Interpretação de um Teste sob a Visão Epidemiológica. Eficiência de um Teste

Takao Kawamura

Araçatuba, SP

A evolução do raciocínio na interpretação dos fenômenos naturais, através dos tempos, trouxe, como conseqüência, as bases matemáticas do pensamento científico. Na medicina não foi diferente: a observação dos fenômenos biológicos, a procura de soluções para diminuir o impacto das doenças e a necessidade de se provar, cientificamente, a eficácia de métodos propedêuticos e de procedimentos terapêuticos abriu as portas para o que, hoje, se denomina medicina baseada em evidências.

Thomas Bayes, um matemático inglês do século XVII legou-nos o seu teorema que estabeleceu que a probabilidade pós-teste de uma doença era função da sensibilidade e especificidade do exame e da prevalência da doença na população (probabilidade pré-teste). Nós médicos, ao formularmos as nossas hipóteses diagnósticas, ao interpretarmos os exames laboratoriais e ao prescrevermos um tratamento, intuitivamente. utilizamos o teorema de Bayes. Hoje, vivemos a era da alta tecnologia em que as pessoas, freqüentemente, tendem a interpretar a positividade de um exame sofisticado e caro como sinônimo de doença. Não devemos esquecer que todos os exames, sem exceção, desde o corriqueiro exame clínico até uma tomografia computadorizada, estão limitados pela sensibilidade, especificidade e valor preditivo pré-teste.

Defenderemos, nesta apresentação, a introdução de um simples e novo conceito (baseado em antigos e conhecidos conceitos), que, provisoriamente, poderíamos denominá-lo de eficiência de um teste (Ef), como uma arma epidemiológica e propedêutica. Não encontramos nenhuma citação sobre o presente conceito proposto, tanto na literatura nacional como na de língua inglesa. Esta explicação tornarse-á mais didática se relembrarmos esses velhos conhecimentos básicos de epidemiologia clínica (tab. I).

Cardioclínica Araçatuba

Correspondência: Takao Kawamura - Rua Tiradentes, 1301 - 16015-020 Aracatuba, SP - E-mail: takaok@terra.com.br

Recebido para publicação em 10/5/01

Aceito em 5/9/01

Sensibilidade (s) - é a probabilidade de um indivíduo avaliado e doente de ter seu teste alterado (positivo).

s = número de indivíduos doentes e com teste positivo/número total de indivíduos doentes; ou:

$$s = VP/(VP + FN)$$
 (equação 1)

Especificidade (e) - é a probabilidade de um indivíduo avaliado e normal ter seu teste normal (negativo).

e = número de indivíduos normais e com teste negativo/número total de indivíduos normais; ou:

$$e = VN/(VN + FP)$$
 (equação 2)

Prevalência (**p**): é a fração de indivíduos doentes na população total avaliada.

p = número de indivíduos doentes / número de indivíduos da população; ou: **p** = **Do/n** (equação 3) (onde: **Do** = doentes; **n** = população)

Valor preditivo positivo **(VPP):** é a probabilidade de um indivíduo avaliado e com resultado positivo ser realmente doente.

$$VPP = VP/(VP + FP)$$
 (equação 4)

Valor preditivo negativo (VPN): é a probabilidade de um indivíduo avaliado e com resultado negativo ser realmente normal.

$$VPN = VN/(VN + FN)$$
 (equação 5)

A partir dos dados expostos podemos delinear as seguintes fórmulas:

Se:
$$Do = p.n, Sa = (1-p).n, VP = s.Do, VN = e.Sa$$

FP = (1-e) . Sa e FN = (1-s) . Do

Onde **Sa** = Sadios Então: VP = s. p. n (equação 6)

VN = e. (1-p).n (equação 7)

 $FP = (1-e) \cdot (1-p) \cdot n (equação 8)$

 $FN = (1-s) \cdot p \cdot n \text{ (equação 9)}$

Do mesmo modo:

VPP = VP/(VP + FP)

VPP = s.p.n/[s.p.n+(1-e).(1-p).n]

VPP = s.p/[s.p+(1-e).(1-p)] (equação 10)

Do mesmo modo:

VPN = VN/(VN + FN)

VPN = e.(1-p).n/[e.(1-p).n+(1-s).p.n]

VPN = e.(1-p)/[e.(1-p)+(1-s).p] (equação 11)

		Doença	
		Positivo	Negativo
Teste	Positivo	VP	FP
	Negativo	FN	VN

Esta "poluição de números" é indigesta para quem não gosta de matemática, mas é fundamental para o entendimento do raciocínio. O acompanhamento do leitor executando as operações matemáticas junto à leitura será extremamente elucidativo.

A indicação e a valorização de um exame para fins diagnósticos deverão ser regidas pela relação custo-beneficio, levando-se em consideração o valor preditivo pré-teste (igual à prevalência da doença). Assim, a interpretação do seu resultado não pode estar divorciada de uma visão epidemiológica bem alicerçada. Recentemente, foi publicada na revista da SOCESP uma excelente revisão sobre o assunto ¹, da qual citaremos um estudo baseado em autópsias estratificando a prevalência da doença arterial coronariana ². Esses dados (tab. II) serão muito úteis para auxiliar nossos cálculos e aumentarão, significativamente, nosso poder diagnóstico.

Vamos analisar três situações comuns na prática clínica. **Exemplo 1:** mulher de 35 anos portadora de dor torácica não-anginosa é submetida a teste ergométrico para avaliar possibilidade de insuficiência coronariana.

Considerar a sensibilidade e a especificidade do teste ergométrico em 85% e 75%, respectivamente. Consultando a tabela II, poderíamos assumir em 1% a prevalência da doença nesse subgrupo de pacientes. A partir desses dados poderíamos construir o seguinte gráfico: utilizando-se as equações 10 e 11 ou plotando-se os dados diretamente na figura 1, poderemos encontrar os seguintes valores: VPP = 3,3% e VPN = 99,8%, significa que, se o teste ergométrico for positivo, existem 3,3% de chances da paciente realmente ser doente, contra 96,7% (100-3,3%) de ser normal, apesar do resultado ser positivo. Se o teste for negativo, existem 99,8% de chances da paciente ser normal contra 0,81% (100-99,8%) de ser doente, apesar do resultado ser negativo. Em outras palavras: testando-se este tipo de paciente seriam necessários realizar 100 testes para diagnosticar insuficiência coronariana em apenas 3,3 pacientes, representando um

Tabela II - Estratificação de prevalência da doença arterial coronariana. Estimativa de probabilidade pré-teste ²						
Dor torácica Angina atípica não-anginosa			Angina	típica		
Idade	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
40 - 49	14,1±1,3	$2,8\pm 0,7$	$46,1\pm 1,8$	$13,3\pm 2,9$	69,7±3,2 87,3±1,0 92,0±0,6	$55,2\pm6,5$
	, ,	, ,	, ,	, ,	$94,3\pm0,4$	

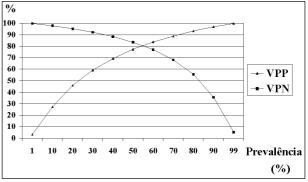


Fig. 1 - Valor preditivo positivo e valor preditivo negativo de um teste (s = 85% e = 75%).

gasto muito grande de recursos e uma relação custo-beneficio muito baixa.

Exemplo 2 - Homem de 65 anos de idade com dor precordial típica é submetido a teste ergométrico para avaliar possibilidade de insuficiência coronariana. Utilizando os mesmos artifícios do exemplo anterior teremos: p = 94%, VPP = 98% e VPN = 24%. Isto significa que se o resultado do exame for positivo, há 98% de chances do indivíduo ser coronariano e 2% (100-98%) de chances de ser normal. Se o teste for negativo, existem 24% de chances do indivíduo ser normal contra 76% (100-24%) de chances de ser coronariano, apesar do resultado ser negativo. Se a decisão de tratar ou de se indicar algum procedimento baseado apenas no resultado do exame, no caso de resultado negativo, existe uma grande chance de se tomar a decisão errada (de não tratá-lo/conduzi-lo como coronariopata).

Exemplo 3-Homem de 55 anos de idade com dor precordial atípica é submetido a teste ergométrico para avaliar possibilidade de insuficiência coronariana. Utilizando-se os mesmos artificios anteriores vamos encontrar os seguintes dados: **p=60%**, **VPP=83,6%** e **VPN=76,9%**. Se o resultado do exame for positivo, há 83,6% de chances do indivíduo ser portador de doença contra 16,4% (100-83,6%); se negativo, há 76,9% de chances do indivíduo ser normal contra 23,1% (100-76,9%). Há uma discriminação bastante evidente, podendo a conduta tomada ser correta se baseada nos resultados do exame.

Dos três exemplos citados, poderíamos ainda extrair mais alguns dados interessantes: a tabela III demonstra claramente que o incremento do valor preditivo positivo pósteste foi maior no indivíduo de probabilidade pré-teste intermediária (exemplo 3). Nos dois extremos, o incremento foi muito pequeno (exemplos 1 e 2).

Existe, pois, uma zona média (de VPP pré-teste) onde o

Tabela III - Análise do incremento no valor preditivo positivo após realização do teste				
Exemplos:	VPP pré-teste (%)	VPP pós-teste (%)	VPP _{pós} - VPP _{pré} (%)	
1	1,0	3,3	2,3	
2	94,0	98,0	4,0	
3	60,0	83,6	23,6	

Arq Bras Cardiol 2002; 79: 437-41.

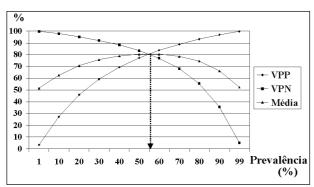
teste tem eficácia máxima e uma melhor relação custo-benefício. As questões seguintes a serem respondidas são: 1) que valores intermediários seriam esses? 2) Esses valores dependem da sensibilidade e especificidade do exame utilizado? 3) A partir de que valores, tanto para **VPP** como para **VPN**, poderíamos assumir como o de um exame eficiente? 4) E a que faixa de prevalência da doença (valor preditivo préteste) corresponderiam esses valores? 5) Essas faixas teriam a mesma extensão para qualquer exame? As respostas a essas questões pedem a definição de um novo conceito, como dito no início desta apresentação, baseado em velhos e conhecidos conceitos, que temporariamente o denominamos de eficiência de um teste (**Ef**).

Como vimos, todo exame tem a sua sensibilidade (s) e especificidade (e) próprias. Para cada conjunto de s e e poderíamos construir um gráfico relacionando VPP e VPN com p (prevalência). No exemplos anteriores de teste ergométrico onde s = 85% e e = 75% poderíamos construir um gráfico onde as curvas VPP e VPN teriam a mesma direção (para a direita), mas orientações diametralmente opostas (de baixo para cima e de cima para baixo, respectivamente). O cruzamento de ambas vai sempre ocorrer num ponto de prevalência média (60% no presente exemplo). Se tirarmos uma média de VPP e VPN obteremos valores que vão gerar uma terceira curva partindo de valores próximos a 50%, subindo até a um pico pouco acima da média de s + e (no nosso exemplo 85 + 75 = 80%) que sempre corresponderá à interseção de VPP com VPN, e descendo depois para os valores iniciais (fig. 2). A esta altura, poderíamos denominar essa terceira curva de eficiência de um teste (Ef).

Assim: $\mathbf{Ef} = (\mathbf{VPP} + \mathbf{VPN})/2$ (equação 12)

Valores de prevalência em torno desse pico definiriam a zona de prevalência de máxima eficácia do exame (**ZEf**).

A questão a ser respondida agora é: como definir o que é eficiente e como calcular essas zonas de eficiência? Utilizando os mesmos artificios utilizados na construção da figura 2, poderíamos fazer um gráfico de $\mathbf{s}=50\%$ e $\mathbf{e}=50\%$. O resultado seria a figura 3: visualmente podemos facilmente constatar que um exame com esta característica não acrescenta nada aos valores preditivos pré-teste, tendo portanto eficiência nula. Por outro lado, um exame hipotético (e pouco provável de existir algum dia) de $\mathbf{s}=100\%$ e $\mathbf{e}=100\%$ geraria a figura 4 onde a eficiência seria a máxima. Poderíamos, por



 $Fig.\ 2 - C\'alculo\ da\ eficiência\ de\ um\ teste:\ m\'edia\ de\ VPP\ e\ VPN\ (s=85\%\ e=75).$

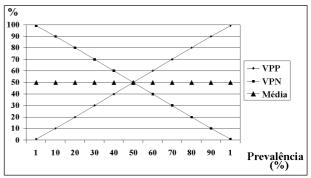


Fig. 3 - Cálculo de eficiência de um teste de s = 50%^e = 50%.

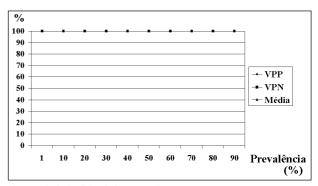


Fig. 4 - Cálculo de eficiência de um teste de s=100% e = 100%. As três curvas (VPP VPN e Média) estão superpostas.

exemplo, assumir o valor 75% como valor de corte para estabelecer o que é eficiente, porém outros poderiam achar que 60% é aceitável e outros, mais prudentes, poderiam exigir 80%. Seria aconselhável então subscrever essa sigla com um valor (Ef₇₅, Ef₆₀, Ef₈₀) de conveniência individual.

Vamos traduzir essas idéias na figura 5. Utilizando-se as fórmulas 10, 11 e 12 é possível construir-se as curvas do gráfico: curva de **VPP** (valor preditivo positivo) ascendente com concavidade para cima, **VPN** (valor preditivo negativo) descendente e com concavidade para cima e curva **Ef** (eficiência de um teste) que é a média das outras duas anteriores: parte do valor pouco acima de 50%, sobe progressivamente até um pico pouco acima da média de sensibilidade e especificidade ([s+e]/2), 85% no nosso exemplo ([90+80]/

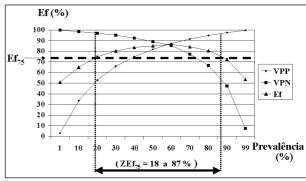


Fig. 5 - Cálculo da zona de prevalência de eficiência máxima (ZEf_{75}) de Ef_{75} de um teste com s=90% e=80%.

2=85) e desce para os valores próximos dos iniciais. Escolhendo-se **Ef=75%**, traçaremos uma reta nesse nível, paralelamente ao eixo de prevalência (**p**). Nos pontos de interseção com curva de eficiência e projetando-se esses pontos no eixo de prevalência (**p**), estaremos delimitando a zona de prevalência de máxima eficiência (**ZEf**₇₅) para esse exame (**ZEf**₇₅=11 a 87%) tendo como referência o nível escolhido para eficiência de 75% (**Ef**₇₅). Por esses dados podemos inferir que se trata de um exame bastante acurado (níveis altos de sensibilidade e especificidade) e eficiente, possuindo um espectro de eficiência bastante largo, ou seja, cobre populações de baixa/média até média/alta prevalência (11 a 87%).

Dispondo de simples informações (sensibilidade, especificidade e prevalência) podemos avaliar o grau de confiabilidade e eficiência do exame e dispor de dados para comparar o alcance de diferentes exames. Para se ter uma melhor idéia comparativa do poder de cada exame, vamos criar a figura 6 com várias curvas de **Ef**(exames com sensibilidade e especificidade variáveis) e usar os mesmos artificios utilizados no gráfico anterior:

As curvas A, B, C, D, E e F correspondem respectivamente a Ef de testes de médias de sensibilidade e especificidade ([s+e]/2) de 50, 60, 70, 75, 80 e 90%. Traçamos uma reta paralela ao eixo de prevalência (p) pois assumimos 75% como ponto de corte para Ef (Ef₇₅). Os pontos de interseção dessa reta com as várias curvas, projetados no eixo p, vão definir as zonas de máxima eficiência (ZEf) da população p para os exames correspondentes no ponto de corte escolhido (ZEf₇₅). Fica fácil concluir que exames A, B e C estão abaixo da reta 75% e são, portanto, pouco ou nada eficientes. O exame D seria eficiente apenas no pico onde ele toca a reta de corte limitando a sua ZEf a um ponto (50%) de p. O exame E teria uma faixa maior de eficiência (ZEf=26 a 74%) e o exame F, uma faixa maior ainda (ZEf=11 a 89%).

Em termos clínicos e epidemiológicos, qual a importância desses conceitos? Ao estipularmos um nível de eficiência, podemos excluir de nossa prática clínica exames e procedimentos pouco eficientes. Por outro lado, poderemos alicerçar nossas condutas e decisões, interpretando corretamente o significado de um resultado de exame. Quanto maior a eficiência de um teste, maior será a população (zona de prevalência) que se beneficiará do mesmo. Decisões ba-

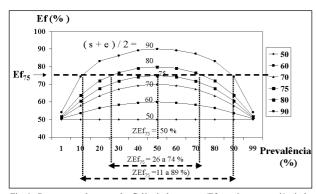


Fig. 6 - Comparação de curvas de eficiência de um teste (Ef) em situações variáveis de se de e (de 50 a 90%). Determinação de zonas de máxima eficiência (Zef).

seadas em um raciocínio bem elaborado e ajustado em números confiáveis, seguramente, resultarão num grau maior de acerto, otimizando a relação custo-benefício. Acreditamos que todos os exames deveriam ter, nas suas conclusões, estipuladas a sua eficiência, a sua zona de melhor eficiência e a chance do paciente avaliado ter ou não doença de acordo com o resultado obtido. Se não for possível determinar a prevalência (valor preditivo pré-teste), então um gráfico com curvas de VPP, VPN e Ef deverá ilustrar o exame para que hipoteticamente o médico assistente possa calcular, de acordo com os dados clínicos disponíveis (por exemplo, no caso de coronariopatia: idade, sexo e característica da dor torácica), as chances de doença e normalidade.

Para o cálculo de prevalência (**p**) realizado na tabela anterior, as fórmulas utilizadas foram: Ac = (VP + VN)/n (equação 13) onde: Ac = acurácia e **n** = número total de exames ou população.

Utilizando-se as equações 6 e 7 e desenvolvendo a equação 13 teremos: $Ac = e + p \cdot (s - e)$ (equação 14) ou: p = (Ac - e)/(s - e) (equação 15).

Se a sensibilidade (s) for igual à especificidade (e), então, para qualquer ponto de p (prevalência), a acurácia (Ac) será constante e igual ao valor de e.

Igualmente, se **s < e**, então Ac será máxima para os menores valores possíveis de p e cairá segundo uma reta descendente até atingir os menores valores para os máximos valores de p.

Por outro lado, se $\mathbf{s} > \mathbf{e}$, então Ac será mínima para os menores valores possíveis de p e subirá segundo uma reta ascendente até atingir os maiores valores para os máximos valores de p (fig. 7).

Sabendo-se os valores de **s**, **e** e **Ac** é possível calcular a prevalência **p** de um determinado estudo (tab. IV).

Sob essa nova óptica também os livros de medicina estão sendo reescritos: conceitos, prevalências e tratamentos de doenças exigem hoje o rigor das evidências. É esta a impressão transmitida pela leitura dos primeiros capítulos de um tradicional livro de cardiologia em sua última edição³. Da mesma forma, números atribuídos para sensibilidade e especificidade dos habituais exames laboratoriais inexoravelmente terão de ser recalculados. Como isso tem sido feito até hoje? O Consenso da *American College of Cardiology* e *American Heart Association* para tomografia computadorizada por emissão de elétrons (*EBCT: electron-beam computed tomography*) selecionou 16 trabalhos de onde

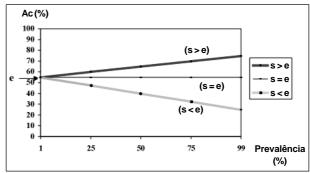


Fig. 7 - Acurácia (Ac), sensibilidade (s) e especificidade (e) de um teste e prevalência (p).

Tabela IV - Sensibilidade, especificidade e acurácia dos 16 trabalhos selecionados no Consenso ACC/AHA para CT de emissão de elétrons (EBCT) ⁴ . Cálculo de prevalência					
Estudo	Sensibilidade (%)	Especificidade (%)	Acurácia (%)	Prevalência (%)	
A	88,4	100	90,7	80,2	
В	100	28,0	40,9	17,9	
C	100	47,2	72,0	47,0	
D	82,5	85,0	83,8	48,0	
E	92,8	66,7	82,5	60,5	
F	96,7	41,3	65,0	42,8	
G	91,0	90,0	90,4	40,0	
Н	99,4	25.7	81,7	76,0	

72.7

43,8

71,0

58,6

21,4

58.3

30,5

66,7

49.2

39.9

92,5

94,6

70,0

87,5

94,7

99.1

95,6

67,9

90.5

I

J

K

L

M

Ν

O

Total

Média Pond

88.2

74,4

70,9

70,4

51,1

94.9

58,4

67,3

69,6

59,1

78.3

60,2

10,0

40.8

40,5

89.7

42,9

50,0

49.4

47,4

extraiu valores médios de 90,5% e 49,2% para sensibilidade e especificidade, respectivamente ⁴. Alguns desses dados estão ordenados na tabela IV que mostra uma grande variabilidade para sensibilidade (de 67,9 a 100%) e de especificidade (de 21,4 a 90%) dos trabalhos envolvidos. Quem está certo ou menos errado? Na realidade esses números traduzem diferentes níveis de corte (porcentagem de cálcio nas artérias coronárias) para "positivo" ou "negativo" e principalmente diferentes populações avaliadas (prevalências de 10 a 89,7%). Isso leva também a resultados bastante variados e consequentemente a números errados.

Recentemente, um folheto direcionado a cardiologistas enaltecia as qualidades de um determinado procedimento diagnóstico para insuficiência coronariana, conferindolhe uma acurácia de pelo menos 90%. Já vimos que acurácia não mede apenas a qualidade de um teste, mas depende também da população avaliada. Para este exame, admite-se valores aproximados de 90% e 50% para sensibilidade e especificidade, respectivamente. Utilizando-se as equações 14 ou 15 teremos:

$$Ac = e + p.(s - e)$$

 $90 = 50 + p.(90 - 50)$ assim: $p = 100\%$

Ou seja: apenas para uma população extremamente selecionada a afirmação é verdadeira. Se um paciente de p = 50% (por exemplo: homem de 50 anos com dor precordial atípica) fosse submetido a esse exame o resultado seria dife-

$$Ac = 50 + 0.5 \cdot (90 - 50)$$

 $Ac = 70\%$

O exame continua sendo bom e acurado, mas nós não podemos nos embriagar com a sedução dos números.

A exemplo dos atuais ensaios randômicos e multicêntricos para procedimentos terapêuticos, a medicina (medicina legal, inclusive) também exigirá números mais concisos e confiáveis para procedimentos diagnósticos e a epidemiologia clínica, com certeza, desenvolverá metodologia própria para recalcular esses números. A trilha a ser seguida deverá passar por um consenso de especialistas onde se determinarão a normatização de um nível de corte para "positivo" e "negativo", a escolha do padrão-ouro e a execução de um determinado número de exames (distribuição multicêntrica) em todas as camadas representativas de uma população.

Vivemos a década do consumidor, do aprimoramento das relações interpessoais e institucionais, da cobrança de resultados e de transparência no destino dos finitos recursos comuns e, muito provavelmente, exigências nesse sentido ocorrerão; creio que devemos nos antecipar e ir ao encontro dos anseios da nossa sociedade e desvendar as revelações do nosso tempo. Isso é justo, prático e muito prazeroso e contemplará quem praticar a boa medicina baseada em evidências.

Referências

- Lagudis S. A probabilidade pré-teste e o resultado de exames complementares. Rev Soc Cardiol Estado de São Paulo 2001: 11: 15-20.
- Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. N Engl J Med 1979; 300: 1350-8.
- Braunwald E, Zipes DP, Libby P. Heart Disease. A Textbook of Cardiovascular Medicine. 6th edition. Philadelphia: WB Saunders Co., 2001.
- O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus Document on Electron-Beam Computed Tomography for the Diagnosis and Prognosis of Coronary Artery Disease. J Am Coll Cardiol 2000; 36: 326-40.