

Capítulo 5

Artigos sobre Testes Diagnósticos

Podemos definir os testes diagnósticos de acordo com a finalidade, em três tipos clínicos distintos: a) rastreamento de doença em pessoas assintomáticas; b) diagnóstico de doença em pessoa com suspeita (sintomas ou sinais de doença); c) avaliar a progressão da doença ou resposta ao tratamento.

Podemos ainda defini-los, de acordo com o tipo de resultado fornecido, em:

a) *Quantitativo*: quando expresso em resultados numéricos (por medidas) podendo ser subdivididos em:

a.1) Resultados dicotômicos

a.2) Resultados contínuos

b) *Qualitativos*: quando apresentam resultados descritivos e que dependem da percepção humana como, por exemplo, os exames de imagem.

No entanto, mesmo testes qualitativos podem ser submetidos a escores e fornecer resultados numéricos. Um bom exemplo é a classificação Bi-RADS para mamografia. Segundo Orel *et al*⁽¹⁾ a negatividade de biópsia para a detecção de câncer de mama foi de 87% na categoria 0, 100% nas categorias 1 e 2, 98% na categoria 3, 70% na categoria 4 e de 3% na categoria 5.

A apresentação de resultados numéricos permite a realização de estudos de performance do teste, facilitando a compreensão de suas qualidades e limitações e das situações nas quais está indicado.

5.1 Performance dos testes diagnósticos

São cinco as principais características de performance dos testes diagnósticos com resultados numéricos: sensibilidade, especificidade, valor preditivo (positivo e negativo), acurácia e razão de verossimilhança (positiva e negativa)⁽²⁾.

Para um melhor entendimento destes conceitos usaremos uma tabela 2x2 conforme descrito abaixo:

Tabela 5.1 - Modelo tabela 2x2

Condição do paciente	Doentes	Não-doentes	Total
Resultado do teste			
Positivo	a	b	a+b
Negativo	c	d	c+d
Total	a+c	b+d	a+b+c+d (N)

Sensibilidade: é a probabilidade de resultado positivo nos doentes (verdadeiro positivo) e é calculado como: $a/(a+c)$.

Especificidade: é a probabilidade de resultado negativo nos não-doentes (verdadeiro negativo) e é calculado como: $d/b+d$.

Valor preditivo positivo: é a probabilidade da presença da doença quando o teste é positivo e é calculado como: $a/a+b$.

Valor preditivo negativo: é a probabilidade da ausência de doença quando o teste é negativo: $d/c+d$.

Acúrcia: é a probabilidade do teste fornecer resultados corretos, ou seja, ser positivo nos doentes e negativo nos não doentes. Expresso de outra forma é a probabilidade dos verdadeiros positivos e verdadeiros negativos como uma proporção de todos os resultados e é calculado como: $(a+d)/(a+b+c+d)$, também expresso como $(a+d)/N$.

Razão de verossimilhança (RV): é a probabilidade de um determinado resultado em alguém com a doença dividida pela probabilidade do mesmo resultado em alguém sem a doença e também pode ser positiva ou negativa.

$$a) \text{ RV}_{\text{positiva}} = \frac{\text{probabilidade do resultado positivo nos doentes}}{\text{probabilidade do resultado positivo nos sadios}} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

Verificando nas fórmulas acima, podemos reescrever a RV positiva como

$$\frac{\text{Sensibilidade}}{(1-\text{Especificidade})}$$

$$b) \text{ RV}_{\text{negativa}} = \frac{\text{probabilidade do resultado negativo nos doentes}}{\text{probabilidade do resultado negativo nos sadios}} = \frac{\frac{c}{a+c}}{\frac{d}{b+d}}$$

que pode, também, ser escrito como

$$\frac{(1-\text{Sensibilidade})}{\text{Especificidade}}$$

Sensibilidade e especificidade descrevem a proporção do resultado positivo ou negativo em quem, sabidamente está, ou não, doente. Por este motivo é necessário outro exame, considerado como padrão-ouro, na diferenciação entre doente e não doente. Como, na prática, não sabemos quem realmente está doente antes de realizarmos o teste, estas características são de pouco valor clínico. São fundamentais, no entanto, para a avaliação inicial de um novo teste. Como são valores não influenciados pela prevalência da doença, podem ser utilizados em diferentes populações e, também, para comparar o potencial diagnóstico de testes diferentes. Valores preditivos positivos e negativos contêm informações sobre o poder do teste (sensibilidade e especificidade) e da população a ser examinada (prevalência da doença), sendo uma medida de maior utilidade clínica. Por outro lado por serem dependentes da prevalência da doença, não podem ser generalizados para pacientes com perfil diferente daquele do estudo clínico inicial e não permitem a comparação entre diferentes testes diagnósticos⁽³⁾.

Razão de verossimilhança, por tratar-se de uma razão entre a probabilidade de um determinado resultado do teste em pessoas doentes sobre o mesmo resultado em pessoas sadias, fornece seu resultado sob a forma de um valor relativo, ou seja, um número sem unidade, permitindo então, sua utilização em fórmulas de cálculo de probabilidade de doença após combinação de testes. Outra característica da RV é poder ser utilizada em testes diagnósticos com

resultados contínuos. Permite, ainda, o cálculo de tabelas e nomogramas que auxiliam o refinamento diagnóstico, levando-se em consideração características próprias de cada paciente. Sua principal desvantagem é a dificuldade de ser calculado, principalmente em situações clínicas.

A acurácia é uma característica principalmente utilizada quando se deseja transformar um teste com resultados numéricos contínuos em testes dicotômicos, através da determinação de um ponto de corte.

5.2 Aplicação prática

A nossa necessidade clínica é entender a utilidade de um teste diagnóstico para o rastreamento das populações sadias, para o diagnóstico de uma determinada doença em pessoas com suspeita ou para avaliar a progressão ou resposta a um determinado tratamento.

Quando a intenção for o diagnóstico de uma doença, como nos casos de exames de rastreamento, o melhor teste é aquele com alta especificidade porque terá mais impacto no valor preditivo positivo. Ou seja, se o teste der resultado positivo é muito pouco provável que a pessoa não esteja, de fato, doente.

Quando a intenção for afastar o diagnóstico de uma doença ou condição, como por exemplo, em paciente suspeito de recidiva ou progressão, considera-se que o melhor teste deve ter alta sensibilidade porque terá mais impacto no valor preditivo negativo. Ou seja, se o teste der resultado negativo é muito pouco provável que a pessoa esteja, de fato, doente⁽³⁾.

Em relação aos testes com resultados quantitativos contínuos existem duas formas de utilização clínica.

1º- Transformá-los em teste dicotômicos através da escolha de um ponto de corte que defina os resultados positivos e negativos, o que é feito geralmente através de utilização da curva ROC (*Receiving Operator Characteristic*) que tem este nome porque foi construída, inicialmente, para a análise de sinais de radares para a detecção de aviões, na segunda guerra mundial.

2º- O uso da RV na detecção da probabilidade da presença de uma doença, de acordo com o resultado atual do teste.

5.2.1 A curva ROC

A curva ROC também denominada de *relative operating characteristics curve*, define as características operacionais de um teste com valores numéricos contínuos.

Para melhor entender o conceito e sua importância prática utilizaremos um exemplo fictício.

Tomemos o teste ilustrado na figura 5.1 que pode assumir valores entre 0,3 a 4,8. Consideremos que, em comparação com outro teste considerado como padrão-ouro para detecção de uma determinada doença, podemos afirmar, com toda certeza, que pessoas com valores ≤ 1 são normais e aquelas com valores $> 4,0$ estão doentes. Para valores entre 1,0 e 4,0 há incidência variável de doentes e não doentes.

Se por exemplo, desejarmos transformar estes valores contínuos em dicotômicos e elegermos o valor 2,0 como ponto de corte para positividade ou negatividade do teste veremos, conforme assinalado na figura 5.1, que entre os doentes 10% serão falso-negativo e entre os sadios 50% serão falso-positivos.

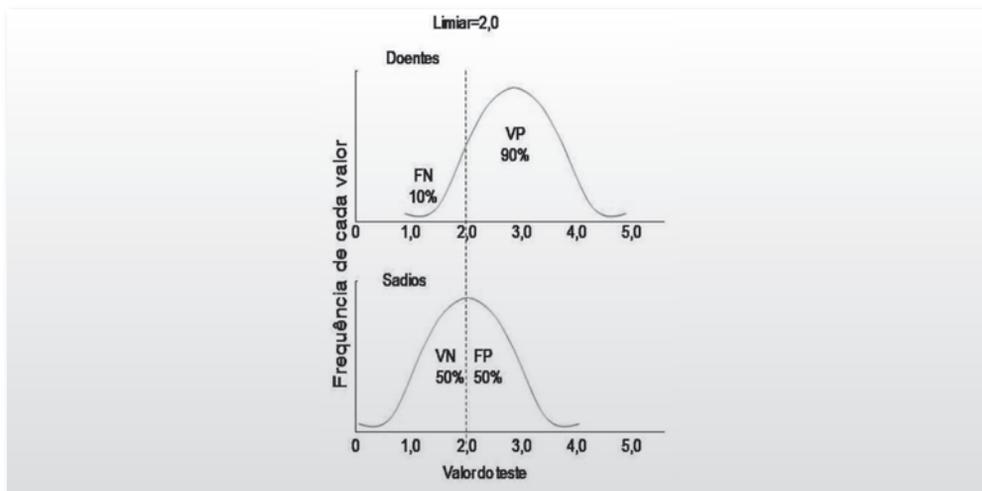


FIGURA 5.1 - Curva de frequência de resultado do teste fictício em populações sabidamente sadia e doente.

A curva ROC é traçada colocando-se no eixo das abscissas (x) as probabilidades de ocorrência de resultados falso-positivos, que também pode ser expresso como 1-especificidade. No eixo das ordenadas (y) colocamos as probabilidades de ocorrência de resultados verdadeiro-positivos que também podem ser expressos como a sensibilidade do teste para aquele valor. Na figura 5.2 uma curva ROC foi construída baseada no exemplo fictício assinalado na figura 5.1.

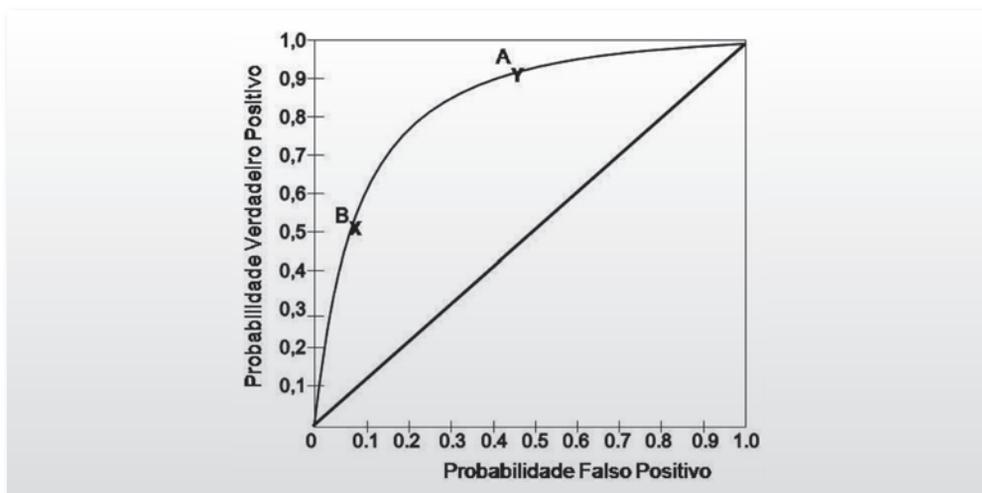


FIGURA 5.2 - Curva ROC do teste fictício descrito na figura 5.1

Neste tipo de representação, a área sob a curva define a probabilidade do teste em detectar os verdadeiro-positivos e os verdadeiro-negativos, ou seja, expressa a acurácia do teste. A linha traçada em diagonal, a 45 graus, expressa a acurácia de 50% e significa que o teste é tão bom quanto qualquer escolha ao acaso, como por exemplo, jogar uma moeda tipo cara ou coroa.

Na linha arqueada situada à esquerda encontra-se uma acurácia de 85%, que significa que em 85% das vezes em que for utilizado, o teste discriminará os verdadeiro-positivos e os verdadeiro-negativos; isto também significa que em 15% das vezes fornecerá um falso resultado.

O ponto de maior acurácia do teste é aquele situado a maior distância da linha dos 45 graus.

Consideremos novamente a figura 5.2; se decidirmos que o ponto de corte está localizado no ponto A, que corresponde ao valor de 2,0 do teste, estaremos aptos a detectar 90% dos doentes, mas incluiremos nesta categoria 50% dos sadios. Em uma doença aguda ou de evolução rápida com alto risco de morte ou lesão irreversível, é fundamental diagnosticar a maioria dos doentes mesmo à custa de alta porcentagem de falso-positivo. Chamamos a este limiar de brando, porque ele maximiza a descoberta de indivíduos doentes, às custas de alto número de falso-positivos.

Se considerarmos, o ponto de corte no valor 3,0 do teste, localizado no ponto B da curva da figura 5.2, estaremos aptos a considerar como sadios 90% das pessoas realmente não doentes. Por outro lado, detectaremos apenas 50% dos doentes. Este limiar é denominado de severo porque reduz o número de falso-positivos às custas de não detectar muitos indivíduos realmente doentes. Em uma patologia de evolução lenta ou sem risco imediato para o paciente, mas cujo tratamento seria de alto risco (como por exemplo, quimioterapia em caso de recidiva de câncer), este deve ser o critério a ser usado na determinação do ponto de corte do teste para a definição da positividade ou negatividade.

5.2.2 Razão de verossimilhança (RV)

Raramente um teste é aplicado sem uma justificativa para seu uso, ou seja, há sempre a probabilidade da existência de uma doença. Nestas circunstâncias devem ser usadas as propriedades da teoria da probabilidade. Thomas Bayes descreveu regras para o uso de probabilidades condicionais, ou seja, como a variação na probabilidade pré-teste pode influenciar no diagnóstico final, mas um estudo mais aprofundado do tema não é objetivo deste texto.

Como vimos anteriormente, a RV pode ser muito útil neste refinamento do diagnóstico em situações clínicas. Para facilitar o entendimento destes conceitos utilizaremos um exemplo citado por Deeks⁽³⁾ e modificado para servir aos nossos propósitos.

Na tabela 5.2 é reproduzida a probabilidade de diagnóstico de câncer baseado na espessura do endométrio conforme medido por ultrassonografia transvaginal em pacientes na pós-menopausa, com sintoma recente de sangramento vaginal e que foram submetidos, em seguida, a curetagem uterina.

Tabela 5.2 - Diagnóstico anatomopatológico de câncer do endométrio em relação à espessura do endométrio medida por ultrassonografia transvaginal

Diagnóstico de Câncer de endométrio	Espessamento do endométrio (mm)							Total
	≤ 4,0	5,0	6,0-10	11-15	16-20	21-25	≥ 26	
Sim	0	2	13	28	27	17	27	114
Não	518	86	232	117	38	17	16	1024
Total	518	88	245	145	65	34	43	1138

A RV pode ser calculada para cada valor de espessura do endométrio percebido pelo exame. Na tabela 5.3 foi construída uma escala de RV positiva baseado nestes dados. Na realidade esta informação não tem muita utilidade clínica. Sua importância está na sua capacidade de ser incluída em fórmulas que permitam um diagnóstico mais preciso na presença de características próprias de cada paciente.

Tabela 5.3 - Cálculo da razão de verossimilhança para espessura endometrial no diagnóstico de câncer de endométrio

Espessura endométrio (mm)	Ca de endométrio		RV+
	Sim (%)	Não (%)	
≤4,0	0(0)	518(50,6)	0/50,6 = 0
5,0	2(1,8)	86(8,4)	1,8/8,4 = 0,2
6,0-10	13(11,4)	232(22,7)	11,4/22,7 = 0,5
11-15	28(24,6)	117(11,4)	24,6/11,4 = 2,2
16-20	27(23,7)	38(3,7)	23,7/3,7 = 6,4
21-25	17(14,9)	17(1,7)	14,9/1,7 = 8,8
≥26	27(23,7)	16(1,6)	23,7/1,6 = 14,8
Total	114(100)	1024(100)	

A aplicação do teorema de Bayes aos testes diagnósticos pode ser expressa pela seguinte fórmula⁽³⁾:

$$\text{chance atual da doença} = \text{chance pré-teste} \times RV$$

Ou seja, a chance de presença da doença depende da chance pré-teste que é função indireta da prevalência da doença e de características próprias do paciente multiplicada pelo valor da RV do resultado do teste efetivamente encontrado naquele paciente.

Como vimos anteriormente, chance (ou *odds*) não é um valor de fácil apreensão imediata. No entanto, ela pode ser transformada em probabilidade segundo a seguinte, e simples, fórmula:

$$\text{Probabilidade} = \frac{\text{Chance}}{\text{Chance} + 1}$$

Usando estes conceitos e os dados fornecidos pela tabela 5.3, construímos a tabela 5.4. Nesta população a chance pré-teste foi calculada como sendo 1/9 por ser esta a relação entre o número de pacientes com diagnóstico de câncer de endométrio (114) dividido pelo número de pacientes sem diagnóstico de câncer do endométrio (1024).

Tabela 5.4 - Chance e probabilidade de câncer de endométrio de acordo com espessura endometrial à ultrassonografia

Espessura endométrio (mm)	Chance pré-teste	RV	Chance pós-teste	Probabilidade da doença
≤4,0	1/9	0	0	0%
5,0	1/9	0,2	0,22=1:45	2,15%
6,0-10	1/9	0,5	0,55=1:18	5,21%
11-15	1/9	2,2	0,25=1:40	20,0%
16-20	1/9	6,4	0,71=1:1,4	41,5%
21-25	1/9	8,8	1,00=1:1,00	50,0%
≥26	1/9	14,8	1,64=5:3	

Notas: Chance da doença nesta população: 1/9 (114 com doença sobre 1024 sem doença)

Para exemplificar como podemos refinar o diagnóstico diante de um caso clínico específico e utilizando os conhecimentos obtidos da literatura, vejamos as seguintes situações.

Qual a probabilidade de câncer de endométrio em uma mulher de 62 anos de idade, com história de sangramento vaginal, com passado de terapia de reposição hormonal (TRH) com estrógeno, sem uso concomitante de progesterona, por mais de 5 anos e com ultrassonografia endovaginal mostrando espessura endometrial de 22mm?

A primeira resposta seria consultar a tabela 5.4, e concluir que ela tem cerca de 50% probabilidade do diagnóstico de câncer de endométrio.

No entanto, podemos refinar este diagnóstico e para isto, em primeiro lugar, necessitamos saber qual é a chance pré-teste, ou seja, qual a prevalência da doença nesta faixa etária. Como não temos este dado com precisão em relação à população brasileira, usaremos uma chance de referência.

Segundo Weiderpass *et al*⁽⁴⁾ a chance de uma mulher na pós-menopausa que recebeu terapia hormonal com estrógeno isolado, por mais de 5 anos, de desenvolver câncer de endométrio é de 5,60 vezes maior do que uma mulher que não fez a reposição. Na população do estudo, constituída por mulheres suecas com idade entre 50 e 74 anos, a chance do câncer de endométrio foi de 687 casos em 3270 controles sadios (0,21 ou 1:4,8).

Se a nossa paciente pertencer a uma população semelhante à do estudo, podemos concluir que seu risco de estar com câncer de endométrio, considerando o resultado da ultrassonografia, é de 91%.

Como podemos ver, a probabilidade da presença da doença é bem maior do que à primeira impressão.

A sequência de cálculos que fornecem esta resposta é de simples execução.

Consideremos a chance pré-teste de 1:4,8.

Consideremos, também, o aumento de risco devido à TRH de 5,6 vezes; a chance desta paciente, antes da realização do teste, pode ser calculada como a chance média (1:4,8) multiplicada pelo aumento de risco pela TRH (5,6) → $1:4,8 \times 5,6 = 1,17$. A chance atual, após o teste, para câncer de endométrio é encontrada multiplicando-se a chance pré-teste (1,17) pela RV do resultado da ultrassonografia (8,8) que é igual a 10,3.

A probabilidade atual da doença é fornecida pela transformação de chance em probabilidade, ou seja, $10,3/(10,3+1)=0,91$ que corresponde a 91%.

E se esta mesma paciente apresentasse espessura endometrial de 8,8 mm? Consulta à tabela 5.4 nos faria supor que a probabilidade do diagnóstico de câncer seria de 5,21%. No entanto, seguindo os mesmos cálculos realizados anteriormente, concluímos que a probabilidade atual da presença da neoplasia é de aproximadamente 37%.

Estes exemplos nos alertam para a importância da análise das várias características de performance dos testes diagnósticos.

É pouco provável, contudo, que em situações clínicas tenhamos tempo para cálculos desta natureza. No entanto, existem tabelas e nomogramas que permitem uma apreciação mais rápida da questão. Para um estudo mais aprofundado recomendamos a leitura do artigo de Fagan⁽⁵⁾.

Referências

1. Orel SG, Kay N, Reynolds C, Sullivan DC. BI-RADS categorization as a predictor of malignancy. *Radiology*. 1999 Jun;211(3):845-50.
2. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ*. 1997 Aug 30;315(7107):540-3. Review.
3. Deeks JJ. Using evaluations of diagnostic tests: understanding their limitations and making the most of available evidence. *Ann Oncol*. 1999 Jul;10(7):761-8. Review.
4. Weiderpass E, Adami HO, Baron JA, Magnusson C, Bergström R, Lindgren A, Correia N, Persson I. Risk of endometrial cancer following estrogen replacement with and without progestins. *J Natl Cancer Inst*. 1999 Jul 7;91(13):1131-7.
5. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5):257.