

PRÁTICAS CIRÚRGICAS BASEADAS EM EVIDÊNCIAS: TOMOGRAFIA COMPUTADORIZADA HELICOIDAL NO DIAGNÓSTICO DA APENDICITE AGUDA

EVIDENCE-BASED SURGICAL PRACTICES: HELICAL COMPUTED TOMOGRAPHY IN ACUTE APPENDICITIS DIAGNOSIS

Carlos Alberto Guimarães, TCBC-RJ¹; Bráulio dos Santos²;
Grupo de Estudos de Revisão Sistemática do Rio de Janeiro

RESUMO: Os cirurgiões são instados a não somente ler os artigos dos periódicos, mas também a compreendê-los e analisá-los criticamente quanto à validade. Eles cuidam melhor dos seus pacientes se são capazes de analisar criticamente a literatura e aplicar os resultados a sua prática. Este é o segundo artigo de uma série - Práticas Cirúrgicas Baseadas em Evidência – que tem por objetivo apresentar a avaliação crítica de um teste diagnóstico (*Rev. Col. Bras. Cir.* 2008; 35(4): 269-273).

Descritores: Medicina baseada em evidências; Educação médica continuada; Diagnóstico; Estudos transversais; Apendicectomia; Apendicite; Ultra-sonografia; Tomografia; Tomografia Computadorizada por Raios X.

INTRODUÇÃO

A elaboração de um diagnóstico é um exercício cognitivo que envolve o raciocínio lógico e o reconhecimento de padrões. Inicialmente se enumeram as possibilidades diagnósticas e estimam-se as suas probabilidades relativas. No segundo passo do processo diagnóstico, incorporam-se novas informações para modificar as probabilidades relativas, para afastar algumas possibilidades e, finalmente, para escolher o diagnóstico mais provável. Assim, passa-se de uma probabilidade – a probabilidade pré-teste – para outra probabilidade – a probabilidade pós-teste.

No processo do diagnóstico diferencial, se considerarmos todas as hipóteses como igualmente prováveis e solicitarmos testes diagnósticos simultâneos, certamente o paciente será submetido a exames desnecessários. Por outro lado, o cirurgião experiente vai considerar primeiro aquelas doenças mais prováveis ou as que são potencialmente mais graves ou as que são mais responsivas ao tratamento.

O objetivo deste segundo artigo da série Práticas Cirúrgicas Baseadas em Evidências é prover os cirurgiões com um guia para a análise da validade dos estudos que exploram as propriedades de um teste diagnóstico.

CENÁRIO CLÍNICO

Um jovem de 18 anos foi admitido no Serviço de Emergência, queixando-se de dor na projeção do hipogástrico que já durava três dias. O exame físico mostrou dor moderada à descompressão súbita no quadrante inferior direito do ab-

dome. A temperatura axilar era de 38,8° C e o exame de sangue revelou leucocitose e aumento da proteína C-reativa.

O cirurgião-chefe solicitou ao residente que providenciasse uma ultra-sonografia (US) com compressão gradual, mas esse último questionou se a tomografia computadorizada helicoidal não teria uma melhor acurácia diagnóstica. O cirurgião-chefe respondeu que desconhecia as evidências científicas da superioridade da tomografia computadorizada (TC) no diagnóstico da apendicite e deu ao residente a tarefa de buscar a literatura para que voltassem a discutir o assunto no dia seguinte.

BUSCA BIBLIOGRÁFICA

O residente, que já havia realizado um breve treinamento sobre busca em bases bibliográficas, resolveu acessar a base MEDLINE pela interface PubMed (<http://www.pubmed.gov>). Na base MeSH, ele selecionou os descritores relevantes e acabou elaborando a seguinte estratégia: (“Tomography, Spiral Computed”[Mesh] AND “Appendicitis”[Mesh]). Com essa estratégia, ele chegou a 41 artigos e selecionou, pelo título, um que comparava a tomografia computadorizada com a ultra-sonografia, era de acesso livre e podia ser obtido no endereço <http://www.sma.org.sg/smj/4801/4801a12.pdf>.

AValiação Crítica do Artigo ENCONTRADO

A avaliação crítica do artigo foi baseada na leitura do texto “How to Use an Article About a Diagnostic Test”².

1. Professor Titular de Clínica Cirúrgica do Curso de Graduação em Medicina da Fundação Educacional Serra dos Órgãos (Teresópolis, RJ); Professor Adjunto do Departamento de Cirurgia – UFRJ.

2. Médico da Coordenação de Ensino e Pesquisa (Unidade de Epidemiologia Clínica) do Instituto Nacional de Cardiologia, Ministério da Saúde.

Recebido em: 15/01/2008

Aceito para publicação em: 20/02/2008

Quando se decide que um artigo sobre um teste diagnóstico é particularmente relevante (título e resumo sugerem que a informação seja importante para o problema do nosso paciente), podemos fazer as mesmas três perguntas mencionadas no primeiro artigo desta série³.

Essas três questões serão respondidas, utilizando o artigo de Gamanagatti *et al.* que comparou a acurácia da tomografia computadorizada helicoidal (TCH) e da ultrasonografia no diagnóstico da apendicite aguda.

OS RESULTADOS DO ESTUDO SÃO VÁLIDOS?

Foi realizada uma comparação independente (cega) do teste diagnóstico com o padrão-ouro?

A avaliação do desempenho de um teste requer a sua comparação com algum procedimento que assegure o diagnóstico de interesse. A esse procedimento, seja ele um exame laboratorial, patológico, funcional ou de imagem, achado cirúrgico ou mesmo seguimento clínico, chamamos padrão de referência ou padrão-ouro. Embora o padrão-ouro quase sempre seja imperfeito, ele deve ser uma das formas consensualmente aceitas pela comunidade científica para certificar o diagnóstico em questão. Sempre que possível, o teste estudado e o padrão-ouro devem ser interpretados de forma independente, ou seja, o profissional que avalia o teste não deve saber do resultado do padrão-ouro e vice-versa. No estudo em questão, os dois testes foram realizados pelo mesmo radiologista, mas foram depois avaliados independentemente por dois outros profissionais mais experientes, e apenas essa última avaliação foi considerada na análise. No entanto, o padrão-ouro para o diagnóstico de apendicite foi o achado operatório e o cirurgião conhecia os resultados dos exames de imagem. O mascaramento dos resultados dos exames resultaria numa avaliação mais correta de seu desempenho diagnóstico, mas seria eticamente inaceitável, já que no melhor interesse dos pacientes, os cirurgiões devem ter acesso a todas as informações pré-operatórias disponíveis.

A amostra dos doentes utilizada no teste diagnóstico incluiu o espectro encontrado na prática clínica?

A amostra ideal de um estudo de acurácia diagnóstica deve ser representativa da população submetida ao teste na prática clínica. Amostras que contenham apenas pacientes com probabilidade pré-teste muito alta ou muito baixa de doença, só são aceitáveis em situações específicas (por exemplo, testes de rastreamento freqüentemente são aplicados a populações de baixo risco de doença; exames invasivos em geral só são realizadas em populações de risco alto). Gamanagatti *et al.* incluíram pacientes consecutivos que chegaram ao hospital com dor abdominal aguda e achados clínicos e laboratoriais sugestivos de apendicite. Portanto, a amostra foi representativa da população à qual, na prática, os testes serão aplicados.

O resultado do teste que está sendo avaliado influenciou na decisão de realizar o teste padrão-ouro?

Para uma avaliação adequada do desempenho diagnóstico de um teste, todos os pacientes incluídos devem submeter-se ao teste e ao padrão-ouro. Quando o padrão-ouro é um procedimento invasivo, a decisão de realizá-lo é freqüente

e compreensivelmente influenciada pelo resultado do teste sob investigação. Os autores relatam que o cirurgião sabia dos resultados dos exames de imagem, mas essa informação não foi utilizada para a decisão de operar os pacientes. Como o artigo não descreve como isso foi alcançado (uma possibilidade seria o cirurgião só ter acesso aos resultados da US e da TC após decisão de operar ou não o paciente), é legítimo que haja dúvida quanto à correção do procedimento.

Seis dos 58 pacientes examinados com a US e a TC não foram submetidos à intervenção cirúrgica; a melhor solução seria incluir todos os pacientes na análise estatística e utilizar as conclusões extraídas do seguimento clínico como substitutas do padrão-ouro. No entanto, os autores optaram por excluí-los, o que pode ter viciado as suas estimativas.

Foi realizada uma descrição detalhada do teste diagnóstico para permitir sua reprodução?

A maior parte dos testes diagnósticos é sujeita a variações técnicas que influenciam o seu desempenho diagnóstico. Portanto, os estudos de acurácia devem incluir descrições operacionais suficientemente detalhadas para que o leitor possa julgar se os testes foram realizados com técnica adequada e para que ele possa reproduzi-los em sua prática. Na nossa opinião, esse item foi abordado corretamente no artigo analisado.

QUAIS SÃO OS RESULTADOS DO ESTUDO?

As razões de verossimilhança para os resultados do teste são apresentadas ou os dados estão disponíveis para calculá-las?

A utilidade de um teste diagnóstico se baseia na sua capacidade de modificar a estimativa prévia de probabilidade de doença. Essa probabilidade prévia, ou probabilidade pré-teste, pode ser obtida de várias maneiras: pela prevalência populacional da doença, por meio de regras de predição clínica ou simplesmente pela tradução numérica do grau de suspeição do diagnóstico pelo médico.

A base para a interpretação de testes diagnósticos é a tabela de contingência teste x doença (tabela 1). Das quatro casas da tabela, duas correspondem a acertos do teste (teste positivo e doença, teste negativo e ausência da doença) e duas a erros (teste positivo e ausência da doença, teste nega-

Tabela 1 - Contingência teste x doença.

	D +	D -
T +	VP	FP
T -	FN	VN

D + Doente

D - Não doente

T + Teste positivo

T - Teste negativo

VP Verdadeiro positivo

VN Verdadeiro negativo

FP Falso positivo

FN Falso negativo

tivo e doença). A essas casas damos, respectivamente, o nome de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN). Há várias formas de expressarmos as propriedades de um teste. Uma delas é a acurácia, ou seja a proporção de acertos:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}$$

A acurácia, isoladamente, é insuficiente para a estimativa das probabilidades pós-teste.

A sensibilidade e a especificidade são propriedades mais úteis. A sensibilidade (S) é a proporção dos doentes com testes positivos:

$$S = \frac{VP}{VP + FN}$$

e a especificidade (E) a proporção dos não doentes com testes negativos:

$$E = \frac{VN}{VN + FP}$$

A sensibilidade e a especificidade, em conjunto, permitem calcular probabilidades pós-teste a partir de probabilidades pré-teste, mas as expressões mais simples da capacidade de o teste modificar a plausibilidade de doença são as razões de verossimilhança. A razão de verossimilhança (RV) do resultado de um teste é a razão entre a proporção de doentes e a proporção de não doentes que apresentam esse resultado, ou seja:

$$RV+ = \frac{P(T+|D+)}{P(T+|D-)} = \frac{\frac{VP}{VP + FN}}{\frac{FP}{VN + FP}} = \frac{S}{1 - E}$$

$$RV- = \frac{P(T-|D+)}{P(T-|D-)} = \frac{\frac{FN}{VP + FN}}{\frac{VN}{VN + FP}} = \frac{1 - S}{E}$$

Antes de prosseguirmos, é necessário introduzir um conceito auxiliar ao de probabilidade, a chance. Vamos supor um baralho de 52 cartas e que dele retiremos uma aleatoriamente. A **probabilidade** (P) de que essa carta seja de copas será o número de copas dividido pelo número total de cartas do baralho: 13/52, ou seja, 1/4. A **chance** (C) de que a carta seja de copas é o número de copas dividido pelo número de cartas que não são copas: 13/39, ou seja, 1/3. A probabilidade é, portanto, o número de resultados favoráveis dividido pelo total

de resultados possíveis, enquanto a chance é o número de resultados favoráveis dividido pelo seu complemento (número de resultados desfavoráveis). A probabilidade e a chance podem ser facilmente convertidas entre si:

$$C = \frac{P}{1 - P} \quad e \quad P = \frac{C}{1 + C}$$

A razão de verossimilhança positiva é a razão entre as chances de doença pós e pré-teste quando o teste é positivo e a razão de verossimilhança negativa a razão entre as chances pós e pré-teste quando o teste é negativo. Dito de outra forma:

- quando o teste é positivo, $C_{pós} = C_{pré} \times RV+$
- quando o teste é negativo, $C_{pós} = C_{pré} \times RV-$

Quanto maior for a razão de verossimilhança positiva, maior a sua utilidade (maior aumento da chance pós-teste). A razão de verossimilhança negativa será tanto mais útil quanto mais próxima de zero (maior redução da chance pós-teste). Uma razão de verossimilhança igual a um significa que o resultado do exame não modifica a estimativa pré-teste. Esses cálculos podem ser realizados manualmente ou por meio de programas de computador, mas uma forma prática de estimar a probabilidade pós-teste a partir da probabilidade pré-teste e da razão de verossimilhança é o nomograma de Fagan⁴, que pode ser encontrado facilmente na internet (por exemplo em http://www.cche.net/usersguides/figures/diagnosis_f1.jpg).

Um princípio básico na análise de estudos de acurácia é **sempre construir as tabelas 2x2** a partir da descrição dos resultados e conferir se as estimativas do desempenho diagnóstico coincidem com as do texto.

Dessa forma, construímos as tabelas para a US (tabela 2) e TC (tabela 3). Os intervalos de confiança da sensibilidade e especificidade foram calculados pelo método de Wilson⁵

Tabela 2 - Contingência para a ultra-sonografia com compressão gradual.

	D+	D-	Total
T+	32	1	33
T-	16	3	19
Total	48	4	52

	Valor	IC 95%
S	66,67	(52,54 a 78,32)
E	75,00	(30,06 a 95,44)
RV+	2,67	(0,91 a 14,73)
RV-	0,44	(0,26 a 1,18)

S Sensibilidade

E Especificidade

RV+ Razão de verossimilhança positiva

RV- Razão de verossimilhança negativa

e os das razões de verossimilhança pelo método de escore^{6,7}, usando o programa CIA versão 2.0.0 (Trevor Bryant. Universidade de Southampton, 2000). O conceito de intervalo de confiança já foi discutido no primeiro artigo da série³.

O nosso cálculo não coincide com a sensibilidade e especificidade da US descritas pelos autores, que correspondem à seguinte tabela (tabela 4).

Ou seja, na Tabela da TC há quatro pacientes com apendicite, mas na da US apenas três! Provavelmente, a diferença se deve ao paciente com apêndice macroscopicamente normal à cirurgia, que ainda assim foi retirado devido à forte suspeita clínica de apendicite. Esse paciente deve ter sido considerado doente na análise da US e não na da TC devido a uma distração dos autores (e omissão do revisor do periódico, que certamente não conferiu os cálculos). Como o padrão-ouro estabelecido pelos próprios autores é o aspecto macroscópico do apêndice e não a decisão quanto à sua retirada, consideraremos corretos os cálculos do Quadro 2, e não os do quatro.

A TC apresentou razões de verossimilhança positiva e negativa melhores que as da US. Os intervalos de confiança de 95% das razões de verossimilhança da US passam pelo valor nulo (um), ou seja, com o tamanho amostral do estudo

Tabela 3 - Contingência para a tomografia computadorizada helicoidal.

	D+	D-	Total
T+	46	1	47
T-	2	3	5
Total	48	4	52

	Valor	IC 95%
S	95,83	(86,02 a 98,85)
E	75,00	(30,06 a 95,44)
RV+	3,83	(1,36 a 21,04)
RV-	0,06	(0,01 a 0,24)

Tabela 4 - Contingência para a ultra-sonografia com compressão gradual (sic).

	D+	D-	Total
T+	33	0	33
T-	16	3	19
Total	49	3	52

	Valor	IC 95%
S	67,35	(53,38 a 78,79)
E	100,00	(43,85 a 100,00)
RV+	∞*	(1,17 a ∞*)
RV-	0,33	(0,37 a ∞*)

* ∞ significa que a estimativa não é calculável (divisão por zero).

nem mesmo temos um nível razoável de confiança de que a US contribua para o diagnóstico de apendicite.

OS RESULTADOS IRÃO AJUDAR NO CUIDADO DOS MEUS DOENTES?

Os resultados dos testes são reprodutíveis e a interpretação é possível no local onde trabalho?

O valor de qualquer teste diagnóstico depende da sua capacidade de dar o mesmo resultado, quando aplicado novamente no mesmo paciente. Uma baixa reprodutibilidade pode ser devida a variações biológicas dos pacientes, variações dos equipamentos ou variações dos indivíduos que realizam e interpretam os exames.

Um artigo sobre um teste diagnóstico deve mencionar aquilo que se espera da sua reprodutibilidade. Isso é particularmente importante quando uma expertise é necessária para realizar ou interpretar o teste.

A ultra-sonografia é simples, barata, não expõe o paciente à radiação, porém é muito dependente do operador, especialmente quando feita com a técnica da compressão gradual. Ademais, o excesso de gases intestinais dificulta muito a interpretação. Por outro lado, a tomografia computadorizada helicoidal é supostamente operador-independente, não requer preparo do paciente nem administração de contraste, mas expõe o paciente à radiação ionizante.

No artigo de Gamanagatti *et al.*, não há menção à reprodutibilidade dos testes. Tanto as ultra-sonografias com a técnica da compressão gradual como as tomografias computadorizadas helicoidais foram realizadas por médicos residentes, mas na análise estatística só foram consideradas as avaliações dos dois radiologistas seniores.

Os resultados são aplicáveis aos meus doentes?

A questão é saber se o teste terá a mesma acurácia entre os meus pacientes e os mencionados no artigo. S e as condições da minha prática clínica são semelhantes às do estudo, e se os meus pacientes preenchem os critérios de inclusão e de exclusão, posso imaginar que os resultados sejam diretamente aplicáveis.

Os pacientes incluídos no artigo de Gamanagatti *et al.* constituem uma amostra predominantemente de adultos (90%) que em um Serviço de Emergência apresentaram quadro clínico-laboratorial sugestivo de apendicite aguda. Como o paciente do cenário clínico preenche os critérios de elegibilidade do estudo, posso aplicar os resultados desse teste diagnóstico.

Os resultados poderão mudar a minha conduta?

A TC apresentou razão de verossimilhança positiva de 3,83 e negativa de 0,06. Como o teste positivo aumenta a chance de doença em cerca de quatro vezes e o teste negativo reduz a chance de doença em cerca de 17 vezes, seria razoável pensar que a TC é muito mais útil para excluir do que para confirmar o diagnóstico de apendicite, mas a verdade é um pouco mais complexa.

Sugerimos que o leitor imprima o nomograma de Fagan (ou faça as contas, se tiver inclinação matemática) e confira os

resultados seguintes, usando as razões de verossimilhança da TC helicoidal no estudo.

1. Paciente com probabilidade pré-teste de 5%:
 - probabilidade de apendicite após TC positiva: 16,8%;
 - probabilidade de apendicite após TC negativa: 0,3%.
2. Paciente com probabilidade pré-teste de 50%:
 - probabilidade de apendicite após TC positiva: 79,3%;
 - probabilidade de apendicite após TC negativa: 5,7%.
3. Paciente com probabilidade pré-teste de 90%:
 - probabilidade de apendicite após TC positiva: 97,2%;
 - probabilidade de apendicite após TC negativa: 35,1%.

A utilidade do exame depende não apenas das razões de verossimilhança, mas também da probabilidade pré-teste, do limiar de diagnóstico (probabilidade abaixo da qual o médico afasta o diagnóstico) e do limiar de tratamento (probabilidade acima da qual o médico considera o diagnóstico confirmado e inicia um procedimento terapêutico). No nosso primeiro exemplo (probabilidade pré-teste de 5%), o exame negativo traria pouca informação adicional e o exame positivo resultaria numa probabilidade pós-teste intermediária, na qual não nos decidiríamos pela cirurgia imediata nem afastaríamos o diagnóstico. No último exemplo (probabilidade pré-teste de 90%), o exame positivo não modificaria a conduta e o negativo colo-

caria novamente o paciente numa probabilidade intermediária. O segundo exemplo (probabilidade de 50%) é aquele em que o teste se mostraria mais útil, já que um exame positivo seria suficiente para apoiar a decisão pela cirurgia e um exame negativo seria suficiente para optarmos por outras alternativas diagnósticas.

Os doentes se beneficiarão com os resultados dos testes?

O último critério para a utilidade de um teste diagnóstico é o julgamento se a informação adicional que ele fornece resulta, finalmente, em algum benefício para o paciente.

Em algumas situações clínicas, o teste diagnóstico pode ter boa acurácia, mas a sua utilidade será limitada, se a modificação da conduta não tiver impacto sobre os desfechos clínicos importantes. Por outro lado, não há dúvida do valor de um teste quando a doença, se não diagnosticada, tem prognóstico ruim, quando o teste tem riscos aceitáveis e quando existe um tratamento efetivo, e esse é o caso no estudo em questão.

RESOLUÇÃO DO CENÁRIO CLÍNICO

Apesar de alguns problemas metodológicos e do pequeno tamanho amostral, o cirurgião-chefe e o residente concordaram que o artigo apoiava a superioridade da TC helicoidal sobre a US com compressão gradual e decidiram, até a avaliação de novas evidências, usar a TCH, como exame de imagem de escolha, nos pacientes com dor abdominal aguda e suspeita de apendicite.

ABSTRACT

Surgeons are told that they need not only read journal articles, but also understand them and make a critical appraisal of their validity. They offer better care if they are able to appraise critically the original literature and apply the results to their practice. This is the second article of a series – Evidence-Based Surgical Practices – which focus on critical appraisal of a diagnostic test.

Key words: *Evidence-based medicine; Education, medical, continuing; Diagnosis; Cross-sectional studies; Appendectomy; Appendicitis; Ultrasonography; Tomography; Tomography, X-Ray Computed.*

REFERÊNCIAS

1. Gamanagatti S, Vashisht S, Kapoor A, Chumber S, Bal S. Comparison of graded compression ultrasonography and unenhanced spiral computed tomography in the diagnosis of acute appendicitis. *Singapore Med J.* 2007;48(1):80-7.
2. Centre for Health Evidence [homepage on the Internet]. Edmonton: Centre for Health Evidence; c2001. Jaeschke R, Guyatt GH, Sackett DL; Evidence Based Medicine Working Group. How to Use an article about a diagnostic test. [update 2007 Aug 15; cited 2008 apr 13]; [about 16 screens]. Available from: <http://www.cche.net/text/usersguides/diagnosis.asp>
3. Santos B, Guimarães CA; Grupo de Estudos de Revisão Sistemática do Rio de Janeiro. Práticas cirúrgicas baseadas em evidências: apendicectomia laparoscópica versus a céu aberto. *Rev Col Bras Cir.* 2008;35(1):56-60.
4. Fagan TJ. Nomogram for Bayes theorem [letter]. *N Engl J Med.* 1975;293(3):257.
5. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc.* 1927;22:209-12.
6. Gart JJ, Nam JM. Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics.* 1988;44(2):323-38.
7. Nam JM. Confidence limits for the ratio of two binomial proportions based on likelihood scores: non-iterative method. *Biom J.* 1995;37(3):375-9.

Como citar este artigo:

Guimarães CA, Santos B; Grupo de Estudos de Revisão Sistemática do Rio de Janeiro. Práticas cirúrgicas baseadas em evidências: tomografia computadorizada helicoidal no diagnóstico da apendicite aguda. *Rev Col Bras Cir.* [periódico na Internet] 2008; 35(4). Disponível em URL: <http://www.scielo.br/rcbc>

Endereço eletrônico: GERS-Rio@yahoo.com