

# Testes diagnósticos 3: Como utilizá-los na prática clínica

## *Diagnostic tests 3: How to use them in medical practice*

Denise Duprat Neves<sup>1</sup>, Ricardo Marques Dias<sup>2</sup>,  
Antônio José Ledo Alves da Cunha<sup>3</sup>

**Descritores:** Teste diagnóstico, razões de verossimilhança.

**Key words:** Diagnostic test, likelihood ratio.

### Introdução

Se o estudo é válido e o teste apresenta um bom rendimento, passamos a avaliar: 1) se os resultados podem ser generalizados, ou mais especificamente, se o teste terá o mesmo rendimento na nossa prática, sendo aplicável aos nossos pacientes; 2) o quão freqüente trará informações importantes, ou se estas serão superiores àquelas fornecidas pela história e exame físico ou pelos exames de rotina; 3) se o teste será de mais fácil execução ou terá menor custo ou será menos invasivo que os já em uso; 4) se irá alterar a maneira de conduzir estes casos, beneficiando os pacientes, principalmente quando identifica uma doença para a qual exista um tratamento<sup>1, 2</sup>. Testes eficazes não garantem o melhor desfecho de um caso clínico, mas certamente são definidores da melhor conduta a ser adotada, o que trará benefícios para o paciente e economia para os serviços de saúde.

Poucos testes são capazes de discriminar completamente os doentes dos não doentes, não existindo teste 100% simultaneamente sensível e específico. Devido à complexidade e diversidade clínica e laboratorial, a correta identificação de uma determinada doença dificilmente dependerá de uma única característica. O diagnóstico é, portanto, um procedimento complexo de

associação de idéias, incluindo dados positivos e negativos, e o conhecimento e a experiência do profissional. Em determinadas enfermidades, apesar de alguns exames complementares serem úteis, estas podem, e devem, ser tratadas apenas com base na história e exame físico – como é o caso, por exemplo, da hipertensão arterial e da asma. Para outras, é necessário seguir uma rotina de testes – como é o caso da investigação de um nódulo pulmonar solitário e das doenças intersticiais difusas. Os testes que precedem o procedimento mais invasivo podem não ser definitivos, mas auxiliam no encaminhamento do diagnóstico.

Como não existe teste diagnóstico perfeito, é bem provável que a probabilidade de doença esteja entre 10 a 90% após o resultado de um único teste. Muitas vezes, necessitamos de maior certeza para afastar ou definir um diagnóstico. Quando a doença é grave ou o tratamento é agressivo, tóxico ou mutilante, precisamos aumentar a chance de certeza de 90% para 99,99%, ou diminuí-la de 10% para 0,1%. Nestes casos, os passos finais são tão importantes como os intermediários, ou seja, pensamos em escala logarítmica onde pequenas diferenças numéricas próximas de 0 e 100% podem ser tão importantes como grandes diferenças numa faixa intermediária.

<sup>1</sup> Profª. Adjunta da UNIRIO, <sup>2</sup> Prof. Titular da UNI-RIO, <sup>3</sup> Prof. Adjunto da UFRJ.

**Endereço para correspondência:** Denise Duprat Neves. Rua Mariz e Barros, 775, Hospital Universitário Gaffrée e Guinle, DEMESP, Pneumologia. Tijuca, Rio de Janeiro. Brasil. CEP 20270-004. E-mail: dduprat@unirio.br Tel: 55 021 2569 7610 – ramal 304  
Artigo recebido para publicação no dia 07/11/2003 e aceito no dia 11/12/2003, após revisão.

São poucos os médicos que compreendem com clareza as propriedades dos testes diagnósticos<sup>3</sup>. Menos ainda são aqueles que têm capacidade de avaliar a combinação de testes. É importante, portanto, que os pesquisadores levem seu enfoque além da acurácia do teste, estabelecendo uma estratégia prática para seu uso.

### **Testes múltiplos, em paralelo ou em série**

Quando realizamos diversos testes e todos são positivos ou negativos, a interpretação dos resultados é óbvia; contudo, nem sempre é assim que ocorre na prática. Na realização de testes múltiplos, podemos avaliar os resultados de duas formas: em paralelo, simultaneamente, ou seja, todos os resultados ao mesmo tempo, ou em série, seqüencialmente, um resultado após o outro. É importante saber para que se quer e o que podemos esperar de um teste, ou conjuntos destes, para a escolha da estratégia ideal de investigação diagnóstica.

Testes em paralelo são solicitados quando necessitamos de uma abordagem rápida: ambulatorialmente, com testes de baixo custo, em pacientes graves como os internados em unidades de tratamento intensivo, ou naqueles que são realizados em material obtido em um procedimento invasivo. Também são úteis em casos para os quais dispomos de dois ou mais testes pouco sensíveis para o diagnóstico. Por exemplo, se temos dois testes totalmente independentes, com sensibilidade de 60% e 80%, e estes são utilizados em paralelo, a sensibilidade sobe para 92%<sup>4, 5</sup>.

Os testes em série são mais indicados quando alguns deles são de custo elevado ou possuem risco, só sendo indicados depois de outros terem sugerido a presença da doença. Esta forma maximiza a especificidade e valor preditivo positivo, em detrimento da sensibilidade e valor preditivo negativo, assegurando que um resultado positivo represente a presença de doença<sup>4, 5, 6</sup>. Se a estratégia for utilizar testes em série, o mais eficiente seria utilizar primeiramente o de maior especificidade, pois teremos o mesmo número de acertos realizando um menor número de exames. Esta conduta, além de ter menor custo, diminuiria o número de fatores a serem considerados pelo médico em sua análise de decisão, diminuindo a possibilidade de falsos resultados<sup>6</sup>.

Em resumo, podemos dizer que quando utilizamos o critério de um ou outro teste positivo aumentamos a chance de acertar e de errar, aumentando a sensibilidade à custa de diminuição da especificidade. Quando estabelecemos o critério de ambos os testes positivos, para confirmar o diagnóstico temos uma chance maior de o caso ser realmente doente. Assim, ganhamos em especificidade com perda da sensibi-

lidade, a não ser que ambos os testes sejam muito sensíveis<sup>4, 5</sup>. A estratégia adotada deve considerar o mínimo de risco e de desconforto para o paciente, o menor custo da realização dos testes e a maximização da capacidade diagnóstica<sup>7</sup>.

### **A utilização da razão de verossimilhança**

O ponto de partida para qualquer diagnóstico é a sua suspeita, baseada na história clínica, na prevalência da doença na região ou em outros exames. É a partir destas informações que criamos as hipóteses diagnósticas, que serão confirmadas ou não pelos exames complementares. A suspeita da presença de uma determinada doença é definida como probabilidade pré-teste. Esta probabilidade irá se modificar em razão do resultado de testes diagnósticos indicados. O teorema de Bayes é uma fórmula que mostra como uma crença existente (a hipótese diagnóstica) é modificada em função de novas informações (testes diagnósticos) para o diagnóstico de certeza<sup>8, 9, 10</sup>.

A dificuldade em estimar a probabilidade pré-teste, ou da hipótese diagnóstica, tem sido um limitador na utilização desta metodologia na prática clínica. Ainda não existem dados disponíveis na literatura sobre a propriedade de cada informação clínica que nos permita estimar com segurança uma probabilidade pré-teste, mas isto já vem sendo pesquisado<sup>11</sup>. Quando desconhecemos estas propriedades podemos lançar mão de um consenso de opiniões de especialistas ou utilizar a prevalência da doença na população<sup>1, 2</sup>.

As propriedades do teste diagnóstico, sensibilidade e especificidade exercem uma grande influência na probabilidade pós-teste, determinando a direção e a magnitude desta mudança. As razões de verossimilhança podem ser utilizadas, de modo fácil, neste processo de análise Bayesiana. A maior desvantagem da razão de verossimilhança (LR) é ser expressa como chance, dificultando sua interpretação, uma vez que estamos acostumados a lidar com probabilidades. A chance pós-teste de uma doença estar presente é calculada multiplicando-se a chance pré-teste pelo valor da LR<sup>4, 10, 12</sup>, ou seja,

Chance pré-testes = prevalência / 1 - prevalência

Chance pós-testes = chance pré-teste \* LR<sub>1</sub> \* LR<sub>2</sub> \*...\* LR<sub>n</sub>

Probabilidade pós-testes = chance pós-teste / 1 + chance pós-teste

Diversos exemplos de sua utilização em pneumologia já foram descritos<sup>13, 14, 15, 16, 17</sup>. Vamos exemplificar com a utilização da LR no diagnóstico da tuberculose pleural<sup>17</sup>. A prevalência da tuberculose em derrames puncionados em adultos, em áreas de

alta prevalência, está em torno de 50%. Isso nos permite calcular uma chance pré-teste da presença de tuberculose igual à unidade, uma vez que na transformação de probabilidade em chance temos:  $0,5 / (1 - 0,5)$ . A associação de dois testes diagnósticos, a ADA maior do que 39U/l (com  $LR+ = 5,56$ ) e linfócitos maior do que 80% ( $LR+ = 2,11$ ), permite calcular uma chance pós-testes de 11,73 (multiplicando-se  $1 * 2,11 * 5,56$ ). A probabilidade pós-testes será, então, de 92%, calculada por:  $11,73 / (1 + 11,73)$ .

Apesar de o cálculo não ser difícil, pode ser cansativo. Um nomograma (figura 1), proposto por Fagan<sup>18</sup>, pode simplificar estas conversões, indo-se diretamente da probabilidade pré-teste, na primeira coluna, para a pós-teste, a terceira coluna, com a reta que passe pela LR calculada, na coluna do meio. No exemplo observado na figura 1, a linha pontilhada mostra a utilização de um teste com  $LR+$  de 7, numa população com prevalência de 30%. A probabilidade da presença de doença, neste caso, é de 75%. Esta probabilidade pós-teste pode ser utilizada como probabilidade pré-teste para um outro teste diagnóstico, exemplificada pela linha cheia. Este, por exemplo, possui uma  $LR+$  de 10 e nos permite calcular uma probabilidade pós-testes maior do que 95%. Nem sempre seu cálculo é tão preciso como nas fórmulas, mas, geralmente, a probabilidade pré-teste também é aproximada.

Geralmente, os testes diagnósticos são mais necessários quando a probabilidade pré-teste se encontra entre 30 e 70%<sup>4, 19</sup>. Nestas condições, sabemos que testes, ou associação destes, com LR maior do que 10 ou menor do que 0,1 levam a uma grande variação, geralmente definitiva, da probabilidade pré-teste para a pós-teste. Se entre 5 e 10 ou 0,1 a 0,2, o efeito é moderado e dependente da probabilidade pré-teste. Valores entre 2 e 5 ou 0,5 a 0,2 podem ser úteis mas seu efeito é pequeno. Quando a LR está entre 1 e 2 ou 0,5 a 1, altera pouco a probabilidade e raramente são importantes<sup>12</sup>.

A LR vem sendo considerada a melhor medida da utilidade de um teste, pois indica como e quanto determinado resultado muda a probabilidade da presença de doença, dada uma probabilidade pré-teste da condição sob investigação<sup>10, 15, 20, 21</sup>. As chances pré-teste contêm a mesma informação que a probabilidade pré-teste ou prevalência; as razões de verossimilhança, a mesma que a sensibilidade e especificidade; e as chances pós-teste, as mesmas que os valores preditivos (probabilidades pós-teste).

A diferença entre a probabilidade pré e pós-teste é uma das maneiras de se avaliar a utilidade de um

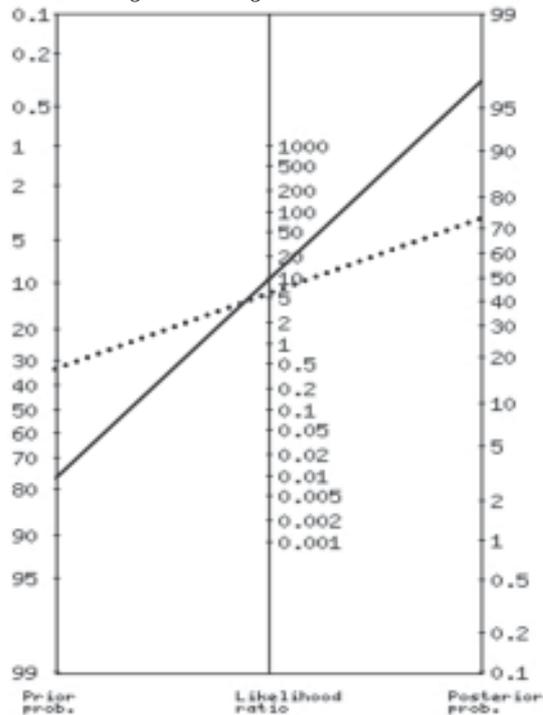
teste diagnóstico<sup>12</sup>. A utilização das LR possui ainda outras vantagens: 1) ser facilmente compreensível e utilizada, na prática, para confirmar ou excluir um diagnóstico; 2) pode ser utilizada seqüencialmente, sendo que a probabilidade pós-teste, após a utilização de um teste, pode ser utilizada como probabilidade pré-teste de outro ou pode ser utilizada em testes combinados, em paralelo, por meio da fórmula descrita; 3) pode ser calculada para diferentes valores discriminatórios obtidos com um teste cujo resultado é expresso em valores contínuos<sup>4, 20</sup>. Esta abordagem permite ainda “personalizar” a regra de predição por intermédio da escolha de diferentes probabilidades pré-teste e de variáveis relevantes.

Deve-se destacar que nem sempre valores mais elevados de um teste estão correlacionados com as maiores LR. Como exemplo podemos citar os valores da LR da adenosina desaminase (ADA) no diagnóstico da tuberculose pleural. O valor da  $LR+$  é máximo (de 5,5 a 7,5) entre 60 a 90 U/l, mas depois começa a diminuir e para valores acima de 120 U/l está em torno de 2<sup>17</sup>. Quando observamos valores muito elevados da ADA no líquido pleural, devemos pensar na possibilidade de resultado falso positivo, principalmente empiema e linfoma.

Os resultados da maioria dos exames complementares não nos permitem uma conclusão definitiva da presença ou ausência de determinada enfermidade. Apenas aumentam ou diminuem a nossa crença, em grau variado, sobre um determinado diagnóstico. Apesar desta conduta informal, intuitiva, ter se mostrado bastante acurada, ela pode variar entre profissionais. Utilizando a metodologia Bayesiana podemos quantificar esta incerteza<sup>10</sup>. Para qualquer doença existe uma probabilidade tal de sua presença que dispensa testes adicionais para a confirmação do diagnóstico, assim como há outra em que sua presença é afastada e outros testes são dispensados. Se a probabilidade encontra-se entre estas duas, novos testes diagnósticos estão indicados.

O valor da probabilidade que separa a ação de se continuar investigando (“limite teste”) ou iniciar a terapêutica (“limite tratamento”) depende de inúmeros fatores, quer genéricos, quer individuais, e ainda não está definido para a maioria das doenças. Novos testes diagnósticos só serão efetivamente implementados quando ações forem tomadas baseadas em seu resultado, alterando a conduta em determinado paciente<sup>22</sup>. A definição dessas probabilidades é importante ainda para avaliarmos se os testes disponíveis são ou não suficientes para a conduta clínica adequada.

Fig. 1 – Nomograma de Fagan.



No exemplo, a linha pontilhada mostra o cálculo da probabilidade pós-teste (de aproximadamente 75%) utilizando um teste com  $LR+ = 7$  e partindo-se de uma prevalência de 30%. A linha cheia mostra o cálculo partindo-se de uma probabilidade pré-teste de 75% e utilizando um teste com  $LR+ = 10$ . Neste caso, a probabilidade da presença da doença é maior do que 95%.

## REFERENCIAS BIBLIOGRÁFICAS

1. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA 1994;271(5):389-91.
2. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? JAMA 1994;271(9):703-7.
3. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA 1995;274:645-51.
4. Fletcher RH, Fletcher SW, Wagner EH. Epidemiologia Clínica: elementos essenciais. 3 ed. Porto Alegre: Artes Médicas; 1996.
5. Medronho RA, Carvalho DM, Bloch KV, Luiz RR, Werneck GL. Epidemiologia. São Paulo: Atheneu; 2002.
6. Doubilet PM, Cain KC. The superiority of sequential over simultaneous testing. Med Decis Making 1985; 5(4):447-51.
7. Castro F, Caccamo LP, Carter KJ, Erickson BA, Johnson W, Kessler E, Ritchey NP, Ruiz CA. Sequential test selection in the analysis of abdominal pain. Med Decis Making 1996;16(2):178-83.
8. Sterne JAC, Davey Smith G. Sifting the evidence: what's wrong with significance tests? Br Med J 2001;322:226-23.
9. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to bayesian methods in health technology assessment. Br Med J 1999;319:508-512.
10. Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis. Evaluation of diagnostic procedures. Br Med J 2002;324:477-80.
11. Sackett DL, Rennie D. The science and art of the clinical examination. JAMA 1992;267:2650-2.
12. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. Br Med J 1994;308:1552.
13. Cummings SR, Lillington GA, Richard RJ. Estimating the probability of malignancy in solitary pulmonary nodules. Am Rev Respir Dis 1986;134: 449-452.
14. Riantawan P, Chaowalit P, Wongsangiem M, Rojanaraweewong P. Diagnostic value of pleural fluid adenosina deaminase in tuberculous pleuritis with reference to HIV coinfection and bayesian analysis. Chest 1999;116:97-103.
15. Verbon A, Weverling GJ, Kuijper S, Speelman P, Jansen HM, Kolk AHJ. Evaluation of different tests for the serodiagnosis of tuberculosis and the use of likelihood ratios in serology. Am Rev Respir Dis 1993;148:378-84.
16. Heffner JE, Highland K, Brown LK. A Meta-analysis Derivation of Continuous Likelihood Ratios for Diagnosing Pleural Fluid Exudates. Am. J. Respir. Crit. Care Med. 2003;167(12):1591-1599.
17. Neves DD. Modelos de predição no diagnóstico da tuberculose pleural: importância da adenosina desaminase. [Doutorado]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2003.
18. Fagan TJ. Nomogram for Bayes' theorem. N Engl J Med 1975;293:257-61.
19. Greenhalgh T. How to read a paper: Papers that report diagnostic or screening tests. Br Med J 1997;315:540-543.
20. Gambino R. The misuse of predictive value - or why you must consider the odds. Ann Ist Super Sanita 1991;27(3):395-9.
21. Steurer J, Fischer JE, Bachmann LM, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. Br Med J 2002;324:824-6.
22. Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. How to use an article about disease probability for differential diagnosis. JAMA 1999;281 (13):1214-9. ■