

ORLANDO BARRETO ZOCRATTO

**ESTUDO CITOLÓGICO DA GLÂNDULA TIREOIDE: COMPARAÇÃO (INTRA E
INTEREXAMINADOR) DOS RESULTADOS COM O EXAME HISTOLÓGICO
DAS PEÇAS CIRÚRGICAS.**

**BELO HORIZONTE
MINAS GERAIS – BRASIL
2010**

ORLANDO BARRETO ZOCRATTO

ESTUDO CITOLÓGICO DA GLÂNDULA TIREOIDE: COMPARAÇÃO (INTRA E INTEREXAMINADOR) DOS RESULTADOS COM O EXAME HISTOLÓGICO DAS PEÇAS CIRÚRGICAS.

Tese apresentada ao Curso de Pós-graduação em Saúde do Adulto da Faculdade de Medicina da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de Doutor em Saúde do Adulto.

ÁREA DE CONCENTRAÇÃO: Aparelho

Digestivo

ORIENTADOR: Prof. Dr. Paulo Roberto Savassi Rocha

**BELO HORIZONTE
MINAS GERAIS – BRASIL
2010**

ORLANDO BARRETO ZOCRATTO

ESTUDO CITOLÓGICO DA GLÂNDULA TIREOIDE: COMPARAÇÃO (INTRA E INTEREXAMINADOR) DOS RESULTADOS COM O EXAME HISTOLÓGICO DAS PEÇAS CIRÚRGICAS.

Tese apresentada e defendida perante Comissão Examinadora constituída pelos Professores:

Paulo Roberto Savassi Rocha (Orientador)

Belo Horizonte, de de 2010

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Reitor: Prof. Dr. Clélio Campolina Diniz

Vice-Reitora: Profa. Dra. Rocksane de Carvalho Norton

Pró-Reitor de Pós-Graduação: Prof. Dr. Ricardo Santiago Gomez

FACULDADE DE MEDICINA

Diretor: Prof. Dr. Francisco José Pena

Vice-Diretor: Prof. Dr. Tarcizo Afonso Nunes

Coordenador do Centro de Pós-Graduação: Prof. Dr. Carlos Faria Santos Amaral

DEPARTAMENTO DE CLÍNICA MÉDICA

Chefe: Profa. Anelise Impelizeri Nogueira

PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE DO ADULTO ÁREA DE CONCENTRAÇÃO: APARELHO DIGESTIVO

Coordenador: Profa. Dra. Teresa Cristina de Abreu Ferrari

COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE DO ADULTO ÁREA DE CONCENTRAÇÃO APARELHO DIGESTIVO

Prof. Dr. Luiz Gonzaga Vaz Coelho

Prof. Dr. Nilton Alves de Rezende

Profa. Dra. Suely Meireles Rezende

Profa. Dra. Valéria Azeredo Passos

Elizabeth Rosária de Miranda (representante discente)

À minha filha Clara, o amor verdadeiro e incondicional,
por ter revigorado minhas forças, pela força para lutar
contras as injustiças que sofremos e pela renovação da fé.

À minha esposa, Keli, pelo carinho, companheirismo,
compreensão e apoio imprescindíveis, principalmente nos
momentos mais difíceis.

Aos meus pais, Nivalda e Orlando, responsáveis por tudo
que sou e pelo exemplo a ser seguido, minha gratidão e
amor eternos.

AGRADECIMENTOS

A todos os que, de alguma forma, colaboraram na execução deste trabalho e contribuíram para a nossa formação profissional, nosso respeito e gratidão.

Particularmente:

Ao Prof. Dr. Paulo Roberto Savassi Rocha pela orientação segura e precisa deste trabalho, e, principalmente, pela amizade, pela compreensão, pelos exemplos de dignidade e de liderança que tanto influenciaram nossa formação profissional.

Ao Prof. Carlos Alberto Ribeiro, membro do Departamento de Anatomia Patológica da Faculdade de Medicina da UFMG, pela inestimável colaboração na execução deste estudo.

Profa. Dra. Teresa Cristina de Abreu Ferrari, Coordenadora do Curso de Pós-graduação em Saúde do Adulto - Área de Concentração Aparelho Digestivo – Faculdade de Medicina da UFMG, pelo apoio e incentivo à nossa formação acadêmica e titulação.

Às colegas Cynthia Koeppel Berenstein e Paula Piedade Garcia que, pronta e eficazmente, realizaram trabalho árduo e imprescindível, sem o qual este estudo não teria sido concluído.

A todos os colegas que participaram indiretamente na execução deste trabalho, nosso sincero agradecimento pela compreensão e ajuda que tornaram possível a sua realização.

E, em especial, aos meus familiares que souberam entender a importância deste projeto para mim, sabendo que a distância nunca irá nos separar.

Àqueles que, involuntariamente, omitimos.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

ABREVIATURAS

RESUMO

ABSTRACT

INTRODUÇÃO	16
REVISÃO DA LITERATURA	18
OBJETIVOS	34
MÉTODO	35
Casuística	35
Método	35
Características dos patologistas participantes	37
Análise estatística	38
RESULTADOS	40
Geral	40
Reprodutibilidade intraexaminador A1 – A2	48
Reprodutibilidade intraexaminador B1 – B2	49
Reprodutibilidade interexaminador A1 – B1	50
Reprodutibilidade interexaminador A2 – B2	51
Examinador A1	52
Examinador A2	52

Examinador B1	53
Examinador B2	53
DISCUSSÃO	55
CONCLUSÕES	84
PROPOSIÇÕES	85
REFERÊNCIAS	86
ANEXO I	100
ANEXO II	101
ANEXO III	102
ANEXO IV	103
ANEXO V	104
ANEXO VI	106

LISTA DE FIGURAS

Figura 1: Desenho esquemático do estudo (correlação inter e intraexaminador)	37
--	----

LISTA DE TABELAS

Tabela 1: Achados citológicos observados pelos patologistas A e B após as quatro análises (A1, A2, B1 e B2) (n = 74)	40
Tabela 2: Números absolutos e proporcionais de achados insuficientes observados em cada análise (A1, A2, B1 e B2), com os correspondentes diagnósticos histológicos de câncer	41
Tabela 3: Comparação entre os resultados apresentados pelos patologistas A e B em relação aos casos insuficientes, com os respectivos cálculos estatísticos	41
Tabela 4: Números absolutos e proporção de lesões citologicamente diagnosticadas como neoplasia folicular observados pelos patologistas A e B nas quatro análises (A1, A2, B1 e B2) com as correspondentes taxas de malignidade e diagnósticos histológicos de câncer ..	42
Tabela 5: Comparação entre os resultados apresentados pelos patologistas A e B em relação aos casos de neoplasia folicular, com os respectivos cálculos estatísticos	42
Tabela 6: Resultados falso-negativos observados por cada examinador (A1, A2, B1 e B2) com os respectivos diagnósticos histológicos considerando-se a definição estreita de malignidade (DEM) e a definição ampla de malignidade (DAM)	43
Tabela 7: Comparação entre as taxas de falso-negativos apresentadas pelos patologistas A e B, após todas as revisões (A1, A2, B1 e B2), com os respectivos cálculos estatísticos, considerando-se os casos de neoplasia folicular benigna	44
Tabela 8: Comparação entre as taxas de falso-negativos apresentadas pelos patologistas A e B, após todas as revisões (A1, A2, B1 e B2), com os respectivos cálculos estatísticos, considerando-se os casos de neoplasia folicular maligna	44
Tabela 9: Resultados falso-positivos observados por cada examinador (A1, A2, B1 e B2) com os respectivos diagnósticos histológicos, considerando-se a definição estreita de malignidade (DEM) e a definição ampla de malignidade (DAM)	45
Tabela 10: Comparação entre as taxas de falso-positivos apresentadas pelos patologistas A e B, após todas as revisões (A1, A2, B1 e B2), com os respectivos cálculos estatísticos, considerando-se os casos de neoplasia folicular maligna	45
Tabela 11: Casos concordantes (n = 45) nas quatro análises (A1, A2, B1 e B2), realizadas pelos patologistas A e B, com os correspondentes diagnósticos histológicos	46
Tabela 12: Comparação da proporção de tumores malignos observados pelos patologistas A e B inter (A1 – B1 e A2 – B2) e intraexaminador(A1 – A2 e B1 – B2), considerando-se os casos de neoplasia folicular benigna	47

Tabela 13: Comparação da proporção de tumores malignos observados pelos patologistas A e B inter (A1 – B1 e A2 – B2) e intraexaminador(A1 – A2 e B1 – B2), considerando-se os casos de neoplasia folicular maligna	47
Tabela 14: Reprodutibilidade intraexaminador para o patologista A (A1 – A2)	48
Tabela 15: Reprodutibilidade intraexaminador para o patologista B (B1 – B2)	49
Tabela 16: Reprodutibilidade interexaminador para os patologistas A e B na primeira análise de lâminas (A1 – B1)	50
Tabela 17: Reprodutibilidade interexaminador para os patologistas A e B na segunda análise de lâminas (A2 – B2)	51
Tabela 18: Comparação entre as taxas de concordância, inter (A1 - B1 e A2 - B2) e intraexaminador (A1 - A2 e B1 - B2), observadas para os patologistas A e B, obtidas pelo coeficiente <i>Kappa</i> (<i>K</i>) e <i>Kappa ponderado</i> (<i>Kp</i>) com os respectivos critérios fornecidos por Landis & Koch	52
Tabela 19: Taxas de sensibilidade (Se), especificidade (Es), valor preditivo positivo (VPP), valor preditivo negativo (VPN), acurácia (Ac), falso-negativo (FN), falso-positivo (FP), razão de verossimilhança do teste positivo (RVP) e razão de verossimilhança do teste negativo (RVN) observadas para os patologistas A e B usando-se as definições estreita (DEM) e ampla de malignidade (DAM). Os resultados estão expressos em percentagem	54

ABREVIATURAS

ABF – Ambulatório Bias Fortes

AF – adenoma folicular

BC – bócio coloide

CA – carcinoma anaplásico

CF – carcinoma folicular

CI – carcinoma indiferenciado

CM – carcinoma medular

CP – carcinoma papilífero

CPVF – carcinoma papilífero variante folicular

DAM – definição ampla de malignidade

DCO – diagnóstico citológico original

DEM – definição estreita de malignidade

EUA – Estados Unidos da América

FM – Faculdade de Medicina

HC – Hospital das Clínicas

HE – hematoxilina e eosina

IC – intervalo da confiança

K – Kappa

Kp – Kappa ponderado

MGG – May-Grünwald-Giemsa

NF – neoplasia folicular

OMS – Organização Mundial da Saúde

OR – *odds ratio*

PAAF – punção aspirativa com agulha fina

PBAG – punção biópsia com agulha grossa

RR – risco relativo

RVN – razão de verossimilhança do teste negativo

RVP – razão de verossimilhança do teste positivo

SBP – Sociedade Brasileira de Patologia

SPC – Sociedade Papanicolau de Citopatologia

UFMG – Universidade Federal de Minas Gerais

US – ultrassom

VPN – valor preditivo negativo

VPP – valor preditivo positivo

X^2 – Qui-quadrado

RESUMO

A punção aspirativa com agulha fina (PAAF) permite melhor seleção de pacientes candidatos à tireoidectomia propiciando a diminuição do número de pacientes operados e o aumento do número de tumores malignos ressecados. Entretanto, existe ampla variabilidade de resultados, que pode refletir as diferenças de experiência adquirida na sua interpretação. Acredita-se que a eficiência do método esteja diretamente relacionada ao treinamento do profissional. O parâmetro crítico da utilidade clínica de um teste diagnóstico é a sua reprodutibilidade interexaminador, e, principalmente, intraexaminador. Os objetivos deste estudo foram avaliar a influência da experiência do examinador na interpretação dos achados citológicos da tireoide, correlacionar os resultados citológicos emitidos pelos patologistas, com e sem experiência, com os exames histológicos das peças cirúrgicas e avaliar a influência do uso da definição ampla de malignidade (DAM) em detrimento do uso da definição estreita de malignidade (DEM) sobre os cálculos das medidas de validação e predição diagnóstica. Os resultados revelaram consistente reprodutibilidade intraexaminador após o intervalo de seis meses, tanto para o patologista experiente ($K = 0,82$ [0,71 – 0,94]; $Kp = 0,79$; $p = 0,988$) quanto para o não experiente ($K = 0,78$ [0,64 – 0,91]; $Kp = 0,71$; $p = 0,291$). Também não houve diferença para o patologista experiente e inexperiente, respectivamente, em termos de taxa de: material insuficiente ($p = 0,72$; $p = 0,48$), neoplasia folicular (NF) ($p = 1,0$; $p = 0,09$), falso-positivo ($p = 1,0$; $p = 0,72$), falso-negativo usando-se a DEM ($p = 0,55$; $p = 0,60$), falso-negativo usando-se a DAM ($p = 0,72$; $p = 0,72$), proporção de tumores malignos identificados usando-se a DEM ($p = 1,0$; $p = 0,59$) e a DAM ($p = 1,0$; $p = 0,66$). A análise da reprodutibilidade interexaminador mostrou diferença significativa, tanto no primeiro momento ($K = 0,55$ [0,38 – 0,72]; $Kp = 0,50$; $p = 0,011$) quanto no segundo, seis meses após ($K = 0,54$ [0,37 – 0,71]; $Kp = 0,44$; $p = 0,021$). Também houve diferença em relação ao número de resultados insuficientes, tanto na primeira ($p = 0,001$) quanto na segunda análise ($p = 0,021$), e de NF na segunda análise ($p = 0,015$). No entanto, não houve diferença em relação à proporção de tumores malignos identificados na primeira ($p = 0,79$ [DEM] e $p = 0,81$ [DAM]) e na segunda análise ($p = 0,78$ [DEM] e $p = 0,50$ [DAM]), assim como de falso-positivos ($p = 0,62$; $p = 0,21$) e de falso-negativos na primeira análise ($p = 0,55$ [DEM] e $p = 0,72$ [DAM]) e na segunda análise ($p = 0,60$ [DEM] e $p = 0,72$ [DAM]). Conclui-se que a reprodutibilidade intraexaminador na interpretação citológica da PAAF é consistente e independe da experiência com o exame, entretanto a reprodutibilidade interexaminador é influenciada. A experiência do examinador teve influência significativa na proporção de resultados insuficientes e de casos compatíveis com NF. O uso da DEM deve ser evitado quando da interpretação dos casos compatíveis com NF, tendo em vista a elevada taxa de malignidade observada nestas amostras, mesmo quando os laudos são emitidos por patologistas experientes. A indicação cirúrgica de nódulos tireoidianos deve levar em consideração a elevada taxa de resultados falso-negativos observada à PAAF, o que ocorre mesmo quando interpretada por patologista experiente, assim como a alta taxa de resultados insuficientes, principalmente devido as elevadas taxas de malignidade observadas nestas amostras.

Palavras-chave: Reprodutibilidade interexaminador e intraexaminador; experiência; punção aspirativa com agulha fina; nódulo tireoidiano.

ABSTRACT

A fine needle aspiration cytology (FNAC) has allowed a better selection of candidates for thyroidectomy providing a reduction in the number of operated patients and increased number of malignant tumors surgically resected. However, there is a wide variation in results, which may reflect differences in experience gained in its interpretation. It is believed that the efficiency of the method is directly related to training professional. The critical parameter of the clinical usefulness of a diagnostic test is its interobserver variability, and especially intraobserver. The objectives of this study was to evaluate the influence of the examiner's experience in the interpretation of cytological findings of thyroid and to correlate the cytologic findings issued by pathologists with and without experience, with the histological examinations of surgical specimens and to evaluate the influence of a broad definition of malignancy (BDM) instead of using the narrow definition of malignancy (NDM) on the calculations of measures and validation of diagnostic prediction. The results revealed consistent intraobserver reproducibility after the lapse of six months, both for the experienced pathologist ($K = 0.82$ [0.71 to 0.94], $Kp = 0.79$, $p = 0.988$) and for non-experienced ($K = 0.78$ [0.64 to 0.91], $Kp = 0.71$, $p = 0.291$). There was no difference for the inexperienced and experienced pathologist, respectively, in terms of rate: insufficient material ($p = 0.72$, $p = 0.48$), follicular neoplasm (FN) ($p = 1.0$, $p = 0.09$), false-positive ($p = 1.0$, $p = 0.72$), false-negative using the NDM ($p = 0.55$, $p = 0.60$), false-negative using the BDM ($p = 0.72$, $p = 0.72$), proportion of malignant tumors identified using the NDM ($p = 1.0$, $p = 0.59$) and BDM ($p = 1.0$, $p = 0.66$). The analysis of interobserver reproducibility showed a significant difference in both the first moment ($K = 0.55$ [0.38 to 0.72], $Kp = 0.50$, $p = 0.011$) or the second after six months ($K = 0.54$ [0.37 to 0.71], $Kp = 0.44$, $p = 0.021$). There was also a difference in the number of unsatisfactory results in both the first ($p = 0.001$) and second examination ($p = 0.021$), and FN in the second analysis ($p = 0.015$). However, no difference in the proportion of malignant tumors identified in the first ($p = 0.79$ [NDM], $p = 0.81$ [BDM]) and the second analysis ($p = 0.78$ [NDM], $p = 0.50$ [BDM]), as well as false-positive ($p = 0.62$, $p = 0.21$) and false-negative in the first analysis ($p = 0.55$ [NDM], $p = 0.72$ [BDM]) and the second analysis ($p = 0.60$ [NDM], $p = 0.72$ [BDM]). Conclude that the intraobserver reproducibility in the interpretation of FNA cytology is consistent and independent of experience to the survey, however, the interobserver reliability is influenced. The experience of the examiner influences significantly the proportion of unsatisfactory results and cases compatible with FN. The use of the NDM should be avoided when interpreting the case compatible with FN, in view of the high rate of malignancy observed in these samples, even when the reports are issued by experienced pathologists. Surgical indication of thyroid nodules should take into account the high rate of false-negative results observed at the FNAC, which occurs even when performed by experienced pathologist, as well as the high rate of unsatisfactory results, mainly due to the high rates of malignancy observed in these samples.

Key-words: Reproducibility of results; technical expertise; interobserver and intraobserver variability; fine needle aspiration cytology; thyroid nodule.

INTRODUÇÃO

Poucos assuntos em cirurgia têm gerado tantas controvérsias quanto a conduta diante do nódulo tireoidiano. Dentre elas, encontra-se a PAAF, utilizada para a obtenção de células para diagnóstico citopatológico de nódulos da tireoide.^{1,2} Durante os 50 anos de seu uso, os princípios básicos da técnica têm-se mantido quase inalterados e o seu sucesso depende da cuidadosa aspiração e da interpretação criteriosa dos achados citomorfológicos. Analisada de forma isolada ela permanece superior à clínica, à cintilografia e ao ultrassom (US) em termos de especificidade diagnóstica.³ Algumas características, incluindo baixo custo, segurança, acurácia e facilidade técnica, permitiram que, ao longo dos dois últimos decênios, ela se tornasse o principal método propedêutico na abordagem inicial das afecções nodulares da tireoide.⁴⁻¹²⁻¹⁶

Existe consenso de que seu emprego como método propedêutico tem permitido melhor seleção de pacientes candidatos à tireoidectomia e propiciado aumento do número de tumores malignos ressecados cirurgicamente.^{7,17-38} Entretanto, existe ampla variabilidade de resultados, que pode refletir as diferenças de experiência adquirida na punção e na interpretação do esfregaço.^{22,25,29,34,35,39-44} Estudos têm mostrado que melhores resultados são alcançados quando a punção, o preparo da lâmina e a sua interpretação são realizados de forma imediata por citopatologista afeito ao método.^{12,41-45}

Acredita-se que a eficiência do método esteja diretamente relacionada ao treinamento do profissional que o realiza de forma frequente e continuada.^{11,25,41,42} Daí, alguns autores têm recomendado que a PAAF seja restrita a centros terciários, onde todas as etapas técnicas do procedimento estão concentradas em poucos indivíduos, o que resultaria no maior acúmulo de experiência.^{25,41,42} Por outro lado, alguns estudos têm mostrado que profissionais pouco experientes apresentam resultados comparáveis aos observados em grandes séries de centros de referência terciários, em termos de material insuficiente, sensibilidade, especificidade e acurácia diagnóstica.^{7,28,46-49} A importância da curva de aprendizado já foi documentada, de modo que, a medida que ambos, punçador e examinador adquirem maior experiência, há melhora progressiva na confiabilidade do exame.^{39,49,50-52}

As maiores limitações aos potenciais benefícios da PAAF incluem, além da inabilidade do médico que realiza a punção e o preparo do esfregaço, a inexperiência do citopatologista que o interpreta.²² Citopatologistas inexperientes podem apresentar alta proporção de resultados falso-

positivos e falso-negativos, resultados insuficientes e de diagnósticos compatíveis com NF.^{5,11,30,42,53-60} Resultados falso-positivos têm sido causa importante de indicação desnecessária de tireoidectomias.⁶¹ Por outro lado, resultados falso-negativos costumam ser ainda mais desastrosos por deixarem de diagnosticar tumor maligno.⁶²

O valor de qualquer teste diagnóstico depende da capacidade de que seu resultado seja repetido quando aplicado novamente no mesmo paciente. Portanto, o parâmetro crítico da utilidade clínica do teste diagnóstico é a sua reprodutibilidade interexaminador, e, principalmente, intraexaminador, sendo esta considerada um dos componentes mais importantes para a avaliação global da acurácia de determinado teste diagnóstico.^{63,64} A reprodutibilidade histológica e, principalmente, citológica dos tumores da tireoide foi muito pouco investigada, tanto no que diz respeito à concordância interexaminador quanto intraexaminador, daí a importância da atual pesquisa.^{7,18,19,29,57,59,63,65-79}

REVISÃO DA LITERATURA

A exploração de massas superficiais por meio de punção aspirativa remonta ao século XIX. O primeiro relato do uso de agulha para a obtenção de material para avaliação microscópica é atribuído a Kun, de Strasbourg, em 1847, citado por Webb.⁸⁰ Em Londres, em 1853, Paget e Erichsen (citados por Webb⁸⁰), em publicações independentes, foram os primeiros a preconizar a aspiração seguida da microscopia, como técnicas de investigação diagnóstica, especialmente para as doenças da mama. Mannheim, em 1931, citado por Koss et al⁸¹, propôs o emprego de agulha fina (1mm), dando início, na Europa, à escola de biópsia aspirativa, baseada no emprego de agulhas finas.

O diagnóstico de tumores por meio de punção gerou grande controvérsia, o que dificultou sua aceitação até 1934 Hayes Martin e Edward Ellis publicaram o primeiro estudo com maior casuística (n = 65) que incluía tumores malignos diagnosticados por PAAF, como citado por Zakowski.⁵⁰ Eles foram os primeiros a descrever a técnica de biópsia aspirativa por agulha de tumores e os pioneiros na punção de tumores de mama. Preconizaram o uso de agulha grossa (18-gauge) que era introduzida através de pequena incisão da pele, após anestesia subcutânea. O material era preparado sob a forma de esfregaços e corado pela técnica de hematoxilina e eosina (HE). Em 1937, Stewart, citado por Webb⁸⁰, relatou não haver dificuldade em diagnosticar o carcinoma anaplásico (CA) de tireoide. Entretanto, segundo ele, a distinção entre lesões benignas e as demais lesões tumorais não era possível de ser feita. Já, naquela época, enfatizou a importância da técnica adequada de aspiração, do preparo da amostra e da cooperação entre clínicos e patologistas para assegurar maior acurácia ao exame. A incisão da pele, como preconizada por Martin e Ellis, foi dispensada por Sayago (citado por Webb⁸⁰), em 1942.

A PAAF propriamente dita, na qual se utilizam agulhas de 22 a 27-gauge, foi primeiramente introduzida por Söderström (citado por Löwhagen et al¹⁶), em 1952, no Radiumhelmet Hospital de Stockholm, dando o passo inicial para sua aplicação no diagnóstico de nódulos tireoidianos.

Apesar do trabalho pioneiro de patologistas e cirurgiões americanos e de ter sido amplamente utilizada na Suécia desde 1952, foram as publicações de Zajicek, do Instituto Karolinska - Stockholm e de Löwhagen et al (citados por Zakowski⁵⁰, Koss et al⁸¹, McDonald & Yazdi⁸² e Kuma et al⁸³) que tornaram, de fato, a PAAF conhecida internacionalmente. Somente

na década de 1970 ela ganhou aceitação mundial como resultado da publicação de trabalhos que mostraram resultados animadores.^{14,16,50,82,84-86} Dentre os fatores que contribuíram para esse início tardio, pode-se citar o receio da implantação de células neoplásicas no trajeto da agulha e/ou sua disseminação sistêmica e a insegurança quanto ao diagnóstico citológico.^{1,87} Acreditava-se que seria pouco provável que o diagnóstico pudesse ser feito pelo exame citológico de alguns grupos celulares retirados com agulha fina.⁵⁶

Estima-se que, em áreas sem deficiência de iodo, 4,0% a 7,0% da população apresente nódulos tireoidianos palpáveis.^{1,8,14,26,34,61,88} Em áreas endêmicas, sua prevalência pode ultrapassar 25,0%, podendo chegar, em estudos de necropsia, a 50,0%.^{1,87} Aproximadamente 23,0% dos nódulos tireoidianos manifestam-se, clinicamente, como lesões únicas. No entanto, representam, na maioria das vezes, nódulos dominantes localizados em glândula multinodular.⁸⁹ As anormalidades ultrassonográficas detectadas numa população adulta selecionada aleatoriamente são expressivas, especialmente em mulheres de meia idade, podendo chegar a 50,0% e até 70,0%.^{13,90,91}

Sabe-se que a incidência de nódulos tireoidianos aumenta com a idade, dieta contendo substâncias bociogênicas e exposição à radiação ionizante, principalmente na região cérvico-facial.^{13,26} Neste caso, a incidência pode chegar a 2% ao ano, com pico após 15 a 20 anos.¹³ Apesar da maior prevalência em mulheres, a incidência relativa do câncer é maior em homens jovens (menores de 20 anos) e idosos (maiores de 70 anos).¹³ Isto ocorre para todos os tumores malignos, especialmente os tipos papilar, folicular e anaplásico.^{6,13}

Apesar de sua elevada prevalência na população geral, o câncer é relativamente pouco frequente. Estima-se que somente 5,0% a 10,0% dos nódulos atóxicos e palpáveis da tireoide sejam malignos, o que corresponde a aproximadamente 1,0% dos tumores malignos em geral podendo chegar a 38,0% quando se consideram os nódulos removidos cirurgicamente.^{3,8,13-15,21,88,91-94} Logo, a importância do estudo do nódulo tireoidiano se prende à sua alta prevalência na população geral, por ser a única manifestação de muitas doenças tireoidianas, e pela alta proporção de casos benignos em relação aos malignos.⁹⁵ Diante desses fatos e da impossibilidade do diagnóstico de certeza pré-operatório, o maior desafio se encontra no estabelecimento de sua natureza (benigna ou maligna) e, com isso, a seleção mais apropriada dos candidatos à operação.^{34,37,26,96} Segundo Wiersinga (1995)⁹⁷, se por um lado a oportunidade do tratamento não

pode ser perdida, por outro, é preocupante o número de pacientes submetidos a operações diagnósticas (“desnecessárias”) nos casos de doenças benignas da tireoide.

Os tumores malignos da glândula tireoide incluem o carcinoma papilar (CP), o carcinoma folicular (CF), o carcinoma medular (CM), o raro e quase universalmente fatal CA ou indiferenciado (CI), além do linfoma e dos tumores metastáticos.^{1,2,4,9,16,27} Dados epidemiológicos e de coortes prospectivas estimam a prevalência aproximada de 80,0% para o CP, 15,0% para o CF e de células de Hürthle, 3,0% a 4,0% para o CM e 1,0% a 2,0% para o CA.⁶ A incidência do CF parece estar declinando em todo o mundo, provavelmente em consequência da suplementação de iodo na dieta.^{98,99} Os tumores diferenciados (CP e CF) são caracterizados por crescimento indolente e bom prognóstico, por serem, frequentemente, curáveis com o tratamento cirúrgico.^{100,101} No entanto, mesmo tumores pequenos (menos de 1cm no maior diâmetro) podem metastatizar e levar ao óbito.⁸⁷

A incidência do câncer é alta em indivíduos que apresentam sinais e sintomas sugestivos de malignidade nos quais a PAAF de nódulos, clinicamente suspeitos, evidenciou tumor em mais de 70,0% dos casos.¹⁰² A distinção entre lesões benignas e malignas não pode ser baseada somente na apresentação clínica, pois a maioria dos pacientes é assintomática. É importante que existam métodos não cirúrgicos que possibilitem a diferenciação de tais lesões.¹⁰²

O US informa sobre as características morfológicas do nódulo (sólido ou cístico, único ou múltiplo, e a presença ou não de sinais ecográficos sugestivos de malignidade).¹⁰¹

A cintilografia com I¹²³ ou Tc^{99m} classifica a lesão em relação a sua habilidade para captar o iodo, propiciando avaliação indireta do potencial de malignidade.^{13,88} Como a maioria dos nódulos benignos é hipofuncionante (“frios”) e o nódulo “quente” pode ser maligno, a cintilografia identifica muitas lesões benignas como potencialmente malignas, o que poderia resultar em muitas operações diagnósticas, caso a indicação fosse baseada nos seus resultados.^{3,88,101} Essa distinção quando feita por meio do US, da cintilografia e da dosagem sérica de tireoglobulina tem se mostrado pouco sensível e específica.^{8,25,26,103}

A PAAF, por sua vez, é o método propedêutico usado para coletar células para posterior avaliação microscópica de seus aspectos morfológicos, sua quantidade e sua distribuição no esfregaço.⁵⁰ Algumas características permitiram que ela ganhasse popularidade como meio diagnóstico e se tornasse o principal método propedêutico na abordagem inicial das afecções nodulares da tireoide. Estas características incluem: confiabilidade, segurança, baixa morbidade,

baixo custo, execução rápida, simplicidade técnica, facilidade de acesso e acurácia diagnóstica.^{1-3,5,17-19,21-23,54,104,105}

A conduta diante do portador de nódulo tireoidiano tem evoluído da remoção cirúrgica de todos os nódulos solitários (tireoidectomia diagnóstica) para abordagem mais seletiva baseada na avaliação clínica e propedêutica incluindo, especialmente, a PAAF.⁸⁷ Logo, o principal objetivo da PAAF é possibilitar a distinção entre nódulos benignos e malignos, e, com isso, selecionar mais apropriadamente os candidatos ao tratamento cirúrgico.^{10,15,17,29,31,40,62,87,92,95,96,105,106} Inúmeros estudos têm demonstrado a influência da PAAF na diminuição da proporção de pacientes tratados cirurgicamente da ordem de 35,0% a 75,0%, enquanto mantém ou até aumenta o percentual de tumores malignos removidos cirurgicamente, de 5,0% a 15,0% para 25,0% a 50,0%.^{7,14,22,25,28,30,32-38,46,47,103,105,107}

Quando a incidência de tumores malignos em espécimes cirúrgicos excisados de acordo com critérios clássicos (história clínica, palpação, US e cintilografia) é comparada com a incidência de tumores excisados de acordo com indicações citológicas, há ganho considerável no número de tumores diagnosticados em excisões citologicamente orientadas, e apreciável redução nos custos.^{3,7,12,20,27,33,34,46,108} Pepper et al (1989)⁴⁷ obtiveram taxa de malignidade diagnosticada pela PAAF de 38,0%, bem superior aos 12,0% quando a operação foi indicada com base somente na avaliação clínica. O mesmo ocorreu com Guimarães et al (1996)¹⁰³, onde a proporção de neoplasias encontrada à operação passou de 20,0% para 48,0% após a introdução da PAAF em seu serviço. Ng et al (1990)³⁸ observaram redução de 95,0% para 60,0% ($p < 0,001$) na porcentagem de doentes operados e aumento de 18,4% para 26,2% na proporção de tumores malignos diagnosticados. Segundo Shaha (2000)¹, os custos foram reduzidos em aproximadamente 25,0% e a incidência de tumores malignos ressecados aumentou de 14,0% para 29,0%. Outro benefício da PAAF refere-se à redução do número de exames de corte-congelação realizados no per-operatório, já que, diante do resultado citológico sugestivo de malignidade, este seria desnecessário por contribuir pouco para o manejo do paciente.⁸⁷

A técnica usada na PAAF, acrescida de algumas modificações, foi amplamente descrita em publicações prévias.^{17,41,109,110} Ela é indicada na propedêutica de nódulos palpáveis, mas pode ser guiada pelo US, sendo, nessas circunstâncias, indicada em nódulos não palpáveis, múltiplos, mistos ou complexos, cistos com projeção sólida em sua parede e quando não se consegue material adequado na punção dirigida pela palpação.¹¹¹ As lâminas devem ser prontamente

fixadas em álcool a 95,0% com, pelo menos uma, secada ao ar e enviadas ao citopatologista para colorações rotineiras: May-Grünwald-Giemsa (MGG), HE e Papanicolaou.^{41,81}

A interpretação do exame citológico obtido pela PAAF permite a subdivisão dos resultados em quatro principais categorias diagnósticas, a saber: benigno (60,0% a 75,0%);^{13,14,18,21} maligno, que inclui o CP, o CM, o CA, o linfoma e os tumores metastáticos para tireoide (1,0% a 10,0%);^{4,9,13,14,18,21,92} NF ou indeterminado (10,0% a 30,0%);^{9,14,15,21,29,30,66,88,92,112} e insuficiente para análise microscópica (10,0% a 20,0%).^{13,15,17,18,21,30,87,106,113} Nos casos indeterminados (NF), há quantidade mínima, ou mesmo ausência de coloide, associada a abundância de células foliculares, o que impossibilita a distinção entre adenomas e adenocarcinomas bem diferenciados.¹⁵ Cerca de 15,0% a 25,0% destes são tumores malignos, o que representa importante limitação do método.^{13,14,18,29,30,48,49,114,115}

Excluindo-se o exame histológico da glândula removida cirurgicamente (exame padrão), a análise citomorfológica bem feita do material colhido pela PAAF, principalmente quando guiada por US, é atualmente o método mais acurado, mais específico¹⁰⁸, e o que apresenta melhor relação custo-benefício para o diagnóstico de lesões nodulares da tireoide.²⁶ Sua capacidade para prever malignidade é duas a quatro vezes superior a de outros métodos propedêuticos. Ela configura-se como o melhor exame disponível com acurácia que varia de 70,0% a 95,0%, sensibilidade de 68,0% a 98,0% e especificidade de 72,0% a 100,0%, além de ter melhor valor preditivo que qualquer teste atualmente disponível.^{3,12-14,22,25,26,30} Assim sendo, ela tem sido utilizada como triagem inicial de pacientes portadores de nódulos tireoidianos, sendo considerada como pré-requisito essencial para sua abordagem bem sucedida.^{14,19-21,41,46,53,94} No entanto, seu sucesso depende da aspiração cuidadosa e da interpretação criteriosa dos achados citomorfológicos, sempre se mantendo em mente as manifestações clínicas da doença e outros dados relevantes da propedêutica.^{29,67,115}

No estudo de Hamming et al (1990)¹⁰², 169 (32,0%) pacientes foram tireoidectomizados com base no exame clínico, na PAAF, ou ambos. Esses pacientes foram, retrospectivamente, distribuídos em três grupos (sem o conhecimento dos resultados citológicos e histológicos) de acordo com a suspeita clínica de malignidade (alta, moderada ou baixa), e, a seguir, foram comparados com o exame padrão. Os resultados da PAAF foram categorizados em maligno, benigno e incerto, e, em seguida, comparados com o exame padrão. A razão de verossimilhança do teste positivo (RVP) e a razão de verossimilhança do teste negativo (RVN) foram calculadas

para os dois grupos (clínico e PAAF). Para os grupos de suspeita clínica, as RVP foram oito, 0,5 e 0,4 para pacientes com suspeita clínica alta, moderada ou baixa de malignidade, respectivamente. Para os grupos da PAAF, as RVP foram 48,1, 5,7 e 1,1, para os resultados, positivo, incerto e benigno, respectivamente. Em outras palavras, um resultado positivo à PAAF é 48 vezes mais provável de ocorrer num paciente com doença maligna que no paciente sem câncer.

A PAAF da tireoide tem elevado VPN para neoplasias além de permitir, com certa precisão, a distinção entre lesões benignas e malignas. Ela é mais precisa na detecção do CA e do CM, com sensibilidade e especificidade que se aproximam de 100,0% e 90,0%, respectivamente.¹²⁻¹⁴ A sua acurácia para detectar o CP e o CF são de 80,0% a 90,0% para o primeiro e de 40,0% para o segundo.¹²⁻¹⁴ Por outro lado, sua maior limitação é a baixa sensibilidade para, corretamente, distinguir as lesões foliculares encapsuladas, tais como: adenoma folicular (AF), bócio colóide (BC), CF, carcinoma papilífero variante folicular (CPVF) e as lesões de células de Hürthle.^{13,15,18,98,99,116-118} Isto ocorre porque o estudo citológico não permite a discriminação entre as formas benigna e maligna dos nódulos foliculares e de células de Hürthle.^{13,66,106,116} Como os critérios para se distinguir o AF do CF são definidos histologicamente (invasão vascular e/ou capsular) e devido ao considerável potencial de malignidade, a tireoidectomia é, usualmente, indicada para completa avaliação diagnóstica.^{15,62,70,71,114} Como 15,0% a 25,0% dos casos de NF são malignos, mais de 80,0% dos pacientes que apresentam esse diagnóstico serão, desnecessariamente, operados.^{13,30,98,104,116}

Condição inerente ao método refere-se à não preservação da arquitetura histológica do material coletado, o que limita sua capacidade em fazer o diagnóstico definitivo. Apesar do seu sucesso, resultados indeterminados e erros diagnósticos são inevitáveis devido à sobreposição de características citológicas (celularidade elevada e presença de microfóliculos)^{13-18,40,61,72,84,98,99,105,116,117}, particularmente a distinção entre o adenoma microfolicular hipercelular e as lesões microfoliculares hiperplásicas, e entre estas e o CF bem diferenciado. Outra limitação refere-se à elevada taxa de material inadequado para diagnóstico (principalmente devido à má qualidade técnica da coleta e/ou preparo das amostras e à amostragem inadequada) que reflete mais a habilidade do punccionador, e, menos frequentemente, a postura conservadora e/ou inexperiência do citopatologista,^{8,11,18,42,43,50-54,61,105,119-122} Sabe-se, também, que existem riscos em todo procedimento invasivo e que, em qualquer teste diagnóstico envolvendo amostragem e

interpretação, importantes diagnósticos podem passar despercebidos ou serem equivocados (falso-positivos e falso-negativos).^{11-14,20-22,30,40,41,50,53,58-62,65-68,73-75,123-126}

A acurácia refere-se à proporção de acertos do teste. Entretanto, isoladamente, ela é insuficiente para estimar as probabilidades pós-teste da doença. Como não existe teste 100% acurado, pressupõe-se a existência de dois tipos de erros: falso-negativos e falso-positivos.¹²⁷ O primeiro é mais grave porque o tratamento pode não ser indicado; o segundo, por sua vez, poderá resultar numa terapia desnecessária.^{106,127,128}

A acurácia encontra-se diretamente relacionada às características do tumor, como tipo histológico, tamanho, mobilidade em relação aos planos profundos e localização.⁸² Não obstante, ela está diretamente relacionada às características do profissional que realiza o exame, tais como: habilidade técnica na coleta do material, qualidade com que o esfregaço é preparado e experiência na sua interpretação.^{11,25,40,53,65,129,130} Espécimes coletados por indivíduos que utilizam a técnica somente ocasionalmente são, frequentemente, de pior qualidade.^{24,49} A ampla variabilidade de resultados pode refletir a existência de diferentes experiências na sua interpretação.^{12,13,25,39,40,43,44,129} Tem sido sugerido, por diferentes autores, que os melhores resultados são alcançados quando a punção, preparo das lâminas e sua interpretação feitas por citopatologista afeito ao método.^{11,12,20-22,25,29,40-45,51,62,65,82,87,114,129-131} Citopatologistas inexperientes podem apresentar alta proporção de resultados compatíveis com NF e insuficientes e de diagnósticos equivocados. Assim, à medida que ambos, puncionador e examinador adquirem maior experiência, ocorre melhora progressiva nas acurácia e confiabilidade do exame e redução da taxa de esfregaços inadequados.^{5,11,30,42,50,53-55,58,59,62}

A habilidade e experiência do puncionador desempenham papel importante na determinação da adequação do espécime.¹¹ Os cuidados na obtenção da amostra e na realização do esfregaço são essenciais para se obter aspirados representativos da lesão e adequados para a interpretação.^{119,132} Quando realizados por profissional inexperiente existe maior possibilidade de ocorrer resultados insuficientes e falso-negativos.^{5,42,43,51,52,54} Palombini et al (1988)⁵⁵ encontraram menor incidência de esfregaços inadequados e de resultados falso-positivos quando patologistas realizaram a palpação, a aspiração e a interpretação citológica em relação aos esfregaços feitos por clínicos e cirurgiões. No entanto, no estudo realizado por Skaaning (1981)¹³³, a experiência isolada (considerada o número total de anos de realização da PAAF) não demonstrou estar associada, de maneira significativa, a esfregaços mais adequados. A

percentagem de espécimes inadequados foi de 15,0%, independentemente do número de anos de experiência.

Segundo Suen (1988)¹³⁴, a celularidade do esfregaço reflete mais a habilidade técnica do médico que realiza a aspiração que a natureza da lesão em si. A taxa de resultados falso-negativos e de esfregaços inadequados pode ser reduzida quando o puncionador é experiente e habilidoso.^{51,52} Resultados aceitáveis podem ser obtidos se os responsáveis pela coleta e interpretação do exame são proficientes, mas não necessariamente a mesma pessoa.^{51,52} Segundo Ribeiro e Brasileiro (1996)³⁹, à medida que aumenta o número de casos analisados, aumenta também a experiência do citopatologista na interpretação dos exames citológicos.

Resultados insuficientes ocorrem mais comumente na presença de lesões maiores que 4cm, particularmente quando elas são císticas ou parcialmente císticas, focais, ou quando há excessiva necrose (debris), os nódulos são escleróticos ou calcificados ou existe muito sangue, o que obscurece os detalhes celulares.^{20,105} Punções desses nódulos podem fornecer aspirados hemorrágicos e, com frequência, o número de células é escasso.^{26,93,113}

Com o advento da PAAF guiada por US houve melhora no processo de coleta do material, sobretudo nos casos em que fatores não relacionados ao operador e ao citopatologista prejudicam o resultado citológico, como, por exemplo, na presença de lesões císticas^{14,20,82,86,99,103,135} ou de nódulos pequenos.^{136,137} O US assegura o adequado posicionamento da agulha para precisa amostragem da lesão.¹⁵ Isto tem resultado na diminuição do número de casos insuficientes, aumento das taxas de sensibilidade, especificidade, acurácia e do diagnóstico de carcinomas incidentais.^{11,125,135,138-140} Nos indivíduos portadores de bócio multinodular, além da punção do nódulo dominante, o US permite a punção de nódulos que apresentam características sugestivas de malignidade.¹⁴¹

As medidas que oferecem maiores informações sobre o teste diagnóstico são: RVP, RVN, sensibilidade, especificidade, acurácia, VPP, VPN e as taxas de falso-positivos e de falso-negativos.^{14,113,127} A sensibilidade e a especificidade são mais úteis que a acurácia e, em conjunto, permitem calcular a probabilidade pós-teste a partir da probabilidade pré-teste. Todavia, as expressões mais simples da capacidade do teste em modificar a plausibilidade da doença são a RVP e a RVN.^{102,106} Elas possuem duas importantes propriedades: primeiro, a de poderem ser calculadas para cada extrato do teste (positivo ou negativo); segundo, por não serem afetadas pela prevalência da população sob estudo.¹¹³

Diversos estudos foram publicados utilizando, como método, a correlação cito-histológica, com os objetivos de avaliar a precisão diagnóstica da PAAF.^{29,61,76,119,124,126,130,132,142} Os casos diagnosticados como insuficientes normalmente são excluídos dos cálculos, uma vez que não fornecem qualquer informação sobre a precisão diagnóstica. De modo geral, o que se conclui a partir da revisão desses estudos é que a PAAF é, geralmente, bem aceita e tolerada por parte dos pacientes, as complicações são raras e sem gravidade e que ela é eficiente, segura e tecnicamente simples de ser executada, mesmo em Serviços de pequeno porte. A ausência de correlação cito-histológica deve-se a prováveis erros na obtenção da amostra, no preparo dos esfregaços ou na sua interpretação.^{1,15,29,99,105} Entretanto, suas limitações devem ser levadas em consideração, principalmente ao se analisar o resultado negativo, devido à possibilidade de se tratar de falso-negativo.⁴³⁻⁴⁵

Nódulos diagnosticados como benignos pela PAAF requerem seguimento por causa da taxa de resultados falso-negativos que pode variar de 1,0% a 26,0%.^{9,30,50,126,142} Resultados falso-negativos são, usualmente, causados por erros de amostragem (nódulo inadequadamente amostrado), espécimes inadequados (preparação inapropriada do esfregaço) e, em menor grau, por inabilidade do patologista para reconhecer a lesão maligna (erro de interpretação).^{30,53,58,59} Amostragem inadequada ocorre mais comumente na presença de bócios multinodulares, lesões císticas, e/ou menores que 1cm ou maiores que 4cm.^{13,59,128} Amostras insuficientes têm sido citadas como importante causa de resultados falso-negativos.¹⁴³

As taxas de resultados falso-positivos reportados na literatura variam de 0% a 11%.^{15,29,67} Sua principal causa é atribuída a erros de interpretação, como na presença de tireoidite de Hashimoto.⁶⁰ Essa ampla faixa de variabilidade de resultados pode refletir a existência de diferentes experiências acumuladas com o método. Isto se deve, principalmente, à sobreposição de critérios citológicos presentes nas lesões foliculares encapsuladas bem como à presença de esfregaços insuficientes.^{58,59,65}

Segundo a Associação Americana de Endocrinologistas e o Colégio Americano de Endocrinologia, os maiores problemas que limitam o potencial benefício da PAAF são a inabilidade do médico que realiza a punção e o preparo do esfregaço e a inexperiência do citopatologista que o interpreta.²² Os passos do procedimento (exame clínico com localização da lesão alvo, coleta do material, preparo do esfregaço e microscopia) devem ser aprendidos mediante treinamento especializado para se tornar proficiente e permitir seu uso otimizado.^{41,50}

Seu uso em larga escala, principalmente em ambientes extra-hospitalares tem sido dificultado pela relutância por parte de muitos médicos tanto em realizá-la quanto em indicá-la, particularmente para aqueles com limitada experiência. Atribui-se este fato à crença de que grande número de procedimentos deve ser executado, a cada ano, para manutenção da habilidade técnica e apropriado nível de capacitação.^{5,25,41,42,144,145} Existem, na literatura, diversos critérios que definem a presença ou ausência de experiência em microscopia da glândula tireoide. Alguns se baseiam somente na afiliação ao Comitê de Citopatologistas do Colégio Americano de Patologistas, outros no tempo de exercício profissional e/ou no número de exames realizados anualmente. Outros consideram o fato do patologista ser renomado internacionalmente.^{25,42,49,71,129,145,146} Entretanto, acredita-se que a eficiência do método esteja diretamente relacionada ao treinamento do profissional, ou seja, ao número de casos analisados.^{11,25,41,42} O que se verifica é que estes critérios são baseados na experiência vivida por cada autor. Como existem diferentes critérios para se definir o diagnóstico citológico sua interpretação requer considerável treinamento.^{39,50}

Existem dados indicando que o patologista só pode ser considerado experiente após examinar, em média, 150 casos ao ano sob supervisão durante o período de formação, e, a partir daí, examinar, aproximadamente, 100 casos anuais.^{42,144,145}

Para Van Herle et al (1982)²⁵, a ampla variabilidade nas taxas de acurácia reflete a existência de diferenças de habilidade e experiência, entre os médicos que puncionam e interpretam a PAAF. Eles sugeriram que a PAAF deveria ser restrita a centros de referência onde profissionais tivessem feito 100 a 200 exames citológicos, inicialmente, para a familiarização com o método, e pelo menos dez exames por semana (cerca de 500 exames ao ano) para a manutenção da habilidade em nível mínimo. Ljung et al (1996)¹⁴⁵ publicaram resultados que mostraram que o treinamento e a realização de, pelo menos, 200 PAAF ao ano são importantes para assegurar alta qualidade.

Haas et al (1993)⁴⁹ relataram os resultados com a PAAF de tireoide onde são realizados, em média, menos de 50 exames ao ano. Durante os quatro anos do estudo, o número de exames aumentou progressivamente em cada ano, passando de 28 para 42, 54 e, por fim, 65, o que possibilitou a oportunidade de avaliar os efeitos do aumento do número de exames sobre os resultados. A proporção de amostras inadequadas diminuiu, inversamente, a cada ano, de 18,0% para 12,0%, 7,0% e, por fim, 6,0%, enquanto o percentual de intervenções cirúrgicas, na qual foi

encontrado tumor maligno, aumentou de 23,0% para 35,0% e 50,0%, respectivamente, nos três últimos anos do estudo. Observaram que houve aumento dos espécimes adequados e da proporção de cânceres em espécimes cirúrgicos coincidente com o aumento do número de exames. Como todas as punções foram realizadas pelo mesmo médico que, também, sempre utilizou a mesma técnica, pode-se extrapolar que o aumento do número de punções teve influência na melhora dos resultados. Entretanto, houve um fator de confusão em relação aos examinadores, uma vez que dois deles abandonaram o estudo ou reduziram suas atividades e dois novos entraram no estudo. Duas inferências podem ser feitas: 1) possibilidade dos novos patologistas possuírem melhor e mais recente treinamento em citologia em relação aos que saíram; 2) aumento da experiência dos novos patologistas com o método ao longo do tempo. Os dados sugerem que os resultados da PAAF de nódulos tireoidianos tendem a melhorar à medida que tanto o puncionador quanto o examinador adquirem maior experiência com o método.

Alguns autores têm questionado a necessidade de se realizar número elevado de exames para se obter bons resultados.^{7,28,48,126} Dessa maneira, seu uso seria excluído de centros que realizam número pequeno ou mesmo moderado de exames.⁷ Para eles, a PAAF poderia ser praticada em estabelecimentos que realizam número modesto de exames com aceitáveis valores de sensibilidade, especificidade e acurácia.^{7,28,46-49,126}

Silverman et al (1986)⁴⁶ descreveram sua experiência com 309 PAAF de tireoide, realizadas durante cinco anos (média de 62 por ano). As punções foram feitas por um endocrinologista e interpretadas por vários patologistas, que obtiveram as seguintes taxas: sensibilidade (93,0%), especificidade (95,1%), VPP (88,9%) e VPN (96,5%).

Liel et al (1985)⁴⁸ publicaram resultados semelhantes em 183 pacientes, onde a média de exames foi de 36 ao ano; na pesquisa de Dwarakanathan et al (1989)⁷ a média foi de 70 exames anuais; para Pepper et al (1989)⁴⁷, resultados aceitáveis podem ser obtidos por médicos que realizam até mesmo 20 a 35 exames por ano, mas eles recomendam que deve existir boa interação entre clínicos e patologistas.

No estudo de Asp et al (1987)²⁸, 155 PAAF foram realizadas em 3,5 anos com média de 44 ao ano. Os exames foram executados por diferentes médicos (três puncionadores e três examinadores). Os valores de sensibilidade, especificidade e acurácia foram, respectivamente, de 100,0%, 47,4% e 73,0%. Os valores desses índices para o CP e para o CF foram, respectivamente (sensibilidade: 100,0% e 100,0%, especificidade: 100,0% e 63,3% e acurácia: 100,0% e 70,2%).

Observaram baixa especificidade para predizer a presença de CF. Os autores concluíram que valores aceitáveis de sensibilidade, especificidade e acurácia, principalmente para o CP, podem ser obtidos mesmo na ausência de experiência tanto com a punção quanto com a interpretação dos espécimes.²⁸

O que se observa a partir desses estudos é que os autores refutam o argumento da necessidade de centros de referência altamente especializados para interpretar materiais citológicos, uma vez que os patologistas podem se tornar proficientes, sem perda significativa de acurácia, mesmo quando interpretam poucos exames anualmente.^{7,28,46-48}

Embora as descrições histológicas propiciem a base para o diagnóstico e a classificação dos tumores e a despeito do uso de nomenclatura e de critérios padronizados são frequentes as variações na reprodutibilidade dos resultados.^{29,57,71,73,75-79,129,147-150} A celularidade, a presença de coloide e a arquitetura do aspirado são aspectos importantes para o diagnóstico de NF.^{27,68} A presença de microfolículos representa sua característica mais sugestiva; no entanto, existe considerável desacordo em relação às características que definem a presença de microfolículos.¹⁴⁶

A sobreposição de aspectos morfológicos histológicos resulta em importante variabilidade diagnóstica na diferenciação entre o BC e o AF e entre o CF, o AF e o CPVF.⁷³ Neste último, tal fato ocorre principalmente quando os aspectos nucleares (fendas nucleares e pseudo-inclusões) do CP não estão bem desenvolvidos ou se encontram presentes apenas focalmente.^{73,79,151} Segundo Elsheikh et al (2008)¹²⁹ não existem critérios bem estabelecidos ou minimamente reprodutíveis para o diagnóstico do CPVF nos casos limítrofes, o que pode resultar em diagnósticos equivocados. Do mesmo modo, a distinção entre o AF e o CF baseia-se na confirmação histológica de presença ou não de invasão vascular e/ou capsular, que muitas vezes é subjetiva.^{15,117,124} Certamente, o CF não-invasivo (intracapsular) deve existir sendo considerado como AF por definição uma vez que não há critérios estabelecidos à microscopia para se fazer tal distinção.⁷³ Assim sendo uma instituição poderá ter muitas variantes foliculares de CP para o diagnóstico citológico de tumor folicular se houver um viés para malignidade.⁷³

O valor de qualquer teste diagnóstico depende da capacidade de que seu resultado seja repetido quando aplicado novamente ao mesmo paciente. Portanto, parâmetro crítico da utilidade clínica de um teste diagnóstico é a sua reprodutibilidade interexaminador e, particularmente, intraexaminador, sendo esta considerada um dos componentes mais importantes para a avaliação global da acurácia do teste diagnóstico.^{63,64} A reprodutibilidade histológica e, principalmente,

citológica dos tumores da tireoide foi muito pouco investigada, tanto no que diz respeito à concordância interobservador quanto intraobservador.^{57,63,68-79} Revisão feita por Kronz & Westra (2005)¹⁵² mostrou taxas de discrepância entre 1,0% e 53,0% para estudos de patologia cirúrgica e entre 17,0% e 60,0% para estudos de exame citológico para todos os sítios da cabeça e pescoço. Alguns estudos focaram, principalmente, a reprodutibilidade diagnóstica das lesões foliculares encapsuladas ou parcialmente encapsuladas da tireoide.^{73,75} Segundo Visvanathan et al (2006)⁶³, há pouca informação disponível sobre a concordância diagnóstica inter e intraobservador entre exames citológicos obtidos pela PAAF de nódulos tireoidianos.

Saxén et al (1978)⁵⁷ avaliaram a concordância histológica entre patologistas. Para tal, uma única lâmina ou poucas lâminas representativas dos exames histológicos de 696 pacientes, registrados nos Centros Nacionais de Câncer da Finlândia, Islândia, Noruega e Suécia, foram revisadas por cinco patologistas nórdicos. Embora eles fossem bem treinados em patologia de tireoide e usassem as mesmas nomenclatura e classificação histológica, o acordo total no diagnóstico só foi observado em 58,0% dos casos e em 82,0% para três deles. O acordo no diagnóstico do CP e do CF foi de 81,0% e 26,0%, respectivamente. Os valores de desacordo foram: CP (7,0%), CA (18,0%), CM (23,0%) e CF (27,0%). O CP é o subtipo histológico mais facilmente reproduzido. Esse achado é consistente com os resultados observados por Fassina et al (1993)⁷⁸ que relataram taxa de acordo global no diagnóstico histológico de 83,0% ($k = 0,68$), sendo quase perfeito para os CA ($k = 0,85$) e CP ($k = 0,81$), substancial para o CM ($k = 0,80$) e moderado para o CF ($k = 0,54$).

Hirokawa et al (2002)⁷³ compararam os resultados de exames histológicos feitos por quatro patologistas americanos e por quatro japoneses que revisaram, independentemente, 21 casos de lesões foliculares encapsuladas ou parcialmente encapsuladas da tireoide. Todos os examinadores fizeram o mesmo diagnóstico em 10,0% dos casos. A taxa de concordância subiu para 29,0% e 76,0% quando se avaliou o diagnóstico fornecido por sete e seis revisores, respectivamente. O acordo global entre patologistas americanos e japoneses foi de 33,0% e 52,0%, respectivamente. A frequência de diagnósticos de BC entre os patologistas japoneses foi consideravelmente maior (31,0%) que entre os americanos (6%). Em contraste, a frequência de diagnósticos de CP para os patologistas americanos foi de 25,0%, e de 4,0% para os japoneses. Quando as lesões estudadas foram distribuídas em duas categorias mais amplas, como benigno para o BC e o AF, e maligno para o CF e o CPVF, o acordo unânime ocorreu em 13 (62,0%) dos

21 casos, ou seja, elevou de 10,0% para 62,0%. Para os autores, os critérios diagnósticos podem não ser inteiramente satisfatórios ou serem inadequados ou difíceis de serem aplicados na prática. Ainda, a subjetividade e a experiência de cada patologista são importantes ingredientes que influenciam o diagnóstico final.

No estudo de Franc et al (2003)⁷⁵, cinco patologistas obtiveram bom nível de acordo ($k = 0,69$) após a revisão de 41 casos de CF da tireoide. Foi analisada a reprodutibilidade interexaminador e estabelecido, em conferência, diagnóstico final de consenso. Um dos patologistas realizou segunda análise, seis meses depois, para a verificação da concordância intraexaminador, cujo resultado foi bom ($k = 0,68$). Os níveis de acordo entre os resultados individuais e o diagnóstico de consenso, obtidos pela estatística k , foram de 0,69, 0,41, 0,35, 0,28 e 0,11. O patologista com maior número de acordos era o que mais examinava, regularmente, amostras de tireoide em relação aos demais. Porém, não foi mencionado o número de exames realizados, em média, por ele. Acordo unânime foi observado em 13 (54,2%) dos 24 casos de CF diagnosticados no consenso final. Dos 41 casos, em 17 (41,4%) o diagnóstico foi alterado, sendo 11 (26,8%) para lesão benigna e seis (14,6%) para outros tipos de tumores malignos. A reprodutibilidade foi aceitável somente para o grupo de CF com invasão óbvia, ruim para o CF minimamente invasivo e para o adenoma atípico, havendo, nestas situações, tendência ao diagnóstico de malignidade em relação ao diagnóstico de consenso. A concordância quanto ao diagnóstico de benignidade ou de malignidade foi próxima de 50,0% para o CPVF e um pouco menor para as lesões foliculares sem o fenótipo papilar. Os resultados indicam que o diagnóstico do CF é baseado em critérios pouco reprodutíveis, como os aspectos nucleares e as invasões, capsular e vascular.

Elsheikh et al (2008)¹²⁹ avaliaram o grau de concordância histológica, inter e intraexaminador, de lesões foliculares da tireoide com aspectos nucleares limítrofes de CP. Para tal, seis patologistas renomados internacionalmente revisaram 15 casos com diagnóstico original suspeito de CPVF. Acordo unânime ocorreu em somente dois (13,0%) pacientes (um com CPVF e outro com tireoidite linfocítica) e em seis (40,0%) casos para quatro patologistas. Acordo para lesões malignas ocorreu em 27,0% e 53,0% para todos e quatro examinadores, respectivamente. O estudo, também, confirmou a existência de ampla variabilidade de critérios diagnósticos do CPVF, que foi evidenciada pelo número discrepante de diagnósticos entre os examinadores, como, por exemplo, um que diagnosticou apenas dois (13,3%) casos (usou critérios mais estritos)

e outro que diagnosticou 14 (93,3%) casos (usou critérios mais liberais). O acordo intraobservador no diagnóstico de CPVF e de malignidade foi de 17,0% a 100,0% e 60,0% a 100,0%, respectivamente. O período de tempo entre a primeira revisão e a segunda variou de dez a 15 meses.

Conclui-se, a partir da análise dos estudos de concordância histológica que mesmo patologistas experientes apresentam pequena taxa de acordo no diagnóstico dos tumores da tireoide, principalmente para as lesões foliculares encapsuladas ou parcialmente encapsuladas bem como para aquelas que possuem predomínio de coloide e para aqueles que possuem aspectos nucleares incompletos ou limítrofes do CP.^{29,57,71,73-75,78,79,129,147,150} Estes estudos poderiam até mesmo produzir argumentos de defesa em caso de processos médico-legais.¹⁵³ Não se pode esperar perfeita correlação cito-histológica se a própria correlação histo-histológica não é confiável.⁵³ Elevadas taxas de desacordo foram mostradas em vários estudos.^{57,73,75-79,149} Segundo Hall et al (1989)⁵³, este fato é subapreciado, o que resulta na atribuição de ausência de correlação cito-histológica, a priori, a erro de diagnóstico por parte dos citopatologistas.

Clary et al (2005)⁶⁸ estudaram o grau de variabilidade interexaminador na categorização citológica de lesões foliculares e analisaram sua precisão diagnóstica por meio de correlação cito-histológica de 50 exames (21 lesões foliculares e 29 NF). Lâminas representativas de cada caso foram selecionadas e revistas por quatro patologistas da Universidade de Pittsburgh-EUA. A concordância interobservador entre os quatro patologistas foi fraca para substancial ($K = 0,20 - 0,62$). A acurácia do diagnóstico citológico em predizer o resultado, corretamente, para NF foi de 77,0% a 90,0%, e de 53,0% a 74,0% para diagnósticos de lesões benignas. Este estudo deixa patente a dificuldade existente em categorizar citologicamente as lesões foliculares da tireoide, uma vez que mesmo entre patologistas pertencentes à mesma instituição, existe grande desacordo quanto ao seu diagnóstico citológico.

Stelow et al (2005)⁷¹ estudaram a variabilidade interexaminador na interpretação microscópica de lesões foliculares e NF com o BC.^{16,18,85,128} Para tal, 20 casos diagnosticados por citopatologista experiente foram revisados, de maneira cega, por outros quatro. Quando se considera a lesão folicular e a NF como entidades distintas, a variabilidade interexaminador foi alta ($K = 0,35$), com consenso em seis casos. Quando elas foram consideradas como entidade única, a variabilidade diminuiu ($K = 0,57$), com consenso em dez casos. Finalmente, quando foram dicotomizadas em afecções que requerem cirurgia e as que não requerem cirurgia, a

variabilidade diminuiu ainda mais ($K = 0,65$). Para os autores, a NF e lesão folicular deveriam ser consideradas como entidade única, e, ainda assim, existiriam casos cujo diagnóstico seria muito difícil. Entretanto, o número de casos analisados foi pequeno para se extrair conclusão definitiva. Além do mais, participaram do estudo citopatologistas de duas instituições diferentes, o que poderia explicar, em parte, as discrepâncias observadas (efeito escola de ensino), como observado em outros trabalhos.^{73,75,154}

Gerhard & Santos (2007)⁷² publicaram estudo sobre a reprodutibilidade, inter e intraexaminador, de 97 exames citológicos da tireoide. Os critérios citológicos usados para definir os diagnósticos foram baseados nas orientações da Sociedade Papanicolaou de Citopatologia (SPC). A reprodutibilidade interexaminador foi determinada pela comparação dos diagnósticos citológicos emitidos por dois patologistas (sendo um experiente, pois analisava mais de 650 exames ao ano) com o diagnóstico final (este incluiu os resultados concordantes e o diagnóstico de consenso para os casos discordantes). A reprodutibilidade intraexaminador foi obtida por meio da comparação do diagnóstico citológico feito por um dos observadores com o diagnóstico citológico original (DCO), que correspondia ao diagnóstico emitido previamente por ele mesmo. Os autores reportaram taxas substanciais de reprodutibilidade interexaminador ($k = 0,71$) com desacordo em 24 (24,7%) casos, e intraexaminador ($k = 0,66$) com desacordo em 23 (23,7%) casos. A discordância foi maior para as lesões foliculares ($n = 18$), o que representa 78,3% dos casos discordantes ($n = 23$). A precisão diagnóstica foi estabelecida mediante correlação cito-histológica em 79 (81,4%) casos. Os casos insuficientes, as lesões císticas e as lesões foliculares ($n = 18$) foram excluídos da análise. Os valores obtidos foram: sensibilidade (89,1%), especificidade (97,0%), VPP (97,6%), VPN (86,5%) e acurácia (92,4%). Ocorreram cinco (6,3%) resultados falso-negativos e um (1,3%) falso-positivo.

OBJETIVOS

1. Avaliar a influência da experiência do examinador na interpretação dos achados citológicos da tireoide, em termos de reprodutibilidade inter e intraexaminador;
2. Correlacionar os resultados citológicos emitidos pelos patologistas (com e sem experiência) com os exames histológicos das peças cirúrgicas;
3. Avaliar a influência do uso da DAM em relação ao uso da DEM sobre os cálculos das medidas de validação e predição diagnóstica.

MÉTODO

Casuística

Entre agosto de 2002 e janeiro de 2005 foram realizadas 225 PAAF, em 215 pacientes, dos quais 180 (83,7%) foram operados, sendo 18 (10,0%) homens e 162 (90,0%) mulheres. A média de PAAF realizadas no Ambulatório Bias Fortes (ABF) do Hospital das Clínicas (HC) da Universidade Federal de Minas Gerais (UFMG) foi de 7,5 exames por mês (1,7 exames por semana). A tireoidectomia total foi realizada em 105 (58,3%) pacientes, a tireoidectomia parcial em 69 (38,3%) e a tireoidectomia subtotal em seis (3,3%). Em 106 (47,1%) casos, os critérios de inclusão não foram preenchidos. Fizeram parte do estudo 74 (41,1%) pacientes dos 180 iniciais, sendo 5 (6,8%) homens e 69 (93,2%) mulheres (relação homem / mulher de 1:13,8). Destes, em 17 (23,0%) foram identificados tumores malignos ao exame histológico e em 57 (77,0%) lesões benignas

Método

Todas as punções foram realizadas no ABF do HC-UFMG. A interpretação dos exames citológicos foi realizada por residentes de Anatomia Patológica juntamente com professores lotados no Departamento de Anatomia Patológica da Faculdade de Medicina (FM) - UFMG, e correspondeu ao DCO, arquivado no prontuário médico. Em todos os casos, o puncionador foi o médico residente em Patologia (terceiro ano de residência) sempre auxiliado por um cirurgião de cabeça e pescoço, membro do Grupo de Cirurgia de Cabeça e Pescoço do Instituto Alfa de Gastroenterologia do HC - UFMG. A punção guiada por US não foi realizada em nenhum caso.

Os critérios de inclusão utilizados foram:

- Ter sido submetido a PAAF e a tireoidectomia;
- Possuir os resultados dos exames, citológico e anátomo-patológico, arquivados no prontuário médico;
- Possuir, no mínimo, duas lâminas por exame, sendo pelo menos uma fixada em álcool e corada pela HE ou Papanicolaou.

Todos os dados observados foram apresentados de forma agregada, de modo que nenhuma característica de identificação de qualquer participante constou na descrição dos

resultados. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da UFMG cujo número do parecer é ETIC 383/04 (ANEXO I).

Os achados citológicos foram distribuídos em quatro categorias diagnósticas (benigno, maligno, NF e insuficiente para análise) com base no guia da SPC (ANEXO II).¹⁷ Lesões indeterminadas ou suspeitas, NF, neoplasia de células de Hürthle, CPVF e aquelas que apresentavam características atípicas sugestivas, mas não diagnósticas de malignidade, foram incluídas dentro da denominação de NF, como recomendado por Castro e Gharib (2005)³⁴ e Stelow et al (2005)⁷¹.

Lesão folicular e NF foram consideradas como uma única entidade e os sinônimos diagnósticos foram agrupados sob uma única denominação. Nas situações em que havia dois diagnósticos, foi escolhido para designar o caso aquele que requer tratamento mais agressivo (por exemplo: na presença de BC e de NF, foi escolhido o segundo como diagnóstico do caso).

As lâminas dos exames citológicos dos 74 pacientes, obtidas por PAAF, foram analisadas por dois patologistas, um com e outro sem experiência na interpretação dos mesmos. As análises foram feitas de maneira independente, cega e sem o conhecimento prévio da história clínica e dos exames diagnósticos (US, cintilografia, DCO e anátomo-patológico) de cada paciente, como recomendado por alguns autores.^{146,155,156} As revisões ocorreram em duas ocasiões distintas, com intervalo mínimo de tempo de seis meses entre a primeira e a segunda análise. Os diagnósticos feitos por cada examinador foram mantidos em sigilo, de modo que os diagnósticos feitos por um observador permaneceram desconhecidos pelo outro. O método usado na pesquisa e os objetivos pretendidos foram omitidos dos participantes, assim como a instituição de origem dos pacientes. Qualquer sinal presente nas lâminas que pudesse propiciar a identificação foi ocultado.

O patologista experiente foi designado pela letra A, enquanto o patologista sem experiência pela letra B.⁷⁰ As análises das lâminas ocorreram em dois momentos, sendo no primeiro, denominadas por A1 e B1, e no segundo, após seis meses, por A2 e B2, respectivamente, para os patologistas A e B. Posteriormente, foram feitas as análises da correspondência, inter e intraexaminador, entre os laudos emitidos por ambos. A primeira análise compreendeu a interpretação de 74 exames e a segunda análise a interpretação dos mesmos 74 exames, porém renumerados aleatoriamente e ordenados na seqüência a partir do número 75, inclusive (reinício da contagem pelo número 75 e término no 148 para os mesmos 74 exames iniciais), sem que eles tomassem conhecimento da renumeração. O desenho esquemático do

estudo está demonstrado na figura 1. Os critérios usados para a definição dos casos e de insuficiência das amostras são mostrados no ANEXO III.

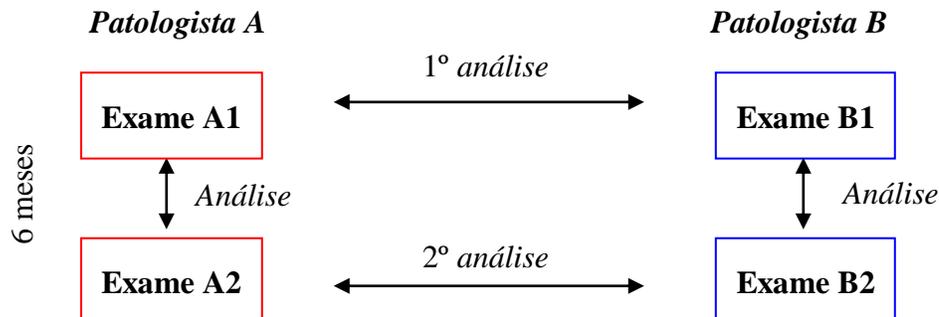


FIGURA 1: Desenho esquemático do estudo (correlação inter e intraexaminador).

A avaliação da reprodutibilidade interexaminador foi feita mediante a comparação dos laudos citológicos emitidos pelos patologistas A e B após a primeira (A1 - B1) e a segunda (A2 - B2) análise. Para a verificação da reprodutibilidade intraexaminador, foram comparados os laudos citológicos emitidos por cada patologista com os laudos emitidos por eles mesmos seis meses após (A1 - A2 e B1 - B2).

Características dos patologistas participantes

Para a escolha dos patologistas, foram adotados critérios que levaram em consideração características que deveriam ser comuns a ambos e critérios que definissem a experiência ou a ausência desta na interpretação de exames citológicos da tireoide. As características comuns a ambos os patologistas foram:

- Graduação na mesma Universidade;
- Especialização em Patologia no mesmo Serviço (mínimo de dois anos);
- Atuação profissional, como patologistas, na mesma cidade;

Critérios de experiência na interpretação de exames citológicos da tireoide:

- Possuir certificação de Especialista pela Sociedade Brasileira de Patologia (SBP);
- Interpretar pelo menos 10 exames por semana (aproximadamente 500 exames por ano);
- Apresentar o mínimo de dez anos de atividade profissional.

Critérios de ausência de experiência na interpretação de exames citológicos da tireoide:

- Máximo de dois anos de conclusão da especialização em Patologia;
- Interpretar menos de 40 exames citológicos da tireoide por ano.

O patologista A graduou-se em Medicina pela UFMG, fez a especialização em Patologia no HC-UFMG e é membro titular da SBP. Ele tem mais de 25 anos de formado e avalia, aproximadamente, 500 exames ao ano, cerca de dez exames semanais. O patologista B graduou-se em Medicina pela UFMG, fez a especialização em patologia no HC-UFMG e não é membro da SBP. Ele tinha três anos de formado e um ano de conclusão da especialização. Avaliou, em média, 40 exames por ano nos últimos dois anos. Ambos atuam como patologistas na cidade de Belo Horizonte.

Análise estatística

O programa EPI-INFO, versão 6.04 (<http://www.cdc.gov>), foi utilizado para montagem do banco de dados e para a realização dos cálculos estatísticos. Para a verificação das reprodutibilidades, inter e intraexaminador, de variáveis qualitativas, foram utilizados os testes *Kappa* (K) e *Kappa ponderado* (Kp), como descritos por Cohen, respectivamente, em 1960 e 1968 e citados por Posner et al (1990)¹⁵⁸, Franc et al. (2003)⁷⁵ e Viera & Garrett (2005)¹⁵⁹.

Como o valor de K usado para representar bom acordo é arbitrário, utilizamos para a interpretação do grau de concordância (inter e intraexaminador) a classificação que contém os valores sugeridos por Landis & Koch (1977)^{160,161} (ANEXO IV). A verificação da existência ou não de diferença estatística entre os resultados foi feita mediante a obtenção do valor p e o cálculo estatístico do teste qui-quadrado (X^2) (com hipóteses bilaterais e um ou três graus de liberdade). Foi estabelecido o nível de significância de 5% ($p < 0,05$) para os resultados encontrados. Para o *odds ratio* (OR) e para o K foram calculados os intervalos de confiança (IC).¹⁶²

Para a verificação da precisão diagnóstica da PAAF na detecção de tumores malignos da tireoide foi realizada correlação cito-histológica, considerando-se a PAAF como o exame diagnóstico a ser testado. A presença da doença foi estabelecida pelo exame padrão (exame anátomo-patológico da glândula tireoide). Tal procedimento padrão foi considerado como representativo do estado verdadeiro da doença (se presente ou ausente). Os casos diagnosticados como insuficiente para o diagnóstico foram excluídos das análises de validação e predição

(sensibilidade, especificidade, acurácia, VPP, VPN, RVP, RVN, falso-positivos e falso-negativos), uma vez que não fornecem qualquer informação diagnóstica. Também foram calculadas as proporções de resultados insuficientes, de tumores malignos diagnosticados e de casos compatíveis com NF. Os cálculos foram feitos considerando-se os resultados maligno, benigno e NF, sendo o último (NF) considerado de duas maneiras distintas, como sugerido por vários autores:^{3,7,14,15,25,28,53,61,106,126,157}

- Benignos → definição estreita de malignidade (DEM);
- Malignos → definição ampla de malignidade (DAM).

RESULTADOS

Geral

Os achados citológicos observados pelos patologistas A e B após todas as análises (A1, A2, B1 e B2) são mostrados na tabela 1. Estes achados com os respectivos diagnósticos histológicos são apresentados no ANEXO V. As lesões císticas relatadas pelo patologista B (B2) foram consideradas benignas, uma vez que havia a descrição clara de “ausência de células malignas compatível com cisto colóide”. Tais lesões foram consideradas pelo patologista A como insuficientes para o diagnóstico.

TABELA 1: Achados citológicos observados pelos patologistas A e B após as quatro análises (A1, A2, B1 e B2) (n = 74).

Achado citológico	Examinador			
	A1	A2	B1	B2
Maligno	9	9	10	8
Benigno	37	39	51	44
Neoplasia Folicular	2	2	4	10
Insuficiente	26	24	9	12
Total	74	74	74	74

O número absoluto e proporcional de achados insuficientes observado em cada análise pelos patologistas A e B, assim como os correspondentes diagnósticos histológicos de câncer e as taxas de malignidade nestas amostras são mostrados na tabela 2. A comparação desses resultados e os respectivos cálculos estatísticos estão expressos na tabela 3.

TABELA 2: Números absolutos e proporcionais de achados insuficientes observados em cada análise (A1, A2, B1 e B2), com os correspondentes diagnósticos histológicos de câncer.

Examinador	Insuficiente		Tumor maligno		
	N	%*	n	%**	Diagnóstico histológico
A1	26	35,1	4	15,4	3 carcinomas papilares 1 carcinoma folicular
A2	24	32,4	2	8,3	2 carcinomas papilares
B1	9	12,2	1	11,0	1 carcinoma papilar
B2	12	16,2	1	8,3	1 carcinoma papilar

* % em relação ao total de casos (n = 74);

** % em relação ao total de casos insuficientes.

TABELA 3: Comparação entre os resultados apresentados pelos patologistas A e B em relação aos casos insuficientes, com os respectivos cálculos estatísticos.

Grupos de comparação	p valor	Odds ratio	Intervalo de confiança
A1 – A2	0,72	1,13	0,54 – 2,36
B1 – B2	0,48	0,72	0,26 – 1,98
A1 – B1	0,001	3,91	1,57 – 9,97
A2 – B2	0,021	2,48	1,06 – 5,88

Teste X^2 com um grau de liberdade.

Com relação aos casos de NF, o patologista A observou dois (2,7%) casos tanto na primeira (A1) quanto na segunda (A2) revisão. Ambos foram concordantes e corresponderam, no exame histológico, a dois casos de câncer (um CP e um CF). O patologista B identificou quatro (5,4%) casos na primeira revisão (B1) e dez (13,5%) na segunda (B2). Os quatro diagnósticos concordantes corresponderam a: BC (n = 2), CP (n = 1) e CF (n = 1). Os outros seis casos observados em B2 corresponderam a: CP (n = 4), AF (n = 1) e BC (n = 1). O número global e a proporção de lesões citologicamente diagnosticadas como NF e a taxa de malignidade nestas amostras, com os correspondentes diagnósticos histológicos de câncer são mostrados na tabela 4.

Os valores dos cálculos estatísticos feitos para as comparações, interexaminador (A1 - B1) e (A2 - B2) e intraexaminador (A1 - A2) e (B1 - B2) são mostrados na tabela 5.

TABELA 4: Números absolutos e proporção de lesões citologicamente diagnosticadas como neoplasia folicular observados pelos patologistas A e B nas quatro análises (A1, A2, B1 e B2) com as correspondentes taxas de malignidade e diagnósticos histológicos de câncer.

Examinador	Neoplasia folicular	Tumor maligno	
	n (%) [*]	n (%) ^{**}	Diagnóstico histológico
A1	2 (2,7)	2 (100)	1 carcinoma papilar 1 carcinoma folicular
A2	2 (2,7)	2 (100)	1 carcinoma papilar 1 carcinoma folicular
B1	4 (5,4)	2 (50)	1 carcinoma papilar 1 carcinoma folicular
B2	10 (13,5)	6 (60)	5 carcinomas papilares 1 carcinoma folicular

*% em relação ao total de casos (n =74);

**% em relação ao total de casos de neoplasia folicular.

TABELA 5: Comparação entre os resultados apresentados pelos patologistas A e B em relação aos casos de neoplasia folicular, com os respectivos cálculos estatísticos.

Grupos de comparação	<i>p</i> valor	<i>Odds ratio</i>	Intervalo de confiança
A1 - A2	1,0	1,0	0,10 – 10,27
B1 – B2	0,09	2,73	0,74 – 10,96
A1 - B1	0,40	0,49	0,06 – 3,22
A2 - B2	0,015	5,63	1,09 – 38,71

Teste X^2 com um grau de liberdade.

O número absoluto e a proporção de resultados falso-negativos, assim como os resultados da comparação entre essas taxas, obtidas por cada examinador (A e B) com os respectivos cálculos estatísticos, estão apresentados nas tabelas 6, 7 e 8.

TABELA 6: Resultados falso-negativos observados por cada examinador (A1, A2, B1 e B2) com os respectivos diagnósticos histológicos considerando-se a definição estreita de malignidade (DEM) e a definição ampla de malignidade (DAM).

Examinador	DEM			DAM		
	n (%) [*]	Diagnóstico citológico	Diagnóstico histológico	n (%) [*]	Diagnóstico citológico	Diagnóstico histológico
A1	5 (38,5)	3 BC ^{**} 2 NF [†]	3 CP ^{***} 1 CP, 1 CF ^{††}	3 (23,1)	3 BC	3 CP
A2	7 (46,7)	5 BC 2 NF	4 CP, 1 CF 1 CP, 1 CF	5 (33,4)	5 BC	4 CP, 1 CF
B1	7 (43,8)	5 BC 2 NF	4 CP, 1 CF 1 CP, 1 CF	5 (31,3)	5 BC	4 CP, 1 CF
B2	9 (56,3)	3 BC 6 NF	2 CP, 1 CF 5 CP, 1 CF	3 (18,8)	3 BC	2 CP, 1 CF

^{*} % em relação ao total de tumores malignos excluindo-se os resultados insuficientes (A1 = 13), (A2 = 15), (B1 = 16) e (B2 = 16);

^{**} BC: bócio colóide;

^{***} CP: carcinoma papilífero;

[†] NF: neoplasia folicular;

^{††} CF: carcinoma folicular.

TABELA 7: Comparação entre as taxas de falso-negativos apresentadas pelos patologistas A e B, após todas as revisões (A1, A2, B1 e B2), com os respectivos cálculos estatísticos, considerando-se os casos de neoplasia folicular benigna.

Grupos de comparação	<i>p</i> valor	<i>Odds ratio</i>	Intervalo de confiança
A1 – A2	0,55	0,69	0,18 – 2,60
B1 – B2	0,60	0,75	0,24 – 2,38
A1 – B1	0,55	0,69	0,18 – 2,60
A2 – B2	0,60	0,75	0,24 – 2,38

Teste X^2 com um grau de liberdade.

TABELA 8: Comparação entre as taxas de falso-negativos apresentadas pelos patologistas A e B, após todas as revisões (A1, A2, B1 e B2), com os respectivos cálculos estatísticos, considerando-se os casos de neoplasia folicular maligna.

Grupos de comparação	<i>p</i> valor	<i>Odds ratio</i>	Intervalo de confiança
A1 – A2	0,72	0,58	0,11 – 2,95
B1 – B2	0,72	1,71	0,34 – 9,47
A1 – B1	0,72	0,58	0,11 – 2,95
A2 – B2	0,72	1,71	0,34 – 9,47

Teste X^2 com um grau de liberdade.

O número absoluto e a proporção de resultados falso-positivos observados por cada examinador (A e B) são mostrados na tabela 9. A comparação entre as taxas de falso-positivos apresentadas pelos patologistas A e B, usando-se a DAM, com os respectivos cálculos estatísticos, é mostrada na tabela 10.

TABELA 9: Resultados falso-positivos observados por cada examinador (A1, A2, B1 e B2) com os respectivos diagnósticos histológicos, considerando-se a definição estreita de malignidade (DEM) e a definição ampla de malignidade (DAM).

	DEM			DAM		
	n (%) [*]	Diagnóstico citológico	Diagnóstico histológico	n (%) [*]	Diagnóstico citológico	Diagnóstico Histológico
A1	1 (3,0)	CI ^{**}	BC ^{**}	1 (3,0)	CI	BC
A2	1 (3,0)	CI	BC	1 (3,0)	CI	BC
B1	1 (2,0)	CM [†] (CI?)	BC	3 (6,0)	CM (CI?) 2 NF ^{††}	3 BC
B2	1 (2,0)	CM (CI?)	BC	5 (10,9)	CM (CI?) 3 NF NF	BC 3 BC AF ^{†††} / BC

* % em relação ao total de casos benignos excluindo-se os resultados insuficientes (A1 = 35), (A2 = 35), (B1 = 49) e (B2 = 46);

** CI: carcinoma indiferenciado; *** BC: bócio coloide; † CM: carcinoma medular; †† NF: neoplasia folicular; ††† AF: adenoma folicular;

TABELA 10: Comparação entre as taxas de falso-positivos apresentadas pelos patologistas A e B, após todas as revisões (A1, A2, B1 e B2), com os respectivos cálculos estatísticos, considerando-se os casos de neoplasia folicular maligna.

Grupos de comparação	<i>p</i> valor	<i>Odds ratio</i>	Intervalo de confiança
A1 – A2	1,0	1,0	0,0 – 37,40
B1 – B2	0,72	1,71	0,34 – 9,47
A1 – B1	0,62	3,08	0,28 – 78,85
A2 – B2	0,21	5,29	0,58 – 122,74

Teste X^2 com um grau de liberdade.

Ocorreram 29 (39,2%) resultados divergentes em pelo menos uma das quatro análises (A1, A2, B1 e B2) e 45 (60,8%) resultados concordantes em todas as análises. Os casos concordantes com os correspondentes diagnósticos histológicos são mostrados na tabela 11.

TABELA 11: Casos concordantes (n = 45) nas quatro análises (A1, A2, B1 e B2), realizadas pelos patologistas A e B, com os correspondentes diagnósticos histológicos.

Citologia concordante	N	%*	Correspondente histológico
Insuficientes	9	20,0	Bócio coloide 8 Carcinoma papilar 1
Bócio colóide	27	60,0	Bócio coloide 21 Adenoma folicular 4 Carcinoma papilar 2
Carcinoma papilar	4	8,9	Carcinoma papilar 4
Carcinoma medular	2	4,5	Carcinoma medular 2
Tireoidite de Hashimoto	2	4,5	Tireoidite de Hashimoto 2
Neoplasia folicular	1	2,1	Carcinoma folicular 1
Total	45	100,0	45

*% em relação ao total de casos (n = 74).

Foram confirmados 17 (23,0%) tumores malignos ao exame anátomo-patológico, a saber: CP (n = 13; 76,4%), CM (n = 2; 11,8%) e CF (n = 2; 11,8%). Destes, em seis (35,3%) houve acordo unânime nas quatro revisões (A1, A2, B1 e B2) e corresponderam acertadamente aos seguintes diagnósticos histológicos: quatro CP (30,8% dos CP) e dois CM (100,0% dos CM). Em quatro casos, apesar de ter havido concordância citológica nas quatro revisões, ocorreu divergência em relação ao resultado histológico, a saber: um CF diagnosticado como NF e três CP diagnosticados como BC (n = 2) e insuficiente (n = 1). Nos sete casos restantes (6 CP e 1 CF), houve pelo menos um achado citológico discordante após as quatro análises.

O número absoluto e a proporção de tumores malignos, em relação ao total de cânceres (n = 17), diagnosticados após cada análise, considerando-se a DEM foi: A1 (n = 8; 47,1%), A2 (n = 8; 47,1%), B1 (n = 9; 59,2%) e B2 (n = 7; 41,2%). As seguintes alterações ocorreram quando se

usou a DAM: A1 (n = 10; 58,8%), A2 (n = 10; 58,8%), B1 (n = 11; 64,7%) e B2 (n = 13; 76,5%). A comparação entre os resultados da proporção de tumores malignos observada pelos patologistas A e B (inter e intraexaminador) com os respectivos cálculos estatísticos está demonstrada nas tabelas 12 e 13.

TABELA 12: Comparação da proporção de tumores malignos diagnosticados pelos patologistas A e B inter (A1 – B1 e A2 – B2) e intraexaminador (A1 – A2 e B1 – B2), considerando-se os casos de neoplasia folicular benigna.

Conjuntos de observações	<i>p</i> valor	<i>Odds ratio</i>	Intervalo de confiança
A1 - B1	0,79	0,88	0,29 – 2,67
A2 - B2	0,78	1,16	0,36 – 3,81
A1 - A2	1,0	1,0	0,34 – 2,97
B1 – B2	0,59	1,33	0,42 – 4,24

Teste X^2 com um grau de liberdade.

TABELA 13: Comparação da proporção de tumores malignos diagnosticados pelos patologistas A e B inter (A1 – B1 e A2 – B2) e intraexaminador (A1 – A2 e B1 – B2), considerando-se os casos de neoplasia folicular maligna.

Conjuntos de observações	<i>p</i> valor	<i>Odds ratio</i>	Intervalo de confiança
A1 - B1	0,81	0,89	0,32 – 2,47
A2 - B2	0,50	0,73	0,27 – 1,95
A1 - A2	1,0	1,0	0,37 – 2,70
B1 – B2	0,66	0,82	0,31 – 2,14

Teste X^2 com um grau de liberdade.

Dos 13 CP diagnosticados à histologia (17,6% do total e 76,5% dos tumores malignos) houve acordo no diagnóstico interexaminador, após a primeira (A1 - B1) análise em cinco (38,5%) casos e após a segunda (A2 - B2) em quatro (30,8%) casos ($p = 0,68$; OR = 1,41 [0,21 –

9,66]). Para as análises intraexaminadores, foi observada concordância em seis (46,2%) casos para o patologista A e em quatro (30,8%) casos para o patologista B ($p = 0,42$; $OR = 1,93$ [0,30 – 13,14]). Os dois casos de CM foram corretamente identificados em todas as análises.

Reprodutibilidade intraexaminador A1 – A2

Houve concordância em 66 (89,2%) dos 74 casos ($K = 0,82$ [0,71 – 0,94]; $Kp = 0,79$; $p = 0,988$) (Tabela 14). O acordo foi maior para os casos benignos ($n = 34$; 45,9% do total e 59,6% dos casos benignos), dos quais 31 (86,1%) foram confirmados à histologia (BC ($n = 26$) e AF ($n = 5$)) enquanto três (13,9%) foram compatíveis com CP (falso-negativos). A NF foi acordada em duas situações correspondendo a um CP e um CF. Vinte e um achados foram insuficientes e concordantes; eles foram diagnosticados, histologicamente, como: BC ($n = 16$), BC associado à tireoidite de Hashimoto ($n = 1$), CP ($n = 2$) e AF ($n = 2$). Ocorreram nove (52,9% dos casos malignos) acordos para malignidade (CM ($n = 2$), CP ($n = 6$) e CI ($n = 1$)); o CI foi falso-positivo, uma vez que o diagnóstico final foi compatível com BC. Nos demais houve confirmação diagnóstica.

Oito (10,8%) dos 74 casos foram discordantes, com cinco casos considerados insuficientes por A1, porém benignos por A2. Três casos foram interpretados como insuficientes por A2, porém, benignos por A1.

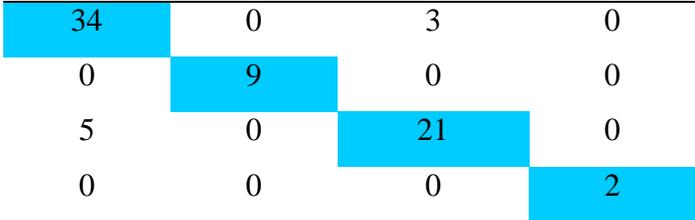
TABELA 14: Reprodutibilidade intraexaminador para o patologista A (A1 – A2).

Exame A1	Exame A2				Total
	Benigno	Maligno	Insuficiente	NF*	
Benigno	34	0	3	0	37
Maligno	0	9	0	0	9
Insuficiente	5	0	21	0	26
NF	0	0	0	2	2
Total	39	9	24	2	74

$K = 0,82$ [0,71 – 0,94]; $Kp = 0,79$

$X^2 = 0,13$ (3 graus de liberdade); $p = 0,988$

* NF: neoplasia folicular

 resultados concordantes

Reprodutibilidade intraexaminador B1 – B2

Houve 65 (87,8%) resultados concordantes ($K = 0,78$ [0,64 – 0,91]; $Kp = 0,71$; $p = 0,291$), sendo a maioria para benigno ($n = 44$; 59,5% do total e 77,2% dos casos benignos) (Tabela 15). As quatro lesões císticas identificadas por B2 foram consideradas benignas (BC). Dos 44 casos, 41 (93,2%) foram confirmados à histologia (34 BC e 7 AF) e três (6,8%) foram divergentes (2 CP e 1 CF). Nove (12,2%) achados foram concordantes com amostras insuficientes e corresponderam, histologicamente, a oito casos de BC e um de CP. Oito (47,1% dos casos malignos) casos de câncer (4 CP e 4 CM) foram concordantes, sendo que em um deles (CM), na segunda revisão (B2), foi questionada a possibilidade de CI. Todos os diagnósticos foram concordantes, exceto dois casos de CM que corresponderam, histologicamente, a BC ($n = 1$) e CP ($n = 1$). Houve quatro (5,4%) acordos para NF, revelados pela histologia como BC ($n = 2$), CP ($n = 1$) e CF ($n = 1$).

Nove (12,2%) resultados foram discordantes sendo dois (2,7%) casos considerados por B1 como CP e por B2 como NF e sete (9,5%) casos compatíveis com benigno por B1 considerados por B2 como: NF ($n = 4$) e insuficiente ($n = 3$).

TABELA 15: Reprodutibilidade intraexaminador para o patologista B (B1 – B2).

Exame B1	Exame B2				Total
	Benigno	Maligno	Insuficiente	NF*	
Benigno	44	0	3	4	51
Maligno	0	8	0	2	10
Insuficiente	0	0	9	0	9
NF	0	0	0	4	4
Total	44	8	12	10	74

$K = 0,78$ [0,64 – 0,91]; $Kp = 0,71$

$X^2 = 3,74$ (3 graus de liberdade); $p = 0,291$

 resultados concordantes

* NF: neoplasia folicular

Reprodutibilidade interexaminador A1 – B1

A concordância foi observada em 54 (73,0%) casos ($K = 0,55$ [0,38 – 0,72]; $Kp = 0,50$; $p = 0,011$), sendo a maioria para benigno ($n = 36$; 48,6% do total e 63,2% dos casos benignos) (Tabela 16). Dos 36 casos benignos, 33 (91,7%) foram confirmados no exame histológico (BC ($n = 28$) e AF ($n = 5$)) enquanto três (8,3%) corresponderam ao CP. Nove (12,2%) achados foram concordantes com laudo insuficiente e corresponderam, histologicamente, a BC ($n = 8$) e a CP ($n = 1$). Ocorreram oito (47,1% dos casos malignos) acordos para malignidade (6 CP e 2 CM). Todos foram concordantes com os respectivos laudos histológicos. Houve um acordo para a NF (1,4%), confirmado como CF no exame histológico.

Vinte (27,0%) resultados foram discordantes, sendo 17 laudos reportados por A1 como insuficientes e por B1 como: benigno ($n = 13$), NF ($n = 2$), CP ($n = 1$) e CM ($n = 1$). Um caso de CP e um de NF diagnosticados por A1 foram considerados benignos por A2. Um caso relatado por A1 como benigno foi identificado como NF por B1.

TABELA 16: Reprodutibilidade interexaminador para os patologistas A e B na primeira análise de lâminas (A1 – B1).

Exame A1	Exame B1				Total
	Benigno	Maligno	Insuficiente	NF*	
Benigno	36	0	0	1	37
Maligno	1	8	0	0	9
Insuficiente	13	2	9	2	26
NF	1	0	0	1	2
Total	51	10	9	4	74

$K = 0,55$ [0,38 – 0,72]; $Kp = 0,50$

$X^2 = 11,20$ (3 graus de liberdade); $p = 0,011$

 resultados concordantes

* NF: neoplasia folicular

Reprodutibilidade interexaminador A2 – B2

Houve concordância em 53 (71,6%) casos ($K = 0,54$ [0,37 – 0,71]; $Kp = 0,44$; $p = 0,021$), sendo a maioria compatível com benigno ($n = 33$; 44,6% do total e 57,9% dos casos benignos) (Tabela 17). Dos 33 casos, 30 (91,0%) foram confirmados no exame histológico, incluindo cinco casos de AF; três (9%) divergiram do exame histológico (CP ($n = 2$) e CF ($n = 1$)). Ocorreram 11 (14,9%) acordos quanto à insuficiência para a interpretação citológica, que corresponderam, no exame histológico, a um caso de CP e dez de BC. A concordância para malignidade ocorreu em sete (41,2% dos casos malignos) casos, a saber: CP ($n = 4$), CM ($n = 2$) e CI ($n = 1$). Em todos, o diagnóstico foi confirmado à histologia, exceto o CI que foi compatível com BC (falso-positivo). Houve dois (2,7%) acordos no diagnóstico de NF, que foram confirmados histologicamente como CF ($n = 1$) e CP ($n = 1$).

Vinte e um (28,4%) resultados foram discordantes. O patologista A (A2) reportou 13 laudos como insuficientes que foram relatados por B2 como: BC ($n = 11$), NF ($n = 1$) e CM ($n = 1$); A2 identificou seis casos como benignos, que foram considerados por B2 como NF ($n = 5$) e insuficiente ($n = 1$). Dois CP encontrados por A2 foram diagnosticados por B2 como NF.

TABELA 17: Reprodutibilidade interexaminador para os patologistas A e B na segunda análise de lâminas (A2 – B2).

Exame A2	Exame B2				Total
	Benigno	Maligno	Insuficiente	NF*	
Benigno	33	0	1	5	39
Maligno	0	7	0	2	9
Insuficiente	11	1	11	1	24
NF	0	0	0	2	2
Total	44	8	12	10	74

$K = 0,54$ [0,37 – 0,71]; $Kp = 0,44$

$X^2 = 9,69$ (3 graus de liberdade); $p = 0,021$

 resultados concordantes

* NF: neoplasia folicular

A tabela 18 apresenta os valores obtidos para os cálculos estatísticos feitos após a comparação das proporções de acordo, inter e intraexaminador, feitas para os patologistas A e B.

TABELA 18: Comparação entre as taxas de concordância, inter (A1 - B1 e A2 - B2) e intraexaminador (A1 - A2 e B1 - B2), observadas para os patologistas A e B, obtidas pelo coeficiente *Kappa* (*K*) e *Kappa ponderado* (*Kp*) com os respectivos critérios fornecidos por Landis & Koch.¹⁶¹

Grupos de comparação	<i>K</i>	Landis e Koch	<i>Kp</i>	Landis e Koch	X^2	<i>p</i> valor
A1 → B1	0,55	Moderado	0,50	Moderado	11,20	0,011
A2 → B2	0,54	Moderado	0,44	Moderado	9,69	0,021
A1 → A2	0,82	Muito bom	0,79	Bom	0,13	0,988
B1 → B2	0,78	Bom	0,71	Bom	3,74	0,291

Examinador A1

Em 48 (65,0%) casos foi atribuído diagnóstico que foi dicotomizado em benigno ou maligno de acordo com a DEM (benigno (n = 39; 81,3%) e maligno (n = 9; 18,7%)), e DAM (benigno (n = 37; 77,1%) e maligno (n = 11; 22,9%)). Usando-se a DEM, o câncer foi diagnosticado em oito (61,5%) dos 13 casos (6 CP e 2 CM) e cinco (38,5%) casos foram falso-negativos. Usando-se a DAM, dez (77,0%) tumores malignos foram identificados (7 CP, 2 CM e 1 CF) e três (23,0%) casos foram falso-negativos. Os resultados são mostrados no anexo VI e os valores obtidos para os testes de validação e predição apresentados na tabela 19.

Examinador A2

Em 50 (67,6%) casos foi atribuído diagnóstico que foi dicotomizado em benigno ou maligno de acordo com a DEM (benigno (n = 41; 82,0%) e maligno (n = 9; 18,0%)), e DAM (benigno (n = 39; 78,0%) e maligno (n = 11; 22,0%)). Usando-se a DEM, o câncer foi diagnosticado em oito (53,3%) dos 15 casos (6 CP e 2 CM) e sete (46,7%) casos foram falso-

negativos. Usando-se a DAM, dez (66,6%) tumores malignos foram identificados (7 CP, 2 CM e 1 CF) e cinco (33,4%) casos foram falso-negativos. Os resultados são mostrados no anexo VI e os valores obtidos para os testes de validação e predição são apresentados na tabela 19.

Examinador B1

Em 65 (87,8%) casos foi atribuído diagnóstico que foi dicotomizado em benigno ou maligno de acordo com a DEM (benigno (n = 55; 84,6%) e maligno (n = 10; 15,4%)), e DAM (benigno (n = 51; 78,5%) e maligno (n = 14; 21,5%)). Usando-se a DEM, o câncer foi diagnosticado em nove (6 CP e 3 CM; 56,2%) dos 16 casos; em um caso, no entanto, o diagnóstico citológico de CM foi alterado para CP. Sete (43,8%) resultados foram falso-negativos. Usando-se a DAM, onze (68,7%) tumores malignos foram identificados (8 CP, 2 CM e 1 CF) e cinco (31,3%) casos foram falso-negativos. Os resultados são mostrados no anexo VI e os valores obtidos para os testes de validação e predição são apresentados na tabela 19.

Examinador B2

Em 62 (83,8%) casos foi atribuído diagnóstico que foi dicotomizado em benigno ou maligno de acordo com a DEM (benigno (n = 54; 87,1%) e maligno (n = 8; 22,9%)), e DAM (benigno (n = 44; 70,1%) e maligno (n = 18; 29,9%)). Usando-se a DEM, o câncer foi diagnosticado em sete (43,8%) dos 16 casos (4 CP e 3 CM); sendo que em um caso (CM) o diagnóstico foi alterado para CP. Nove (56,2%) resultados foram falso-negativos. Usando-se a DAM, treze (81,2%) tumores malignos foram identificados (10 CP, 2 CM e 1 CF) e três (18,8%) casos foram falso-negativos. Os resultados são mostrados no anexo VI e os valores obtidos para os testes de validação e predição apresentados na tabela 19.

TABELA 19: Taxas de sensibilidade (Se), especificidade (Es), valor preditivo positivo (VPP), valor preditivo negativo (VPN), acurácia (Ac), falso-negativo (FN), falso-positivo (FP), razão de verossimilhança do teste positivo (RVP) e razão de verossimilhança do teste negativo (RVN) observadas para os patologistas A e B usando-se as definições estreita (DEM) e ampla de malignidade (DAM). Os resultados estão expressos em percentagem.

Observador	Se	Es	VPP	VPN	FP	FN	RVP	RVN	Ac
A1 (DEM)	61,5	97,0	89,0	87,0	3,0	38,5	22,0	0,40	87,5
A1 (DAM)	77,0	97,0	91,0	92,0	3,0	23,0	27,5	0,24	91,6
A2 (DEM)	53,3	97,0	89,0	83,0	3,0	46,7	18,9	0,48	84,0
A2 (DAM)	66,6	97,0	91,0	87,0	3,0	33,4	23,8	0,34	88,0
B1 (DEM)	56,2	98,0	90,0	87,3	2,0	43,8	28,2	0,45	87,7
B1 (DAM)	68,7	94,0	78,6	90,2	6,0	31,3	11,3	0,33	87,7
B2 (DEM)	43,8	98,0	87,5	83,3	2,0	56,2	20,8	0,57	83,9
B2 (DAM)	81,2	89,1	72,2	93,2	10,9	18,8	7,5	0,20	93,2

DISCUSSÃO

A relação homem/mulher observada foi de 1/13,8, sendo superior a de outras séries.^{15,28,53,61,99,105} Apesar da maior incidência relativa do câncer de tireoide em indivíduos do sexo masculino^{4,9,13,14}, nenhum dos cinco homens que fizeram parte do estudo apresentava câncer, sendo este encontrado em 17 (24,0%) das 69 mulheres. A idade variou de 14 a 88 anos, com média de 54,3 e mediana de 49,0. Estes dados são semelhantes aos descritos por Tseng et al (2008)¹⁰⁵, que encontraram 111 (15,1%) homens e 626 (84,9%) mulheres, com idade entre 13 e 92 (média de 51,8 anos). Na série de El Hag et al (2003)⁹⁹ a idade variou de 5 a 90 anos e a mediana foi de 36 anos.

Neste estudo, a proporção de tireoidectomias em relação ao número de PAAF realizadas foi de 83,7%, superior à descrita por outros autores.^{14,28,32,33,53,46,105} Possível explicação para tal fato encontra-se na indicação cirúrgica por razões clínicas, tais como: presença de bócio volumoso e suspeita clínica de malignidade; e/ou insegurança com o laudo citológico, resultando na necessidade de esclarecimento diagnóstico histológico.

Quando se compara a proporção de tumores malignos (17/74; 23,0%), confirmados histologicamente, em relação ao número de operações, observa-se que ela foi inferior à descrita por Asp et al (64,0%)²⁸, Silverman et al (72,0%)⁴⁶ e Gerhard & Santos (53,6%)⁷². No entanto, essa taxa (23,0%) foi superior ou semelhante às descritas por outros autores.^{1,14,15,33,99,102,116} Por outro lado, a proporção de CP (13/17; 76,5%) em relação aos demais tumores malignos foi superior à descrita por Gerhard & Santos (40,2%)⁷² e Gharib et al (53,3%)¹¹⁶ e menor que a reportada na casuística de de Vos tot Nederveen Cappel et al (81,8%)¹⁵ e de Blansfield et al (89,0%)¹²⁶.

A amostra ideal de um estudo de acurácia diagnóstica deve ser representativa da população submetida ao teste na prática clínica. Portanto, pode-se questionar, neste trabalho, o tamanho da amostra, uma vez que, para alguns autores¹⁰⁶, a amostra suficiente deve conter pelo menos 100 casos para garantir adequada representatividade de estudos sobre testes diagnósticos. No entanto, o número de PAAF realizadas no ABF - UFMG é modesto (média de 7,5 exames ao mês ou 1,7 por semana) se comparado ao de outros Serviços.^{15,22} Além do mais, 106 (47,1%) exames não preencheram os critérios de inclusão. Tamanhos de amostra semelhantes foram reportadas por Gerhard & Santos (n = 79)⁷², Van Herle et al (n = 74)²⁵ e Hall et al (n = 72)⁵³.

O HC–UFMG, por se tratar de centro de referência, pode resultar em um viés de seleção devido ao recrutamento exclusivo, o que pode gerar maior representação de determinada afecção. Entretanto, quando se trata de doenças relativamente pouco frequentes, como o câncer de tireoide, essa maior representação pode ser vantajosa, pois assegura que o teste seja aplicado a grande número de doentes.¹⁰⁶ Tal fato pode também explicar a grande proporção de operações. É possível que esta amostra corresponda ao espectro adequado de pacientes para os quais o teste diagnóstico é aplicado na prática clínica. Por último, esse viés proporciona falsa elevação da RVP e falsa redução da RVN, como citado por outros autores.^{28,33,49}

Apesar do desenho deste estudo ser prospectivo, o material para análise já estava disponível. Deste modo, não foi possível controlar a realização de outros métodos diagnósticos (US, cintilografia, etc), que foram utilizados na avaliação e conduta destes pacientes. Como as amostras de lâminas já se encontravam prontas, não foi possível o reconhecimento de fatores outros que não a PAAF, como causa para as eventuais discrepâncias apresentadas. Entretanto, deve-se levar em conta que esses mesmos métodos também estavam, igualmente, disponíveis para os indivíduos portadores de nódulos tireoidianos, submetidos ou não ao exame padrão.

O número de aspirações por nódulo não foi alvo de investigação desta pesquisa. Sabe-se que este número é variável. Para inclusão no estudo, cada caso deveria possuir, pelo menos, duas lâminas por nódulo puncionado, como descrito por Stelow et al (2005)⁷¹, cujo número variou de duas a 18 lâminas. Esta abordagem encontra-se, também, dentro da recomendação feita por Ribeiro & Brasileiro (1996)³⁹. Eles avaliaram a taxa de suficiência das amostras, em relação ao número de lâminas por paciente, e observaram que o número de amostras insuficientes não dependia nem da quantidade de lâminas examinadas nem do processo patológico. Concluíram que o número de lâminas por punção não precisava ser maior que três. Para eles, mais importante que o número de lâminas é o número de punções realizadas em cada nódulo, o que poderia reduzir o número de resultados falso-negativos. O número recomendado de aspirações por nódulo tem variado de uma a oito.^{3,13,143} No nosso estudo, o número de lâminas por paciente variou de duas a 16 com média de 7,4. Isto deve-se, provavelmente, ao fato do HC-UFMG ser instituição de ensino (maior número de punções e esfregaços como meio de treinamento).

Método diferente foi adotado por Saxén et al (1978)⁵⁷, Hirokawa et al (2002)⁷³ e Lloyd et al (2004)⁷⁹ que selecionaram apenas uma ou poucas lâminas representativas para análise. Entretanto, deve-se ressaltar que tais estudos focaram a concordância histológica e não a

citológica. Se fossem selecionadas somente as lâminas representativas, os casos insuficientes seriam automaticamente excluídos da análise. Optamos por manter todas as lâminas para as análises, o que permitiu: a avaliação de todos os resultados citológicos possíveis (maligno, benigno, NF e insuficiente), a correlação da taxa de resultados insuficientes com a experiência de cada examinador (A e B) e a identificação da taxa de malignidade nestas amostras.

Ao avaliar a reprodutibilidade diagnóstica, é importante se ter em mente as definições de acordo e reprodutibilidade. Acordo refere-se ao número global ou proporcional de casos com o mesmo diagnóstico, inclusive aqueles que podem ser atribuídos à chance. Ele é calculado como a proporção de vezes em que os participantes fazem o mesmo diagnóstico. A reprodutibilidade, por sua vez, refere-se à parte do acordo que não pode ser explicada puramente pela chance.

Estudos de reprodutibilidade devem envolver índice estatístico (como o coeficiente K) que leva em consideração o fato de que os observadores poderão concordar ou discordar simplesmente por acaso. Este é o indicador de concordância ajustada, pois leva em consideração a concordância atribuída ao fator chance.¹⁵⁶⁻¹⁵⁹ As estimativas de reprodutibilidade de variáveis categóricas (dados qualitativos) são medidas, apropriadamente, pela estatística K .^{75,159} O K é apropriado para atributos binários, tais como positivo ou negativo e presença ou ausência, enquanto o Kp é o índice estatístico utilizado para calcular a reprodutibilidade quando as variáveis são ordinais e os resultados expressos em duas ou mais categorias.^{158,159} Ele leva em consideração que os atributos ordinários podem apresentar variação da resposta e que alguns desacordos podem ser mais significativos que outros, fornecendo maior peso numérico a tal desacordo.¹⁵⁸⁻¹⁶¹ Até 2002, foram poucos os estudos que utilizaram a estatística K para a verificação da reprodutibilidade que, antes de 1990, era quase sempre expressa em porcentagens de acordo.^{57,73,76-78,144} Neste estudo, foram efetuados os cálculos de reprodutibilidade por meio da obtenção do Kp e do K , o primeiro por ser considerado o coeficiente mais apropriado, e o segundo (K) por ser mais amplamente conhecido e mais usado nos artigos publicados, o que poderia facilitar as comparações.

Para a comparação dos resultados de reprodutibilidade, não foi possível utilizar a versão baseada no X^2 do teste de McNemar, uma vez que este só pode ser usado para cálculos que envolvem tabelas 2×2 com um grau de liberdade. Como, neste estudo, os cálculos de reprodutibilidade envolviam tabelas do tipo 4×2 (três graus de liberdade) foi usado o teste do X^2

padrão. Não se obteve o IC nestas circunstâncias pelo fato de se estar trabalhando com variáveis categóricas nominais (benigno, maligno, insuficiente e NF) e não variáveis contínuas.

Para quase todos os objetivos a respeito da verificação do grau de concordância entre examinadores, valores maiores que 0,75 podem ser tomados como excelente acordo enquanto valores menores que 0,40 devem ser interpretados como acordo ruim.¹⁴⁴ Contudo, no presente estudo, usamos a classificação recomendada, e, amplamente aceita, por Landis & Koch (1977)¹⁶¹, o que poderia facilitar as comparações.

Quando se avalia o teste diagnóstico deve-se sempre mencionar se os resultados encontrados influenciaram ou não a decisão de realizar o exame padrão. No presente estudo, assim como nos demais estudos envolvendo a correlação cito-histológica de afecções da tireoide, existe um viés de verificação. Isto ocorre porque, de acordo com o resultado da PAAF, há influência direta na indicação ou não do exame padrão (exame histológico da peça ressecada cirurgicamente). Pacientes cuja suspeita clínica e/ou exame citológico sugerem malignidade são mais prováveis de serem submetidos à intervenção cirúrgica do que aqueles com resultados negativos ao exame clínico ou ao estudo citológico. Estes últimos são acompanhados clinicamente não sendo candidatos à operação, o que, conseqüentemente, impede a aplicação do exame padrão.^{14,127} Acrescido a isto, sabe-se que muitos tumores malignos podem não se manifestarem clinicamente. Como resultado, as taxas de verdadeiro-negativo e de falso-negativo do exame citológico são pouco confiáveis.^{50,53,106,155} Por outro lado, pacientes com suspeita clínica ou achados citológicos sugestivos de malignidade são submetidos à operação, o que permite a comparação com o exame histológico (padrão). O mesmo ocorre em relação aos casos indeterminados (NF), nos quais a tireoidectomia, normalmente, também é indicada, o que permite mais acurada determinação das taxas de verdadeiro-positivo e de falso-positivo.^{53,106,155} Vale ressaltar que esse viés de verificação é inevitável no caso da PAAF, já que ela se tornou prática-padrão na tomada de decisão quanto a indicação cirúrgica de nódulos tireoidianos.

O nível de reprodutibilidade depende não só do tipo de evento em questão, mas também de numerosos fatores relacionados ao examinador, ao procedimento em teste e ao ambiente, onde as observações são feitas. O examinador, ao repetir o exame, deve ignorar resultados prévios, obtidos por ele mesmo ou por outro examinador, sob pena de ser influenciado por este conhecimento e deturpar a avaliação, mesmo que involuntariamente.¹⁵⁶ Para elevar a reprodutibilidade são essenciais a definição clara do evento, as regras inequívocas de mensuração

e a presença de esquemas de classificação apropriados, de modo que todos os casos tenham, sem ambigüidade, um local para serem alocados. Para que a reprodutibilidade aumente, basta diminuir o número de categorias diagnósticas (por exemplo, valores positivos e negativos no lugar de muito alto, alto, médio, baixo e muito baixo). Isto pode ser evidenciado no trabalho de Hirokawa et al (2003)⁷³, que ao distribuírem os casos em duas categorias mais amplas (benigno para BC e AF e maligno para CF e CPVF) elevaram o acordo unânime de 10,0% para 62,0%. O estudo de Stelow et al (2005)⁷¹, por sua vez, ao considerar a lesão folicular e a NF como entidades distintas, resultou em grande variabilidade interexaminador ($K = 0,35$). Por outro lado, quando elas foram consideradas como entidade única, a variabilidade diminuiu ($K = 0,57$). Estes autores demonstraram, ainda, que quando foi feita a dicotomização entre afecções que requerem cirurgia e as que não requerem, a variabilidade (interobservador) diminuiu ainda mais ($K = 0,65$). Nesta pesquisa não foram estabelecidas regras de mensuração e o ambiente onde as observações foram feitas correspondeu ao próprio laboratório de trabalho dos participantes. O objetivo era de que as condições da pesquisa fossem exatamente as mesmas dos seus ambientes rotineiros de trabalho.

Foram utilizadas as quatro categorias diagnósticas padronizadas e amplamente aceitas (benigno, maligno, NF e insuficiente).^{1,13,15,18,47,61,66,87,88,112} O uso da classificação limitada em categorias teve como objetivos: apresentação dos resultados de forma categorizada, aplicação dos cálculos de validade e acurácia, comparação com os de outros estudos e a possibilidade de que os índices obtidos possam ser extrapolados para a prática clínica.^{50,61} Assim, clínicos e cirurgiões podem utilizar os resultados de forma categorizada, o que facilita a conduta diante do portador de nódulo tireoidiano, especialmente no que se refere à necessidade ou não de tratamento cirúrgico.

Para a definição das categorias diagnósticas para alocação dos casos e a definição de adequácia das amostras, os patologistas A e B se basearam nos parâmetros citomorfológicos apresentados no ANEXO III. Observa-se que os critérios foram exatamente os mesmos. Tal fato pode ser explicado por serem ambos provenientes da mesma escola de ensino (FM-UFGM). Em relação aos critérios de suficiência das amostras, eles encontram-se próximos aos descritos por outros autores.^{18,29,92,97,114,118,126}

Bakshi et al (2003)¹¹⁴ correlacionaram resultados citológicos inconclusivos de 45 casos de NF com o diagnóstico histológico. O critério de suficiência das amostras (pelo menos oito a dez grupamentos bem preservados em pelo menos duas lâminas cada um contendo no mínimo de oito a dez células) foi reavaliado a fim de determinar se era demasiado exigente. Diferentes

características citológicas (celularidade, arquitetura celular, pleomorfismo nuclear, inclusão/ranuras, cromatina, alterações das células de Hürthle, linfócitos, macrófagos, coloide e células gigantes multinucleadas) foram pontuadas e comparadas com o diagnóstico cirúrgico final; dos 45 casos, apenas um (2,2%) era maligno (CF invasivo) e 44 (97,8%) eram benignos. Para os autores, o critério adotado de adequação é pertinente. Entretanto, recomendam acompanhamento cuidadoso para os pacientes com exames citológicos insuficientes.

Com relação aos cistos, houve diferença em relação aos laudos emitidos pelos patologistas envolvidos no estudo. O patologista A não forneceu nenhum laudo compatível com lesão cística, uma vez que, em sua descrição, estas foram consideradas insuficientes para a interpretação. Conduta semelhante é adotada por outros autores.^{7,61} O patologista B, na segunda análise (B2), identificou quatro (5,4%) lesões císticas, porém as considerou compatíveis com benignidade (cistos coloides). O DCO continha seis (8,1%) laudos compatíveis com lesões císticas, também consideradas como benignas (BC). Os relatos de cistos observados no estudo apresentam proporção menor que a observada na literatura, onde são descritas taxas de 15,0% a 25,0%²⁶. Elas podem incluir cistos coloides simples, nódulos coloides hemorrágicos, mas 15,0% são CP necróticos e 30% são adenomas hemorrágicos.^{13,113} Aproximadamente 25,0% das lesões císticas maiores que 4cm são malignas, porém, frequentemente, fornecem número insuficiente de células para sua interpretação.^{13,26,113} A SPC considera satisfatória a amostra proveniente de lesões císticas, desde que contenha numerosos macrófagos e hemossiderina, mesmo com escassez ou ausência de células foliculares. Entretanto, junto ao laudo, deve ser acrescida a observação de que a interpretação foi limitada devido à escassez ou ausência de tais células.¹⁷ O US tem sido usado com o propósito de distinguir nódulo sólido de cístico com a crença de que, no primeiro, o risco de malignidade é maior.¹⁰⁶ Por outro lado, alguns estudos, como o de McHenry et al (1999)¹⁶³, evidenciaram que o risco de câncer é similar ou até maior em nódulos císticos. Neste estudo, todos os casos diagnosticados como cistos foram confirmados como benignos à histologia.

A avaliação do desempenho de um teste requer sua comparação com outro procedimento que assegure o diagnóstico de interesse. A esse procedimento, seja ele exame laboratorial, de imagem, achado cirúrgico ou mesmo seguimento clínico, denomina-se padrão de referência ou procedimento padrão, e é ele que estabelece o diagnóstico de certeza da presença ou ausência da doença. Embora o procedimento padrão possa ser frequentemente imperfeito, ele deve ser,

consensualmente, aceito pela comunidade científica para a certificação do diagnóstico. No caso de afecções da glândula tireoide, ele corresponde ao exame histológico da peça cirúrgica, uma vez que somente ele propicia a análise tecidual adequada.^{63,106,155,156}

Este estudo focou somente a experiência na interpretação citológica dos esfregaços de nódulos tireoidianos obtidos pela PAAF, mais precisamente sua reprodutibilidade. Procedeu-se, neste estudo, do mesmo modo que Saxén et al (1978)⁵⁷ e Renshaw et al (2006)¹⁴⁶, ou seja, os examinadores não tiveram conhecimento dos aspectos clínicos, dos exames propedêuticos e das interpretações citológicas feitas pelo outro examinador. Além disso, todos os aspectos do método empregado na pesquisa foram omitidos dos participantes, de modo que todas as avaliações foram cegas e independentes. Método alternativo foi descrito por Stelow et al (2005)⁷¹, onde todos os casos a serem revisados foram acompanhados de história clínica fictícia e genérica. A ausência de qualquer dado da história clínica evita o viés da informação clínica com a consequente falsa elevação da acurácia, uma vez que o patologista poderia ser influenciado por ela quando fosse emitir o diagnóstico. Deste modo, esta conduta indica, objetivamente, a acurácia do método isolado.^{50,71,73,102,146,160} Assim, nenhum dos observadores participou do processo de coleta e preparo do material. O que se pretendeu avaliar foi a experiência do patologista em interpretar exames citológicos e não a habilidade em coletar o material. Deve ficar claro que estas características são distintas, mas interrelacionadas e imprescindíveis para a obtenção de bons resultados.^{25,28,42,55,119,145}

Para a avaliação intraexaminador, as análises dos exames não devem ser muito próximas umas das outras, de modo que a primeira avaliação não influencie a seguinte.⁴⁶ No atual estudo, foi adotado o intervalo de seis meses entre as avaliações, como utilizado por Franc et al (2003)⁷⁵ em seu estudo sobre a reprodutibilidade histológica do CF da tireoide. Deste modo, garante-se que as avaliações sejam independentes umas das outras e cegas, o que permite, em parte, a validação dos resultados. No entanto, pode-se questionar se o período de tempo adotado (seis meses) entre as interpretações teria sido suficiente para que não houvesse influência nos resultados. Entretanto, acreditamos que o número de casos e de lâminas analisados foram suficientemente grandes para tornar a memorização improvável, principalmente quando se compara com o estudo realizado por Elsheikh et al (2008)¹⁴¹ que, apesar de adotarem o intervalo de dez a 15 meses, focaram a concordância histológica de apenas 15 casos, além do número de lâminas analisadas ter sido de uma ou duas. Além do mais, foi utilizado o artifício da

renumeração aleatória das lâminas (75 a 148), sem o conhecimento dos participantes, como empregado por Elsheikh et al (2008)¹²⁹ e por O’Sullivan et al (1996)¹⁶⁰.

Quando se discute a reprodutibilidade diagnóstica, a literatura recomenda que deve ser considerado o “efeito do observador” ao se analisar o “efeito do objeto”.¹⁵⁴ Como exemplo do primeiro pode-se citar o trabalho de Hirokawa et al (2002)⁷³, onde dois dos observadores envolvidos na pesquisa eram provenientes da mesma Instituição em relação aos demais. Infelizmente, os autores não investigaram esse aspecto (influência do “efeito escola de ensino” no resultado diagnóstico). Isto é relevante quando da avaliação de reprodutibilidade, uma vez que a possibilidade de acordo (ou desacordo) no comportamento diagnóstico poderia ser influenciada pela escola de formação. O mesmo ocorreu no trabalho de Stelow et al (2005)⁷¹ sobre a variabilidade interexaminador de exames citológicos da tireoide, onde foram empregados citopatologistas de Instituições diferentes, o que poderia explicar, em parte, as discrepâncias observadas. Por outro lado, quando os observadores são oriundos da mesma Instituição há maior propensão a acordarem no diagnóstico.¹⁶⁴

O objetivo dos patologistas apresentarem características comuns foi o de eliminar efeitos que pudessem influir na análise, como o efeito “escola de ensino”. Portanto, ambos os patologistas (A e B) frequentaram todo o curso de graduação em medicina na UFMG e se especializaram em Patologia no HC-UFMG. Além disso, ambos exerciam sua atividade profissional na mesma cidade (Belo Horizonte). Este último fato é relevante levando-se em conta as diferentes prevalências das doenças da tireoide.¹⁵⁴ Áreas onde há deficiência de iodo apresentam maior prevalência de BC em relação à áreas não carentes.¹⁶⁴ Com isso, patologistas provenientes de áreas ricas em iodo poderiam ser menos experimentados em diagnosticar o BC. Como os examinadores são oriundos da mesma região, o efeito da prevalência das doenças, e, portanto, a maior ou menor experiência em diagnosticar determinada afecção é minimizado.¹⁶⁴

Como critério de definição de experiência foi escolhida a proposta sugerida por Van Herle et al (1982)²⁵, pelo fato de ser uma das mais exigentes (interpretar 500 exames ao ano). Este número fica um pouco aquém do proposto por Stelow et al (2005)⁷¹, que recomendam mais de 600 exames ao ano. No entanto, fica bem acima do indicado por Ljung et al (1996)¹⁴⁵ e (2001)⁴² que estabelecem, como critério, cerca de 200 exames anuais. Com relação ao tempo de atividade profissional, consideramos, como proposto por Lloyd et al (2004)⁷⁹ e Fonseca et al (1998)¹⁴⁴ o mínimo de dez anos. Franc et al (2003)⁷⁵ selecionaram cinco patologistas experientes usando,

como critério, o tempo de 20 a 25 anos de experiência profissional. No presente estudo, o patologista experiente (A) apresentava 25 anos de exercício profissional. Desta maneira, além de preencher o critério de tempo mínimo de dez anos como proposto por alguns autores,^{79,144} preenchia, também, o critério usado por Franc et al.⁷⁵

Por outro lado, para a escolha do patologista inexperiente (B), este não deveria possuir certificação de especialista em patologia pela SBP, como sugerido por Fonseca et al (1998)¹⁴⁴, e ter concluído a especialização em Patologia em, no máximo dois anos. Com relação ao número de exames realizados, foi escolhido o critério de 40 exames ao ano (máximo três exames por mês), como proposto por Haas et al (1993)⁴⁹. O patologista B não possuía a certificação exigida, apresentava três anos de conclusão do curso médico, terminara, recentemente, sua especialização em Patologia no HC-UFMG e examinava 40 exames, em média, anualmente.

Os critérios histológicos utilizados para o diagnóstico de muitas afecções tireoidianas, como as lesões foliculares encapsuladas, não são precisos e apresentam considerável variabilidade (baixa reprodutibilidade).^{57,73,75-79,150} A interpretação destas lesões é, frequentemente, muito difícil, o que torna o diagnóstico histológico amplamente variável, mesmo entre patologistas experientes.^{73,79,129} Outro fato que ocorre e colabora para que ocorram essas discrepâncias, refere-se à utilização de diferentes classificações como, por exemplo, a inclusão do CPVF na categoria de CP. Adotando este critério, Ron et al (1986)⁷⁶, elevaram a proporção do CP de 52% para 85% do total de tumores presentes em sua série. Estes autores também revisaram os casos originalmente diagnosticados como benignos, e, em nenhum deles, ocorreu a mudança de diagnóstico para maligno. Isto sugere que, diante de casos limítrofes, o patologista tende para a classificação que requer tratamento mais “agressivo”, isto é, tende para o diagnóstico de câncer. Por outro lado, o patologista quando revisa lâminas em caráter de pesquisa, encontra-se mais confortável para emitir o diagnóstico, em conformidade com a classificação padrão, uma vez que não há envolvimento direto com os pacientes e o cirurgião e cujo laudo influenciará na conduta a ser seguida.

No presente estudo não foi constatada diferença entre o número absoluto e a proporção de tumores malignos identificados pelos patologistas, experiente (A1 e A2) e inexperiente (B1 e B2), tanto inter quanto intraexaminador, independentemente do uso da DEM ou da DAM, como está demonstrado nas tabelas 12 e 13. No entanto, considerando-se apenas o CP evidencia-se baixa proporção de acordos nas duas situações.

O CP apresenta critérios citológicos e histológicos bem definidos para o diagnóstico, o que torna seu diagnóstico mais reprodutível, com índices de acordo superiores a 80,0%.^{16,57,72,73,77} Este aspecto não foi confirmado em nosso estudo uma vez que, dos 13 (76,5%) casos de CP presentes na amostra, houve acordo no diagnóstico citológico interexaminador em cinco (38,5%) na primeira análise (A1 - B1) e em quatro (30,8%) na segunda (A2 - B2). Não houve diferença significativa entre o número de acordos presentes na primeira análise em relação à segunda ($X^2 = 0,17$, $p = 0,68$; OR = 1,41 [0,21 – 9,66]). A reprodutibilidade intraexaminador foi evidenciada em seis (46,2%) casos para o patologista A (A1 - A2) e em quatro (30,8%) para o patologista B (B1 - B2), sem haver diferença significativa entre ambos ($X^2 = 0,65$, $p = 0,42$; OR = 1,93 [0,30 – 13,14]). O acordo unânime nas quatro análises ocorreu em quatro casos (30,8%). Esses percentuais são inferiores aos descritos por Gerhard & Santos (2007)⁷², que relataram acordo em 30 (93,7%) dos 32 casos de CP na análise interobservador e em 100,0% na análise intraobservador. Diante desses dados, podemos supor que tenha ocorrido a inclusão de alguns casos de CPVF na categoria de CP, o que explica a baixa reprodutibilidade e a elevada proporção de CP na observadas amostra. Isto foi observado por Ron et al (1986)⁷⁶ em sua pesquisa. alguns diagnósticos histológicos de CP possam estar equivocados e se tratem de outros tipos, principalmente o CF e o CPVF. Sabe-se que os critérios histológicos também são pouco reprodutíveis e que as lesões foliculares encapsuladas, como o CF e o CPVF, são mais difíceis de terem seus resultados reproduzidos.^{57,73,75-79,108,129,147,150} Sabe-se também que o carcinoma papilífero apresenta características morfológicas subjetivas, o que pode explicar a baixa reprodutibilidade observada.

Há necessidade de que sejam estabelecidos critérios morfológicos mais precisos, tanto qualitativa quanto quantitativamente, como, por exemplo, a presença de alterações nucleares claras para o diagnóstico do CP, maior rigor na aplicação dos critérios já estabelecidos, ou ambos, com o intuito de melhorar a consistência do diagnóstico histológico, e, conseqüentemente, melhorar sua reprodutibilidade.^{73,129} Os critérios mínimos a serem reproduzidos devem ser estabelecidos por especialistas, em conferência, consenso ou por formulação de classificação internacionalmente aceita e que reflita o potencial biológico e o comportamento clínico dos tumores. Até que o diagnóstico histológico se torne também mais reprodutível, o entendimento acerca dos casos indeterminados da citologia permanecerá limitado.¹⁰⁸

Optamos por manter todos os laudos para a análise de reprodutibilidade, inclusive as lesões císticas (reportadas por B2) uma vez que todos continham a observação de que se tratavam de “lesão benigna, compatível com cisto colóide”. A análise da comparação intraexaminador para o patologista A mostrou taxas de acordo ($K = 0,82$ e $Kp = 0,79$) muito boas para o K e para o Kp pelos critérios fornecidos por Landis & Koch.¹⁶¹ Entretanto, não foi observada diferença entre os resultados ($X^2 = 0,13$; $p = 0,988$), o que evidencia a alta reprodutibilidade entre eles. Observa-se que o valor do X^2 é muito inferior a 7,81 e o valor p quase próximo ao valor um (1), o que permite a conclusão de que os resultados são praticamente iguais. O mesmo foi observado na análise intraexaminador dos resultados para o patologista B, que também mostrou taxas substanciais de acordo ($K = 0,78$ e $Kp = 0,71$). Portanto, os resultados apresentados em B1 e em B2 são semelhantes ($X^2 = 3,74$; $p = 0,291$). Estes dados encontram-se de acordo com os apresentados por Gerhard & Santos (2007)⁷², que encontraram taxas substanciais de reprodutibilidade intraxaminador ($k = 0,66$), ou seja, existência de similaridade dos resultados reportados. Contudo, o valor p relatado é contraditório ($p < 0,001$), mostrando que houve diferença significativa. No nosso estudo, observamos considerável reprodutibilidade tanto para o patologista A ($Kp = 0,79$) quanto para o B ($Kp = 0,71$) além de valores de p maiores que 0,05, o que confirma a semelhança dos resultados.

A análise da comparação interexaminador feita após a primeira análise (A1 – B1) mostrou taxa moderada e significativa de desacordo ($K = 0,55$ e $Kp = 0,50$; $X^2 = 11,20$; $p = 0,011$). O mesmo foi observado na análise interexaminador após a segunda análise (A2 – B2) dos resultados ($K = 0,54$, e $Kp = 0,44$; $X^2 = 9,69$; $p = 0,021$). Observa-se consenso entre o valor do X^2 , do p , do K e do Kp , ou seja, há diferença entre os resultados apresentados pelos examinadores com e sem experiência em ambas as análises (A1 - B1) e (A2 – B2). Os resultados encontram-se em desacordo com os reportados por Gerhard & Santos (2007)⁷², que reportaram taxa substancial ($k = 0,71$) de reprodutibilidade. Contudo, o valor p reportado ($p < 0,001$) nos parece contraditório, pois era de se esperar um valor elevado para o p .

A análise dos resultados de reprodutibilidade ilustra, claramente, a importância e a necessidade de uma avaliação abrangente e consistente das concordâncias, intra e interexaminador, mesmo para métodos diagnósticos amplamente aceitos como a PAAF da glândula tireoide. Os resultados observados sugerem a presença de consistência interna (análise intraobservador) e ausência de consistência externa (análise interobservador) para a referida

interpretação. Uma potencial abordagem para melhorar a confiabilidade e a reprodutibilidade inclui o treinamento adicional e/ou a modificação dos critérios diagnósticos existentes.⁶³

Segundo Franc et al (2003)¹⁶⁴, a análise intraexaminador é muito importante porque possibilita a verificação por si do “efeito do objeto”, uma vez que ela permite destacar o nível de experiência de cada patologista em campo, caso a comparação dos resultados intraobservadores detecte alguma discrepância. Por meio desta análise pode-se inferir que, indiretamente, também estava sendo avaliado se o exercício profissional, durante os seis meses de intervalo entre as análises, e, conseqüentemente, a aquisição de maior experiência poderiam influenciar nos resultados dos exames. A ocorrência de considerável consistência (intraexaminador) entre os resultados emitidos por A e B, independentemente da experiência de cada um, permite a conjectura de três suposições: primeira, de que o tempo (seis meses) foi insuficiente para propiciar maior habilidade na interpretação dos aspirados; segunda, a possibilidade de memorização dos exames (menos provável); terceira, o uso de critérios próprios com consistência ao longo do tempo.¹⁵⁹

A baixa reprodutibilidade entre observadores pode ser devida a variações biológicas dos pacientes, variações dos equipamentos ou variações dos indivíduos que realizam e interpretam os exames. Como os dois primeiros não se aplicam a este estudo (isto é, as amostras analisadas foram exatamente as mesmas e não se utilizou nenhum tipo de equipamento) pode-se concluir que a diferença encontrada pode ser atribuída às variações dos avaliadores usualmente relacionada com a experiência de cada um na interpretação de esfregaços de tireoide.⁷³ Por outro lado, quando o estudo sobre reprodutibilidade conclui que o teste é altamente reprodutível, duas possibilidades são prováveis: primeira, de que o teste é bastante simples e fácil de ser aplicado aos pacientes; segunda, de que os participantes do estudo são altamente qualificados no referido teste.

Nesta pesquisa, não foi estabelecido diagnóstico final de consenso para os casos discordantes como feito por Gerhard & Santos (2007)⁷². Como o próprio nome diz, ele representa apenas o consenso, e não o diagnóstico definitivo que é histológico. O principal objetivo desta pesquisa foi comparar os resultados e avaliar o grau de acordo (inter e intraexaminador) entre dois patologistas com diferentes níveis de experiência na PAAF. Com isso, o que se pretendia avaliar era se a experiência pode de fato ser responsabilizada pela obtenção de melhores resultados, uma vez que os laudos, reportados por cada um deles, foram correlacionados ao

exame histológico. As lâminas dos exames histológicos não foram interpretadas por um terceiro patologista, já que não estava sendo avaliada a acurácia e a reprodutibilidade do exame anátomo-patológico. Deve ficar claro que na existência de diagnóstico equivocado, pode-se atribuir o desacordo, equivocadamente, ao citopatologista quando se procederam os cálculos de acurácia.⁵³

Os valores encontrados para a RVP para os avaliadores A e B (Tabela 19) foram consistentemente conclusivos em fornecer o diagnóstico considerável para todas as observações, exceto na segunda análise do patologista B (B2) quando foi usada a DAM (RVP = 7,5). Isto é, para o patologista A, quando procederam os cálculos considerando-se a DEM, na primeira (A1) e na segunda (A2) análise, um resultado positivo à PAAF foi, respectivamente, 22 e 18,9 vezes mais provável de ocorrer no paciente com doença maligna que no paciente sem câncer. Para a DAM, esses valores foram, respectivamente, 27,5 e 23,8 vezes (estes dados são conclusivos).^{102,155,165,166} Para o patologista B, quando se procederam os cálculos considerando-se a DEM na primeira (B1) e na segunda (B2) análise, um resultado positivo à PAAF foi, respectivamente, 28,2 e 20,8 vezes mais provável de ocorrer no paciente com doença maligna que no paciente sem câncer. Para a DAM, esses valores foram, respectivamente, 11,3 e 7,5 vezes. Estes dados são conclusivos no primeiro e podem, ocasionalmente, ser clinicamente relevantes no segundo.^{102,155,165,166}

Observa-se a ocorrência de comportamento discrepante para os dois observadores com relação à RVP, quando foram usadas as duas definições distintas de malignidade (DEM e DAM). Para o patologista A, houve aumento da RVP quando se usou a DAM em relação à DEM, uma vez que a especificidade ficou inalterada (97,0%) mas a sensibilidade aumentou; para uma mesma especificidade, quanto maior a sensibilidade maior será a RVP. O contrário se deu com o patologista B, que apresentou redução importante da RVP (para 20,8% e para 7,5%). Isto pode ser explicado pelo fato de ter havido redução da especificidade (mais expressiva em B2) apesar do aumento concomitante da sensibilidade; para uma mesma sensibilidade, quanto maior a especificidade maior a RVP. O fato de ter sido mais expressiva em B2 é explicado pelo número excessivo de casos de NF (n = 10) diagnosticados por aquele patologista, o que resultou numa redução maior da especificidade.

Os valores encontrados para a RVN para os avaliadores A e B estão demonstrados na tabela 19. Observa-se que, em nenhuma situação, os resultados foram consistentes em fornecer valor diagnóstico considerável (estabelecido na literatura como menor que 0,1)^{102,155,166}. A

segunda análise do patologista B (B2), ao usar a DAM, obteve RVN de 0,2, o que pode significar ser pouco provável observar exame negativo em indivíduo doente. Contudo, os valores encontrados devem ser analisados com cautela, pois somente em B2 (DEM), cuja RVN observada foi de 0,57 é que se pode concluir que eles, de fato, alteraram, em pequeno grau e sem significância clínica, a probabilidade da doença (RVN maior que 0,5).^{102,155,166} Nota-se que ocorreu redução do valor da RVN quando foi usada a DAM em relação à DEM, dado que pode ser explicado pelo aumento da sensibilidade, ao mesmo tempo em que a especificidade alterou pouco ou não se alterou (isto se deve à inclusão dos casos compatíveis com NF na categoria de malignos).

A utilidade do exame propedêutico depende não apenas das razões de verossimilhança, mas também da probabilidade pré-teste, do limiar de diagnóstico (probabilidade abaixo da qual se afasta o diagnóstico) e do limiar terapêutico (probabilidade acima da qual se confirma o diagnóstico e se inicia o tratamento). Com relação à probabilidade pré-teste para a aplicação dos testes diagnósticos, o ideal seria usá-lo numa faixa intermediária de prevalência (entre 25,0% e 65,0%). Isto porque em situações de baixa prevalência (5,0% por exemplo), o exame negativo traria pouca informação adicional e o exame positivo resultaria numa probabilidade pós-teste não conclusiva, na qual não se decide pela cirurgia imediata nem se afasta o diagnóstico. Esta é exatamente a situação do câncer de tireoide, que apresenta baixa prevalência. Assim sendo, o exame negativo contribui pouco para o manejo do paciente. Por outro lado, em situações de alta prevalência (90,0% por exemplo), o exame positivo não modifica a conduta e o negativo coloca o paciente numa probabilidade intermediária. Quando a prevalência é intermediária, o teste se mostra mais útil, pois o exame positivo é suficiente para apoiar a decisão cirúrgica enquanto o negativo é suficiente para se optar por outro método diagnóstico, antes de tomar a decisão pela operação. Conclui-se que as razões de verossimilhança são similares aos valores preditivos, mas sem a vulnerabilidade das variações das prevalências das doenças.¹⁶⁶

As taxas de acurácia relatadas nesta pesquisa foram altas, com variação de 83,9% a 93,2% para todos os observadores em todas as revisões (A1, A2, B1 e B2). Entretanto, larga variação na acurácia diagnóstica (50,0% a 97,0%) é descrita, o que poderia refletir a eficácia da técnica para o sítio particular que está sendo amostrado e a *expertise* do profissional que coleta, prepara e examina a amostra.^{12-14,25,39,40,53,65,130} Esta ampla variação torna o procedimento insatisfatório quando da tomada da decisão terapêutica. O que se questiona são as razões para essas

discrepâncias.⁴¹ Alguns autores recomendam que todos os passos desse processo devem ser realizados por profissionais experientes e afeitos ao método.¹⁶⁷⁻¹⁶⁹ No entanto, existe, entre os médicos, o conceito equivocado de que as falhas do método resultam da falta de experiência do citopatologista em interpretar o material amostrado. Obviamente, isto é verdade para algumas situações, mas, na experiência de Zarbo et al (1991)¹⁷⁰ e Ljung et al (1996)¹⁴⁵, a maioria dos erros cito-diagnósticos resulta da análise de material insatisfatório para o diagnóstico. Isto se deve, provavelmente, à falta de confiança em interpretar os nódulos benignos. Segundo a Associação Americana de Endocrinologistas e o Colégio Americano de Endocrinologia, citopatologistas inexperientes sentem-se inseguros ao afirmar que determinada lesão é compatível com benignidade ou malignidade, isentando-se da responsabilidade do diagnóstico. Assim sendo são frequentes laudos indeterminados (NF) para os casos limítrofes.²²

A PAAF é considerada, tecnicamente, fácil de ser executada, fato que determina a confusão entre "simplicidade técnica" e "facilidade de execução". Como não há necessidade de equipamentos sofisticados, o método é descrito como "simples". Assim, muitos profissionais passaram a realizá-la sem treinamento específico o que tem gerado resultados enganosos e condutas equivocadas. Apesar da facilidade técnica, a PAAF requer considerável treinamento, acúmulo de experiência e conhecimento de citologia e de patologia cirúrgica.^{43,50} Por ser tecnicamente fácil, tem havido muito desinteresse no treinamento supervisionado, o que tem acarretado sua realização por profissionais sem habilitação adequada.

Amostras insuficientes são aquelas nas quais não foi possível atribuir qualquer diagnóstico, portanto, aquelas que não preencheram os critérios específicos de adequácia previamente estabelecidos, ou seja, não apresentavam número mínimo de células foliculares bem preservadas necessárias para análise citológica segura.^{18,43} O percentual de resultados insatisfatórios relatado varia de 10,0% a 20,0%.^{13,17,30,113} Algumas publicações têm reportado índices menores que 8,0%^{8,54,72,93}, principalmente quando a aspiração é feita pelo patologista, quando a taxa de material adequado pode chegar a 97,0%⁹⁹, e quando ela é guiada pelo US.³⁰ Por outro lado, algumas publicações têm mostrado índices superiores a 25,0%.^{11,118}

Em relação às amostras insuficientes, o patologista A reportou 26 (35,1%) casos na primeira análise (A1) e 24 (32,4%) na segunda (A2). A análise intraexaminador (A1 – A2) mostrou não haver diferença (OR = 1,13 [0,54 – 2,36], $p = 0,72$). O patologista B reportou nove (12,2%) casos na primeira análise (B1) e 12 (16,2%) na segunda (B2). A análise intraexaminador

(B1 – B2) também mostrou não haver diferença (OR = 0,72 [0,26 – 1,98], $p = 0,48$). No entanto, a primeira análise interexaminador (A1 - B1) mostrou diferença significativa (OR = 3,91 [1,57 – 9,97], $p = 0,001$), o mesmo ocorrendo após a segunda (A2 - B2) análise (OR = 2,48 [1,06 – 5,88], $p = 0,021$). Portanto, a probabilidade de se encontrar resultados insuficientes foi maior para o patologista A em relação ao patologista B. Este dado confirma a afirmação de que a taxa de inadequácia pode ser afetada pelo limiar com que a PAAF é considerada inadequada. É provável que citopatologistas sejam mais propensos a considerarem o espécime insatisfatório quando desconhecem a história clínica do paciente. Isto se reflete na postura conservadora por parte dos patologistas em relação ao citodiagnóstico, uma vez que amostras contendo poucas células ou compostas somente por células degenerativas espumosas são consideradas não diagnósticas, embora haja o senso comum de que elas não possuem evidência de malignidade. Nestes casos, o espécime é interpretado como insuficiente, não havendo comprometimento do patologista com o laudo, seja ele benigno ou maligno.^{43,51,52,55} Por outro lado, é conhecido o fato de patologistas experientes tenderem a apresentar menor taxa de resultados insuficientes, uma vez que há tendência a alocar mais casos em categorias diagnósticas.^{25,26}

Amostras insuficientes jamais devem ser consideradas como benignas (esfregaço insatisfatório não é o mesmo que resultado negativo), mas, simplesmente, indicativas de tireoidectomia ou de que a PAAF deverá ser repetida, uma vez que a celularidade é escassa para análise.^{42,53,61,93,105} A repetição do exame pode fornecer material adequado em 85,0% a 90,0% dos casos. Entretanto, em 10,0% a 15,0% das vezes, o espécime permanecerá inadequado.¹⁸ Elas podem estar associadas à probabilidade variável de câncer (0,0% a 8,5%)^{18,33,85,114,138,171,172}, e têm sido citadas como a causa mais comum de resultados falso-negativos.^{13,18,30,116,138,143} Nota-se que elas são inferiores às encontradas na nossa pesquisa: A1 (15,4%), A2 (8,3%), B1 (11,0%), B2 (11,0%) e DCO (22,2%). Pode-se conjecturar que elas foram elevadas devido à grande proporção de tireoidectomias (83,7%), cuja causa deve-se provavelmente à indicação cirúrgica por razões clínicas ou por insegurança com o laudo citológico, resultando na necessidade de esclarecimento diagnóstico histológico. É provável que, em outras pesquisas, a indicação cirúrgica tenha se baseado mais nos achados citológicos de malignidade, uma vez que a proporção de tireoidectomias foi menor (9,0% a 50,0%).^{17,18,32,33,28,53}

Embora tenha havido a tentativa de definir critérios mais quantitativos para a adequácia do exame, a SPC tem adotado abordagem mais pragmática e sugere que critérios muito rígidos

não deveriam ser usados, mas sim critérios mais qualificados e baseados na interpretação individual de cada caso.¹⁷ Problema que pode ser grave é a descrição, nos laudos de exames citológicos, de “ausência de células malignas” ou de “células malignas não visualizadas” em espécimes que são hipocelulares ou acelulares, o que resulta numa falsa conclusão de benignidade.

Como os valores do X^2 e de p só informam sobre a existência ou não de diferença estatística entre os resultados comparados, foi calculado, também, o OR, uma vez que ele mede a força de associação entre as variáveis (chance), porém não mede o risco (incidência) da variável estudada. Portanto, se o OR é interpretado como o risco relativo (RR), que é usado para estimar a incidência das doenças para estudos do tipo coorte, ele tende a superestimar o tamanho do efeito.¹⁶² No entanto, quando a doença tem baixa prevalência, o OR pode ser utilizado como boa estimativa do RR. Outro aspecto a respeito do OR é que o tamanho da amostra não afeta o seu valor, mas sim seu IC;¹⁶² quanto maior a amostra em estudo menor será o IC. A interpretação para os valores de OR encontrados será de associação entre variáveis, mas o risco não será estimado, como recomendado por Davies et al (1998)¹⁶². Por exemplo, não se pode inferir que o risco de se encontrar resultado insuficiente pelo patologista A seja 3,9 vezes maior em relação ao B, na primeira análise; pode-se concluir apenas que a probabilidade de se encontrar resultados insuficientes, na primeira análise, foi maior para o patologista A em relação ao patologista B, com OR de 3,91 [1,57 – 9,97].

A designação citológica de tumor folicular pode corresponder à área nodular adenomatosa em bócio, tireoidite, adenoma ou carcinoma.^{108,125} O uso dos termos, lesão folicular e NF, de maneira distinta, representa tentativa de distinguir, à citologia, espécimes que apresentam achados mais sugestivos de malignidade, em relação a outros que seriam mais “indeterminados”.^{71,72} Eles podem ser considerados como duas categorias distintas, como recomendado por Gerhard & Santos (2007)⁷², que enfatizam que o termo NF só deveria ser usado quando a suspeita de malignidade for alta. A maioria dos patologistas não adota tal conduta e usam estes termos indistintamente para representar o mesmo diagnóstico. Além do mais, ambos os diagnósticos são indicativos de operação e podem apresentar extensa sobreposição de achados citológicos e histológicos, facilitando a análise estatística. Portanto, não nos parece razoável a utilização destes termos de maneira distinta, já que além de não trazerem benefício evidente aos

pacientes podem gerar mais confusão. No atual estudo não foi feita distinção entre lesões foliculares e NF.

Citopatologistas inexperientes podem apresentar alta proporção de resultados indeterminados (NF).^{5,11,50,59,60} Este dado ficou evidenciado neste estudo, onde o patologista A reportou dois casos (2,7%) tanto na primeira (A1) quanto na segunda (A2) análise, enquanto o patologista B observou quatro (5,4%) casos na primeira (B1) e dez (13,5%) na segunda (B2) análise. A comparação dos resultados inter e intraexaminador mostrou que houve diferença significativa apenas na segunda (A2 - B2) análise interexaminador ($p = 0,015$), o que permite a conclusão de que a probabilidade do patologista B (B2) emitir laudo compatível com NF é maior que a do patologista A (A2) com um OR de 5,63 [1,09 – 38,71] (Tabela 5). Este dado é corroborado na literatura e sua explicação pode estar, provavelmente, na falta de confiança, por parte dos patologistas inexperientes em interpretar nódulos que seriam benignos, sendo frequentes os laudos indeterminados para os casos limítrofes.^{22,56} Também, como reflexo da insegurança, o patologista inexperiente tende para o diagnóstico que requer o tratamento mais agressivo, como descrito por Ron et al (1986)⁷⁶. Entretanto, deve ser destacado que quando se realizou a comparação dos achados citológicos com os resultados histológicos, verificou-se em B2 que dos dez casos reportados como NF seis corresponderam a lesões malignas (5 CP e 1 CF). Se considerarmos que a operação deve ser indicada em todos os casos de NF, na primeira (B1) análise do patologista B, dois tumores (1 CP e 1 CF) teriam sido identificados enquanto dois BC teriam sido operados desnecessariamente. No caso do patologista A, tanto em A1 quanto em A2, os mesmos dois tumores malignos observados na primeira análise de B1 teriam sido identificados, e nenhuma operação desnecessária teria sido feita. Porém, ao comparar com B2, quatro CP não teriam sido descobertos pelo patologista A.

O patologista B reportou maior número de NF em B2 em relação à B1, com um OR de 2,73 [0,74 – 10,96], porém sem significância estatística ($p = 0,09$). Também não houve diferença quando se compararam os achados de cada análise separadamente (A1, A2, B1 e B2) com o DCO, cujos valores de p foram respectivamente: 0,15, 0,27, 0,05 e 0,22.

Apesar das taxas de sensibilidade (65,0% a 99,0%) e especificidade (72,0% a 100,0%), descritas na literatura, é bem conhecido o fato do câncer de tireoide ser detectado por meio da PAAF com alta sensibilidade e baixa especificidade.^{3,7,13,14,23,28,30,46,47,49,61,72,140,154,173} Isto se deve, em parte, à inabilidade em diferenciar as lesões foliculares encapsuladas.^{73-75,171} Todavia, no

presente estudo, os valores de especificidade superiores a 94,0% foram encontrados em todas as análises independentemente do examinador (Tabela 19). As taxas de sensibilidade observadas (variação de 43,8% a 81,2%) foram inferiores às descritas na literatura,^{14,18,23,28,30,61,72-174} principalmente quando os casos de NF foram considerados benignos (DEM). O pequeno número de lesões foliculares (13,5%), histologicamente confirmadas (CF (n = 2; 2,7%), AF (n = 8; 10,8%) e nenhum caso de CPVF), presentes na amostra pode explicar o valor elevado da especificidade e o baixo valor da sensibilidade em relação aos descritos na literatura.^{18,49} Para o DCO a sensibilidade foi especialmente baixa (30,7%), e, mesmo quando foi usada a DAM, a sensibilidade foi baixa (53,8%). A baixa sensibilidade observada para o DCO pode ser explicada pela existência de elevado número de avaliadores que interpretaram as amostras, fato também verificado na série de Dixon et al (2005)¹⁷⁵, onde a PAAF de mama, no primeiro momento, quando realizada por múltiplos profissionais, obteve sensibilidade de apenas 66,0% para malignidade. Todavia, no segundo período, quando foi feita por um único profissional a sensibilidade passou para 99,0%. No entanto, Silverman et al (1986)⁴⁶ haviam, previamente, descrito elevadas taxas de sensibilidade (93,0%) e especificidade (95,1%) a partir de 309 punções de tireoide feitas por um único endocrinologista mas interpretadas por vários patologistas.

A especificidade é considerada baixa quando seu valor é inferior a 60,0%²⁵; quando isto ocorre, o teste ainda permite que muitos pacientes portadores de doenças benignas sejam submetidos à operação diagnóstica. Observa-se, neste estudo, que as taxas de especificidade obtidas no DCO foram muito elevadas, tanto para a DEM (100,0%) quanto para a DAM (97,6%). Sabe-se que baixas taxas de especificidade estão associadas a maior indicação cirúrgica. Assim sendo, parece contraditório o fato de se ter elevada especificidade associada à elevada proporção (83,7%) de tireoidectomias em relação ao número de PAAF. Esta proporção, como já mencionado, foi bem superior à de outras séries.^{14,15,28,33,46,53,105} Esperava-se que essa proporção tivesse sido menor, tendo em vista a elevada taxa de especificidade observada no DCO. Este dado foi surpreendente, já que o principal objetivo da PAAF é reduzir o número de tireoidectomias ao selecionar mais apropriadamente os pacientes para a operação.^{14,15,22,28,92,101,105,111}

Quando se utiliza a DAM a sensibilidade aumenta, assim como aumenta o número de falso-positivos, mas, em contrapartida, a especificidade e os resultados falso-negativos diminuem.^{25,87} Essa estratégia parece ser aceitável pois tende a diminuir a perda de casos malignos, mas carrega o ônus da operação ser indicada para muitos pacientes portadores de

afecções benignas. A DEM é utilizada para melhorar a especificidade, uma vez que nela os casos interpretados como NF são incluídos dentro da categoria de benignos, o que eleva a especificidade e a taxa de falso-negativo enquanto diminui a sensibilidade.^{14,25,126} Isso pode ser observado no trabalho de Ashcraft e Van Herle (1981)³ que, ao empregarem a DAM, obtiveram valores de 92,0%, 74,0% e 2,5% para as taxas de sensibilidade, especificidade e falso-negativo, respectivamente. A percentagem de tumores malignos encontrados à operação foi de 45,0%. Usando-se a DEM os autores encontraram valores de 73,0%, 99,0% e 6,0% para as taxas de sensibilidade, especificidade e falso negativo, respectivamente. A proporção de tumores malignos encontrados à operação foi de 95,0%. A DAM também foi usada nos trabalhos de Alves et al (2002)¹¹⁵ e El Hag et al (2003)⁹⁹, que obtiveram sensibilidade e especificidade, respectivamente, de 94,0% e 97,0% para o primeiro e de 85,7% e 97,6% para o segundo, sendo que, neste último, a especificidade passou para 100,0% quando os casos de NF foram excluídos da análise. Estes resultados são semelhantes aos descritos por Goellner et al (1987)¹⁸, que encontraram especificidade de 99,0% quando consideraram somente os casos malignos (DEM), de 47,0% quando consideraram os casos malignos e as NF (DAM) e de 27,0% quando consideraram apenas as NF. Haas et al (1993)⁴⁹ reportaram baixa especificidade (36,0%), o que corresponde a 13 dos 36 casos confirmados de câncer. Entretanto, se considerarmos que os outros 23 (64,0%) casos corresponderam a lesões foliculares (22 adenomas e um tumor de células de Hürthle), a especificidade passa para 100,0% (em termos de lesões que requerem tireoidectomia para excluir malignidade).

Em nosso estudo, os VPP e VPN, tanto para o patologista A quanto para o B, foram muito semelhantes. Quando se utilizou a DAM, os valores referentes ao VPN aumentaram em todas as situações (A1, A2, B1 e B2), o que é facilmente explicável, ao serem considerados malignos os casos compatíveis com NF. Os valores de VPN calculados neste trabalho (variação de 82,6% a 93,2%) foram semelhantes aos descritos por outros autores.^{46,47,130} Idealmente, no teste de triagem, o VPN deveria ser otimizado (menos resultados falso-negativos), mesmo com o ônus do aumento de resultados falso-positivos.⁵³ Em contrapartida, o uso da DAM diminuiu o VPP, como pode ser observado na tabela 19, exceto para o patologista A nas duas análises (A1 e A2), onde os valores passaram de 89,0% para 91,0%. Isto ocorreu porque o VPP é mais influenciado pela especificidade. Apesar do patologista B, tanto em B1 quanto em B2, apresentar aumento da

sensibilidade, houve redução da especificidade quando se utilizou a DAM, o que explica a diminuição do VPP observada para o mesmo.

A análise dos cálculos de predição diagnóstica confirma as afirmações de que a RVP e o VPP são mais influenciados pela especificidade, enquanto a RVN e o VPN são mais influenciados pela sensibilidade.^{53,102,106,155,166}

Os valores de VPP (variação de 72,2% a 100,0%) foram semelhantes aos descritos por Silverman et al (88,9%)⁴⁶, mas superiores aos descritos por Tabaqchali et al (65,5%)¹³⁰ e Pepper et al (38,0%)⁴⁷. No estudo de Pepper et al⁴⁷, das 102 PAAF, 84 (82,0%) corresponderam à benignidade e 18 (18,0%) à NF, não sendo observado nenhum tumor maligno. Somente 21 pacientes foram operados (inclusive um caso falso-negativo), sendo oito do grupo benigno e 13 do grupo de NF, com cinco (28,0%) casos confirmados de malignidade (CP (n = 3) e CF (n = 2)). A maior crítica àquele estudo refere-se ao pequeno tamanho da amostra (n = 21), o que acarreta ressalvas na sua interpretação.

Na atual pesquisa, o DCO foi feito por patologistas professores lotados no Departamento de Anatomia Patológica da FM–UFMG. Por se tratar de instituição de ensino onde os casos duvidosos costumam ser discutidos, o DCO poderia ser considerado, para efeito de interpretação, como diagnóstico emitido por patologistas experientes. Entretanto, quando se analisa o número de exames por ano, observa-se que, durante o período de 30 meses, foram realizados 225 exames com média de 7,5 procedimentos ao mês. Como existem muitos professores no referido Departamento, pode-se extrapolar que o número de exames efetivamente interpretados por cada membro pode ser ainda menor. No entanto, sabe-se que muitos deles também analisam amostras em outros Serviços, daí ser difícil retirar conclusão adequada sobre a experiência na interpretação do DCO.

A exata incidência de tumores malignos “perdidos” (não diagnosticados) não é conhecida, assim como a taxa de falso-negativo. A taxa de resultados falso-negativos descrita mais comumente na literatura é estimada entre 2,0% e 7,0%, podendo variar entre 1,0% e 26,0%.^{9,14,16,29,49,50,59,61,124,126,132,136,142} Esta ampla faixa de variação pode ser explicada pela impossibilidade de se aplicar o exame padrão a todos os casos citologicamente considerados benignos. Somente pequena percentagem dos pacientes, aproximadamente 10,0%, com diagnóstico citológico compatível com benignidade, é, de fato, operada, tendo como indicação razões clínicas.¹⁴ Deve-se ressaltar o fato de que muitos tumores malignos da glândula tireoide

têm comportamento indolente, de modo que alguns indivíduos não manifestam, clinicamente, o câncer.^{50,102,126} Portanto, a verdadeira frequência de resultados falso-negativos e verdadeiro-negativos é difícil de ser avaliada, pelo fato desses pacientes não terem sido submetidos ao exame padrão.^{14,41,53,106,115} Acreditamos que as taxas de falso-negativo reportadas na literatura são subapreciadas, uma vez que grande parte das operações é baseada nos resultados da PAAF (baixa proporção de tireoidectomias em relação ao número de exames). Com isso, se pacientes com resultados à PAAF negativos, indeterminados ou insuficientes fossem operados, poder-se-ia esperar maior número de tireoidectomias, e, provavelmente, de resultados falso-negativos, como observado no nosso estudo.

Outra explicação para ampla variação nas taxas de falso-negativos deve-se, em parte, à inclusão ou não de microcarcinomas (tumores menores que 1cm no maior diâmetro) na avaliação dos resultados. Estes tumores são, geralmente, diagnosticados incidentalmente em exames histológicos, e, não raramente, são encontrados próximos a nódulos grandes, que provavelmente foram o alvo da punção.^{16,67} Os autores adeptos da exclusão dos incidentalomas alegam que estas lesões não foram originalmente investigadas pela PAAF, pois não correspondiam ao nódulo dominante à palpação ou não foram selecionadas para aspiração durante o exame ultrassonográfico.⁷² O nódulo pode não ser adequadamente amostrado por não ter sido, de fato, puncionado, falha esta conhecida como “perda geográfica”. Entretanto, a abordagem mais correta é considerar tais lesões “perdidas” como erro de amostragem, e, conseqüentemente, como falso-negativas.

No estudo de Blansfield et al (2002)¹²⁶, a sensibilidade e a especificidade obtidas foram de 80,0%; contudo, ao serem excluídos os microcarcinomas, os valores se elevaram para 91,0% e 83,0%, respectivamente. No entanto, consideramos que o diagnóstico final a ser considerado deve ser o mais importante e o mais preciso, isto é, aquele que requer tratamento mais agressivo e que tenha implicações prognósticas, independentemente se a glândula apresenta outras afecções. Dessa maneira procedeu-se neste estudo. Segundo Yang et al (2002)¹⁷⁶, alguns microcarcinomas são, clinicamente, as principais lesões porque eles poderiam ser multifocais e apresentarem doença metastática. Portanto, se a glândula possui um tumor maligno, independentemente do fato de ter sido ou não amostrado, ele deve ser considerado como o diagnóstico final do paciente. Caso a PAAF não o tenha amostrado, o resultado deverá ser computado como falso-negativo.^{58,59,87,105,128,177,178}

Na maioria das séries, os resultados falso-negativos são usualmente causados por erros de amostragem, tais como espécimes inadequados (amostragem inadequada do nódulo, como ocorre, por exemplo, em caso de lesões focais) ou preparação inapropriada do esfregaço.^{1,15,18,29,43,44,49-52,61,72,99,105,126} Em menor grau, os falso-negativos resultam da interpretação equivocada dos aspectos citológicos, o que se traduz na inabilidade do patologista para reconhecer a lesão maligna (erro de interpretação).^{1,30,53,58,59,99,105,142} Segundo Bakhos et al (2000)⁶⁵, a principal causa de diagnóstico falso-negativo consistiu de material inadequado e subdiagnóstico do CP associado à degeneração cística; por outro lado, os resultados falso-positivos estavam relacionados ao diagnóstico de NF. Segundo os autores, a PAAF pode apresentar altas sensibilidade e especificidade quando realizada por clínicos ou patologistas experientes. A grande variabilidade observada nas taxas de falso-negativos pode refletir, também, diferentes níveis de experiência, já que ela pode ser reduzida quando a PAAF é realizada pelo mesmo médico.^{51,52} Atribui-se este fato a maior habilidade e experiência do puncionador. Como a PAAF é considerada procedimento de triagem, deve-se tentar minimizar os diagnósticos falso-negativos, mesmo às custas do aumento de resultados falso-positivos.¹⁷ O que torna menos grave os resultados falso-negativos é o fato de que em algumas ocasiões a operação é realizada, tendo como indicações os dados da história clínica.¹⁰² O mesmo ocorre com pacientes cujas amostras são insuficientes que, usualmente, são operados; no entanto, nesta situação, são descritas baixas taxas de malignidade¹²⁶, como pode ser verificado no estudo de Löwhagen et al (1979)¹⁷⁹ que encontraram baixas taxas de falso-positivo (0,0%) e de falso-negativo (2,2%).

Os casos compatíveis com NF não foram excluídos dos cálculos de predição (VPP, VPN, RVP e RVN) e validação de testes diagnósticos (sensibilidade, especificidade e acurácia)^{61,72}, já que eles não representam amostra inadequada, mas sim a presença de casos nos quais não se pode, seguramente, determinar a natureza benigna ou maligna da lesão. Eles foram considerados de duas maneiras distintas de acordo com a definição de malignidade (DEM ou DAM), como sugerido por vários autores.^{3,7,14,15,25,28,53,61,126,157} Outros, como de Vos Tot Nederveen Cappel et al (2001)¹⁵, Mahar et al (2006)⁶¹ e Gerhard e Santos (2007)⁷² excluíram desses cálculos tanto os aspirados insuficientes como aqueles compatíveis com NF, por considerarem que ambos não fornecem informação diagnóstica conclusiva.

As taxas reportadas de falso-negativo na atual pesquisa variaram de 38,5% a 56,2% quando foi utilizada a DEM, e de 18,8% a 33,4% quando foi usada a DAM. Portanto, elas foram

muito maiores que as relatadas por Grant et al (0,7%)¹⁷², Hall et al (1,3%)⁵³ e Ashcraft e Van Herle (1,7%)³. Por outro lado, elas foram semelhantes às descritas por Ko et al (21,5%)⁶⁷ e Sidawy et al (19,0%)²⁹ quando foi usada apenas a DAM para o patologista A na primeira análise (A1) e para o patologista B na segunda análise (B2). O que se observa é que elas foram muito elevadas em relação à literatura, principalmente quando utilizada a DEM. Isto é explicado pelo fato da DEM não incluir os casos compatíveis com NF e que 20,0% a 25,0% destes correspondem a tumores malignos.^{13,18,30,116,179} A prevalência de afecções foliculares confirmadas histologicamente foi de 10 (13,5%) casos, portanto, dentro do percentual descrito de 10,0% a 30,0% na literatura.^{9,14,15,21,29,30,68,88,92,112} No entanto, na nossa pesquisa as taxas de malignidade variaram entre 50,0% e 100,0%. Além do mais, sabe-se que o CP representa um dos tumores mais facilmente diagnosticados à citologia e dos que apresentam o maior índice de reprodutibilidade.^{16,57,77,78} Quando se avalia sua prevalência na amostra nota-se que ela foi relativamente alta (n = 13; 76,5% do total de câncer e 17,6% dos 74 casos). Portanto, era de se esperar taxas menores de resultados falso-negativos. Provável explicação encontra-se na elevada taxa de tireoidectomia (83,7%) em relação às descritas na literatura.^{14,28,32,33,46,53,105} Podemos conjecturar que a indicação cirúrgica baseada em outras razões (história clínica, US, etc) permitiu a identificação de parcela de tumores malignos que se fossem operados apenas com base no resultado maligno à PAAF não teriam sido descobertos. Outro fato observado é de que a probabilidade de ocorrência de resultados falso-negativos foi semelhante para ambos os patologistas, com e sem experiência, tanto na primeira (A1 e B1) quanto na segunda análise (A2 e B2), como evidenciado nas tabelas 7 e 8.

Um fator que determina a qualidade do aspirado é a falta de consciência a respeito dos aspectos técnicos básicos envolvidos na preparação e fixação das lâminas incluindo demora na fixação ou má técnica, o que resulta na presença de artefatos e células esmagadas. Um esfregaço mal fixado ou indevidamente corado é inadequado para avaliação, mesmo para os mais experimentados citopatologistas. Portanto, a acurácia diagnóstica da citologia aspirativa depende também do adequado manuseio e preparo de amostras representativas.^{1,29,43,50} Belfiore e La Rosa (2001)²⁷ recomendam que cada centro deve alcançar e manter alto padrão de perícia em todos os passos do procedimento bem como implantar guias clínicos objetivando melhores resultados. Isto pode explicar a existência de algumas discrepâncias observadas nos resultados.^{42,43,52,53,61,119}

Alternativamente, pode-se realizar a punção biópsia com agulha grossa (PBAG) em detrimento da agulha fina, o que poderia propiciar fragmento de tecido adequado para histologia convencional. Porém, a PBAG é mais invasiva que a PAAF além de, potencialmente, causar maior incidência de complicações (hemorragia, punção traqueal e lesão do nervo laríngeo recorrente) e causar mais desconforto devido o grosso calibre da agulha.^{3,87} Até o momento, não há nenhuma evidência de que a PBAG tenha qualquer vantagem sobre a PAAF na investigação clínica de nódulos tireoidianos, mesmo na distinção entre tumores foliculares benignos e malignos.^{13,180} Mehotra et al (2006)¹⁸⁰ publicaram os resultados comparando a PBAG guiada por US versus PAAF da tireoide guiada pela palpação. Mesmo com a vantagem do US, os autores não conseguiram demonstrar vantagem clara da PBAG, que apesar de ter menor número de resultados insuficientes (fato que pode ser atribuído ao uso do US) apresentou menor especificidade para o diagnóstico de malignidade. Nos casos de PBAG, com material insuficiente, em 25,0% dos casos a investigação posterior demonstrou malignidade, fato que aconteceu em 11,0% dos casos de material insuficiente pela PAAF. Estes resultados reforçam a idéia de que todos os casos de PAAF com material insuficiente necessitam serem reavaliados. Deve-se ainda levar em conta a possibilidade de ocorrência de alterações teciduais secundárias ao uso da agulha grossa, como no caso da mama, que podem prejudicar a análise histológica final.¹⁰⁸

A condução de pacientes portadores de nódulos múltiplos ainda é controversa. Em primeiro lugar, cerca da metade dos pacientes com nódulos clinicamente solitários possuem, de fato, múltiplos nódulos à operação.¹²² Em segundo lugar, a presença de múltiplos nódulos não pode ser considerada fator preditivo de benignidade.⁸⁶ Alguns estudos têm demonstrado não haver diferença significativa em relação à incidência de câncer de tireoide, quando se comparam glândulas uninodulares com as multinodulares.^{92,122,181,182} Barroeta et al (2006)¹⁸³ objetivaram avaliar se existe relação entre o risco de malignidade e o número de nódulos tireoidianos. A taxa de malignidade foi de 39,0% para pacientes portadores de nódulos únicos, de 41,0% quando havia dois nódulos e de 21,0% quando havia três nódulos. Porém, deve-se destacar que 153 pacientes portadores de microcarcinoma papilar foram excluídos das análises. Concluíram que o risco de malignidade é similar para pacientes com um ou dois nódulos, mas diminui na presença de três ou mais nódulos; conclusão semelhante foi corroborada por outros autores.^{141,177}

É comum que somente o nódulo dominante seja puncionado, ficando os demais sem serem amostrados, possibilitando que o câncer passe despercebido.¹⁷⁷ Portanto, no caso de bócios

multinodulares, existe a dúvida de qual ou quais devem ser puncionados, com a preferência recaindo, geralmente, sobre aqueles suspeitos clínica e ultrassonograficamente (hipoecóicos, limites indefinidos, presença de microcalcificações, etc). Como esses sinais não são patognomônicos de malignidade e sabendo-se que nem sempre o nódulo dominante corresponde à lesão tumoral, além do que a incidência do câncer pode ser até maior que a encontrada em bócios uninodulares, o ideal seria puncionar todos os nódulos.^{86,92,122,141,177,181-183}

Vários estudos têm demonstrado que os erros de amostragem são a principal causa de perda diagnóstica, daí a necessidade de que todos os profissionais que realizam a PAAF apresentem adequado treinamento em todos os passos do processo.^{41,145,170} Para Werga et al (2000)⁴¹, um dos maiores problemas do método é que a PAAF parece simples, tanto que muitos médicos a realizam de maneira casual, o que dificulta o acúmulo de experiência com o procedimento. A qualidade dos serviços prestados deve ser mantida em alto nível. Segundo Singh et al (2003)⁴³, a qualidade com que a PAAF é realizada deveria ser auditada periodicamente, principalmente em relação à taxa de inadequácia e aos aspectos técnicos do método.

No presente estudo, as taxas de falso-positivo variaram de 2,0% a 10,9% e foram semelhantes às reportadas na literatura, onde existem relatos que variam de 0,0% a 11,0% (Tabela 19).^{16,31,58-60,98,136,179,184} As maiores taxas foram observadas para o patologista B quando se usou a DAM para o cálculo (B1 = 6,0% e B2 = 10,9%). O mesmo ocorreu no estudo de Asp et al (1987)²⁸, onde dez casos foram falso-positivos, todos com diagnóstico de NF e considerados malignos (DAM). O caso falso-positivo apresentado pelo patologista A deveu-se ao diagnóstico equivocado de CI em caso de BC; o mesmo ocorreu para o patologista B que identificou equivocadamente um caso de CM, que também se tratava de BC. O maior número de falso-positivo apresentado por este patologista é reflexo do maior número de diagnósticos compatíveis com NF, quando a DAM foi considerada. No entanto, a comparação entre as taxas de falso-positivo, observadas para cada examinador, mostra que elas foram semelhantes (Tabela 19), principalmente quando a DEM foi utilizada. Para o patologista A, elas foram idênticas nas duas análises independentemente do uso da DEM ou da DAM. Entretanto, a análise interexaminador, mostrou que apesar de não ser significativa no primeiro momento (A1 – B1) ($p = 0,62$), a probabilidade do patologista B reportar resultado falso-positivo foi maior em relação ao patologista A com um *OR* de 3,08 e [0,28 – 78,85]. No segundo (A2 – B2) momento, não houve diferença significativa ($p = 0,21$) apesar da probabilidade do patologista B reportar resultado

falso-positivo tenha sido maior com *OR* de 5,29 e [0,58 – 122,74]. Deve ser salientada a presença de IC muito amplos, o que indica que a amostra usada foi pequena para análise estatística mais adequada.

Goellner et al (1987)¹⁸ apresentaram somente um caso (0,5%) falso-positivo em 221 de sua série que correspondeu ao paciente com tireoidite de Hashimoto; esta taxa é semelhante à descrita por Bottles et al (1986)³¹ que encontraram valor menor que 0,1%. A sua principal causa é atribuída a erros de interpretação, principalmente na presença de tireoidite de Hashimoto.⁶⁰ Quase todos esses erros resultam da confusão de nódulos hiperplásicos e lesões de células de Hürthle com o CP e o CF.^{58,59} Os resultados falso-positivos têm sido causa importante de indicação excessiva de tireoidectomias.⁶¹

Para Roskel & Buley (2004)⁶², o citopatologista deveria informar o grau de certeza quando libera o laudo citológico e deixar claro quando a correlação com os dados clínicos e imaginológicos são essenciais. Por sua vez, os clínicos e cirurgiões que recebem o resultado do exame citológico, deveriam reconhecer quando as possibilidades diagnósticas são reduzidas ou mesmo confirmadas, principalmente quando se refere ao tumor maligno. Portanto, além da experiência na interpretação citológica da PAAF, a estreita correlação com a história clínica e os dados da propedêutica são essenciais para a melhora da sua acurácia e para a condução do paciente.^{4,17,50}

Programas que assegurem a garantia de qualidade e de melhoria da qualidade deveriam ser parte integrante da prática profissional relacionada à PAAF. Todo laboratório e cada equipe médica devem documentar seu desempenho, monitorizar seus resultados e compará-los com os relatados na literatura (controle de qualidade continuado).¹⁶⁵

Apesar dos benefícios evidentes incluindo o de apresentar elevado VPN para neoplasias e permitindo, com certa precisão, a distinção entre lesões benignas (BC e tireoidites) e malignas (CP, CF e CA), a PAAF apresenta importantes limitações que incluem: incapacidade de distinguir, de forma inequívoca, as lesões foliculares encapsuladas;^{14,18,41,105} presença de resultados equivocados (falso-negativo e falso-positivo)⁸⁷; dificuldade em se detectar o microcarcinoma papilar e erros de coleta, preparo e interpretação dos esfregaços.¹⁰⁵ Suas limitações e as ocasionais complicações da técnica têm sido relatadas como justificativa para abandono da PAAF, porém seus resultados precisam ser comparados com dados mais robustos de modalidades diagnósticas alternativas antes de se cogitar seu abandono.⁶²

Na nossa opinião, a PAAF, isoladamente, não pode determinar a escolha entre tratamento clínico ou cirúrgico das lesões nodulares da glândula tireoide, como corroborado por outros autores.²³ Em hipótese alguma ela deve substituir o exame histológico, principalmente nos casos onde existem razões, clínicas ou ultrassonográficas, para a suspeição de malignidade.^{87,97,165} Devido ao considerável potencial de malignidade observado nos casos de NF e em amostras insuficientes, a política do autor, assim como a de outros, é indicar a operação em todos os pacientes que apresentem estes diagnósticos, ou pelo menos repetir a PAAF no caso de resultados insuficientes. Caso assim persista, a operação deve ser indicada.^{1,23,47,87,165,185} A indicação cirúrgica de nódulos tireoidianos deve levar em consideração a elevada taxa de resultados falso-negativos observada à PAAF, o que ocorre mesmo quando interpretada por patologista experiente, assim como a alta taxa de resultados insuficientes, principalmente devido as elevadas taxas de malignidade observadas nestas amostras.

Deve ser enfatizado que a PAAF é apenas um exame propedêutico complementar que pode ser valioso quando usado apropriadamente na condução de pacientes portadores de nódulos tireoidianos. Muitos estudos têm tentado encontrar testes preditivos de malignidade para o grupo suspeito (NF), mas os resultados são discordantes.^{112,186} Microscopia eletrônica, citometria de fluxo, imunohistoquímica e marcadores genéticos têm sido avaliados com o objetivo de determinar se eles são capazes de melhorar a confiabilidade diagnóstica nestes casos. Embora os resultados globais tenham sido desapontadores, dois desses marcadores (*human bone marrow endothelial cell [HBME-1]* e *galectin-3*) têm mostrado resultados promissores com relação à habilidade de prever malignidade.¹⁸⁷

Profissionais que utilizam a PAAF precisam ter abordagem realística do que é alcançável com o método. Como qualquer teste diagnóstico envolvendo amostragem e interpretação, diagnósticos importantes podem não ser feitos.⁶² O impacto negativo da reafirmação de um resultado equivocado, no contexto da detecção ou não do câncer de tireoide, consolida a necessidade de se ter elevado nível de confiabilidade, inter e intraexaminador, para procedimentos diagnósticos, principalmente para aqueles que são operador-dependente, tal como a PAAF da glândula tireoide.⁶³ Para Hamady et al (2005)¹⁵⁰ o tratamento dos tumores de tireoide deveria ser responsabilidade de equipe multidisciplinar e especializada que inclua um patologista experiente. É imperativo que clínicos, cirurgiões e citopatologistas tenham entre si uma comunicação clara para se assegurarem de que o procedimento proposto é o mais apropriado para

a questão levantada.⁶² Assim como existem critérios estritos para estabelecer se a amostra é adequada, parâmetros de treinamento do citopatologista e do profissional que realiza a punção devem ser estabelecidos e avaliados no momento de se considerar o laudo citológico.

Estudos adicionais envolvendo principalmente séries com maior número de pacientes são necessários para determinar se a reprodutibilidade (inter e intraexaminador) em citologia da tireoide pode ser melhorada por mais extensivo treinamento, mudanças na classificação citológica das lesões ou pelo aumento do número de células coletadas.⁴² Treinamento formal na obtenção do material e na leitura das lâminas está associado à melhora da sensibilidade, da especificidade e da acurácia da PAAF.^{12,28,43,45,51,52,87,175} Por fim, o que se espera para os próximos decênios é de que se avance mais no conhecimento sobre os fatores biológicos de prognóstico, que poderiam ser obtidos pela própria PAAF, o que poderia sedimentar a sua indicação e sua importância nas afecções da glândula tireoide.¹

CONCLUSÕES

1. A reprodutibilidade intraexaminador na interpretação citológica da PAAF é consistente e independe da experiência com o exame. Já a reprodutibilidade interexaminador é influenciada pela experiência, evidenciada, principalmente, pelas diferenças nas taxas de resultados insuficientes e de casos compatíveis com NF.
2. A baixa reprodutibilidade interexaminador associada à elevada reprodutibilidade intraexaminador pode sugerir o uso de critérios diagnósticos próprios com consistência ao longo do tempo.
3. A utilização da DAM em detrimento da DEM na avaliação de NF eleva a acurácia da PAAF.

PROPOSIÇÕES

1. Avaliar os motivos que resultaram na baixa taxa de reprodutibilidade, tanto inter quanto intraexaminador, observada com o CP.
2. Análise das lâminas presentes neste estudo (n = 74) por um terceiro ou mesmo quarto examinador (com e sem experiência ou ambos com ou sem experiência) para comparação com os resultados observados neste estudo.
3. Análise das lâminas de histologia para a verificação da acurácia e reprodutibilidade do exame anátomo-patológico.
4. Aumentar o tamanho da amostra para pelo menos 100 casos para a verificação da sua influência nos resultados.
5. Realizar estudo comparativo entre a PAAF e a PBAG.
6. Realizar estudo semelhante, porém utilizando intervalo de tempo superior a seis meses, para avaliar se a experiência adquirida terá influência nos resultados, principalmente, sobre a reprodutibilidade intraexaminador.
7. Investigar porque a PAAF foi pouco utilizada no HC-UFMG e porque ela é pouco usada para orientar a indicação de tireoidectomias.

REFERÊNCIAS*

1. Shaha AR. Controversies in the management of thyroid nodule. *Laryngoscope* 2000;110:183-93.
2. Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi (AACE/AME). Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract* 2006;12:63-102.
3. Ashcraft MW, Van Herle AJ. Management of thyroid. II. Scanning techniques, thyroid suppressive therapy, and fine needle aspiration. *Head Neck Surg* 1981;3:297-322.
4. Guidelines of the Papanicolaou Society of Cytopathology for the examination of fine-needle aspiration specimens from thyroid nodules. The Papanicolaou Society of Cytopathology Task Force on Standards of Practice. *Diagn Cytopathol* 1997;17:239-47.
5. Gupta DK, Mooney EE, Layfield LJ. Fine-needle aspiration cytology: a survey of current utilization in relationship to hospital size, surgical pathology volume, and institution type. *Diagn Cytopathol* 2000;23:59-65.
6. Hundahl SA, Cady B, Cunningham MP et al. Initial results from a prospective cohort study of 5583 cases of thyroid carcinoma treated in the United States during 1996. *Cancer* 2000;89:202-17.
7. Dwarakanathan AA, Ryan WG, Staren ED et al. Fine needle aspiration biopsy in the thyroid. *Arch Intern Med* 1989;149:2007-9.
8. Ramacciotti CE, Pretorius HT, Chu EW et al. Diagnostic accuracy and use of aspiration biopsy in the management of thyroid nodules. *Arch Inter Med* 1984;144:1169-73.
9. The American Thyroid Association Guidelines Taskforce: management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2006;16:1-33.

*Referências bibliográficas normalizadas de acordo com International Committee of Medical Journal Editors.

Uniform requirements for Manuscripts submitted to Biomedical Journals. *Ann Intern Med* 1997; 126:36-47. Updated May 2008 available from: URL: <http://www.icmje.org>.

10. Ashcraft MW, Van Herle AJ. Management of thyroid nodules. I: History and physical examination, blood tests, X-ray tests, and ultrasonography (review). *Head Neck Surg* 1981;3:216-27.
11. Nasuti JF, Gupta PK, Baloch ZW. Diagnostic value and cost-effectiveness of on-site evaluation of fine-needle aspiration specimens: review of 5,688 cases. *Diagn Cytopathol* 2002;27:1-4.
12. Castro MR, Gharib H. Thyroid fine-needle aspiration biopsy: progress, practice, and pitfalls. *Endocr Pract* 2003;9:128-36.
13. Mazzaferri EL. Management of a solitary thyroid nodule. *N Engl J Med* 1993;329:553-9.
14. Gharib H, Goellner JR. Fine-needle aspiration biopsy of the thyroid: an appraisal. *Ann Intern Med* 1993;118:282-9.
15. de Vos tot Nederveen Cappel RJ, Bouvy ND, Bonjer HJ et al. Fine needle aspiration cytology of thyroid nodules: how accurate is it and what are the causes of discrepant cases? *Cytopathology* 2001;12:399-405.
16. Löwhagen T, Willens JS, Lundell G et al. Aspiration biopsy cytology in the diagnosis of thyroid cancer. *World J Surg* 1981;5:61-73.
17. The Papanicolaou Society of Cytopathology Task Force on Standards of Practice. Guidelines of the Papanicolaou Society of Cytopathology for the examination of fine-needle aspiration specimens from thyroid nodules. *Mod Pathol* 1996;9:710-5.
18. Goellner JR, Gharib H, Grant CS et al. Fine needle aspiration cytology of the thyroid, 1980 to 1986. *Acta Cytol* 1987;31:587-90.
19. Ravetto C, Colombo L, Dottorini ME. Usefulness of fine needle aspiration in the diagnosis of thyroid carcinoma: a retrospective study in 37895 patients. *Cancer* 2000;90:357-63.
20. Burch HB, Burman KD, Reed HL et al. Fine needle aspiration of thyroid nodules: determinants of insufficiency rate and malignancy yield at thyroidectomy. *Acta Cytol* 1996;40:1176-83.
21. Gharib H, Goellner JR, Johnson DA. Fine needle aspiration cytology of the thyroid. A 12-year experience with 11,000 biopsies. *Clin Laboratory Med* 1993;13:699-709.
22. AACE clinical practice guidelines for the diagnosis and management of thyroid nodules. Thyroid Nodule Task Force. *Endocr Pract* 1996;2:78-84.

23. Parmeggiani U, Bove F, Calabria M et al. Needle aspiration cytology in thyroid surgery (our experience in 104 operated cases). *Ann Ital Chir* 1994;65:59-67.
24. Thyroid Carcinoma Task Force. AACE/AAES medical / surgical guidelines for clinical practice: management of thyroid carcinoma. *Endocr Pract* 2001;7:202-20.
25. Van Herle AJ, Rich P, Ljung BE et al. The thyroid nodule. *Ann Intern Med* 1982;96:221-32.
26. Rojeski MT, Gharib HN. Nodular thyroid disease. Evaluation and management. *N Engl J Med* 1985;313:428-36.
27. Belfiore A, La Rosa GL. Fine-needle aspiration biopsy of the thyroid. *Endocrinol Metab Clin North Am* 2001;30:361-400.
28. Asp AA, Georgitis W, Waldron EJ et al. Fine needle aspiration of the thyroid: use in an average health care facility. *Am J Med* 1987;83:489-93.
29. Sidawy MK, Del Vecchio DM, Knoll SM. Fine-needle aspiration of thyroid nodules: correlation between cytology and histology and evaluation of discrepant cases. *Cancer* 1997;81:253-9.
30. Caruso D, Mazzaferrri EC. Fine-needle aspiration biopsy in the management of thyroid nodules. *Endocrinol* 1991;1:194-202.
31. Bottles K, Miller TR, Coehn MB et al. Fine-needle aspiration biopsy. Has its time come? *Am J Med* 1986;81:525-31.
32. Miller JM, Hamburger JI, Kini SR. The impact of needle biopsy on the preoperative diagnosis of thyroid nodules. *Henry Ford Hosp Med J* 1980;28:145-8.
33. Hamberger B, Gharib H, Melton LJ III et al. Fine needle aspiration biopsy of the thyroid nodules: impact on thyroid practice and cost of care. *Am J Med* 1984;73:381-4.
34. Castro MR, Gharib H. Continuing controversies in the management of thyroid nodules. *Ann Intern Med* 2005;142:926-31.
35. Galloway JW, Sardi A, DeConti RW et al. Changing trends in thyroid surgery. 38 years' experience. *Am Surg* 1991;57:18-20.
36. Julian JS, Pittman CE, Accettullo L et al. Does fine needle aspiration biopsy really spare patients thyroidectomy? *Am Surg* 1989;55:238-42.
37. Leonard N, Melcher DH. To operate or not to operate? The value of fine-needle aspiration cytology in the assessment of thyroid swellings. *J Clin Pathol* 1997;50:941-3.

38. Ng EH, Tan SK, Nambiar R. Impact of fine needle aspiration cytology on the management of solitary thyroid nodules. *Aust N Z J Surg* 1990;60:463-6.
39. Ribeiro CA, Brasileiro Filho G. Histocytological correlation in fine needle aspiration biopsy of the thyroid gland. *Rev Assoc Med Bras* 1996;42:73-8.
40. Suen KC. Fine-needle aspiration biopsy of the thyroid. *CMAJ* 2002;167:491-5.
41. Werga P, Wallin G, Skoog L et al. Expanding role of fine-needle aspiration cytology in thyroid diagnosis and management. *World J Surg* 2000;24:907-12.
42. Ljung BM, Drejet A, Chiampi N et al. Diagnostic accuracy of fine-needle aspiration biopsy is determined by physician training in sampling technique. *Cancer* 2001;93:263-8.
43. Singh N, Ryan D, Berney D et al. Inadequate rates are lower when FNAC samples are taken by cytopathologists. *Cytopathology* 2003;14:327-31.
44. Wu M, Burstein DE, Yuan S et al. A comparative study of 200 fine needle aspiration biopsies performed by clinicians and cytopathologists. *Laryngoscope* 2006;116:1212-5.
45. Padel AF, Coghill SB, Powis SJ. Evidence that sensitivity is increased and the inadequacy rate decreased when pathologists take aspirates for cytodiagnosis. *Cytopathology* 1993;4:161-5.
46. Silverman JF, West RL, Larkin EW et al. The role of fine needle aspiration biopsy in the rapid diagnosis and management of thyroid neoplasm. *Cancer* 1986;57:1164-70.
47. Pepper GM, Zwickler D, Rosen Y. Fine-needle aspiration biopsy of the thyroid nodule: results of a start-up project in a general teaching hospital setting. *Arch Inter Med* 1989;149:594-6.
48. Liel Y, Zirkin HJ, Sobel RJ. Fine needle aspiration of the thyroid. Five years' experience with 183 patients. *Isr J Med Sci* 1985;21:719-21.
49. Haas S, Trujillo A, Kunstle J. Fine needle aspiration of thyroid nodules in a rural setting. *Am J Med* 1993;94:357-61.
50. Zakowski MF. Fine-needle aspiration cytology of tumors: diagnostic accuracy and potential pitfalls. *Cancer Investigation* 1994;12:505-15.
51. Lee KR, Foster RS, Papillo JL. Fine needle aspiration of the breast. Importance of the aspirator. *Acta Cytol* 1987;31:281-4.
52. Dixon JM, Lamb J, Anderson TJ. Fine needle aspiration of the breast: importance of the operator. *Lancet* 1983;2:564.

53. Hall TL, Layfield LJ, Philippe A et al. Sources of diagnostic error in fine needle aspiration of the thyroid. *Cancer* 1989;63:718-25.
54. Frable WJ, Frable MA. Thin needle aspiration biopsy of the thyroid gland. *Laryngoscope* 1980;90:1619-25.
55. Palombini L, Fulciniti F, Vetrani A et al. Fine needle aspiration biopsy of breast masses: a critical analysis of 1956 cases in 8 years (1977-1984). *Cancer* 1988;61:2273-7.
56. Hamburger JI. Extensive personal experience. Diagnosis of thyroid nodules by fine needle biopsy: use and abuse. *J Clin Endocrinol Metab* 1994;79:335-9.
57. Saxén E, Fransilla K, Bjarnason O et al. Observations in the classification of thyroid cancer. *Acta Pathol Microbiol Scand [A]* 1978;86:483-6.
58. Boey J, Hsu C, Collins RJ. False-negative errors in fine-needle aspiration biopsy of dominant thyroid nodules. A prospective follow-up study. *World J Surg* 1986;10:623-30.
59. Hsu C, Boey J. Diagnostic pitfalls in the fine-needle aspiration of thyroid nodules. A study of 555 cases in chinese patients. *Acta Cytol* 1987;31:699-704.
60. Meko JB, Norton JA. Large cystic/solid thyroid nodules: a potential false-negative fine-needle aspiration. *Surgery* 1995;118:996-1004.
61. Mahar SA, Husain A, Islam N. Fine needle aspiration cytology of thyroid nodule: diagnostic accuracy and pitfalls. *J Ayub Med Coll Abbottabad* 2006;18:26-9.
62. Roskell DE, Buley ID. Fine needle aspiration cytology in cancer diagnosis. *BMJ* 2004;329:244-5.
63. Visvanathan K, Santor D, Syed ZA. In Soon H, Davidson NE, Helzlsouer KJ. The importance of cytologic intrarater and interrater reproducibility: the case of ductal lavage. *Cancer Epidemiol Biomarkers Prev* 2006;15:2553-6.
64. Svanholm H, Starklint H, Gundersen HJ et al. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. *Apmis* 1989;97:689-98.
65. Bakhos R, Selvaggi SM, DeJong S et al. Fine-needle aspiration of the thyroid: rate and causes of cytohistopathologic discordance. *Diagn Cytopathol* 2000;23:233-7.
66. Gharib H. Current evaluation of thyroid nodules. *Trends Endocrinol Metab* 1994;5:365-9.
67. Ko HM, Jhu IK, Yang SH et al. Clinicopathologic analysis of fine needle aspiration cytology of the thyroid. A review of 1,613 cases and correlation with histopathologic diagnoses. *Acta Cytol* 2003;47:727-32.

68. Clary KM, Condel JL, Liu Y et al. Interobserver variability in the fine needle aspiration biopsy diagnosis of follicular lesions of the thyroid gland. *Acta Cytol* 2005;49:378-82.
69. Baloch ZW, Hendreen S, Gupta PK et al. Interinstitutional review of thyroid fine-needle aspirations: impact on clinical management of thyroid nodules. *Diagn Cytopathol* 2001;25:231-4.
70. Smith J, Cheifetz RE, Schneidereit N et al. Can cytology accurately predict benign follicular nodules? *Am J Surg* 2005;189:592-5.
71. Stelow EB, Bardales RH, Crary GS et al. Interobserver variability in thyroid fine-needle aspiration interpretation of lesions showing predominantly colloid and follicular groups. *Am J Clin Pathol* 2005;124:239-44.
72. Gerhard R, Santos GC. Inter and intraobserver reproducibility of thyroid fine-needle aspiration cytology: an analysis of discrepant cases. *Cytopathology* 2007;18:105-11.
73. Hirokawa M, Carney A, Goellner J, et al. Observer variation of encapsulated follicular lesions of the thyroid gland. *Am J Surg Pathol* 2003;27:1177-9.
74. Lloyd RV, Erickson LA, Sebo TJ. Diagnosis of follicular variant of papillary carcinoma by a panel of endocrine pathologists. *Mod Pathol* 2003;16:476A.
75. Franc B, de la Salmoniere P, Lange F et al. Interobserver and intraobserver reproducibility in the histopathology of follicular thyroid carcinoma. *Hum Pathol* 2003;34:1092-100.
76. Ron E, Griffel B, Liban E et al. Histopathologic reproducibility of thyroid disease in an epidemiologic study. *Cancer* 1986;57:1056-9.
77. Holm LE, Lowhagen T, Silfversward C. The reliability of malignant thyroid tumor diagnosis in the Swedish Cancer Registry. Review of 200 cases. *Acta Pathol Microbiol Scand* 1980;88:251-4.
78. Fassina AS, Montesco MC, Ninfo V et al. Histological evaluation of thyroid carcinomas: reproducibility of the "WHO" classification. *Tumori* 1993;79:314-20.
79. Lloyd RV, Erickson LA, Casey MB et al. Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. *Am J Surg Pathol* 2004;28:1336-40.
80. Webb AJ. Early microscopy: history of fine needle aspiration (FNA) with particular reference to goitres. *Cytopathology* 2001;12:1-6.
81. Koss LG, Woyke S, Olszewski W. Biopsia por aspiración; interpretación citológica y basis histológicas. Buenos Aires: Panamericana 1988;15-34,69-136.

82. MacDonald L, Yazdi HM. Nondiagnostic fine needle aspiration biopsy of the thyroid gland: a diagnostic dilemma. *Acta Cytol* 1996;40:423-8.
83. Kuma K, Matsuzuka F, Yokozawa T. Fate of untreated benign thyroid nodules: results of long-term follow-up. *World J Surg* 1994;18:495-8.
84. Franklyn JA, Sheppard MC. Aspiration cytology of thyroid. *BML* 1987;295: 510-1.
85. Miller JM, Hamburger JI, Kini SR. Diagnosis of thyroid nodules. Use of fine-needle aspiration and needle biopsy. *JAMA* 1979;241:481-4.
86. Cochand-Priollet B, Guillausseau PJ, Chagnon S et al. The diagnostic value of fine-needle aspiration biopsy under ultrasonography in nonfunctional thyroid nodules: a prospective study comparing cytologic and histologic findings. *Am J Med* 1994;97:152-7.
87. Wong CKM, Wheeler MH. Thyroid nodules: rational management. *World J Surg* 2000;24:934-41.
88. Giuffrida D, Gharib H. Controversies in the management of cold, hot, and occult thyroid nodules. *Am J Med* 1995;99:642-50.
89. Walsh RM, Watkinson JC, Franklin J. The management of the solitary thyroid nodule: a review. *Clin Otolaryngol* 1999;24:388-97.
90. Ezzat S, Sarti DA, Cain DR et al. Thyroid incidentalomas. Prevalence by palpation and ultrasonography. *Arch Intern Med* 1994;154:1838-40.
91. Brander AE, Viikinkoski VP, Nickels JI et al. Importance of thyroid abnormalities detected at US screening: a 5-year follow-up. *Radiology* 2000;215:801-6.
92. Souza DAT, Freitas HMP, Muzzi M et al. Ultrasound-guided fine needle aspiration biopsy of thyroid nodules: study of 63 cases. *Radiol Bras* 2004;37:333-5.
93. Sclabas GM, Staerkel GA, Shapiro SE et al. Fine-needle aspiration of the thyroid and correlation with histopathology in a contemporary series of 240 patients. *Am J Surg* 2003;186:702-10.
94. La Rosa GL, Belfiore A, Giuffrida D et al. Evaluation of the fine needle aspiration biopsy in the preoperative selection of cold thyroid nodules. *Cancer* 1991;67:2137-41.
95. Meurisse M, Detroz B, Hamoit E et al. Prediction of malignancy in solitary thyroid nodules: value of preoperative explorations. *Acta Chir Belg* 1994;94:25-9.
96. McHenry CR, Walfish PG, Rosen IB. Non-diagnostic fine needle aspiration biopsy: A dilemma in management of nodular thyroid disease. *Am Surg* 1993;59:415-9.

97. Wiersinga WN. Is repeated fine-needle aspiration cytology indicated in (benign) thyroid nodules? *Eur J Endocrinol* 1995;132:661-2.
98. Goldstein RE, Nettekville JL, Burkey B et al. Implications of follicular neoplasms, atypia and lesions suspicious for malignancy diagnosed by fine needle aspiration of thyroid nodules. *Ann Surg* 2002;235:656-64.
99. El Hag IA, Kollur SM, Chiedozi LC. The role of FNA in the initial management of thyroid lesions: 7-year experience in a district general hospital. *Cytopathology* 2003;14:126-30.
100. Saaman NA, Maheshwari YK, Nader S et al. Impact of therapy for differentiated carcinoma of the thyroid: an analysis of 706 cases. *J Clin Endocrinol Metab* 1983;56:1131-8.
101. Ashcraft MW, Van Herle AJ. Management of thyroid nodules. I: history and physical examination, blood tests, X-ray tests, and ultrasonography (review). *Head Neck Surg* 1981;3:216-27.
102. Hamming JF, Goslings BM, van Steenis GJ et al. The value of fine-needle aspiration biopsy in patients with nodular thyroid disease divided into groups of suspicion of malignant neoplasms on clinical grounds. *Arch Intern Med* 1990;150:113-6.
103. Guimarães EM, Morais DM, Silva SJ et al. Impacto da citologia aspirativa da tireoide com agulha fina: estudo da precisão diagnóstica. *Rev Ass Med Bras* 1996;42:2-6.
104. Lawrence W Jr, Kaplan BJ. Diagnosis and management of patients with thyroid nodules. *J Surg Oncol* 2002;80:157-70.
105. Tseng CE, Wei CK, Kuo CS et al. Fine needle aspiration cytology of thyroid nodules: evaluation of diagnostic accuracy. *Tzu Chi Med J* 2008;4:296-303.
106. Archibald S, Bhandari M, Thomas A. Users' guides to the surgical literature: how to use an article about a diagnostic test. *Can J Surg* 2001;44:17-23.
107. Hamburger JI. Consistency of sequential needle biopsy findings for thyroid nodules. Management implications. *Arch Intern Med* 1987;147:97-9.
108. Schmitt FC. Punção aspirativa por agulha fina da tireoide: a precisão e solidez de um método diagnóstico. *Rev Portug Endocrinol Diabetes e Metabol* 2007;2:5-7.
109. Nguyen GK, Lee MW, Ginsberg J et al. Fine-needle aspiration of the thyroid: an overview. *Cytojournal* 2005;2:12.

110. Matesa N, Dabelic N, Tabain I et al. Fine needle aspiration of the thyroid. *Acta Clin Croat* 2002;41:123-31.
111. Boutin P, Bozorg GA, Terrada C et al. Results of fine needle aspiration biopsy, frozen section diagnosis and definite histological results in thyroid pathology. Report of 163 cases. *Rev Laryngol Otol Rhinol* 2003;124:59-63.
112. Tuttle RM, Lemar H, Burch HB. Clinical features associated with an increased risk of thyroid malignancy in patients with follicular neoplasia by fine-needle aspiration. *Thyroid* 1998;8:377-83.
113. de Los Santos ET, Keyhani-Rofagha SK, Cunningham JJ et al. Cystic thyroid nodules: the dilemma of malignant lesions. *Arch Intern Med* 1990;150:1422-7.
114. Bakshi Na, Mansoor I, Jones BA. Analysis of inconclusive fine-needle aspiration of thyroid follicular lesions. *Endocr pathol* 2003;14:167-75.
115. Alves MLD, Maciel RMB, Valeri FV et al. Predictive value of clinical examination, scintigraphy, aspirative cytology and serum thyroglobulin in the nontoxic solitary thyroid nodule: prospective study of 110 operated patients. *Arq Bras Endocrinol Metab* 2002;46:648-53.
116. Gharib H, Goellner JR, Zinsmeister AR et al. Fine-needle aspiration biopsy of the thyroid. The problem of suspicious cytologic findings. *Ann Intern Med* 1984;101:25-8.
117. Gardner HA, Ducatman BS, Wang HH. Predictive value of fine needle aspiration of thyroid in the classification of follicular lesions. *Cancer* 1993;71:2598-603.
118. Goellmer JR, Johnson DA. Cytology of cystic papillary carcinoma of the thyroid. *Acta Cytol* 1982;26:797-9.
119. Oertel YC, Oertel JE. Diagnosis of benign thyroid lesions: fine-needle aspiration and histopathologic correlation. *Ann Diagn Pathol* 1998;2:250-63.
120. Kini SR, Miller JM, Hamburger JI. The cytopathology of thyroid nodule by fine needle aspiration. *Acta Cytol* 1978;22:605-6.
121. Petersen SV, Greisen O. Fine needle aspiration biopsy of the thyroid gland. *Otolaryngol Head Neck Surg* 1984;92:295-7.
122. McCall A, Jarosz H, Lawrence AM et al. The incidence of thyroid carcinoma in solitary cold nodules and in multinodular goiters. *Surgery* 1986;100:1128-31.
123. Frable WJ. Fine-needle aspiration biopsy: a review. *Hum Pathol* 1983;14:9-28.

124. Baloch ZW, Sack MJ, Yu GH et al. Fine-needle aspiration of thyroid: an institutional experience. *Thyroid* 1998;8:565–9.
125. Schmitt FC. Thyroid cytology: FNA is still the best diagnostic approach. *Cytopathology* 2006;17:211-2.
126. Blansfield JA, Sack MJ, Kukora JS. Recent experience with preoperative fine-needle aspiration biopsy of thyroid nodules in a community hospital. *Arch Surg* 2002;137:818-21.
127. Atallah AN. Avaliação crítica dos testes diagnósticos e suas aplicações. *Rer Bras Patol Clin* 1989;25:60-3.
128. Jayaram G. Fine needle aspiration cytologic study of the solitary thyroid nodule. Profile of 308 cases with histologic correlation. *Acta Cytol* 1985;29:967-73.
129. Elsheikh TM, Asa SL, Chan JK et al. Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. *Am J Clin Pathol* 2008;130:736-44.
130. Tabaqchali MA, Hanson JM, Johnson SJ et al. Thyroid aspiration cytology in Newcastle: a six year cytology/histology correlation study. *Ann R Coll Surg Engl* 2000;82:149-55.
131. Welker MJ, Orlov D. Thyroid nodules. *Am Family Physician* 2003;67:559-66.
132. Oertel YC, Oertel JE. Diagnosis of malignant epithelial thyroid lesions: fine needle aspiration and histopathologic correlation. *Ann Diagn Pathol* 1998;2:377-400.
133. Skaaning K. Fine needle biopsy as a routine examination: a histologically verified material of tumors of the head and neck submitted to fine needle biopsy. *Ugeskr Laeger* 1982;143:814-7 (abstract).
134. Suen KC. How does one separate cellular follicular lesions of the thyroid by fine needle aspiration biopsy? *Diagn Cytopathol* 1988;4:78-81.
135. Danese D, Sciacchitano S, Farsetti A, Andreoli M, Pontecorvi A. Diagnostic accuracy of conventional versus sonography-guided fine-needle aspiration biopsy of thyroid nodules. *Thyroid* 1998;8:15–21.
136. Liel Y, Ariad S, Barchana M. Long term follow up of patients with initially benign thyroid fine needle aspiration. *Thyroid* 2001;11:775-8.
137. Safirullah, Naeem M, Akbar K. Role of fine needle aspiration cytology (FNAC) in the diagnosis of thyroid swellings. *J Postgrad Med Inst* 2004;18:196-201.

138. Chow SM, Law SC, Au SK et al. Changes in clinical presentation, management and outcome in 1348 patients with differentiated thyroid carcinoma: experience in a single institute in Hong Kong, 1960-2000. *Clin Oncol* 2003;15:329-36.
139. Liebeskind A, Sikora AG, Komisar A et al. Rates of malignancy in incidentally discovered thyroid nodules evaluated with sonography and fine-needle aspiration. *J Ultrasound Med* 2005;24:629-34.
140. Carmeci C, Jeffrey RB, McDougall IR, Nowels KW, Weigel RJ. Ultrasound-guided fine-needle aspiration biopsy of thyroid masses. *Thyroid* 1998;8:283-9.
141. Papini E, Guglielmi R, Bianchini A et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-doppler features. *J Clin Endocrinol Metab* 2002;87:1941-6.
142. Ylagan LR, Farkas T, Dehner LP. Fine needle aspiration of the thyroid: A cytohistologic correlation and study of discrepant cases. *Thyroid* 2004;14:35-41.
143. Hamburger JI, Husain M, Nishiyama R et al. Increasing the accuracy of fine needle for thyroid nodules. *Arch Pathol Lab Med* 1989;113:1035-41.
144. Fonseca MT, Camargos PA, Ferreira CS et al. Interobserver agreement in assessing plain radiographs of maxillary sinus. *Int J Pediatr Otorhinolaryngol* 1998;45:41-6.
145. Ljung BM, Drejet A, Chiampi N. The factors important for successful fine-needle aspiration biopsies: a comparative study of breast diagnoses at two hospitals. *Pathology* 1996;4:493-8.
146. Renshaw AA, Wang E, Wilbur D et al. Interobserver agreement on microfollicles in thyroid fine-needle aspirates. *Arch Pathol Lab Med* 2006;130:148-52.
147. Cook IS, McCormick D, Poller DN. Referrals for second opinion in surgical pathology: implications for management of cancer patients in the UK. *Eur J Surg Pathol* 2001;27:589-94.
148. Kronz JD, Westra WH, Epstein JI. Mandatory second opinion surgical pathology at a large referral hospital. *Cancer* 1999;86:2426-35.
149. Westra WH, Kronz JD, Eisele DW. The impact of second opinion surgical pathology on the practice of head and neck surgery: a decade of experience at a large referral hospital. *Head Neck* 2002;24:684-93.

150. Hamady ZZ, Mather N, Lansdown MR et al. Surgical pathological second opinion in thyroid malignancy: impact on patient's management and prognosis. *Eur J Surg Oncol* 2005;31:74-7.
151. Williams ED. Two proposals regarding the terminology of thyroid tumors. *Int J Surg Pathol* 2000;8:181-3.
152. Kronz JD, Westra WH. The role of second opinion pathology in the management of lesions of the head and neck. *Curr Opin Otolaryngol Head Neck Surg* 2005;13:81-4.
153. LiVolsi VA. Can we agree to disagree? *Hum Pathol* 2003;34:1081-2.
154. Cramer SF. Interobserver variability in surgical pathology. *Adv Pathol* 1996;9:3-82.
155. Jaeschke R, Guyatt GH, Sackett DL. The evidence-based medicine working group. User's guides to the medical literature. III. How to use an article about a diagnostic test. B: what are the results and will they help me in caring for my patients? *JAMA* 1994;271:703-7.
156. Fitzmaurice G. Statistical methods for assessing agreement. *Nutrition* 2002;18:694-6.
157. Hoffman HW. Diagnostic accuracy of fine needle aspiration biopsy in the diagnosis of thyroid malignancy. *Pathologist* 1986;40:9-14.
158. Posner KL, Sampson PD, Caplan RA et al. Measuring interrater reliability among multiple raters: an example of methods for nominal data. *Statistics in Medicine* 1990;9:1103-15.
159. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360-3.
160. O'Sullivan JP, Ismail SM, Barnes WSF et al. Inter- and intra-observer variation in the reporting of cervical smears: specialist cytopathologists versus histopathologists. *Cytopathology* 1996;7:78-89.
161. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
162. Davies HTO, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ* 1998;316:989-91.
163. McHenry CR, Slusarczyk SJ, Khiyami A. Recommendations for management of cystic thyroid disease. *Surgery* 1999;126:1167-72.
164. Franc B. Observer variation of lesions of the thyroid. Letters to the editor. *Am J Surg Pathol* 2003;27:1177-8.

165. Sanders LE, Rossi R, Harboldt S et al. Handling of thyroid diseases: usefulness and limitations of rapid and fine-needle percutaneous biopsy. *Rev Med Chil* 1996;124:688-93.
166. Gallagher EJ. Clinical utility of likelihood ratios. *Ann Emerg Med* 1998;31:391-7.
167. Burch HB. Evaluation and management of the solid thyroid nodule. *Endocrinol Metab Clin North Am* 1995;24:663-710.
168. Abele JS, Wagner LT, Miller TR. Fine-needle aspiration of the breast: cell counts as an illusion of adequacy. A clinical cytopathologist's point of view. *Cancer* 1998;84:319-23.
169. Al-Kaisi N. The spectrum of the "Gray Zone" in breast cytology: a review of 186 cases of atypical and suspicious cytology. *Acta Cytol* 1994;38:898-908.
170. Zarbo RJ, Howanitz PJ, Bachner P. Interinstitutional comparison of performance in breast fine-needle aspiration cytology. A Q-probe quality indicator study. *Arch Pathol Lab Med* 1991;115:743-50.
171. Kini SR, Miller JM, Hamburger JI et al. Cytopathology of follicular lesions of the thyroid gland. *Diagn Cytopathol* 1985;1:123-32.
172. Grant CS, Hay ID, Gough IR et al. Long term follow-up of patients with benign fine-needle aspiration cytologic diagnosis. *Surgery* 1989;106:980-6.
173. Caraway NP, Sneige N, Samaan N. Diagnostic pitfalls in thyroid fine needle aspiration: a review of 394 cases. *Diagn Cytopathol* 1993;9:345-50.
174. Yang J, Schnadig V, Logrono R et al. Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer* 2007;111:306-15.
175. Dixon JM, Anderson TJ, Lamb J et al. Fine needle aspiration cytology, in relationships to clinical examination and mammography in the diagnosis of a solid breast mass. *Br J Surg* 2005;71:593-6.
176. Yang GC, LiVolsi VA, Baloch ZW. Thyroid microcarcinoma: fine-needle aspiration diagnosis and histologic follow-up. *Int J Surg Pathol* 2002;10:133-9.
177. Marqusee E, Benson CB, Frates MC et al. Usefulness of ultrasonography in the management of nodular thyroid disease. *Ann Intern Med* 2000;1339:696-700.
178. Perez LA, Gupta PK, Mandel SJ, LiVolsi VA, Baloch ZW. Thyroid papillary microcarcinoma. Is it really a pitfall of fine needle aspiration cytology? *Acta Cytol* 2001;45:341-6.

179. Löwhagen T, Granberg PO, Lundell G et al. Aspiration biopsy cytology (ABC) in nodules of the thyroid gland suspected to be malignant. *Surg Clin North Am* 1979;59:3–18.
180. Mehotra P, Viswanathan H, Johnson SJ et al. Ultrasound guidance improves the adequacy of our pre-operative thyroid cytology but not its accuracy. *Cytopathology* 2006;17:137-44.
181. Baloch MN, Aslam T, Maher M. Rationale for near total thyroidectomy in patients with multinodular goitre. *Pak J Surg* Jul 2007;23:177-9.
182. Zeppa P, Benincasa G, Lucariello A et al. Association of different pathologic processes of the thyroid gland in fine needle aspiration samples. *Acta Cytol* 2001;45:347–52.
183. Barroeta JE, Wang H, Shiina N. Is fine-needle aspiration (FNA) of multiple thyroid nodules justified? *Endocr Pathol* 2006;17:61-5.
184. Layfield LJ Reichman A, Bottlesk, Giuliano A. Clinical determinants for the management of thyroid nodules by fine needle aspiration cytology. *Arch Otolaryngol Head Neck Surgery* 1992;118:717-21.
185. Yeh MW, Demircan O, Ituarte P et al. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. *Thyroid* 2004;14:207–15.
186. Schlinkert RT, van Heerden JA, Goellner JR et al. Factors that predict malignant thyroid lesions when fine-needle aspiration is “suspicious for follicular neoplasm”. *Mayo Clin Proc* 1997;72:913-6.
187. Raphael SJ. The meanings of markers: ancillary techniques in diagnosis of thyroid neoplasia. *Endocr Pathol* 2002;13:301-11.

ANEXO I

ANEXO II

Distribuição das lesões pertencentes a cada categoria diagnóstica empregadas no estudo e seus correspondentes diagnósticos citológicos baseados no guia da Sociedade Papanicolaou de Citopatologia.¹⁷

Categorias diagnósticas	Correspondência citológica
Benigno	Bócio coloide Tireoidite de Hashimoto Tireoidite subaguda Doença de Graves Cistos*
Maligno	Carcinoma papilífero Carcinoma medular Carcinoma indiferenciado
Neoplasia folicular (ou indeterminado)	Neoplasia folicular Neoplasia de células de Hürthle Carcinoma papilar variante folicular
Inadequada para análise diagnóstica	Ausência de critérios de adequácia: <ul style="list-style-type: none"> ▪ Poucas células ▪ Material hemorrágico ▪ Fluido cístico Cistos**

* Lesões císticas foram reportadas somente pelo patologista B (não experiente), porém ele as descreveu como lesões benignas, compatíveis com bócio coloide;

** As lesões císticas identificadas pelo patologista A (experiente) foram reportadas como inadequadas para análise diagnóstica.

ANEXO III

CrITÉrios usados pelos patologistas A e B para a definiÇão dos casos:

Categorias diagnósticas	Descrição microscópica
Bócio Coloide	presença de coloide fluido (às vezes espesso) ao fundo; agrupamentos em monocamada ou em “favo de mel” de células foliculares de tamanho pequeno a intermediário, com cromatina fina, estas podendo também estar isoladas; presença de núcleos nus e macrófagos, que podem apresentar fagocitose de hemossiderina
Carcinoma Papilar	presença de células foliculares geralmente de tamanho intermediário a grande, com núcleo de aspecto em “vidro fosco” ou com clareamento, podendo haver nucléolo evidente, presença de frequentes fendas e pseudo-inclusões nucleares (principalmente mais de três inclusões); as células podem apresentar núcleos alongados e se arranjam em agrupamentos sinciciais, tridimensionais ou esboçando papila; presença de calcificações do tipo psamomatosas e células gigantes multinucleadas, o coloide geralmente é escasso e espesso e o esfregaço geralmente apresenta alta celularidade (mas este último não é critério fidedigno, pois um bócio bem puncionado ou um bócio hiperplásico também podem ser hipercelulares)
Neoplasia Folicular	células foliculares agregadas, geralmente com “monotonia” celular (mas pode haver variação de tamanho dos núcleos), frequentemente esboçando microfóliculos, podendo haver coloide espesso impactado no centro. As células mostram cromatina regularmente distribuída, podendo haver hiper cromasia nuclear; o nucléolo pode ser evidente e podem ser encontradas fendas; o coloide ao fundo é escasso e o esfregaço geralmente é celular (vale aqui a observação citada acima)
Carcinoma Medular	células geralmente de tamanho grande, citoplasma eosinofílico, com núcleos excêntricos (aspecto plasmocitoide) e nucléolo proeminente. As células são vistas isoladas ou formando agregados frouxos, pode haver células gigantes multinucleadas com núcleos hiper cromáticos; depósito de material amorfo (amiloide) ao fundo, sem material coloide. Há uma variante com células alongadas, fusiformes (os dois aspectos podem aparecer em um mesmo tumor)
Carcinoma anaplásico	fundo com material necrótico, debris granulares e células inflamatórias contendo células de tamanho grande, pleomórficas, com núcleos bizarros ou multinucleadas e nucleolos evidentes
Cisto *	Presença de sangue e coloide ao fundo contendo macrófagos, por vezes com fagocitose de hemossiderina; presença de esparsos núcleos nus (não se identificam agrupamentos de células foliculares)
Adequácia do esfregaço	Presença de, no mínimo, dez grupamentos de células, contendo, cada um, pelo menos, dez células foliculares bem preservadas

*Descrito somente pelo patologista B.

ANEXO IV

Correlação entre os valores da estatística *Kappa* e a classificação fornecida por Landis & Koch (1977) para a avaliação do grau de concordância intra e interexaminador.

Estatística <i>Kappa</i>	Grau de concordância
< 0,00	Ausência de concordância
0,00 - 0,20	Ruim
0,21 – 0,40	Fraca
0,41 – 0,60	Moderada
0,61 – 0,80	Boa / substancial
0,81 – 1,00	Muito boa / quase perfeita

ANEXO V

Histologia	DCO*	Resultados dos exames citológicos							
		Nº	A1	Nº	A2	Nº	B1	Nº	B2
BC/T. HAS	INSUF	1	INSUF	75	INSUF	1	INSUF	75	INSUF
BC	BC	2	BC	76	BC	2	BC	76	INSUF
BC	BC	3	BC	77	BC	3	BC	77	BC
BC	BC	4	BC	78	BC	4	BC	78	BC
BC	CISTO	5	BC	79	BC	5	BC	79	BC
BC	INSUF	6	INSUF	80	INSUF	6	BC	80	INSUF
BC	INSUF	7	INSUF	81	INSUF	7	INSUF	81	INSUF
GRAVES	BC	8	BC	82	INSUF	8	BC	82	BC
AF	BC	9	BC	83	BC	9	BC	83	BC
BC	CISTO	10	BC	84	BC	10	BC	84	BC
CP	INSUF	11	INSUF	85	INSUF	11	CM	85	CM
BC	BC	12	INSUF	86	INSUF	12	BC	86	BC
BC	BC	13	INSUF	87	INSUF	13	BC	87	CISTO
BC	BC	14	CI	88	CI	14	CM (CI?)	88	CM (CI?)
BC	CISTO	15	INSUF	89	INSUF	15	BC	89	CISTO
BC	CISTO	16	INSUF	90	INSUF	16	BC	90	INSUF
CP	BC	17	NF	91	NF	17	BC	91	NF
BC	BC	18	BC	92	BC	18	BC	92	BC
BC	BC	19	BC	93	BC	19	BC	93	BC
BC	CISTO	20	BC	94	INSUF	20	BC	94	BC
BC	CISTO	21	INSUF	95	INSUF	21	BC	95	BC
CP	CP	22	CP	96	CP	22	BC	96	NF
AF	INSUF	23	BC	97	BC	23	BC	97	BC
BC	BC	24	BC	98	BC	24	BC	98	NF
CM	INSUF	25	CM	99	CM	25	CM	99	CM
CM	INSUF	26	CM	100	CM	26	CM	100	CM
BC	BC	27	BC	101	BC	27	BC	101	BC
BC	NF	28	INSUF	102	INSUF	28	NF	102	NF
CP	CP	29	CP	103	CP	29	CP	103	CP
CP	NF	30	INSUF	104	INSUF	30	INSUF	104	INSUF
CP	BC	31	BC	105	BC	31	BC	105	BC
BC	BC	32	BC	106	BC	32	BC	106	BC
T. HAS	BC	33	T. HAS	107	T. HAS	33	T. HAS	107	T. HAS
T. HAS	BC	34	BC	108	T. HAS	34	BC	108	BC
CP	BC	35	BC	109	BC	35	BC	109	BC
BC	BC	36	BC	110	BC	36	BC	110	BC
AF	BC	37	BC	111	BC	37	BC	111	BC
CP	CP	38	CP	112	CP	38	CP	112	CP
BC	BC	39	BC	113	BC	39	BC	113	BC
BC	BC	40	BC	114	BC	40	BC	114	BC
BC	INSUF	41	INSUF	115	BC	41	BC	115	BC
AF	BC	42	INSUF	116	INSUF	42	BC	116	CISTO
AF	BC	43	INSUF	117	INSUF	43	BC	117	CISTO
AF	BC	44	BC	118	BC	44	BC	118	BC
T. HAS	T. HAS	45	T. HAS	119	T. HAS	45	T. HAS	119	T. HAS
BC	BC	46	BC	120	BC	46	BC	120	BC

BC	INSUF	47	INSUF	121	INSUF	47	INSUF	121	INSUF
BC	INSUF	48	INSUF	122	INSUF	48	BC	122	BC
BC	INSUF	49	INSUF	123	INSUF	49	INSUF	123	INSUF
AF	BC	50	BC	124	BC	50	BC	124	BC
BC	INSUF	51	INSUF	125	INSUF	51	INSUF	125	INSUF
T. HAS	BC	52	BC	126	BC	52	BC	126	T. HAS/BC
BC	BC	53	BC	127	BC	53	BC	127	BC
BC	BC	54	BC	128	BC	54	BC	128	BC
BC	BC	55	BC	129	BC	55	BC	129	BC
CP	INSUF	56	BC	130	BC	56	NF	130	NF
CF	BC	57	INSUF	131	BC	57	BC	131	BC
CP	BC	58	INSUF	132	BC	58	CP	132	NF
BC	BC	59	BC	133	INSUF	59	BC	133	BC
BC	INSUF	60	INSUF	134	INSUF	60	INSUF	134	INSUF
BC	BC	61	BC	135	BC	61	BC	135	BC
CP	CP	62	CP	136	CP	62	CP	136	CP
CP	BC	63	CP	137	CP	63	CP	137	CP
CP	NF	64	CP	138	CP	64	CP	138	NF
BC	BC	65	BC	139	BC	65	BC	139	BC
BC	INSUF	66	INSUF	140	INSUF	66	INSUF	140	INSUF
BC	INSUF	67	INSUF	141	INSUF	67	BC	141	BC
BC	BC	68	BC	142	BC	68	BC	142	BC
CF	NF	69	NF	143	NF	69	NF	143	NF
BC/T. HAS	INSUF	70	BC	144	BC	70	BC	144	BC
BC	INSUF	71	INSUF	145	INSUF	71	INSUF	145	INSUF
BC	BC	72	BC	146	BC	72	BC	146	BC
BC	BC	73	INSUF	147	BC	73	NF	147	NF
AF/BC	BC	74	INSUF	148	BC	74	BC	148	NF

* DOC: diagnóstico citológico original

BC: bócio colóide

INSUF: insuficiente

AF: adenoma folicular

CA: carcinoma anaplásico

CI: carcinoma indiferenciado

CP: carcinoma papilífero

CF: carcinoma folicular

CM: carcinoma medular

NF: neoplasia folicular

ANEXO VI

DEM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	8	1	9
	-	5	34	39
	Total	13	35	48

DAM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	10	1	11
	-	3	34	37
	Total	13	35	48

Resultados fornecidos por A1 de acordo com as definições estreita (DEM) e ampla (DAM) de malignidade.

DEM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	8	1	9
	-	7	34	41
	Total	15	35	50

DAM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	10	1	11
	-	5	34	39
	Total	15	35	50

Resultados fornecidos por A2 de acordo com as definições estreita (DEM) e ampla (DAM) de malignidade.

DEM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	9	1	10
	-	7	48	55
	Total	16	49	65

DAM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	11	3	14
	-	5	46	51
	Total	16	49	65

Resultados fornecidos por B1 de acordo com as definições estreita (DEM) e ampla (DAM) de malignidade.

DEM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	7	1	8
	-	9	45	54
	Total	16	46	62

DAM	Câncer (diagnóstico histológico)			
PAAF		+	-	Total
	+	13	5	18
	-	3	41	44
	Total	16	46	62

Resultados fornecidos por B2 de acordo com as definições estreita (DEM) e ampla (DAM) de malignidade.