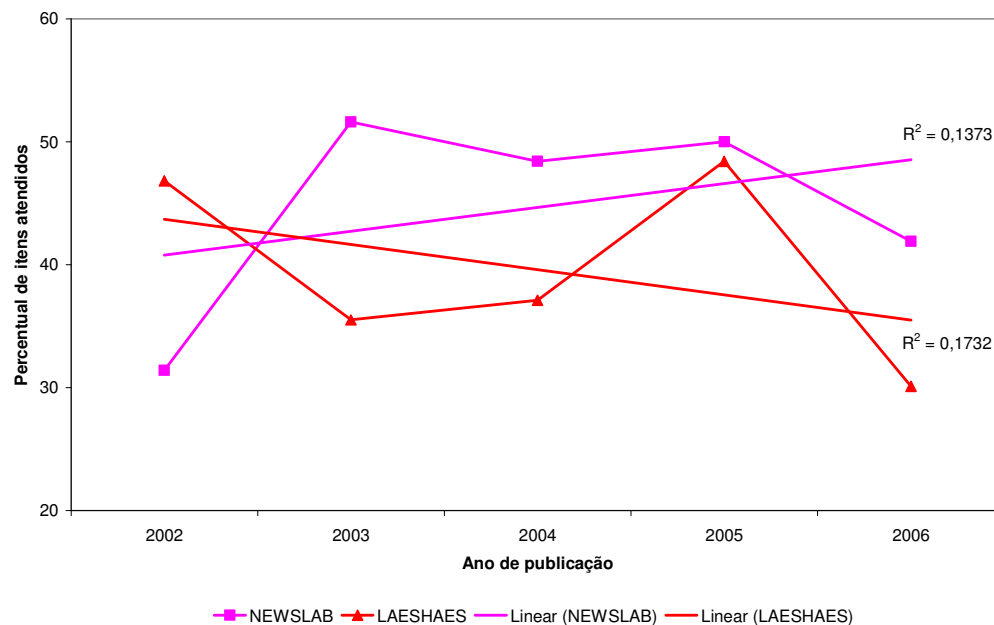
### 5.5 Análise da variação dos itens atendidos do STARD e QUADAS no período analisado

As proporções de itens atendidos pelo STARD e pelo QUADAS, obtidas para cada ano do período analisado foram comparadas. Esta comparação foi realizada para verificar se houve melhora na qualidade de redação ano a ano. Para tal, foi utilizada a análise de qui-quadrado ( $\chi^2$ ) Armitage para tendências lineares.

Utilizando as proporções de vezes que os itens do roteiro STARD foram atendidos através dos cinco anos analisados, o teste de  $\chi^2$  Armitage para tendências lineares demonstrou tendência linear estatisticamente significativa para a Revista Brasileira de Análises Clínicas ( $P=0,0021$ ) e revista NewsLab ( $P=0,0324$ ). Através das linhas de tendência com seus respectivos coeficientes de determinação ( $R^2$ ) a Revista Brasileira de Análises Clínicas apresentou essa característica de forma marcante com  $R^2 = 0,8772$ . Os gráficos com os percentuais de itens atendidos com a aplicação do STARD aos diferentes periódicos e seus coeficientes de determinação são apresentados nas Figuras 13, 14.

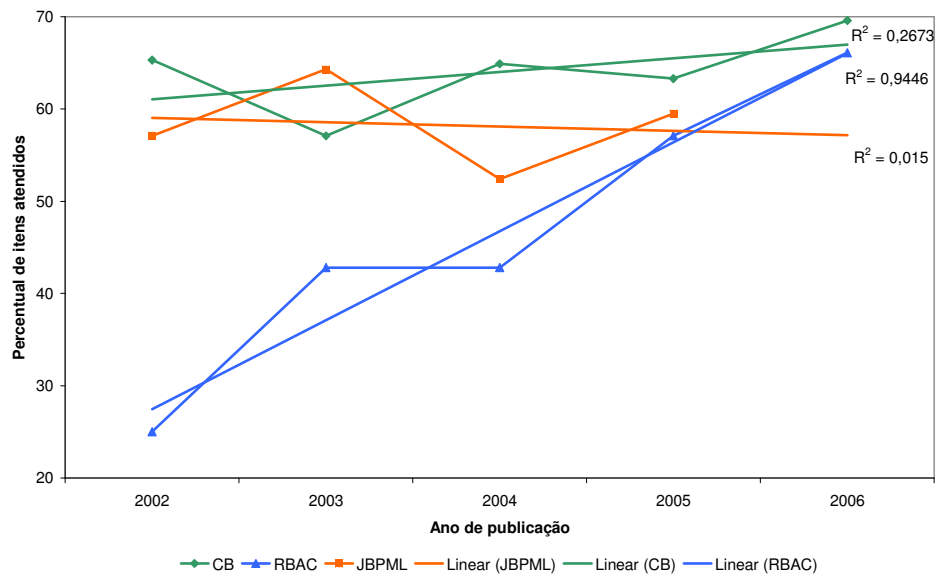


**Figura 13.** Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro STARD nos diferentes anos analisados nos periódicos *Clinical Biochemistry*, RBAC e JBPML.

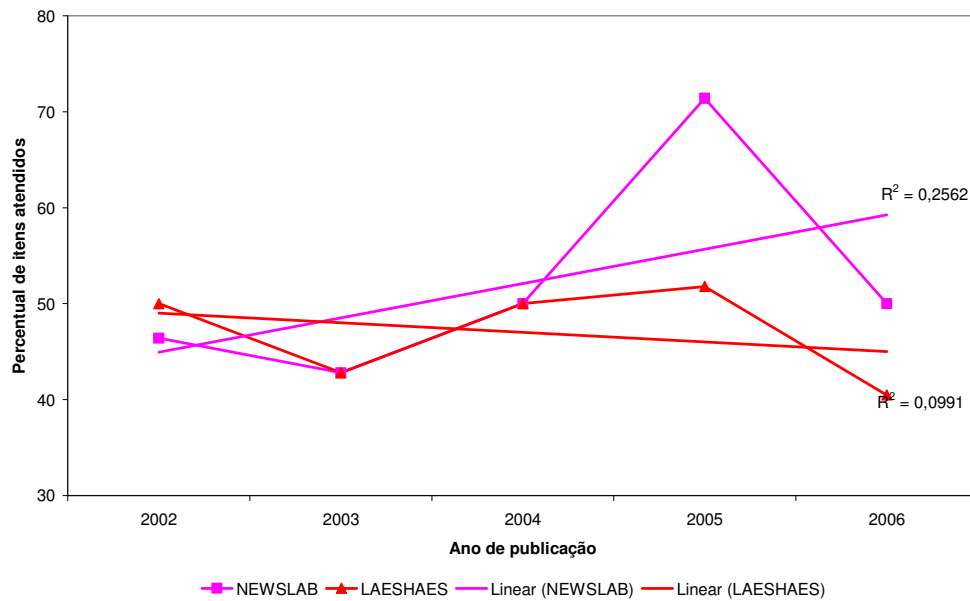


**Figura 14.** Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro STARD nos diferentes anos analisados nos periódicos NewsLab e LAES&HAES.

Utilizando as proporções de vezes que os itens do instrumento QUADAS foram atendidos através dos cinco anos analisados, o teste de  $\chi^2$  Armitage para tendências lineares demonstrou tendência linear estatisticamente significativa para a Revista Brasileira de Análises Clínicas ( $P=0,0002$ ) com coeficiente de determinação de 0,9446. Os gráficos com os percentuais de itens atendidos com a aplicação do QUADAS aos diferentes periódicos e seus coeficientes de determinação são apresentados nas Figuras 15 e 16.



**Figura 15.** Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro QUADAS nos diferentes anos analisados nos periódicos *Clinical Biochemistry*, RBAC e JBPML.



**Figura 16.** Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro QUADAS nos diferentes anos analisados na revista NewsLab e LAES&HAES.

## **5.6 Avaliação dos itens do STARD e QUADAS concernentes à validade interna e externa dos artigos avaliados**

Os itens do STARD e QUADAS foram divididos naqueles relacionados à validade interna e validade externa. Validade interna refere-se ao grau em que as conclusões de um estudo são corretas para a amostra de indivíduos investigados e validade externa está relacionada ao grau de veracidade dos resultados de uma observação em outros cenários, ou seja, a capacidade de generalização do estudo. Para o roteiro STARD os itens relacionados à validade interna são 4, 5, 6, 7, 9, 10, 11, 13, 14, 16, 17, 19, 21, 22, 23 e 24. Aqueles relacionados à validade externa são 3, 15, 18 e 25. Para o instrumento QUADAS àqueles ligados à validade interna são os itens 3, 4, 5, 6, 7, 10, 11, 12 e 14 e à validade externa são os itens 1 e 2.

### **5.6.1 *Clinical Biochemistry***

#### **5.6.1.1 STARD**

A frequência com que os itens relacionados à validade interna foram atendidos variou de 3,8% a 98,1% (questão 13b e 21, respectivamente). A média de itens atendidos foi de 57,6% ( $\pm 10,6$ ). Os itens atendidos por mais de 90% dos artigos foram o 4, 6, 7, 9a, 16, 21. Referem-se à descrição do modo de recrutamento dos participantes do estudo (questão 4), à descrição da coleta de dados como um estudo prospectivo ou retrospectivo (questão 6), à descrição do padrão de referência (questão 7), à definição das unidades, limites superiores de positividade e/ou categorias dos resultados do teste em avaliação (questão 9a), à descrição do número de participantes nas diferentes etapas do estudo (questão 16), à apresentação da avaliação da acurácia diagnóstica (questão 21). Os itens atendidos por menos de 10% dos artigos foram 10a, 13b, que referem-se respectivamente à descrição do número, treinamento e especialidade das pessoas que executaram ou interpretaram o teste padrão de referência e à descrição de métodos para a reprodutibilidade do teste padrão de referência (Tabela 7).

Quanto aos itens relacionados à validade externa, a média de itens atendidos foi de 91% ( $\pm 14,8$ ) e a frequência com que os itens foram atendidos variou de 77,3% a 100% (questão 18 e 25, respectivamente). Referem-se à descrição da distribuição da severidade da

doença nos pacientes com a patologia estudada ou outros diagnósticos em pacientes sem a condição de interesse (Tabela 7).

### **5.6.1.2 QUADAS**

A frequência com que os itens relacionados à validade interna foram atendidos variou de 1,9% a 88,7%. A média de itens atendidos foi de 56,8% ( $\pm 14,5$ ). Para as respostas “não claro” os itens relacionados variaram entre 1,9% e 90,5%, sendo a média 37,5% ( $\pm 14,0$ ). Para as respostas “não” os itens variaram de 0% e 13,2%.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 14).

Quanto aos dois itens relacionados à validade externa, estes obtiveram 94,3% (questão 1) e 73,6% (questão 2) de artigos com estes itens atendidos. Referem-se, à representatividade do conjunto de pacientes que receberiam o teste na prática clínica e à descrição clara dos critérios de seleção dos pacientes (Tabela 14).

## **5.6.2 Jornal Brasileiro de Patologia e Medicina Laboratorial**

### **5.6.2.1 STARD**

A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 100%. A média de itens atendidos foi de 52,3% ( $\pm 15,4$ ). Os itens atendidos por mais de 90% dos artigos foram o 6, 9a, 21. Referem-se à descrição da coleta de dados como um estudo prospectivo ou retrospectivo (questão 6), à definição das unidades, limites superiores de positividade e/ou categorias dos resultados do teste em avaliação (questão 9a), à apresentação da avaliação da acurácia diagnóstica (questão 21). Os itens atendidos por menos de 10% dos artigos foram 10b, 11a, 11b, 13a, 13b, 23, 24a, 24b que se referem respectivamente à descrição do número, treinamento e especialidade das pessoas que executaram ou interpretaram o teste padrão de referência, à avaliação cegada dos leitores dos

testes para os resultados do teste em avaliação e padrão de referência, à descrição de métodos para cálculo da reprodutibilidade do teste em avaliação e do padrão de referência, à descrição da variabilidade da acurácia diagnóstica entre subgrupos, revisores ou centros, à avaliação da reprodutibilidade do teste em avaliação e do teste padrão de referência (Tabela 8).

Quanto aos itens relacionados à validade externa, a média de itens atendidos foi de 59,1% ( $\pm 28,0$ ) e a frequência com que os itens foram atendidos variou de 36,4% a 100% (questão 15 e 25, respectivamente). Referem-se à descrição das características clínicas e demográficas da população estudada e à discussão da aplicabilidade clínica dos resultados do estudo (Tabela 8).

#### 5.6.2.2 QUADAS

A frequência com que os itens relacionados à validade interna foram atendidos variou de 18,2% a 91%. A média de itens atendidos foi de 56,5% ( $\pm 14,4$ ). Para as respostas “não claro” os itens relacionados variaram entre 0% e 72,7%, sendo a média 30,3% ( $\pm 18,6$ ). Para as respostas “não” os limites superiores e inferiores foram respectivamente 0% e 45,4% com média 13,1 ( $\pm 10,9$ ).

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 15).

Quanto aos dois itens relacionados à validade externa, estes obtiveram 91% (questão 1) e 27,3% (questão 2) de artigos com estes itens atendidos. Referem-se, à representatividade do conjunto de pacientes que receberiam o teste na prática clínica e à descrição clara dos critérios de seleção dos pacientes (Tabela 15).

### 5.6.3 LAES&HAES

#### 5.6.3.1 STARD

A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 86,7%. A média de itens atendidos foi de 44,3% ( $\pm 14,9$ ). Os itens atendidos por mais de 80% dos artigos foram o 6, 16, 21. Referem-se à descrição da coleta de dados como um estudo prospectivo ou retrospectivo (questão 6), à descrição do número de participantes nas diferentes etapas do estudo (questão 16), à apresentação da avaliação da acurácia diagnóstica (questão 21). Os itens atendidos por menos de 10% dos artigos foram 10a, 10b, 13b, 24b que se referem respectivamente à descrição do número, treinamento e especialidade das pessoas que executaram ou interpretaram o teste em avaliação e do teste padrão de referência, à descrição de métodos para cálculo da reprodutibilidade do padrão de referência, à avaliação da reprodutibilidade do teste padrão de referência (Tabela 9).

Quanto aos itens relacionados à validade externa, a média de itens atendidos foi de 35,0% ( $\pm 24,6$ ) e a frequência com que os itens foram atendidos variou de 6,7% a 86,7% (questão 15 e 25, respectivamente). Referem-se à descrição das características clínicas e demográficas da população estudada e à discussão da aplicabilidade clínica dos resultados do estudo. As demais questões referentes à validade externa foram reportadas por 33,3% dos artigos analisados (questão 3) e por apenas 13,3% dos artigos analisados (questão 18) (Tabela 9).

#### 5.6.3.2 QUADAS

A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 93,3%. A média de itens atendidos foi de 54,1% ( $\pm 12,5$ ). Para as respostas “não claro” os itens relacionados variaram entre 0% e 100%, sendo a média 42,2% ( $\pm 12,7$ ). Para as respostas “não” os limites superiores e inferiores foram respectivamente 0% e 13,3% com média 13,1 ( $\pm 10,9$ ).

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à independência na interpretação do teste avaliado e do teste padrão (questões 10 e



11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 16).

Quanto aos dois itens relacionados à validade externa, estes obtiveram 20% (questão 1) e 6,7% (questão 2) de artigos com estes itens atendidos. Referem-se, à representatividade do conjunto de pacientes que receberiam o teste na prática clínica e à descrição clara dos critérios de seleção dos pacientes. A questão 1 obteve 80% de respostas “não claro” e o item 2 apresentou 86,7% de respostas “não” (Tabela 16).

## 5.6.4 NewsLab

### 5.6.4.1 STARD

A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 100%. A média de itens atendidos foi de 50,7% ( $\pm 13,0$ ). Os itens atendidos por mais de 80% dos artigos foram o 6, 9a, 16, 17, 21 e 22. Referem-se à descrição da coleta de dados como um estudo prospectivo ou retrospectivo (questão 6), à definição das unidades, limites superiores de positividade e/ou categorias dos resultados do teste em avaliação (questão 9a), à descrição do número de participantes nas diferentes etapas do estudo (questão 16), à descrição do intervalo de tempo entre o teste avaliado e o teste padrão de referência ou qualquer tratamento administrado entre eles (questão 17), à apresentação da avaliação da acurácia diagnóstica (questão 21) à descrição do manejo dos resultados indeterminados, perdidos e *outliers* (questão 22). Os itens atendidos por menos de 10% dos artigos foram 5, 10a, 10b, 11a, 11b, 13b, 24b que se referem respectivamente à descrição da amostragem dos participantes, especificando como os pacientes foram selecionados, à descrição do número, treinamento e especialidade das pessoas que executaram ou interpretaram o teste sob avaliação e o teste padrão de referência, à avaliação cegada dos leitores dos testes para os resultados do teste em avaliação e padrão de referência, à descrição de métodos para cálculo da reprodutibilidade do teste padrão de referência, à avaliação da reprodutibilidade do teste padrão de referência (Tabela 10).

Quanto aos itens relacionados à validade externa, a média de itens atendidos foi de 25,0% ( $\pm 25,0$ ) e a frequência com que os itens foram atendidos variou de 11,1% (questões 3, 15 e 18) a 66,7% (questão 25). Referem-se à descrição da população estudada, critérios de inclusão e exclusão e local onde os dados foram coletados, à descrição das características

clínicas e demográficas da população estudada, à descrição da severidade da doença nos pacientes com a patologia estudada e em participantes sem a condição de interesse e à discussão da aplicabilidade clínica dos resultados do estudo (Tabela 10).

#### **5.6.4.2 QUADAS**

A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 88,9%. A média de itens atendidos foi de 59,2% ( $\pm 13,6$ ). Para as respostas “não claro” os itens relacionados variaram entre 0% e 100%, sendo a média 38,3% ( $\pm 14,8$ ). Para as respostas “não” os limites superiores e inferiores foram respectivamente 0% e 11,1%.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 17).

Quanto aos dois itens relacionados à validade externa, estes obtiveram 44,4% (questão 1) e 0% (questão 2) de artigos com estes itens atendidos. Referem-se, à representatividade do conjunto de pacientes que receberiam o teste na prática clínica e à descrição clara dos critérios de seleção dos pacientes. A questão 1 obteve 55,6% de respostas “não claro” e o item 2 apresentou 66,7% de respostas “não” (Tabela 17).

### **5.6.5 Revista Brasileira de Análises Clínicas**

#### **5.6.5.1 STARD**

A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 100%. A média de itens atendidos foi de 44,5% ( $\pm 15,1$ ). Os itens atendidos por mais de 70% dos artigos foram o 6, 9a, 21 e 22. Referem-se à descrição da coleta de dados como um estudo prospectivo ou retrospectivo (questão 6), à definição das unidades, limites superiores de positividade e/ou categorias dos resultados do teste em avaliação (questão 9a), à apresentação da avaliação da acurácia diagnóstica (questão 21) à descrição do manejo dos resultados indeterminados, perdidos e outliers (questão 22). Os itens atendidos por menos de 10% dos artigos foram 5, 10a, 10b, 11a, 11b, 13a, 13b, 24a, 24b que se referem

respectivamente à descrição da amostragem dos participantes, especificando como os pacientes foram selecionados, à descrição do número, treinamento e especialidade das pessoas que executaram ou interpretaram o teste sob avaliação e o teste padrão de referência, ao cegamento dos leitores dos testes para os resultados do teste em avaliação e do teste padrão de referência, à descrição de métodos para cálculo da reprodutibilidade do teste em avaliação e do teste padrão de referência, à apresentação da avaliação da reprodutibilidade do teste em avaliação e do teste padrão de referência (Tabela 11).

Quanto aos itens relacionados à validade externa, a média de itens atendidos foi de 53,8% ( $\pm 24,7$ ) e a frequência com que os itens foram atendidos variou de 23% (questão 18) a 100% (questão 25) de artigos com estes itens atendidos. Referem-se, à descrição da distribuição da severidade da doença nos pacientes com a patologia estudada e em participantes sem a condição de interesse e à discussão da aplicabilidade clínica dos resultados do estudo (Tabela 11).

#### **5.6.5.2 QUADAS**

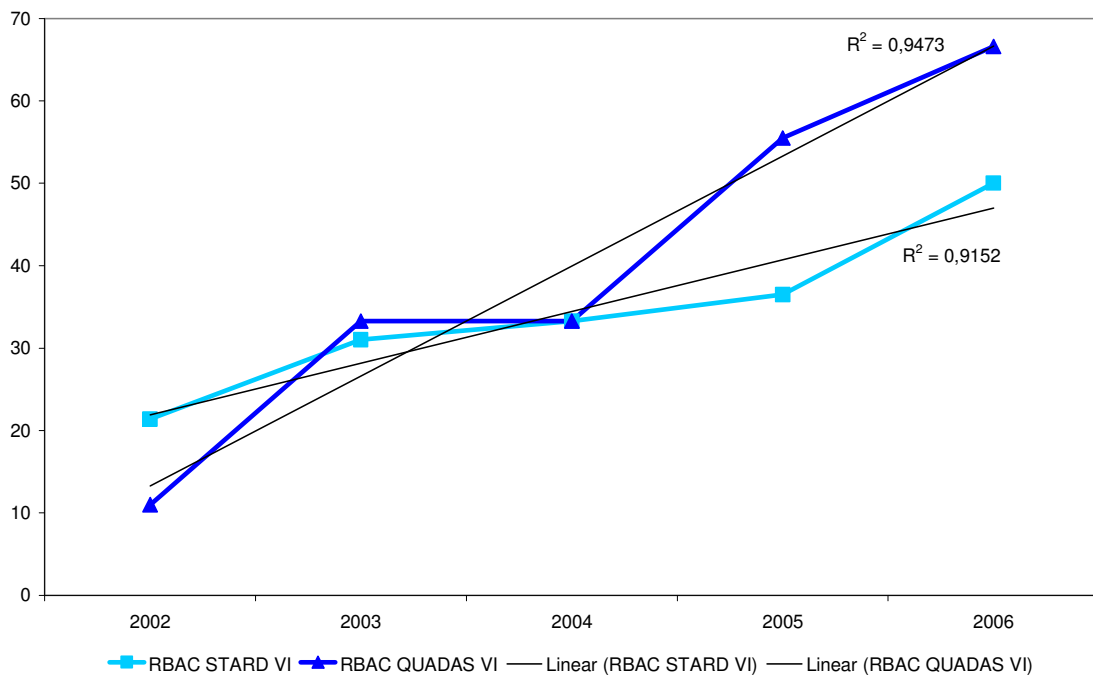
A frequência com que os itens relacionados à validade interna foram atendidos variou de 0% a 72,7%. A média de itens atendidos foi de 45,3% ( $\pm 23,8$ ). Para as respostas “não claro” os itens relacionados variaram entre 0% e 92,3%, sendo a média 46,1% ( $\pm 21,2$ ). Para as respostas “não” os limites superiores e inferiores foram respectivamente 0% e 23%.

Nenhuma questão obteve respostas “sim” em todos os artigos. As questões 10, 11 e 12 apresentaram os maiores percentuais de resposta “não claro”. Respectivamente elas estão associadas à independência na interpretação do teste avaliado e do teste padrão (questões 10 e 11 respectivamente) e ao conhecimento dos dados clínicos do paciente para interpretação do teste em avaliação (questão 12) (Tabela 18).

Quanto aos dois itens relacionados à validade externa, estes obtiveram 69,2% (questão 1) e 23% (questão 2) de artigos com estes itens atendidos. Referem-se, à representatividade do conjunto de pacientes que receberiam o teste na prática clínica e à descrição clara dos critérios de seleção dos pacientes. A questão 1 obteve 23% de respostas “não claro” e o item 2 apresentou 46% de respostas “não” (Tabela 18).

### 5.7 Análise da variação dos itens concernentes à validade interna e externa atendidos do STARD e QUADAS no período analisado

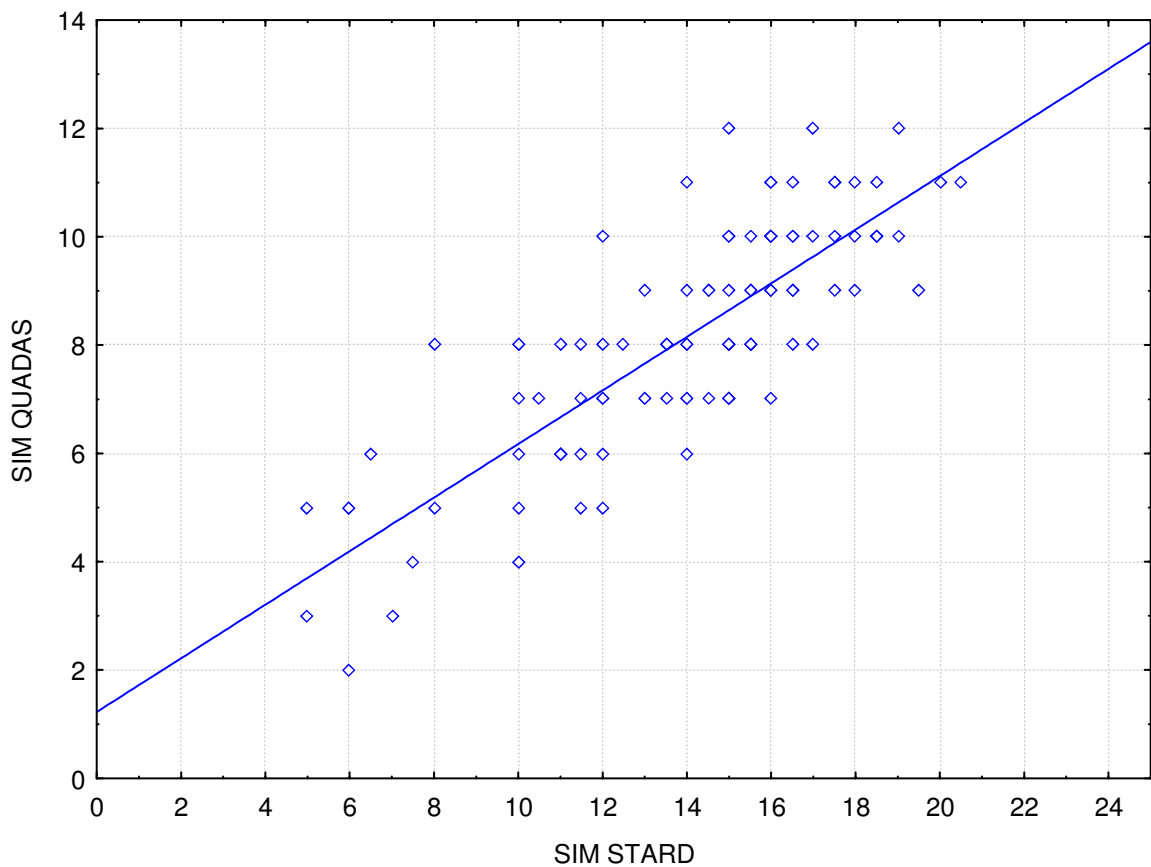
Utilizando as proporções de vezes que os itens concernentes à validade interna e externa do instrumento QUADAS e do roteiro STARD foram atendidos através dos cinco anos analisados, o teste de  $\chi^2$  Armitage para tendências lineares demonstrou tendência linear estatisticamente significativa apenas para a Revista Brasileira de Análises Clínicas na aplicação do roteiro STARD e do instrumento QUADAS nos itens relacionados à validade interna em ambos ( $P=0,0012$  e  $P<0,0001$ , respectivamente). Os gráficos com os percentuais de itens atendidos com a aplicação do QUADAS e STARD e seus coeficientes de determinação são apresentados nas figuras 17.



**Figura 17.** Gráfico para avaliação da presença de tendência linear pela aplicação do roteiro STARD e QUADAS concernentes às questões relacionadas à validade interna na RBAC.

## 5.8 Correlação entre STARD e QUADAS

Com o objetivo de analisar o comportamento conjunto dos resultados da aplicação do roteiro STARD e do instrumento QUADAS e verificar se existe relação entre as duas variáveis foi construído um gráfico de dispersão. A simples observação da distribuição de pontos, indica existir uma correlação entre as respostas “sim” do QUADAS e do STARD. Através do coeficiente de correlação linear de Pearson obteve-se um resultado de  $r=0,8133$ , indicando uma relação linear positiva forte entre as duas variáveis. O diagrama de dispersão com os pares de valores pode ser observado na Figura 18.



**Figura 18.** Diagrama de dispersão das variáveis respostas afirmativas STARD e QUADAS na totalidade de artigos avaliados (n=101).

## 6. Discussão

A busca manual realizada em nosso estudo localizou e recuperou 101 artigos de acurácia diagnóstica nos seis periódicos avaliados. Do total recuperado, a distribuição dos artigos entre os periódicos foi: *Clinical Biochemistry* (52,5%), LAES&HAES (14,8%), RBAC (12,9%), JBPML (10,9%), NewsLab (8,9%), e RBCF (0%). Pode-se notar que mais de 50% dos artigos foram publicados na única publicação internacional analisada, indexada no MEDLINE e com fator de impacto 2,072 (ELSEVIER, 2008). Smidt *et al* (2005, 2006a, 2006b) analisando periódicos com fator de impacto maior ou igual a 4 constataram que a maior parte dos estudos de acurácia diagnóstica são publicados em periódicos direcionados a doenças ou áreas do conhecimento específicas como é o caso desta publicação, o que poderia explicar parcialmente a distribuição desigual encontrada. No estudo publicado por Wilczynski (2008), avaliando a qualidade de redação de estudos de acurácia diagnóstica através da aplicação do STARD se verificou uma média de mais de 60 estudos de acurácia diagnóstica publicados anualmente em periódicos da área de radiologia contra uma média de apenas 6 estudos em periódicos médicos de clínica geral. Em estudo piloto realizado em 18 volumes de nossos periódicos nacionais e 12 volumes do periódico *Clinical Biochemistry* selecionados de forma aleatória entre os anos 2002 e 2006, encontramos um número aproximadamente três vezes maior de artigos avaliando o desempenho técnico quando comparados a estudos clínicos avaliando o desempenho diagnóstico, demonstrando que estes periódicos não publicam com regularidade artigos de acurácia diagnóstica. Aliado a estes resultados está a diferença estatística significativa pelo teste de  $\chi^2$  encontrada entre a proporção de artigos de acurácia diagnóstica publicados nos periódicos nacionais quando comparados ao periódico internacional. Portanto, esta diferença pode estar relacionada ao nível de complexidade exigido por este tipo de investigação diagnóstica, atingido em menor proporção pelos periódicos nacionais analisados ou ainda, que estes periódicos nacionais da área das Análises Clínicas não são direcionados ao tema.

Com o objetivo de facilitar a identificação dos artigos de acurácia diagnóstica, é recomendado pela iniciativa STARD e pelo QUADAS, que seja utilizado o termo “acurácia diagnóstica” no título ou resumo de um artigo que compare os resultados de um ou mais testes com os resultados de um padrão de referência. Também objetivando facilitar a recuperação de estudos de acurácia diagnóstica, a base de dados MEDLINE da *National Library of Medicine* introduziu, em 1991, um descritor específico ou termo MeSH (*Medical Subject Headings*) para estudos diagnósticos: “sensibilidade e especificidade” (BOSSUYT *et al.*, 2003a).

Contudo, nosso processo de busca e recuperação dos artigos de acurácia diagnóstica nos diferentes periódicos foi dificultado por falhas na identificação objetiva dos mesmos. Aproximadamente 18% do total de artigos recuperados não apresentavam os termos acurácia diagnóstica, sensibilidade ou especificidade no título ou resumo. Este percentual foi bastante variado entre os periódicos analisados (Figura 1), alcançando até 44% dos artigos do periódico NewsLab. Nestes casos, foi necessária a completa leitura dos artigos para caracterizá-los como estudos de acurácia diagnóstica. Problema semelhante foi encontrado por Deville, Bezemer e Bouter (2000) ao utilizarem estes descritores no processo de busca em periódicos indexados no MEDLINE entre os anos de 1992 e 1995 identificando, corretamente, apenas 51% de todos os estudos de acurácia diagnóstica. Rama, Poovali e Apsingi (2006) utilizando os descritores “sensibilidade e especificidade” puderam identificar 73% dos artigos de acurácia diagnóstica incluídos em seu estudo, mas obtiveram 135 falsos positivos, ou seja, sua busca obteve um valor preditivo positivo de apenas 16,7%. Estudos que procuram determinar a utilidade de filtros metodológicos nas estratégias de busca para estudos diagnósticos em bases de dados eletrônicas também pontuam a baixa sensibilidade e especificidade destas estratégias e a falta de informações em seus resumos como causa de omissão de um número considerável de estudos relevantes que não são identificados corretamente e que, portanto, não são incluídos em revisões sistemáticas de métodos diagnósticos (DOUST *et al.*, 2005; LEEFLANG *et al.*, 2006). Além disso, a ausência total de resumo foi constatada em 6% dos artigos do periódico LAES&HAES e 33% no periódico NewsLab, enquanto a ausência de resumo estruturado variou entre 4% e 100%, dificultando a procura e recuperação dos estudos (Tabela 5).

Os artigos recuperados apresentaram heterogeneidade quanto à distribuição entre as áreas das Análises Clínicas. O periódico *Clinical Biochemistry* apresentou um predomínio de artigos da área de bioquímica compatível com o escopo da publicação. O mesmo predomínio foi observado no JBPML, assim como verificado no estudo de Lumbreras-Lacarra, Ramos-Rincon e Hernández-Aguado (2004). Imunologia foi a área mais freqüente nos artigos dos periódicos LAES&HAES, NewsLab e RBAC, correspondendo a 61%, 23% e 32% respectivamente. A importância dos índices de acurácia como sensibilidade e especificidade para os estudos na área de imunologia pode explicar o predomínio desta área entre os artigos de acurácia diagnóstica, pois é uma área altamente dependente destes índices para que possam ser utilizados na substituição de tecnologias, testes de triagem ou como testes adicionais em investigações diagnósticas (BOSSUYT *et al.*, 2006). O número de artigos na área de histoquímica foi bastante significativo no periódico JBPML, representando 27% dos artigos

de acurácia diagnóstica. Este achado aparentemente está de acordo com os objetos de estudo desta publicação que inclui, além da medicina laboratorial, a patologia clínica, a patologia e a citopatologia. Os artigos dos periódicos LAES&HAES e NewsLab abrangeram maior número de áreas das Análises Clínicas, demonstrando também maior homogeneidade entre as proporções das áreas. Estes dados são compatíveis com a proposta dos periódicos de serem publicações gerais na área das Análises Clínicas. O trabalho de Delwiche (2003) demonstra clara associação entre o escopo de periódicos e as áreas laboratoriais apresentadas com maior frequência nos artigos publicados. Contudo, o escopo das publicações analisadas no presente trabalho pode explicar apenas parcialmente esta distribuição, pois o enfoque apenas nos artigos de acurácia diagnóstica não reflete a real distribuição da totalidade de artigos publicados entre as áreas das Análises Clínicas nestes periódicos.

Os resultados de nosso estudo demonstram que a qualidade de redação dos artigos de acurácia diagnóstica obtida a partir da aplicação do roteiro STARD está, em geral, aquém da objetivada pela iniciativa STARD, ou seja, que o delineamento, condução, análise e resultados destes estudos possam ser compreendidos pelos leitores através de uma redação completa e transparente (BOSSUYT *et al.*, 2003c). A média de itens atendidos em nosso estudo foi  $15,9 \pm 2,2$  ( $63,6\% \pm 8,7$ ) para o periódico *Clinical Biochemistry*,  $13,4 \pm 2,9$  ( $53,6 \pm 11,5$ ) para o JBPML,  $12,0 \pm 3,2$  ( $48,1\% \pm 12,6$ ) para a RBAC,  $10,6 \pm 3,6$  ( $42,2\% \pm 14,5$ ) para a revista LAES&HAES e  $11,1 \pm 3,4$  ( $44,4\% \pm 13,8$ ) para a revista NewsLab.

Quanto aos periódicos nacionais analisados, menos de 44% dos artigos publicados atenderam a mais de 50% dos itens do STARD com o mínimo de 5 e o máximo de 19,5 itens atendidos. Estes resultados são próximos aos encontrados por Siddiqui, Azuara-Blanco e Burr (2005) que ao analisarem a qualidade de redação de 16 estudos de acurácia diagnóstica em 5 dos principais periódicos internacionais de oftalmologia publicados em 2002 constataram que o número de itens completamente atendidos variou entre 8 e 19 e que apenas 44% dos artigos atendiam explicitamente mais de 50% dos itens do STARD. No mesmo ano, outro estudo avaliando a qualidade de redação em 12 periódicos médicos com fator de impacto maior que 4 verificaram que nenhum dos 124 artigos avaliados atendeu a mais de 20 itens do STARD e que apenas 41% dos artigos atenderam a mais de 50% dos 25 itens (SMIDT *et al.*; 2005). Resultados semelhantes foram encontrados por Mahoney e Ellison (2007) durante a análise de 52 artigos avaliando a acurácia de monitores de glicemia *point-of-care*. Neste estudo apenas 42% dos artigos analisados atenderam a mais da metade dos itens do STARD.

O periódico *Clinical Biochemistry* apresentou 90% dos artigos analisados com mais de 50% dos itens atendidos, resultado superior aos trabalhos já citados, mas compatível com o



estudo de Rama, Poovali e Apsingi (2006) que ao analisarem a qualidade de redação de 37 artigos de acurácia diagnóstica em três periódicos de ortopedia geral obtiveram uma média de  $15,0 \pm 3,3$  ( $60\% \pm 13,2$ ) itens do STARD e 38% dos artigos atendendo mais de 2/3 dos itens do STARD. O periódico *Clinical Biochemistry*, por sua vez, apresentou 41% dos artigos atendendo a mais de 2/3 dos 25 itens do STARD e uma média de  $15,9 \pm 2,2$  ( $63,6\% \pm 8,7$ ).

De modo geral, as publicações em diversas áreas demonstram uma qualidade de redação abaixo da objetivada pela iniciativa STARD. As áreas avaliadas incluem a oftalmologia (PARANJOTHY; SHUNMUGAM; AZUARA-BLANCO, 2007; JOHNSON; SIDDIQUI; AZUARA-BLANCO; 2007; ZAFAR; KHAN; SIDDIQUI; 2008; SHUNMUGAM; AZUARA-BLANCO; 2006), ginecologia (SELMAN; KHAN; MANN; 2005), biologia molecular e proteômica (LUMBRERAS; JARRÍN; AGUADO; 2006), ortopedia (RAMA; POOVALI; APSINGI; 2006) entre outras. A qualidade de redação de 76 estudos avaliando testes de triagem para retinopatia diabética publicados entre 1995 e 2006 avaliada por Zafar, Khan e Siddiqui (2008) demonstrou uma média de  $8,0 \pm 3,3$  itens atendidos e que apenas 12% dos artigos atendiam explicitamente mais de 50% dos itens do STARD. Resultados semelhantes foram encontrados por Shunmugam e Azuara-Blanco (2006) onde o número e itens completamente atendidos variou entre 5 e 18 e onde apenas 24% dos artigos analisados atendiam explicitamente mais de 50% dos itens do STARD na avaliação da redação de 29 estudos de acurácia diagnóstica utilizando o tomógrafo de retina Heidelberg no diagnóstico de glaucoma. Johnson, Siddiqui e Azuara-Blanco (2007) na avaliação da redação de 30 estudos de acurácia diagnóstica utilizando a tomografia de coerência óptica no diagnóstico de glaucoma constataram que o número de itens completamente atendidos variou entre 5 e 17 e que apenas 26,7% dos artigos analisados atendiam explicitamente mais de 50% dos itens do STARD. No estudo conduzido por Lumbreras, Jarrín e Aguado (2006) na avaliação de 44 estudos de acurácia diagnóstica de testes genéticos, moleculares e proteômica, foi encontrada média de 9,8 itens (CI 95%, 8,8 – 10,6).

Em relação ao atendimento de itens específicos, as falhas mais frequentes na redação dos artigos foram a dos itens 10a, 10b, 11a, 11b, 13a, 13b, 20, 24a e 24b. A maioria das publicações falhou em descrever o número, treinamento e especialidade das pessoas que executaram e interpretaram o teste avaliado e o padrão de referência (item 10). Os resultados foram 6% (3/53) e 13% (7/53) para o periódico *Clinical Biochemistry*, 18% (2/11) e 9% (1/11) para o JBPM, 0% (0/13) e 0% (0/13) para a RBAC, 0% (0/15) e 7% (1/15) para a LAES&HAES, 0% (0/9) e 0% (0/9) na NewsLab. De modo semelhante, este item foi

pobremamente descrito no estudo de Johnson, Siddiqui e Azuara-Blanco (2007) onde apenas um artigo entre 30 analisados atendeu o item para ambos os testes. A maioria dos artigos (65%) analisados por Rama, Poovali e Apsingi não atenderam a este item. A importância de se conhecer o nível de treinamento dos indivíduos que executaram e interpretaram os testes está relacionada ao julgamento pelos leitores da real possibilidade de obterem resultados similares em seus próprios cenários clínicos (SIMEL; RENNIE; BOSSUYT, 2008; STARD STATEMENT, 2008). Deve-se ter em mente também que a experiência profissional, especialização e treinamento prévio melhoram a interpretação dos testes e reduzem a variação interobservador e que os responsáveis pela interpretação dos testes são mais propensos a interpretar como anormais, testes com resultados subjetivos, em locais de alta prevalência da condição de interesse, uma tendência conhecida como viés de contexto (STARD STATEMENT, 2008). Apesar disso, muitos testes laboratoriais exibem medições realizadas automaticamente tornando a probabilidade deste tipo de viés menos frequente (JOHNSON; SIDDIQUI, AZUARA-BLANCO, 2007).

Todos os periódicos analisados apresentaram menos de 50% dos artigos descrevendo se os indivíduos responsáveis pela interpretação do teste sob avaliação e do padrão de referência desconheciam o resultado do outro teste ou se haviam informações clínicas disponíveis para os leitores (item 11). O ato de ocultar essas informações dos indivíduos que interpretam os testes é conhecida como “cegamento” ou “mascaramento” (FLETCHER; FLETCHER, 2006). Este cuidado minimiza vieses de revisão, incluindo viés de revisão de teste, viés de revisão diagnóstica e viés de revisão clínica. Estes devem ser evitados, pois o conhecimento dos resultados do padrão de referência pode influenciar quem interpreta o teste sob avaliação e vice versa, aumentando a concordância entre os resultados do teste avaliado e do padrão de referência e levando a medidas de acurácia diagnóstica que não correspondem à realidade. (BOSSUYT, 2008; STARD STATEMENT, 2008; WHITING *et al.*, 2003). Trata-se de itens pouco atendidos pela maioria das publicações analisando a qualidade de redação de testes diagnósticos. Menos de 50% dos estudos analisados por Siddiqui, Azuara-Blanco e Burr (2005) e apenas 20% dos artigos analisados por Johnson, Siddiqui e Azuara-Blanco (2007) atenderam de forma completa o item 11 do roteiro STARD.

Se possível, os autores devem avaliar a reprodutibilidade dos métodos de ensaio utilizados no seu estudo e informar o procedimento de como fazê-lo. Esta variabilidade que diz respeito à quantidade de variação que ocorre durante a operação de um equipamento é denominado também por outros termos como imprecisão, variação analítica ou ruído analítico (SIMEL; RENNIE; BOSSUYT, 2008; STARD STATEMENT, 2008). A descrição de

métodos para o cálculo da reprodutibilidade do teste em avaliação (item 13a) foi apresentada em apenas 13% (7/53) dos artigos do periódico *Clinical Biochemistry*, 0% (0/11) no JBPML, 0% (0/13) na RBAC, 13% (2/15) na LAES&HAES e por 11% (1/9) dos artigos da NewsLab. Para o padrão de referência (item 13b) foi apresentada em 4% (2/53) dos artigos do periódico *Clinical Biochemistry*, 0% (0/11) no JBPML, 0% (0/13) na RBAC, 7% (1/15) na LAES&HAES e por 0% (0/9) dos artigos da NewsLab. Este item também estava entre os menos atendidos no trabalho de Siddiqui, Azuara-Blanco e Burr (2005) onde apenas 12,5% dos artigos analisados atenderam ao item e no estudo de Zafar, Khan e Siddiqui (2008) onde apenas 8% dos artigos avaliados descreveram métodos para análise da reprodutibilidade.

Outro item pobremente atendido pela maioria dos artigos analisados foi a descrição de qualquer evento adverso pela execução do teste avaliado ou do padrão de referência (item 20). Alguns estudos pontuam este item como “não aplicável”, quando a natureza do teste avaliado e do padrão de referência não é invasiva (JOHNSON; SIDDIQUI; AZUARA-BLANCO, 2007; PARANJOTHY; SHUNMUGAM; AZUARA-BLANCO, 2007; SHUNMUGAM; AZUARA-BLANCO, 2006), porém nosso estudo não se deteve em um teste diagnóstico específico sendo o item avaliado normalmente e demonstrando baixo percentual de atendimento. Este foi atendido em apenas 4% (2/53) dos artigos do periódico *Clinical Biochemistry*, 0% (0/11) no JBPML, 8% (1/13) na RBAC, 0% (0/15) na LAES&HAES e em 0% (0/9) dos artigos da NewsLab. A redação de todos os eventos adversos nos estudos de acurácia diagnóstica pode fornecer informação adicional sobre a utilidade clínica de um teste, pois nem todos os testes são seguros. É importante também saber se o padrão de referência é invasivo ou apresenta riscos aos pacientes submetidos a ele (STARD STATEMENT, 2008).

Apesar de recomendado aos autores que apresentem todas as medidas de reprodutibilidade do teste que foram realizadas durante o estudo, a reprodutibilidade não foi descrita pela maioria dos artigos analisados. A omissão da precisão dos índices de acurácia pode fazer uma diferença considerável na interpretação dos resultados de um estudo. (HARPER; REEVES, 1999). Para métodos analíticos quantitativos deve-se apresentar o coeficiente de variação das concentrações relevantes para o estudo, o número de determinações ou o número de dias de teste para o teste avaliado e do padrão de referência (item 24) (STARD STATEMENT, 2008). Os itens 24a e 24b foram atendidos em apenas 47% (25/53) e 13% (7/53) dos artigos do periódico *Clinical Biochemistry*, 0% (0/11) e 0% (0/11) no JBPML, 0% (0/13) e 0% (0/13) na RBAC, 27% (4/15) e 0% (0/15) na LAES&HAES e por 22% (2/9) e 0% (0/9) dos artigos da NewsLab. Os artigos avaliados por Lumbreras-Lacarra, Ramos-Rincón e Hernández-Aguado (2004) apresentaram a avaliação da

reprodutibilidade com maior frequência (83% em 1996, 81% em 2001 e 68% em 2002) quando comparados aos nossos resultados. Apesar de utilizarem os padrões metodológicos utilizados por Reid, Lachs e Feinstein (1995) o objetivo dos itens é comparável. Os mesmos autores pontuam o elevado percentual de descrição do item ao perfil dos especialistas em testes diagnósticos laboratoriais que geralmente dedicam maior atenção à imprecisão analítica, aspecto não observado na maioria dos artigos analisados no presente estudo.

A falha na redação de alguns itens do STARD e QUADAS pode confundir os leitores, mas não necessariamente invalidar a evidência contida no artigo. Contudo, a falta de mascaramento dos interpretadores dos testes, a presença de viés de verificação, a ausência da descrição de resultados indeterminados, ou a inclusão de pacientes em fase avançada da condição-alvo, quase sempre irá superestimar o valor diagnóstico de um teste (SHUNMUGAM; AZUARA-BLANCO, 2006). De acordo com Lijmer *et al.* (1999), a descrição incompleta da população do estudo ou dos métodos dos estudos de acurácia diagnóstica podem aumentar o valor diagnóstico do teste em 40% e 70%, respectivamente, quando comparados com artigos com detalhes suficientes em sua redação.

Quanto aos itens 8, 9, 10, 11, 13 e 24 que são relacionados tanto ao teste avaliado quanto ao padrão de referência, foi observada uma melhor redação dos itens referentes ao teste avaliado quando comparados ao padrão de referência. Esta característica foi observada por outros autores (COPPUS *et al.*, 2006; RAMA; POOVALI; APSINGI, 2006) e aventamos que possa estar relacionada a um maior enfoque proposital nas características do teste avaliado, na maior parte das vezes menos conhecido e mais recente que o padrão de referência para o diagnóstico, a um maior conhecimento do padrão de referência por parte dos profissionais e portanto relegado a um segundo plano na redação do artigo ou ao espaço exíguo dedicado aos artigos nas publicações, obrigando os autores a optar em descrever em maiores detalhes o teste sob avaliação.

A diferença máxima de média entre os periódicos analisados de aproximadamente 20% ou 5 itens, não evidencia com clareza a diferença na qualidade da redação, mas esta se torna clara durante o exercício de leitura e avaliação entre os periódicos. Apesar de uma análise mais detalhada da qualidade de redação estar além do escopo de nosso trabalho, esta engloba características adicionais como clareza na linguagem utilizada, ausência de ambigüidade e consistência entre o resumo, métodos, resultados e discussão. Além disso, 56% dos artigos analisado no periódico NewsLab e 60% no periódico LAES&HAES não haviam sido originalmente publicados nesses periódicos, por essa razão problemas de tradução e falta de clareza estão entre os maiores problemas enfrentados na avaliação da qualidade de redação

destes artigos. Entre os artigos não-originais, apesar dos problemas de tradução, alguns artigos atingiram percentuais elevados de atendimento aos itens como é o caso de Shirit *et al.* (2005) com 19,5 itens atendidos, o que demonstra que alguns aspectos da qualidade da redação não são avaliados pelo STARD.

Além disso, os periódicos LAES&HAES e NewsLab apresentaram os menores percentuais de artigos (33% e 11% respectivamente) descrevendo a população estudada (item 3), o que dificulta aos leitores analisarem se os pacientes estudados são similares aos que seriam atendidos no cenário clínico de interesse e, portanto, a generalização dos resultados (SIMEL; RENNIE; BOSSUYT, 2008). Quanto à descrição da amostragem dos participantes (item 5) apenas 55% dos artigos no periódico *Clinical Biochemistry*, 27% no JBPML, 20% na LAES&HAES, 0% na NewsLab deixaram explícito se os participantes foram arrolados no estudo de maneira consecutiva ou não, pois teoricamente a qualidade metodológica é maior nos estudos que arrolam pacientes de maneira consecutiva do que naqueles que incluem pacientes de maneira prospectiva não-consecutiva (SIMEL; RENNIE; BOSSUYT, 2008). É também importante para que os leitores conheçam o esquema de amostragem, pois auxilia no julgamento da generalização (validade externa) dos achados do estudo (BOSSUYT *et al.*, 2003a).

A descrição adequada das características clínicas e demográficas dos participantes (item 15) permite ao leitor julgar a possibilidade de viés na seleção dos participantes e da aplicabilidade dos resultados em outros cenários clínicos, dificultando, portanto a generalização dos resultados (PARANJOTHY; SHUNMUGAM; AZUARA-BLANCO, 2007). O item foi adequadamente descrito pela maioria dos artigos do periódico *Clinical Biochemistry* (94%), mas em apenas 36% (4/11) no JBPML, 38% (5/13) na RBAC, 7% (1/15) na LAES&HAES e por 11% (1/9) dos artigos da NewsLab.

De modo geral, os periódicos analisados em âmbito internacional são indexados nas principais bases de dados referenciais do mundo, MEDLINE, EMBASE (Excerpta Medica Database) e BIOSIS (Biological Abstracts-Current Contents) (DANTAS, 2004) e possuem índices bibliográficos elevados. Nossos resultados aparentemente não conseguem discriminar entre a qualidade da redação destes periódicos internacionais e periódicos nacionais nem ao menos indexados como é o caso da revista NewsLab e LAES&HAES. Concordamos com Wilczynski (2008) ao afirmar que muitos itens do STARD são abertos à interpretação e multifacetados. Por essa razão, o processo de avaliação para cada item pode variar entre os estudos, levando a diferenças na frequência do atendimento dos itens. Algumas diferenças entre as médias de itens encontradas neste estudo e nos demais podem estar relacionadas a

diferentes formas de pontuação dos itens do *checklist* STARD por alguns autores (JOHNSON, SIDDIQUI; AZUARA-BLANCO, 2006; SHUNMUGAM; AZUARA-BLANCO, 2006; SIDDIQUI; AZUARA-BLANCO; BURR, 2005; ZAFAR, KHAN; SIDDIQUI, 2008; LÉGARÉ *et al.*, 2007). Enquanto em nosso estudo, os itens do STARD foram avaliados como atendidos ou não atendidos, nos estudos citados anteriormente, houve três possibilidades de caracterização dos itens: completamente, parcialmente e não atendidos. O fato de não pontuarmos os itens como parcialmente atendidos, pode ter gerado médias mais elevadas e, por conseguinte a aparente discrepância de médias apresentadas em nosso trabalho. Por exemplo, o item 3 no trabalho de Shunmugam e Azuara-Blanco (2006) foi pontuado como parcialmente atendido se ao menos um aspecto da população estudada houvesse sido descrita (cenário clínico, local, critérios de inclusão e critérios de exclusão) enquanto em nosso trabalho uma descrição parcial (ao menos 2 aspectos) foi pontuada como atendimento ao item. Apesar de alguns artigos analisados não mencionarem se o desenho do estudo foi prospectivo ou retrospectivo (item 6), estes mencionavam que os pacientes consentiram na realização do estudo, sendo portanto, o item pontuado como atendido. O item 21 também pode ter influenciado os resultados, pois a descrição das estimativas de acurácia diagnóstica sem a imprecisão (por exemplo, intervalo de confiança) poderia ter sido pontuada como parcial, enquanto em nosso estudo, apenas a apresentação dos índices foi pontuada como atendimento ao item. Outra possível causa para essa diferença pode ser a avaliação de áreas do conhecimento distintas, com suas peculiaridades.

A utilização do STARD e QUADAS de modo conjunto, como em nosso trabalho, foi realizada em revisões sistemáticas, onde coube ao STARD a qualidade de redação e ao QUADAS a avaliação da qualidade dos artigos incluídos nestas revisões (LÉGARÉ *et al.*, 2007; ROPOSCH *et al.*, 2006; STENGEL *et al.*, 2005; TRIJFFEL *et al.*, 2005). Apesar de o objetivo do QUADAS ir além da qualidade de redação dos artigos de acurácia diagnóstica, diversos itens possuem correspondentes no *checklist* STARD, por exemplo, ambos os instrumentos se referem ao espectro dos pacientes, a ocultar os resultados do padrão de referência do responsável pela interpretação do teste avaliado e vice-versa e à recomendação da apresentação de um diagrama de fluxo permitindo a transparência quanto à perdas de participantes durante o estudo o que pode influenciar os resultados do mesmo. De forma semelhante à aplicação do *checklist* STARD, a aplicação do QUADAS resultou em uma qualidade em geral mediana considerando o alcance de suas respostas “sim”. A média de itens do QUADAS foi  $9,0 \pm 1,8$  ( $64,3\% \pm 12,6$ ) para a revista *Clinical Biochemistry*,  $8,1 \pm 1,4$

(57,8 ± 10,3) para o JBPML, 7,1 ± 2,7 (50,5% ± 19,2) para a RBAC, 6,3 ± 1,8 (45,2% ± 12,8) para a revista LAES&HAES e 7,3 ± 2,0 (52,4% ± 14,3) para a revista NewsLab.

Coerente com os itens menos atendidos no *checklist* STARD, as questões com menos respostas positivas na aplicação do instrumento QUADAS foram as questões relativas à descrição dos critérios de seleção dos pacientes e à avaliação cegada dos leitores dos testes para os resultados do teste em avaliação e padrão de referência (itens 2, 10, 11 e 12). As falhas foram mais pronunciadas nas publicações nacionais e menos de 40% de todos os artigos analisados evitaram vieses de revisão, corroborando com a pouca descrição de avaliação cegada dos interpretadores através da avaliação pelo roteiro STARD.

Apesar da recomendação pela iniciativa STARD e QUADAS da apresentação de um diagrama de fluxo, este foi apresentado em apenas um estudo no periódico *Clinical Biochemistry*. Os periódicos NewsLab e LAES&HAES apresentaram da mesma forma, apenas um artigo contendo o diagrama, tratando-se de um mesmo artigo não-original publicado em ambas revistas. Sua importância está relacionada à apresentação do teste avaliado e do padrão de referência utilizados, a seqüência em que foram realizados e o número de pacientes que foram submetidos a cada teste, facilitando aos revisores a identificação destes itens na leitura do artigo. Sua apresentação também está relacionada à melhora na qualidade de redação dos ensaios clínicos controlados e randomizados (EGGER; JUNI; BARTLETT, 2001). Apesar destas evidências, nenhum diagrama foi apresentado nos 30 artigos avaliados por Johnson, Siddiqui e Azuara-Blanco (2007) na área da oftalmologia. De forma semelhante, apenas 2 artigos em uma série de 54 artigos avaliados pelos STARD e QUADAS apresentaram um diagrama de fluxo no trabalho de Mann, Hewitt e Gilbody (2008) analisando a qualidade dos estudos diagnósticos utilizando instrumentos psicométricos e somente 8% (6/76) dos estudos avaliados por Zafar, Khan e Siddiqui (2008) apresentaram um diagrama de fluxo.

Devido ao número reduzido de artigos de acurácia diagnóstica nos periódicos nacionais não foi realizada comparação entre a qualidade de redação dos diferentes periódicos. O periódico *Clinical Biochemistry* obteve a maior média na aplicação do STARD, seguido pelo JBPML, RBAC, NewsLab e LAES&HAES. Na aplicação do QUADAS a ordem das médias mudou apenas entre a revista NewsLab e a RBAC. Acreditamos que a qualidade de redação possa ser influenciada pela indexação dos periódicos e seus índices bibliográficos como fator de impacto, SJR e Índice H (Tabela 6). Entre os periódicos analisados, o de maior indexação é o *Clinical Biochemistry* (MEDLINE), seguido pelo JBPML (SciELO e LILACS), RBCF (SciELO) e RBAC (LILACS). Quanto ao fator de

impacto, SJR e Índice H, os únicos avaliados foram o periódico *Clinical Biochemistry* (F.I. 2,072; SJR 0,303; IH 43), o JBPM (F.I. 0,0593; SJR 0,037; IH 1) e o RBCF (SJR 0,046; IH 5). As médias de itens atendidos pelo STARD seguem este padrão em nosso estudo e os artigos com menor qualidade de redação pertencem às duas revistas analisadas não-indexadas LAES&HAES e NewsLab. Apesar da indexação e dos índices bibliográficos, a RBCF não pôde ser avaliada devido à ausência de artigos de acurácia diagnóstica no período analisado. Dentre os periódicos analisados, o *Clinical Biochemistry* foi o único a apoiar a iniciativa STARD em seu ano de publicação através de um editorial (MCQUEEN, 2003) e de um artigo de explanação sobre a iniciativa (BOSSUYT *et al.*, 2003b). Mesmo que o *checklist* STARD não faça parte das instruções aos autores neste periódico, o conhecimento de sua existência e do guia de como aplicar às questões aos manuscritos antes de sua submissão pode ser um fator suficientemente forte para influenciar a média de itens atendidos. O estudo publicado por Smidt *et al* (2005), analisando periódicos médicos internacionais de fator de impacto 4 ou superior não evidencia relação entre a média de itens do STARD atendidos e o respectivo fator de impacto da publicação. Por exemplo, periódicos voltados à áreas específicas do conhecimento como o periódico *Radiology* de fator de impacto 4,1 apresentou média de  $13,2 \pm 2,3$  itens atendidos, enquanto o periódico *Circulation* de fator de impacto 10,9 apresentou média de  $10,3 \pm 3,6$  itens atendidos. Em medicina geral o *New England Journal of Medicine* de fator de impacto 29,5 apresentou média de  $14,3 \pm 2,7$  itens enquanto o *JAMA* de fator de impacto 15,4 apresentou média de  $15,5 \pm 2,3$  itens. Apesar destas evidências, os artigos publicados avaliando a qualidade de redação analisam periódicos indexados, dificultando traçar um paralelo entre o nosso estudo e os demais.

Analisando mudanças nas proporções de vezes que os itens do roteiro STARD e do instrumento QUADAS foram atendidos através dos cinco anos analisados e, portanto se houve melhora no atendimento o teste de  $\chi^2$  Armitage para tendências lineares demonstrou tendência linear estatística significativa apenas para a Revista Brasileira de Análises Clínicas ( $P=0,0021$ ) e revista NewsLab ( $P=0,0324$ ) a partir dos itens do STARD e apenas na RBAC ( $P=0,0002$ ) a partir da avaliação dos itens do QUADAS. A avaliação da qualidade da redação dos estudos de acurácia diagnóstica com a utilização do STARD tem sido realizada por diversos autores em diferentes áreas do conhecimento, porém, poucos investigaram se houve melhora significativa na qualidade de redação desde a publicação do STARD. Smidt *et al* (2006b), analisando 265 artigos de acurácia diagnóstica em 12 periódicos científicos médicos internacionais (6 periódicos de clínica geral e 6 periódicos com foco em uma doença ou área específica do conhecimento) demonstraram estatisticamente uma pequena melhora, onde a



média de itens atendidos pelo STARD por artigo publicado foi de  $11,8 \pm 3,3$  (47%) em 2000 e  $13,6 \pm 3,2$  (54,4%) em 2004. No mesmo estudo, em 2004 se verificou nos periódicos analisados, que 64% dos artigos científicos demonstraram adesão a mais da metade dos 25 itens em comparação a 41% em 2000. Contudo, não houve diferenças entre os periódicos adotantes do STARD, ou seja, aqueles que mencionam a iniciativa STARD nas instruções aos autores e os não-adotantes.

Wilczynski (2008) avaliando a redação de 240 artigos de acurácia diagnóstica em 6 periódicos médicos internacionais que publicaram a declaração STARD em 2003 e outros seis que não o publicaram, criou grupos de artigos pré-STARD (publicados em 2001 e 2002) e pós-STARD (publicados em 2004 e 2005) não verificando diferença nem interação significativa entre estas características. Coppus *et al.* (2006) ao avaliar se a introdução do STARD levou a uma melhora na qualidade de redação em estudos de acurácia diagnóstica publicados em dois dos principais periódicos de medicina reprodutiva nos anos 1999 e 2004 não encontrou melhora significativa entre as médias de itens atendidos, sendo estas respectivamente  $12,1 \pm 3,3$  e  $12,4 \pm 3,2$ . Paranjothy, Shunmugan e Azuara-Blanco (2007) avaliando a acurácia da polarimetria de varredura a laser no diagnóstico de glaucoma, analisaram 20 artigos, sendo os dez primeiros no período de 1997 a 2000 e os últimos de 2004 a 2005 e demonstraram que não houve diferença quando comparadas as publicações do primeiro período e as mais recentes.

Uma possível explicação para a falta de melhora na maioria dos estudos após a publicação do *checklist* STARD é que as avaliações foram conduzidas demasiadamente cedo para que alguma melhora se tornasse evidente. Outro fato poderia estar ligado ao tempo de *peer review*, ou seja, o tempo envolvido entre a submissão de um manuscrito a um periódico e o aceite final para a publicação, não permitindo aos autores incorporar as diretrizes do STARD a seus trabalhos (COPPUS *et al.*; 2006). A utilização do CONSORT para a redação de ensaios clínicos controlados e randomizados evidenciou melhora na redação dos artigos em periódicos que adotaram o CONSORT após 2 anos de sua publicação (MOHER; JONES; LEPAGE, 2001; PLINT *et al.*, 2006). Entretanto, os artigos avaliados no trabalho conduzido por Wilczynski (2008) foram publicados 2 anos após a promoção da iniciativa STARD e não ficou demonstrada melhora significativa na redação. Talvez, como sugerido por alguns autores, os aspectos de delineamento dos ensaios terapêuticos fossem mais bem conhecidos e mais fáceis de aplicar na ocasião de publicação do CONSORT, sendo que este mesmo nível de conhecimento e facilidade não era evidente para os estudos de acurácia diagnóstica quando da publicação do STARD (REEVES, 2005; COPPUS *et al.*, 2006).

Outra possível explicação para a falta de evidências quanto à melhora na redação dos estudos de acurácia diagnóstica pode ser o pouco incentivo e cobrança da utilização do *checklist* por parte dos periódicos que publicaram a iniciativa STARD. Durante a revisão das instruções aos autores nos periódicos adotantes da iniciativa, é difícil determinar o nível de imposição deste tipo de diretriz aos autores ou se esta imposição tenha mudado após 2 anos da publicação do STARD. Autores que procuram diferenças na utilização do STARD entre publicações adotantes e não-adotantes explicam a inexistência desta diferença a uma noção geral dos princípios de delineamento em acurácia diagnóstica por parte dos autores e periódicos. Esta noção pode ter sido atingida pelo fato do STARD ter sido preparado a partir de 33 *checklists* previamente publicados, tornando as diretrizes de redação amplamente disponíveis para todos os autores antes e depois da publicação da iniciativa STARD (WILCZYNSKI, 2008). Assim como Smidt et al (2007), acreditamos que para ocorrer uma melhora significativa na qualidade de redação deve haver uma contínua atenção dos autores dos manuscritos e dos revisores responsáveis pelo processo de revisão por pares.

Nossos resultados constituem evidência de melhora na redação dos artigos de acurácia diagnóstica no periódico RBAC e NewsLab. Não podemos avaliar a influência da publicação da iniciativa STARD na melhoria de redação nesses periódicos, pois o mesmo não foi publicado nem consta nas instruções aos autores nesses periódicos. O crescimento marcante do sistema de pós-graduação na Área de Ciências Farmacêuticas no Brasil onde se registrou, na última década, uma expansão de 100% quanto ao número de programas oferecidos, bem como significativa expansão dos recursos humanos e materiais dos programas (SCHENKEL, 2006), pode ser um dos fatores relacionados à esta melhoria. O crescimento do número de programas é acompanhado de um aumento do número de publicações e este pode estar relacionado ao nível de complexidade dos trabalhos submetidos aos periódicos. Esta melhora pode ainda estar relacionada à indexação na base de dados LILACS do periódico RBAC, que favorece uma maior procura por parte dos autores para publicação ou ainda numa mudança quanto ao processo editorial das revistas, com o incentivo da utilização de diretrizes para tipos específicos de manuscritos científicos submetidos.

De modo geral, o instrumento QUADAS foi relativamente fácil de ser utilizado. Entretanto, sua aplicação foi prejudicada pelas inconsistências de redação encontradas, apesar do tempo decorrido desde a publicação da iniciativa STARD. Embora os autores de estudos de acurácia diagnóstica não considerem a aplicação posterior do QUADAS a seus artigos, todos os itens de ambos os instrumentos são relacionados. Por essa razão, a aderência aos

itens do STARD durante a redação do manuscrito submetido a uma publicação poderia auxiliar os revisores na aplicação posterior do QUADAS durante revisões sistemáticas.

A avaliação dos itens concernentes à validade interna e externa dos artigos demonstrou padrão semelhante à totalidade dos itens, ficando evidenciado um melhor atendimento dos itens pelo periódico internacional avaliado. Enquanto o STARD é uma ferramenta prospectiva utilizada para o desenvolvimento de um estudo bem delineado, o QUADAS é considerado um instrumento retrospectivo utilizado para analisar o rigor metodológico de um estudo que investiga a acurácia diagnóstica de um teste (COOK; CLELAND; HUIJBREGTS, 2007). Por essa razão, apesar de possuírem objetivos distintos, ambos QUADAS e STARD possuem itens em comum. Sendo assim, as principais falhas na validade interna e externa dos artigos refletem-se na aplicação do instrumento QUADAS evidenciados pelo elevado número de respostas “não” e “não claras” aos itens.

Quanto à correlação encontrada entre o número de itens atendidos em ambas as ferramentas, nosso trabalho corrobora com o pensamento de Smidt *et al* (2005) que acreditavam existir uma associação positiva entre a qualidade metodológica de um estudo e sua qualidade de redação, sendo mais fácil redigir um estudo bem realizado do que um estudo com pobre delineamento ou em que ocorreu um grande número de desvios em seu protocolo. Nesse último caso os autores podem ser menos propensos a relatar em detalhes os acontecimentos durante o estudo.

## 7. Conclusões

Apesar da importância dos estudos de acurácia diagnóstica para a avaliação e validação de novos testes diagnósticos, estes são pouco realizados, fato este, demonstrado pela escassez de artigos recuperados, especialmente nos periódicos nacionais.

A qualidade da redação dos estudos de acurácia diagnóstica a partir da análise dos resultados da aplicação do STARD e do QUADAS apresenta, em média, uma baixa qualidade de redação por ambos os instrumentos, sendo esta, mais pronunciada nos periódicos nacionais.

Alguns aspectos concernentes à clareza na linguagem não são avaliados de maneira adequada pelo roteiro STARD e pelo instrumento QUADAS, haja visto que textos com graves problemas de redação alcançaram um total de itens atendidos comparáveis a outros artigos que não apresentavam os mesmos problemas.

Verificou-se melhora na qualidade de redação apenas nos periódicos RBAC e NewsLab. Nos demais periódicos analisados, a qualidade de redação continua similar àquela observada antes das publicações do STARD e QUADAS no ano de 2003.

Apesar dos visíveis esforços, na tentativa de melhoria da redação dos estudos de acurácia diagnóstica, através da criação de diretrizes específicas como o STARD e do seu amplo conhecimento, ao menos em âmbito internacional, o não atendimento da totalidade dos itens por nenhum dos artigos avaliados e as falhas de redação e metodológicas continuamente repetidas, demonstra que estas diretrizes não são utilizadas de maneira correta nem na confecção de um manuscrito, nem no processo editorial de revisão por pares. O processo de melhoria da redação deve passar, obrigatoriamente, pelo conhecimento do roteiro STARD pelos autores que submetem seus manuscritos para publicação e pela orientação e cobrança de uso nas Instruções aos Autores por parte dos periódicos.

A avaliação dos instrumentos utilizados para a análise da qualidade de redação dos estudos no presente trabalho constitui uma perspectiva para futuros trabalhos na área, assim como a continuidade de avaliação da redação de outras publicações na área de Análises Clínicas.

## 8. Referências Bibliográficas

AGEWALL, S. Evaluation of point-of-care test systems using the new definition of myocardial infarction. **Clinical Biochemistry**, Montreal, v. 36, n. 1, p.27-30, 2003.

AGRESTI, A. **Categorical Data Analysis**. 2. ed. EUA: Wiley, 2002.

AMARAL, R. G. *et al.* Revisão rápida de esfregaços cervicais como método de garantia interna de qualidade. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 39, n. 2, p.151-155, 2003.

ATALAY, F. *et al.* Pleural adenosine deaminase in the separation of transudative and exudative pleural effusions. **Clinical Biochemistry**, Montreal, v. 38, n. 12, p.1066-1070, 2005.

BEGG, C. *et al.* Improving the quality of reporting of randomized controlled trials: the CONSORT statement. **Journal of the American Medical Association**, Chicago, v. 276, n. 8, p.637-639, 1996.

BENOS, D. J. *et al.* The ups and downs of peer review. **Advances in Physiology Education**, Bethesda, v. 31, p.145-152, 2006.

BERWANGER, O.; AVEZUM A. ; GUIMARÃES, H. P. Cardiologia Baseada em Evidências: onde buscar evidências. **Arquivos Brasileiros de Cardiologia**, São Paulo, v. 86, n. 1, p.56-60, 2006.

BIANCHI, R. A. *et al.* Utility of base deficit, lactic acid, microalbuminuria, and C-reactive protein in the early detection of complications in the immediate postoperative evolution. **Clinical Biochemistry**, Montreal, v. 37, n. 5, p.404-407, 2004.

BIBLIOTECA VIRTUAL EM SAÚDE (Brasil). **Lançamento do portal SJR enriquece o universo dos indicadores bibliométricos**. Disponível em: <[http://cvirtual-ccs.bvsalud.org/tiki-read\\_article.php?articleId=233](http://cvirtual-ccs.bvsalud.org/tiki-read_article.php?articleId=233)>. Acesso em: 04 out. 2008.

BLACKSELL, S. D. *et al.* A systematic review and meta-analysis of the diagnostic accuracy of rapid immunochromatographic assays for the detection of dengue virus IgM antibodies during acute infection. **Transactions Of The Royal Society Of Tropical Medicine And Hygiene**, London, v. 100, p.775-784, 2006.

BOLANDPARVAZ, S. *et al.* Urinary 5-hydroxy indole acetic acid as a test for a early diagnosis of acute appendicitis. **Clinical Biochemistry**, Montreal, v. 37, n. 11, p.985-989, 2004.

BOSSUYT, P *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. **Clinical Chemistry**, Washington, v. 49, n. 1, p.7-18, 2003a.

BOSSUYT, P. M. *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. **Clinical Biochemistry**, Montreal, v. 26, p.1-7, 2003b.

BOSSUYT, P. *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. **Clinical Chemistry**, Washington, v. 49, n. 1, p.1-6, 2003c.

BOSSUYT, P. M. Comparative accuracy: assessing new tests against existing diagnostics pathways. **British Medical Journals**, London, v. 332, p.1089-1092, 2006.

BOSSUYT, P. M. The quality of reporting in diagnostic test research: getting better, still not optimal. **Clinical Chemistry**, Washington, v. 50, n. 3, p.465-466, 2004.

BOTTINI, P. V. *et al.* Utilização da relação albumina/creatinina no diagnóstico de microalbuminúria. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 41, n. 2, p.99-103, 2005.

BRENNINKMEIJER, E. E. A. *et al.* Diagnostic criteria for atopic dermatitis: a systematic review. **British Journal Of Dermatology**, Nottingham, v. 158, p.754-765, 2007.

BRUDER, R. C. S. *et al.* A importância da contra-imunoeletoforese na detecção de antígenos nucleares extraíveis para o diagnóstico de doenças reumáticas sistêmicas. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 40, n. 1, p.15-19, 2004.

BRUNS, D (Ed.). The STARD initiative and the reporting of studies of diagnostic accuracy. **Clinical Chemistry**, Washington, v. 49, n. 1, p.19-20, 2003.

BRUNS, D. *et al.* Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. **Clinical Chemistry**, Washington, v. 46, n. 7, p.893-895, 2000.

BUHLING, K. J. *et al.* Comparison of point-of-care testing glucose meters with standard laboratory measurement of the 50g-glucose-challenge test (GCT) during pregnancy. **Clinical Biochemistry**, Montreal, v. 36, n. 5, p.333-337, 2003.

CAKIR, E. *et al.* Pleural fluid and serum procalcitonin as diagnostic tools in tuberculous pleurisy. **Clinical Biochemistry**, Montreal, v. 38, n. 3, p.234-238, 2005.

CAPES: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brasil). **Classificação de periódicos, anais, revistas e jornais.** Disponível em: <<http://qualis.capes.gov.br/webqualis/>>. Acesso em: 04 out. 2008.

CARVALHO, F. M. *et al.* Diagnóstico coproparasitológico: estudo comparativo entre os métodos de Faust & cols., Lutz, Baermann & Moraes e Coprotest. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 34, n. 2, p.75-77, 2002.

CARVALHO, P. G. *et al.* Avaliação do emprego de teste combinado antígeno/anticorpo na identificação de infecção recente de HIV-I/II em doadores de sangue. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 37, n. 3, p.169-174, 2005.

CASTRO, A. **Avaliação da qualidade da informação.** Disponível em: <<http://www.metodologia.org/livro/>>. Acesso em: 26 out. 2008.

CASTRO, A. P. W. *et al.* Urea breath test for the detection of *Helicobacter pylori* using a stable isotope. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 40, n. 2, p.63-67, 2004.

CHANG, J. T. *et al.* Increase diagnostic efficacy by combined use of fingerprint markers in mass spectrometry - Plasma peptidomes from nasopharyngeal cancer patients for example. **Clinical Biochemistry**, Montreal, v. 39, n. 12, p.1144-1151, 2006.

CHEN, C. *et al.* The value of biliary amylase and Hepatocarcinoma-Intestine-Pancreas/Pancreatitis-associated Protein I (HIP/PAP-I) in diagnosing biliary malignancies. **Clinical Biochemistry**, Montreal, v. 38, n. 6, p.520-525, 2005.

CHIAPPELLI, F. *et al.* Evidence-based practice in dentistry: benefit or hindrance. **Journal of Dental Research**, Alexandria, v. 82, n. 1, p.6-7, 2003.

COLLINSON, P.O. *et al.* Diagnostic and prognostic role of cardiac troponin I (cTnI) measured on the DPC Immulite. **Clinical Biochemistry**, Montreal, v. 39, n. 7, p.692-696, 2006.

CONCEIÇÃO, G. C. da *et al.* Avaliação do tubo germinativo em secreção vaginal a fresco para triagem de *Candida albicans*: Um teste rápido. **NewsLab**, São Paulo, v. 73, p.106-112, 2005.

COOK, C; CLELAND, J; HUIJBREGTS, P. Creations and critique of studies of diagnostic accuracy: use of the STARD and QUADAS methodological quality assessment tools. **The Journal Of Manual & Manipulative Therapy**, Oregon, v. 15, n. 2, p.93-102, 2007.

COPPUS, S. F. P. J. *et al.* Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. **Fertility And Sterility**, Birmingham, v. 86, p.1321-1329, 5 nov. 2006.

CRUZ, D. A. L. M.; PIMENTA, C. A. M. Prática baseada em evidências, aplicada ao raciocínio diagnóstico. **Revista Latino-Americana de Enfermagem**, São Paulo, v. 13, n. 3, p.415-422, 2005.

DANTAS, P. E. C. (Ed.). Indexação bibliográfica em bases de dados: O que é? Para que serve? Onde estamos? **Arquivos Brasileiros de Oftalmologia**, São Paulo, v. 67, n. 4, p.569-570, 2004.

DE BEM, A. F. *et al.* Valor diagnóstico da determinação da atividade da adenosina deaminase no derrame pleural tuberculoso em indivíduos infectados e não infectados pelo HIV. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 37, n. 1, p.23-26, 2005.

DELWICHE, F. A. Mapping the literature of clinical laboratory science. **Journal Of The Medical Library Association**, Chicago, v. 91, n. 3, p.303-310, 2003.

DEVILLÉ, W. L. *et al.* Conducting systematic reviews of diagnostic studies: didactic guidelines. **BMC Medical Research Methodology**, London, v. 9, n. 2, p.1-13, 2002.

DEVILLÉ, W. L. J. M.; BEZEMER, P. D.; BOUTER, L. M. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. **Journal Of Clinical Epidemiology**, Maastricht, v. 53, p.65-69, 2000.

DORIZZI, R. M. *et al.* Procalcitonin in the diagnosis of inflammation in intensive care units. **Clinical Biochemistry**, Montreal, v. 39, n. 12, p.1138-1143, 2006.

DOUST, J. A. *et al.* Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodological filters and the lack of information in the abstract. **Journal Of Clinical Epidemiology**, Maastricht, v. 58, p.444-449, 2005.



EGGERDING, V. C.; TRISCOTT, M. X.. Auto D-Dímero - um estudo retrospectivo avaliando os valores preditivos negativos, a sensibilidade e a especificidade negativa do teste. **LAES&HAES**, São Paulo, v. 146, p.158-168, 2003.

EGGERDING, V. C.; TRISCOTT, M. X. Auto D-Dímero - um estudo retrospectivo avaliando os valores preditivos negativos, a sensibilidade e a especificidade negativa do teste. **NewsLab**, São Paulo, v. 59, p.184-190, 2003.

EGGER, M.; JÜNI, P.; BARTLETT, C. Value of flow diagrams in reports of randomized controlled trials. **Journal of the American Medical Association**, Chicago, v. 285, n. 15, p.1996-1996, 2001.

EISSA, S. *et al.* Comparison of cytokeratin 20 RNA and angiogenin in voided urine samples as diagnostic tools for bladder carcinoma. **Clinical Biochemistry**, Montreal, v. 37, n. 9, p.803-810, 2004.

EISSA, S. *et al.* HER2/neu expression in bladder cancer: relationship to cell cycle kinetics. **Clinical Biochemistry**, Montreal, v. 38, n. 2, p.142-148, 2005.

EL-MASRY, S. *et al.* Serum laminin assayed by Slot-Blot-ELISA in patients with combined viral hepatitis C and schistosomiasis. **Clinical Biochemistry**, Montreal, v. 39, n. 6, p.652-657, 2006.

EL-SALAHY, E. M. Correlation between polyamines and apoptosis among Egyptian breast cancer patients. **Clinical Biochemistry**, Montreal, v. 35, n. 7, p.555-560, 2002a.

EL-SALAHY, E. M. Evaluation of cytokeratin-19 & cytokeratin-20 and interleukin-6 in Egyptian bladder cancer patients. **Clinical Biochemistry**, Montreal, v. 35, n. 8, p.607-613, 2002b.

ELSEVIER. **Clinical Biochemistry Description**. Disponível em: <[http://www.elsevier.com/wps/find/journaldescription.cws\\_home/525463/description](http://www.elsevier.com/wps/find/journaldescription.cws_home/525463/description)>. Acesso em: 17 out. 2008.

ERBAGCI, A. B. *et al.* Serum prolidase activity as a marker of osteoporosis in type 2 diabetes mellitus. **Clinical Biochemistry**, Montreal, v. 35, n. 4, p.263-268, 2002.

ESKICORAPCI, S. Y. *et al.* Serum ribonuclease activity in the diagnosis of prostate cancer in men with serum prostate-specific antigen levels between 2.5 and 20 ng/mL. **Clinical Biochemistry**, Montreal, v. 39, n. 4, p.363-366, 2006.

FABRIS, C. *et al.* Assessment of liver fibrosis progression in patients with chronic hepatitis C and normal alanine aminotransferase values: The role of AST to the platelet ratio index. **Clinical Biochemistry**, Montreal, v. 39, n. 4, p.339-343, 2006.

FISCHER, R. *et al.* Diagnóstico in vitro de segunda geração para detecção simultânea de norovírus do genótipo I e II. **LAES&HAES**, São Paulo, v. 164, p.148-154, 2006.

FLETCHER, R. H.; FLETCHER, S. W. **Epidemiologia clínica: Elementos essenciais**. 4. ed. Porto Alegre: Artmed, 2006. 288 p.

FLORES, R. E. A medicina baseada em evidências e o diagnóstico laboratorial. **NewsLab**, São Paulo, v. 73, n. , p.92-103, 2005.

GLAS, A. *et al.* The Diagnostic odds ratio: a single indicator of test performance. **Journal Of Clinical Epidemiology**, Maastricht, v. 56, p.1129-1135, 2003.

GONZÁLEZ, C. *et al.* Antinuclear antibodies (ANA) screening by enzyme immunoassay with nuclear HEP-2 cell extract and recombinant antigens: analytical and clinical evaluation. **Clinical Biochemistry**, Montreal, v. 35, n. 6, p.463-469, 2002.

GONZÁLEZ, C. *et al.* Clinical evaluation of a microsphere bead-based flow cytometry assay for the simultaneous determination of anti-thyroid peroxidase and anti-thyroglobulin antibodies. **Clinical Biochemistry**, Montreal, v. 38, n. 11, p.966-972, 2005.

GOODMAN, S. N. *et al.* Manuscript quality before and after peer review and editing at Annals of Internal Medicine. **Annals Of Internal Medicine**, Philadelphia, v. 121, n. 1, p.11-21, 1 jul. 1994.

GUGELMIN, E. S.; TORRES, L. F. B. The accuracy of acetylcholinesterase reaction in rectal suction biopsy in the diagnosis of Hirschsprungs disease. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 41, n. 6, p.431-436, 2005.

GUILLOUX, L.; HAMBERGER, C.. Acesso ao ensaio de IgE específico utilizando o Immutable 2000 da DPC - Diagnostic Products Corporation. **LAES&HAES**, São Paulo, v. 155, p.130-146, 2005.

GÜLTEPE, M. *et al.* Urine methylmalonic acid measurements for the assessment of cobalamin deficiency related to neuropsychiatric disorders. **Clinical Biochemistry**, Montreal, v. 36, n. 4, p.275-282, 2003.

GUYATT, G.; COOK, D.; HAYNES, B. Evidence based medicine has come a long way. **British Medical Journals**, London, v. 329, p.990-991, 2004.

HARPER, R.; REEVES, B. Reporting of precision of estimates for diagnostic accuracy: a review. **BMJ**, London, v. 318, p.1322-1323, 1999.

HATJIMIHAİL, A. T. Quality of diagnostic accuracy studies: the development, use and evaluation of QUADAS. **Evidence Based Medicine**, London, v. 11, p.189, 2006.

HENSCHKE, N.; MAHER, C. G.; REFSHAUGE, K. M. A systematic review identifies five "red flags" to screen for vertebral fracture in patients with low back pain. **Journal Of Clinical Epidemiology**, Maastricht, v. 61, p.110-118, 2008.

HIRSCH, J. E. An index to quantify an individuals scientific research output. **Physics and Society**, Ithaca, v. 5, n. 25, p.1-5, 29 set. 2005.

HONEST, H.; KHAN, K. S. Reporting of measures of accuracy in systematic reviews of diagnostic literature. **BMC Health Services Research**, London, p.1-04, 2002.

HOSHINO-SHIMIZÛ, S. *et al.* Enteroparasitoses: Inovação tecnológica do kit TF-Test para exames parasitológicos. **NewsLab**, São Paulo, v. 55, p.106-108, 2002.

HUTCHINSON, L. M. *et al.* Development of a sensitive and specific enzyme-linked immunosorbent assay for thymosin beta 15, a urinary biomarker of human prostate cancer. **Clinical Biochemistry**, Montreal, v. 38, n. 6, p.558-571, 2005.

INTERNATIONAL COMMITTEE OF MEDICAL JOURNAL EDITORS. Uniform requirements for manuscripts submitted to biomedical journals. **Journal of the American Medical Association**, Chicago. v. 277, p.927-34, 1997.

JOHNSON, Z. K; SIDDIQUI, R; AZUARA-BLANCO, A. The quality of reporting of diagnostic accuracy studies of optical coherence tomography in glaucoma. **Ophthalmology**, Baltimore, v. 114, p.1607-1612, 2007.

KADOKAWA, Y. *et al.* Evaluation of newly developed ELISA using "MESACUP-2 Test Mitochondrial M2" kit for the diagnosis of primary biliary cirrhosis. **Clinical Biochemistry**, Montreal, v. 36, n. 3, p.203-210, 2003.

KASSIM, S. K. *et al.* Vascular endothelial growth factor and interleukin-8 are associated with poor prognosis in epithelial ovarian cancer patients. **Clinical Biochemistry**, Montreal, v. 37, n. 5, p.363-369, 2004.

KIR, Z. Ö. *et al.* Serum prolidase I activity and some bone metabolic markers in patients with breast cancer: in relation to menopausal status. **Clinical Biochemistry**, Montreal, v. 36, n. 4, p.289-294, 2003.

KNOTTNERUS, J. A.; MURIS, J. W. Assessment of the accuracy of diagnostic tests: the cross-sectional study. **Journal Of Clinical Epidemiology**, Maastricht, v. 56, p.1118-1128, 2003.

KOÇAK, H. *et al.* Determination of diagnostic and prognostic values of urinary interleukin-8, tumor necrosis factor- $\alpha$ , and leukocyte arylsulfatase-A activity in patients with bladder cancer. **Clinical Biochemistry**, Montreal, v. 37, n. 8, p.673-678, 2004.

KRAJDEN, M. *et al.* Detecção qualitativa do RNA do vírus da hepatite C: uma comparação da sensibilidade analítica, do desempenho clínico e do fluxo de trabalho do ensaio HCV Amplicor Cobas, versão 2.0 e ensaio qualitativo de amplificação de transcrição mediada do RNA do HCV. **LAES&HAES**, São Paulo, v. 149, p.162-180, 2004.

KUOKKANEN, M. *et al.* Um teste rápido para o diagnóstico endoscópico de hipolactasia de biópsia pós-bulbar duodenal. **NewsLab**, São Paulo, v. 78, p.150-156, 2006.

LAES&HAES (São Paulo). **O que é a revista LAES&HAES?** Disponível em: <<http://www.laes-haes.com.br/>>. Acesso em: 17 out. 2008.

LANDIS, J. R; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, Arlington, v. 33, p.159-174. 1977.

LANDKROON, A P *et al.* Quality assessment of reviewers reports using a simple instrument. **Obstetrics & Gynecology**, Baltimore, v. 108, n. 4, p.979-985, 2006.

LANGLOTS, C. P. Fundamental Measures of Diagnostic Examination Performance: Usefulness for Clinical Decision Making and Research. **Radiology**, Boston, v. 228, p.3-9, 2003.

LEDLEY, R. S.; LUSTED, L. B. Reasoning foundations of medical diagnosis. **Science**, Washington, v. 130, n. 3366, p.9-21, 1959.

LIJMER, J. *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. **Journal of the American Medical Association**, Chicago, v. 282, n. 11, p.1061-1066, 15 set. 1999.

LEEFLANG, M. M. G. *et al.* Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. **Journal Of Clinical Epidemiology**, Maastricht, v. 59, p.234-240, 2006.

LÉGARÉ, F. *et al.* Instruments to assess the perception of physicians in the decision-making process of specific clinical encounters: a systematic review. **BMC Medical Informatics And Decision Making**, London, v. 30, n. 7, p.1-16, 2007.

LIMA, A. J. A. *et al.* Diagnóstico laboratorial das micobactérias isoladas na urina: comparação entre métodos convencionais e a reação em cadeia da polimerase (PCR). **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 38, n. 3, p.141-145, 2006.

LIMA, G.; GROTTTO, H. Laboratory assessment of iron status and reticulocyte parameters in differential diagnosis of iron deficiency anemia and heterozygous  $\beta$ -thalassemia. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 38, n. 4, p.273-280, 2002.

LOBO, T.; FEIJÓ, G.; SIMÕES-BARBOSA, A. Relato de um estudo comparativo de diagnóstico de tricomonose no Distrito Federal. **NewsLab**, São Paulo, p.156-162, 2002.

LUMBRERAS, B. *et al.* QUADOMICS: an adaptation of the Quality Assessment of Diagnostic Accuracy Assessment (QUADAS) for the evaluation of the methodological quality studies on the diagnostic accuracy of “-omics” - based technologies. **Clinical Biochemistry**, Montreal, v. 41, p.1316-1325, 2008.

LUMBRERAS, B; JARRÍN, I.; AGUADO, I. H. Evaluation of the research methodology in genetic, molecular and proteomics tests. **Gaceta Sanitaria**, Barcelona, v. 20, n. 5, p.368-373, 2006.

LUMBRERAS-LACARRA, B.; RAMOS-RINCON, J.; HERNANDEZ-AGUADO, I. Methodology in diagnostic laboratory test research in Clinical Chemistry and Clinical Chemistry and Laboratory Medicine. **Clinical Chemistry**, Washington, v. 50, n. 3, p.530-536, 2004.

LY, T. D. *et al.* Avaliação da sensibilidade e especificidade de seis ensaios HIV combinados antígeno p24 e anticorpos. **LAES&HAES**, São Paulo, v. 159, p.134-156, 2006.

LYON, M. E. *et al.* A preliminary evaluation of the interaction between urine specific gravity and leukocyte esterase results using Bayer Multistix and the Clinitek 500. **Clinical Biochemistry**, Montreal, v. 36, n. 7, p.579-582, 2003.

MABROUK, R. A. W. R.; ALI-LABIB, R. Detection of urokinase plasminogen activator receptor and c-erbB-2 in sera of patients with breast and ovarian carcinoma. **Clinical Biochemistry**, Montreal, v. 36, n. 7, p.537-543, 2003.

MACEDO, H. W. *et al.* Avaliação de testes imunológicos para o diagnóstico da neurocisticercose. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 38, n. 2, p.93-103, 2002.

MAHMOUD, R. A. K.; EL-GENDI, H. I.; AHMED, H. H.. Serum neopterin, tumor necrosis factor-alfa and soluble tumor necrosis factor receptor II (p75) levels and disease activity in Egyptian female patients with systemic lupus erythematosus. **Clinical Biochemistry**, Montreal, v. 38, n. 2, p.134-141, 2005.

MAHONEY, A.; ELLISON, J. Assessing the quality of glucose monitor studies: a critical evaluation of published reports. **Clinical Chemistry**, Washington, v. 53, n. 6, p.1122-1128, 2007.

MANN, R.; HEWITT, C. E.; GILBODY, S. M. Assessing the quality of diagnostic studies using psychometric instruments: applying QUADAS. **Social Psychiatry and Psychiatric Epidemiology**, London, p.1-8, 2008.

MARINHO, V. F. Z. *et al.* Micrometástases de carcinoma da mama em linfonodos axilares: detecção por imunoistoquímica versus hematoxilina e eosina. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 40, n. 2, p.127-132, 2004.

MARSILLACH, J. *et al.* The role of circulating monocyte chemoattractant protein-I as a marker of hepatic inflammation in patients with chronic liver disease. **Clinical Biochemistry**, Montreal, v. 38, n. 12, p.1138-1140, 2005.

MARTINS, R. *et al.* Avaliação de testes de diagnósticos rápidos na detecção de infecção assintomática do líquido amniótico em mulheres gestantes. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 38, n. 3, p.151-154, 2006.

MAUÉS, F. C. J. *et al.* Estudo comparativo dos resultados obtidos pelos métodos de imunocromatografia e o ELISA para detecção de anticorpos IgM e IgG anti-vírus dengue. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 35, n. 1, p.5-9, 2003.

MCQUEEN, M. J. (Ed.). Evidence-based laboratory medicine: addressing bias, generalisability and applicability in studies on diagnostic accuracy: The STARD initiative. **Clinical Biochemistry**, Montreal, v. 36, n. , p.1-7, 2003.

MCQUEEN, M. J. Overview of evidence-based medicine: challenges for evidence-based laboratory medicine. **Clinical Chemistry**, Boston, v. 47, n. 8, p.1536-1546, 2001.

MOHAMMADZADEH, G. S. *et al.* Measurement of glutathione S-transferase and its class-pi in plasma and tissue biopsies obtained after laparoscopy and endoscopy from subjects with esophagus and gastric cancer. **Clinical Biochemistry**, Montreal, v. 36, n. 4, p.283-288, 2003.

MOHER, D. *et al.* Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. **The Lancet**, New York, v. 354, p.1896-1900, 1999.

MOHER, D.; JONES, A.; LEPAGE, L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. **Journal of the American Medical Association**, Chicago, v. 285, n. 15, p.1992-1995, 2001.

MOLINARI, I. *et al.* Fecal chymotrypsin and elastase-1 determination on one single stool collected at random: diagnostic value for exocrine pancreatic status. **Clinical Biochemistry**, Montreal, v. 37, n. 9, p.758-763, 2004.

MOTTA, L. R. *et al.* Avaliação do InstantScreen na rotina, um ensaio ultra-rápido para detecção de anticorpos anti-HIV - 1/2. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 36, n. 2, p.71-72, 2003.

MOURA, L. R. Informação: a essência da qualidade. **Ciência da Informação**, Brasília, v. 25, n. 1, p.1-10, 1996.

MUNRO, W; HEALY, R. The validity and accuracy of clinical tests used to detect labral pathology of the shoulder: A systematic review. **Manual Therapy**, Kidlington, p.1-12, 2008.

NEHMY, R. M. Q; PAIM, I. A desconstrução do conceito de "qualidade da informação". **Ciência da Informação**, Brasília, v. 27, n. 1, p.36-45, 1998.

NEVES, D. A. Ciência da informação e cognição humana: uma abordagem do processamento da informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p.39-44, 2006.

NEWSLAB (São Paulo). **Sobre a NewsLab**. Disponível em: <<http://www.newslab.com.br/newslab/sobre.htm>>. Acesso em: 17 out. 2008.

OLIVEIRA, E. L. *et al.* Avaliação do índice de concordância entre dois ensaios imunoenzimáticos (ELISA) para detecção de anticorpos anti-Toxicara canis. **LAES&HAES**, São Paulo, v. 163, p.182-186, 2006.

OLIVEIRA, M. R. A. A.; CASTRO, M. G.; LANGHI, J. D. M.. Validação de testes sorológicos anti-HBc e anti-HIV para triagem de doadores de sangue. **LAES&HAES**, São Paulo, v. 149, p.148-155, 2004.

OXMAN, A.; SACKETT, D.; GUYATT, G. User's guide to the medical literature. **Journal of the American Medical Association**, Chicago, v. 270, n. 17, p.2093-2095, 1993.

PAIM, I.; NEHMY, R. M. Q; GUIMARÃES, C. G. Problematização do conceito "qualidade da informação". **Perspectivas da Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p.111-119, 1996.

PAL, S. *et al.* Antibodies against 9-o-acetylated sialoglycans: a potent marker to monitor clinical status in childhood acute lymphoblastic leukemia. **Clinical Biochemistry**, Montreal, v. 37, n. 5, p.395-403, 2004.

PANAGEAS, K. S. *et al.* Analysis of Multiple 2x2 Tables with Site-Specific Periodontal Data. **Journal of Dental Research**, Eua, v. 82, n. 7, p.514-517, 2003.

PARANJOTHY, B.; SHUNMUGAM, M.; AZUARA-BLANCO, A. The quality of reporting of diagnostic accuracy studies in glaucoma using scanning laser polarimetry. **Journal of Glaucoma**, Portland, v. 16, n. 8, p.670-675, 2007.

PENTTILÄ, K. *et al.* Myoglobin, creatine kinase MB isoforms and creatine kinase MB mass in early diagnosis of myocardial infarction in patients with acute chest pain. **Clinical Biochemistry**, Montreal, v. 35, n. 8, p.647-653, 2002.

PEREIRA, A. S. *et al.* Avaliação da acurácia de testes laboratoriais para detecção de amostras de *Klebsiella pneumoniae* produtora de betalactamase de espectro estendido. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 39, n. 4, p.301-308, 2003.



PEREIRA, M. G. **Epidemiologia: Teoria e Prática**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2006.

PETERSEN, A. B. *et al.* Performance evaluation of a specific IgE assay developed for the ADVIA centaur immunoassay system. **Clinical Biochemistry**, Montreal, v. 37, n. 10, p.882-892, 2004.

PINHO, A. A.; MATTOS, M. C. F. I. Validade da citologia cervicovaginal na detecção de lesões pré-neoplásicas e neoplásicas de colo de útero. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 38, n. 3, p.225-231, 2002.

PLINT, A. C. *et al.* Does the CONSORT checklist improve the quality of reports of randomized controlled trials? A systematic review. **Medical Journal Of Australia**, Sydney, v. 185, n. 4, p.263-267, 2006.

PORCEL, J. M. *et al.* Soluble oncoprotein 185 HER-2 in pleural fluid has limited usefulness for the diagnostic evaluation of malignant effusions. **Clinical Biochemistry**, Montreal, v. 38, n. 11, p.1031-1033, 2005.

PRICE, C. Evidence-based laboratory medicine: supporting decision-making. **Clinical Chemistry**, Washington, v. 46, n. 8, p.1041-1050, 2000.

QIU, G.; FANG, J.; HE, Y. 5 CpG island methylation analysis identifies the MAGE-A1 and MAGE-A3 genes as potential markers of HCC. **Clinical Biochemistry**, Montreal, v. 39, n. 3, p.259-266, 2006.

RAMA, K. R. B. S.; POOVALI, S.; APSINGI, S. Quality of reporting of orthopaedic diagnostic accuracy studies is suboptimal. **Clinical Orthopaedics And Related Research**, New York, v. 447, n. , p.237-246, 2006.

RAMOS, T. Z.; PIZZOLITTO, E. L.; PIZZOLITTO, A. C. Uso do teste com cloridrato de trifetil tetrazólio (CTT) para detecção de bacteriúria sintomática e assintomática. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 38, n. 3, p.197-199, 2006.

RAO, A. A.. Eficiência em testes de alergia in vitro no Kuwait utilizando os painéis de alergia Immulite. **NewsLab**, São Paulo, v. 52, p.170-174, 2002.

REEVES, B. C. Evidence about evidence. **The British Journal Of Ophthalmology**, London, v. 89, n. 3, p.253-254, 2005.

REID, M.; LACHS, M.; FEINSTEIN, A. Use of methodological standards in diagnostic test research: getting better but still not good. **Journal Of The American Medical Association**, Chicago, v. 274, n. 8, p.645-651, 23 ago. 1995.

REY, E. *et al.* Fasting plasma glucose versus glucose challenge test: screening for gestational diabetes and cost effectiveness. **Clinical Biochemistry**, Montreal, v. 37, n. 9, p.780-784, 2004.

ROPOSCH, A. *et al.* Development dysplasia of the HIP: quality of reporting of diagnostic accuracy for US. **Radiology**, Boston, v. 241, n. 3, p.854-860, 2006.

ROSSATTO, E. R.; SILVA, A. S. Avaliação Comparativa dos testes de PSA Total, PSA Livre/PSA Total, PSA Complexado e PSA Complexado/PSA Total no diagnóstico do câncer de próstata. **NewsLab**, São Paulo, v. 62, p.132-142, 2004.

RUD, B. *et al.* The Osteoporosis Self-Assessment Tool versus alternative tests for selecting postmenopausal women for bone mineral density assessment: a comparative systematic review of accuracy. **Osteoporosis International**, New York, p.1-9, 2008.

SACKETT, D. L. *et al.* Evidence based medicine: what it is and what it isn't. **British Medical journals**, London, v. 312, p.71-72, 1996.

SACKETT, D. How to read clinical journals. **Canadian Medical Association Journal**, Ottawa, v. 124, p.555-558, 1981.

SACKETT, L. D.; HAYNES, R. B. The architecture of diagnostic research. **British Medical Journals**, London, v. 324, p.539-541, 2002.

SADOYAMA, G.; GONTIJO FILHO, P. P. Colonizações do sítio de inserção e da ponta do cateter vascular central: experiência de 96 pacientes no Hospital de Clínicas da Universidade Federal de Uberlândia. **NewsLab**, São Paulo, v. 54, p.160-168, 2002.

SÁEZ-ALQUÉZAR, A. *et al.* Avaliação de um kit (protótipo) ELISA para detecção de anticorpos anti-T. pallidum. **NewsLab**, São Paulo, v. 73, p.122-128, 2005a.

SÁEZ-ALQUÉZAR, A. *et al.* Avaliação de um kit ELISA para detecção de anticorpos anti-T. cruzi. **LAES&HAES**, São Paulo, v. 157, p.136-142, 2005b.

SÁEZ-ALQUÉZAR, A. *et al.* Especificidade de três kits anti-HCV utilizados na triagem sorológica de doadores de sangue. **LAES&HAES**, São Paulo, v. 160, p.90-94, 2006.

SANTOS, F. L. N.; SOARES, N. M. Comparação entre as técnicas de Baermann e Baermann modificado para o diagnóstico da estrogiloidíase. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 37, n. 2, p.93-95, 2005.

SANTOS, I. B. C. *et al.* Avaliação do método clássico e do CHROMagar Candida na identificação de leveduras. **LAES&HAES**, São Paulo, v. 154, p.182-192, 2005.

SANTOS, L. M. *et al.* Validação do método automatizado MB/Bact para o diagnóstico da tuberculose pulmonar. **LAES&HAES**, São Paulo, v. 135, p.178-182, 2002.

SAPIN, R. *et al.* Increased sensitivity of a new assay for anti-thyroglobulin antibody detection in patients with autoimmune thyroid disease. **Clinical Biochemistry**, Montreal, v. 36, n. 8, p.611-616, 2003.

SAUERBREI, A.; WUTZLER, P. Detecção sorológica para o vírus Herpes simples IgG tipo-específica pelo recente ELISA baseada em glicoproteína G recombinante e altamente purificado. **LAES&HAES**, São Paulo, v. 162, p.204-214, 2006.

SCHENKEL, E. P. *et al.* Educação Farmacêutica em nível de Pós-Graduação no Brasil. **Revista Brasileira de Pós-graduação**, Brasília, v. 3, n. 6, p.175-192, 2006.

SCIELO (São Paulo). **Revista Brasileira de Ciências Farmacêuticas - About the Journal**. Disponível em: <<http://www.scielo.br/revistas/rbcf/paboutj.htm>>. Acesso em: 17 out. 2008a.

SCIELO (São Paulo). **Jornal Brasileiro de Patologia e Medicina Laboratorial - About the Journal**. Disponível em: <<http://www.scielo.br/revistas/jbpml/iaboutj.htm>>. Acesso em: 17 out. 2008b.

SEGURA, M. E. A.; MAGALHÃES, A. V. Galactin-3 immunostaining in thyroid neoplasms. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 41, n. 5, p.341-346, 2005.

SEKEROGLU, M. R. *et al.* Diagnostic value of cytokeratin-18 as a tumor marker in bladder cancer. **Clinical Biochemistry**, Montreal, v. 35, n. 4, p.327-331, 2002.

SELMAN, T. J.; KHAN, K. S.; MANN, C. H. An evidence-based approach to test accuracy studies in gynecologic oncology: the "STARD" checklist. **Gynecologic Oncology**, New York, v. 96, p.575-578, 2005.

SÈNE, D. *et al.* Biological markers of liver fibrosis and activity as non-invasive alternatives to liver biopsy in patients with chronic hepatitis C and associated mixed cryoglobulinemia vasculitis. **Clinical Biochemistry**, Montreal, v. 39, n. 7, p.715-721, 2006.

SHAPIRO, D. E. The interpretation of diagnostic tests. **Statistical Methods In Medical Reseach**, Indianapolis, v. 8, p.113-134, 1999.

SHIRIT, D. *et al.* Valor diagnóstico do teste de dímero-D na trombose venosa profunda: Melhoria dos resultados com nova metodologia e utilizando níveis de discriminação. **LAES&HAES**, São Paulo, v. 152, p.138-154, 2004.

SHUNMUGAN, M; AZUARA-BLANCO, A. The quality of reporting of diagnostic accuracy studies in glaucoma using the Heidelberg retina tomograph. **Investigative Ophtalmology & Visual Science**, Rockville, v. 47, n. 6, p.2317-2323, 2006.

SIDDIQUI, M. A. R; AZUARA-BLANCO, A.; BURR, J. The quality of reporting of diagnostic accuracy studies published in ophthalmic journals. **British Journal of Ophthalmology**, London, v. 89, p.261-265, 2005.

SILVA, S. L. Gestão do conhecimento: uma revisão crítica orientada pela abordagem da criação do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p.143-151, 2004.

SILVA, S. L. *et al.* Cristalúria como um método auxiliar de investigação de pacientes litíasicos: uma experiência na Região Metropolitana de Fortaleza. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 36, n. 4, p.233-235, 2004.

SIMEL, D. L.; RENNIE, D.; BOSSUYT, P. M. M. The STARD Statement for reporting diagnostic accuracy studies: Application to the History and Physical Examination. **Journal of General Internal Medicine**, Indianapolis, v. 23, n. 6, p.768-774, 18 mar. 2008.

SMIDT, N *et al.* Endorsement of the STARD Statement by biomedical journals: survey of Instructions for Authors. **Clinical Chemistry**, Washington, v. 53, n. 11, p.1983-1985, 2007.

SMIDT, N. *et al.* Quality of reporting of diagnostic accuracy studies. **Radiology**, Oak Brook, v. 235, p.347-353, maio 2005.

SMIDT, N. *et al.* Reproducibility of the STARD checklist: an instrument to assess the quality of reporting of diagnostic accuracy studies. **BMC Medical Research Methodology**, London, v. 12, n. 6, p.1-10, 15 mar. 2006a.

SMIDT, N. *et al.* The quality of diagnostic accuracy studies since the STARD statement: Has it improved?. **Neurology**, Hagerstown, v. 67, n. , p.792-797, 2006b.

SOCIEDADE BRASILEIRA DE ANÁLISES CLÍNICAS (Rio de Janeiro). **Revista Brasileira de Análises Clínicas**. Disponível em: <<http://www.sbac.org.br/pt/index.html>>. Acesso em: 17 out. 2008

STARD STATEMENT. **Standards for the Reporting of Diagnostic accuracy studies**. Disponível em: <<http://www.stard-statement.org/>>. Acesso em: 12 dez. 2008.

STENGEL, D. *et al.* Association between compliance with methodological standards of diagnostic research and reported test accuracy: meta-analysis of focused assessment of US for trauma. **Radiology**, Boston, v. 236, n. , p.102-111, 2005.

STENMARK, D. The relationship between information and knowledge. **Proceedings Of Iris 24**, Ulvik, v. 24, p.1-11, 2001

STINGHEN, A. E. M.; NASCIMENTO, A. J.; LEONART, M. S. S. Método de Papanicolaou em material cérvico-vaginal para a triagem de infecção por *Candida sp.*, *Trichomonas vaginalis* e *Chlamydia Trachomatis*. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 36, n. 2, p.111-115, 2004.

SUTTON-RILEY, J. M. *et al.* A single serum test for measuring early pregnancy outcome with high predictive value. **Clinical Biochemistry**, Montreal, v. 39, n. 7, p.682-687, 2006.

SUZUKI, T. *et al.* Diagnostic implications of circulating oxidized low density lipoprotein levels as a biochemical risk marker of coronary artery disease. **Clinical Biochemistry**, Montreal, v. 35, n. 5, p.347-353, 2002.

TATSIONI, A. *et al.* Challenges in systematic reviews of diagnostics technologies. **Annals Of Internal Medicine**, Philadelphia, v. 142, n. 12, p.1048-1055, 2005.

TOZZOLI, R. *et al.* Avaliação do sistema analítico Eneasystem III Bioallergy na determinação dos auto-anticorpos nucleares. **LAES&HAES**, São Paulo, v. 163, p.156-162, 2006.

TRENTI, T. Evidence-based laboratory medicine as a tool for continuous professional improvement. **Clinica Chimica Acta**, Gent, v. 333, n. , p.155-167, 2003.

TRIJFFEL E. *et al.* Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. **Manual Therapy**, Kidlington, v. 10, p.256-269, 2005.

UNIVERSIDADE DE SÃO PAULO (Brasil). **O que é fator de impacto?** Disponível em: <[http://biblioteca.igc.usp.br/fator\\_impacto/fator\\_impacto.htm](http://biblioteca.igc.usp.br/fator_impacto/fator_impacto.htm)>. Acesso em: 04 out. 2008.

UTIYAMA, S. R. R. *et al.* Correlação dos anticorpos anti-endomísio e antitransglutaminase coma doença celíaca. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 34, n. 1, p.39-45, 2002.

UZUN, H. *et al.* Serum cystatin C level as a potentially good marker for impaired kidney function. **Clinical Biochemistry**, Montreal, v. 38, n. 9, p.792-798, 2005.

VAN DEN BRUEL, A. *et al.* The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. **Journal Of Clinical Epidemiology**, Maastricht, v. 60, p.1116-1122, 2007.

VANDENBROUCKE, J. P. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. **Annals Internal Medicine**, Philadelphia, v. 147, n. 8, p.163-194, 2007.

VANDERHEYDEN, M. *et al.* Head to head comparison of N-terminal pro-B-type natriuretic peptide and B-type natriuretic peptide in patients with/without left ventricular systolic dysfunction. **Clinical Biochemistry**, Montreal, v. 39, n. 6, p.640-645, 2006.

VASCONCELLOS-SILVA, P. R; CASTIEL, L. D. Proliferação das rupturas paradigmáticas: o caso da medicina baseada em evidências. **Revista de Saúde Pública**, São Paulo, v. 39, n. 3, p.498-506, 2005.

WEBER, M. *et al.* Diagnostic utility of a new immunoassays for the cardiac markers cTnI, myoglobin and CK-MB mass. **Clinical Biochemistry**, Montreal, v. 38, n. 11, p.1027-1030, 2005.

WHITING, P. F. *et al.* Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. **BMC Medical Research Methodology**, London, v. 9, n. 6, p.1-8, 2006.

WHITING, P. *et al.* Sources of variation and bias in studies of diagnostic accuracy. **Annals Of Internal Medicine**, Philadelphia, v. 140, n. , p.189-202, 2004.

WHITING, P. *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. **BMC Medical Research Methodology**, London, v. 25, n. 3, p.1-13, 10 nov. 2003.

WILCZYNSKI, N. L. Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication: Before-and-after study. **Radiology**, Boston, v. 248, n. 3, p.817-823, 2008.

WOO, H. Y. *et al.* A prospective study of a new immunochemical fecal occult blood test in Korean patients referred for colonoscopy. **Clinical Biochemistry**, Montreal, v. 38, n. 4, p.395-399, 2005.

WORSTER A. *et al.* Diagnostic accuracy of BNP and NT-proBNP in patients presenting to acute care settings with dyspnea: A systematic review. **Clinical Biochemistry**, Montreal, v. 41, p.250-259, 2008.

YAGMUR, E. *et al.* Elevation of Ne-(carboxymethyl)lysine-modified advanced glycation end products in chronic liver disease is an indicator of liver cirrhosis. **Clinical Biochemistry**, Montreal, v. 39, n. 1, p.39-45, 2006.

YAMAN, H. *et al.* Elevated urine neopterin levels in nonalcoholic steatohepatitis. **Clinical Biochemistry**, Montreal, v. 38, n. 2, p.187-190, 2005.

YOSHIDA, C. S. *et al.* Análise do desempenho da prova de nitritos das tiras reativas de urina para triagem de infecção bacteriana do trato urinário. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 38, n. 4, p.255-258, 2006.

YOUNG, H. *et al.* Enzimaimunoensaio Enzywell recombinante para o diagnóstico sorológico da sífilis. **LAES&HAES**, São Paulo, v. 136, p.162-172, 2002.

ZAFAR, A.; KHAN, G. I.; SIDDIQUI, R. The quality of reporting of diagnostic accuracy studies in diabetic retinopathy screening: a systematic review. **Clinical & Experimental Ophthalmology**, Auckland, v. 36, p.537-542, 2008.

ZAMBON, C. *et al.* Non-invasive diagnosis of Helicobacter pylori infection: simplified C13-urea breath test, stool antigen testing, or DNA PCR in human feces in a clinical laboratory setting? **Clinical Biochemistry**, Montreal, v. 37, n. 4, p.261-267, 2004.

ZAMBON, C. *et al.* Quantitative PSA mRNA determination in blood: A biochemical tool for scoring localized prostate cancer. **Clinical Biochemistry**, Montreal, v. 39, n. 4, p.333-338, 2006.

ZANINOTTO, M. *et al.* NT-proBNP in the differential diagnosis of acute dyspnea in the emergency department. **Clinical Biochemistry**, Montreal, v. 38, n. 11, p.1041-1044, 2005.

ZHANG, Y. *et al.* Tree analysis of mass spectral urine profiles discriminates transitional cell carcinoma of the bladder from noncancer patient. **Clinical Biochemistry**, Montreal, v. 37, n. 9, p.772-779, 2004.

ZWEIG, M. H.; CAMPBELL, G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. **Clinical Chemistry**, Chicago, v. 39, n. 4, p.561-577, 1993.



ANEXO A –STARD (*Standards for Reporting of Diagnostic Accuracy*)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS			
<i>Participants</i>	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
<i>Test methods</i>	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
<i>Statistical methods</i>	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Methods for calculating test reproducibility, if done.	
RESULTS			
<i>Participants</i>	14	When study was performed, including beginning and end dates of recruitment.	
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	
<i>Test results</i>	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	
<i>Estimates</i>	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	How indeterminate results, missing data and outliers of the index tests were handled.	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

ANEXO B –*QUADAS (Quality Assessment of Diagnostic Accuracy Studies)*

<b>Item</b>		<b>Yes</b>	<b>No</b>	<b>Unclear</b>
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	( )
2 .	Were selection criteria clearly described?	( )	( )	( )
3.	Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	( )
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	( )	( )
6.	Did patients receive the same reference standard regardless of the index test result?	( )	( )	( )
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	( )
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	( )	( )
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	( )
10	Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	( )
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	( )
13.	Were uninterpretable/ intermediate test results reported?	( )	( )	( )
14.	Were withdrawals from the study explained?	( )	( )	( )