

Utilização Racional de Exames Diagnósticos em Cardiologia

Bráulio Luna F^o

São Paulo, SP

*“The problems of disease are more complicated and difficult than any others with which the trained mind has to grapple... **Variability is the law of life.** As no two faces are the same, so no two bodies are alike, and no two individuals react alike and behave alike under the abnormal conditions which we know as disease. This is the fundamental difficulty in the education of the physician, and one which he or she may never grasp...” **Probability is the guide of life.**”*

William Osler¹ 1921

A prática médica é uma atividade que combina arte e ciência. No passado, angariava prestígio pelo talento individual dos seus praticantes, hoje, exige que seus atos sejam também baseados em pesquisas científicas. Embora esta evolução tenha se dado de forma natural e sem grandes conflitos, instados a justificar um diagnóstico, os médicos muitas vezes encontram dificuldades.

Em função do treinamento e da experiência cotidiana, aprendem a estimar graus de certezas diagnósticas, nas mais variadas situações clínicas. Todavia, só ocasionalmente tomam consciência da complexa cadeia de eventos utilizados no raciocínio clínico. Fazem isto quase que intuitivamente e, com o tempo, aprendem a fazê-lo de maneira cada vez mais eficiente. O que de fato poucos percebem é que, o tempo todo, estão quantificando, probabilisticamente, as informações que conduzem aos diagnósticos. A consequência imediata é que, por mais seguro que se julgue em estar certo, há sempre a possibilidade de se estar errado. Surpreendente é que, na maioria das vezes, estão quase sempre certos, mesmo não sabendo como isso realmente acontece.

Embora não seja objetivo deste artigo desvendar a hermenêutica do raciocínio clínico, pretendemos, estudando a lógica dos testes diagnósticos, ajudar os médicos a utilizar, de forma eficiente, este importante instrumento da medicina moderna.

Contexto clínico - Considere a seguinte situação: um homem de 48 anos com dor torácica atípica procurou seu médico. Após história clínica, que não revelou outras manifestações nem fatores de riscos, e exame físico inalterado, foi-lhe solicitado eletrocardiograma (ECG) e teste ergométrico (TE). O resultado do ECG foi normal mas o TE foi positivo. Como poderia ser portador de doença arterial coronária (DAC) recomendou-se a realização de estudo cineangiocoronariográfico. Conversando com seus familiares, o paciente resolveu consultar outro médico, que por sua vez orientou um teste cintilográfico de estresse com tálcio 201 (TTA).

Raciocínio clínico - Já foi destacado que o processo da decisão clínica baseia-se, conscientemente ou não, em probabilidade. Por exemplo, a chance do paciente descrito apresentar uma doença aterosclerótica das artérias coronárias antes da realização do TE era de cerca de 46%. Após um TE positivo esta chance elevou-se para 64%. De maneira que para o primeiro clínico isto já era evidência suficiente para solicitar a realização de um exame invasivo - a cineangiocoronariografia. Já o segundo clínico pode muito bem ter considerado que o TE poderia ter apresentado um resultado falso-positivo, e gostaria de estar mais seguro antes de indicar um exame invasivo. Dentro desta última situação, a chance de um resultado falso-positivo do TE era de 36%, o que poderia ser alto para alguns clínicos. Um teste de estresse com TTA positivo, nesta circunstância, elevaria para 87% a chance de doença coronária, tornando compulsória a continuidade da investigação diagnóstica. Por outro lado, se o TTA tivesse dado um resultado negativo, a chance de DAC no paciente reduzir-se-ia para próximo de 10%. Embora superior aos 5% esperados para indivíduos nessa faixa etária, provavelmente o estudo cineangiocoronariográfico não teria sido indicado pela maioria dos clínicos² (fig. 1).

Este tipo de raciocínio delineado acima é realizado diuturnamente pelos médicos. É provável que muitos não o façam de maneira tão objetiva e considerem outros fatores, seguindo sua intuição. Estará neste momento realizando um ato menos científico, porque não fundamentado em evidências apoiadas em pesquisas; não reproduzível, porque baseado em visão individual ou particular e, possivelmente, não endossável pela maioria dos seus colegas³.

Universidade Federal do Estado de São Paulo - Escola Paulista de Medicina
Correspondência: Bráulio Luna F^o - Laboratório de Cardiologia Experimental -
Disciplina de Cardiologia - UNIFESP-EPM - Rua Leandro Dupret 317 - 04025-011
São Paulo, SP

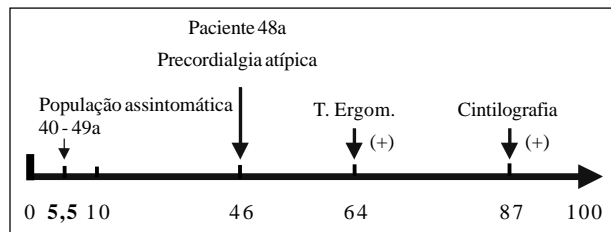


Fig. 1 - Probabilidade de doença arterial coronariana (%).

Princípios básicos dos testes diagnósticos - Fazer diagnóstico é um processo imperfeito que resulta mais em probabilidade de se estar certo do que em certeza. Tanto a prática como a pesquisa clínica, freqüentemente envolvem a avaliação de testes diagnósticos. Por exemplo, a avaliação do nível de proteína C reativa é útil na avaliação prognóstica dos pacientes coronários? Entre os pacientes com hipertensão arterial, a dosagem sérica da renina é útil no diagnóstico de doença renovascular?

Os testes diagnósticos são instrumentos importantes na facilitação das decisões que os médicos são obrigados a realizar como parte intrínseca da sua atividade. Geralmente, obedecem a vários tipos de tomada de decisão, como por exemplo⁴: 1) confirmar a presença de uma doença; 2) avaliar a gravidade do quadro clínico; 3) estimar o prognóstico de uma doença; 4) monitorar a resposta de uma conduta terapêutica.

Lógica interna dos testes diagnósticos - O teste diagnóstico ideal daria sempre respostas corretas - positivo para a presença da doença e negativo para a ausência - seria rápido, seguro, incruento, confiável e barato. Por conseguinte, na prática não existe teste ideal. Os estudos dos testes diagnósticos utilizam desenhos semelhantes aos estudos observacionais, mas seus objetivos e procedimentos estatísticos são diferentes, como veremos.

Dois modelos devem estar na mente do pesquisador quando for analisar ou desenhar um trabalho tipo teste diagnóstico: 1^o) ter como referência de qualidade os estudos tipo ensaio clínico randomizado duplo cego e 2^o) refletir a prática clínica.

Em relação a esses tópicos é necessário avaliar se os pacientes foram escolhidos de forma aleatória para se submeter ao novo teste versus o teste usual? Se os pacientes que realizaram o novo teste tiveram melhor resultado?; o novo teste resultou em menor tempo de internação hospitalar, aumentou a sobrevida ou simplesmente resultou em uma conta hospitalar maior? E, finalmente, o teste foi avaliado em condições similares a sua aplicação no futuro. Estas são indagações que rotineiramente devemos fazer⁵.

Estrutura básica - A estrutura básica dos testes diagnósticos é similar aos outros estudos observacionais, e são desenhados para determinar quão bem um teste discrimina a presença ou ausência de uma doença. Possui uma variável preditora - o resultado do teste - e uma variável resultante - a presença ou ausência da doença.

A variável preditora do teste embora possa ser categórica ou contínua, habitualmente é dicotomizada, definindo os critérios (*cut off*) que confirmam ou refutam a presença da doença. A variável resultante no teste diagnóstico é a presença ou ausência da doença, confirmada por um teste considerado *gold standard*.

Sensibilidade e especificidade (fig. 2) - Quando se avalia um teste diagnóstico, quatro situações são possíveis: a) o teste é positivo e o paciente tem a doença - verdadeiro positivo (VP); b) o teste é positivo mas o paciente não tem a doença - falso positivo (FP); c) o teste é negativo e o paciente tem a doença - falso negativo (FN); d) o teste é negativo e o paciente não tem a doença - verdadeiro negativo (VN). Obviamente, os melhores testes são aqueles que apresentam poucos resultados falso-positivos e falso-negativos.

A avaliação do desempenho de um teste (acurácia) depende da maneira como se sabe que a doença está verdadeiramente presente ou ausente. Para isto, é necessário selecionar o teste que será considerado padrão de referência da verdade - o teste *gold standard*. Deve-se aceitar como verdadeiro o resultado desse teste. Exemplo clássico de teste de referência, aceito sem contestação, é a biópsia na hepatite crônica. Também aceito, mas com alguma restrição, é a análise cinecoronariográfica do grau de obstrução da coronária na angina de peito ou a elevação enzimática da CK-MB no infarto agudo do miocárdio (IAM). Às vezes o pesquisador, por falta de melhor opção, pode utilizar testes de referência questionáveis, como os critérios de Jones para a febre reumática.

Vale destacar que o teste *gold standard* pode ser um exame simples ou complexo, dispendioso, arriscado e, freqüentemente, até não ser verdadeiro. Existem situações em que apenas o acompanhamento do paciente poderá confirmar a presença ou ausência da doença. Outro aspecto não muito estudado é que os médicos geralmente não investi-

		Status da doença		
		Presente	Ausente	Total
Teste	Positivo	Verdadeiro-positivo (A)	Falso Positivo (B)	A+B
	Negativo	Falso-negativo (C)	Verdadeiro-negativo (D)	C+D
Total		A+C	B+D	
Sen = A/A+C		Esp = D/B+D		
Prevalência = A+C/ A+B+C+D		Acurácia = A+D/ A+B+C+D		
VPP = A/A+B		VPN = D/C+D		
RPP = A/A+C		RPN = C/A+C		
B/B+D		D/B+D		

Fig. 2 - Característica e definição dos principais índices utilizados na análise dos testes diagnósticos. SEN- sensibilidade; ESP- especificidade; VPP- valor preditivo positivo; VPN- valor preditivo negativo; RPP- razão de probabilidade positiva; RPN- razão de probabilidade negativa.

gam com muita determinação os resultados negativos dos testes diagnósticos (falso positivo e falso negativo). Por conseguinte, é difícil na literatura precisar qual a frequência e suas principais causas. Quem não se recorda da época em que por falta de critério de verdade adequado houve um exagero no diagnóstico de prolapso da valva mitral?

Como o teste *golden standard* é imperfeito, pode acontecer, na introdução de um novo método, o paradoxo do teste diagnóstico em avaliação apresentar resultado ruim, quando de fato, é melhor. Exemplo clássico deste fenômeno aconteceu na avaliação do ultra-som abdominal no diagnóstico da colelitíase em relação ao colecistograma como exame de referência.

Podemos agora acrescentar que o desempenho dos testes diagnósticos é avaliado pela estimação da sua sensibilidade e especificidade.

A sensibilidade de um teste é definida pela proporção de pessoas com a doença de interesse que têm o resultado do teste positivo. Indica quão bom é o teste para identificar os indivíduos doentes: $Sen = a/a+c$ ou $VP/VP+FN$.

A especificidade de um teste é a proporção de pessoas sem a doença que têm o teste negativo. Indica quão bom é o teste para identificar os indivíduos não doentes: $Esp = d/d+b$ ou $VN/VN+FP$.

Utilizando o exemplo inicial do paciente com dor torácica atípica que se submeteu a vários exames, calculamos a sensibilidade e especificidade do TE para o diagnóstico de doença coronária realizado pela cinecoronariografia naquele contexto clínico, de acordo com a lógica geral dos testes diagnósticos^{4,5} (fig. 3).

Escolha do Limiar diagnóstico do teste

(ponto de corte)

O teste diagnóstico pode apresentar variável categórica, como por exemplo a biópsia miocárdica, com resultado positivo ou negativo para a rejeição do transplante cardíaco, ou contínua, como o infradesnívelamento em mm do segmento ST-T do ECG que positivará o TE. Nesta última circunstância é necessário definir o nível de alteração da variável contínua que positivará o teste. Esta escolha envolve uma decisão entre aumentar a sensibilidade à custa de redução da especificidade ou vice-versa. A maioria dos pesquisadores deve avaliar cuidadosamente a importância rela-

		Cinecoronariografia		
		Presente	Ausente	
Teste Ergom	Positivo	605	340	605 + 340
	Negativo	284	702	284 + 702
	Total	889	1042	1931
Prevalência - $A+C/A+B+C+D = 889/1931 = 46\%$; SEN - $605/889 = 68\%$; ESP - $802/1042 = 77\%$; VPP - $605/945 = 64\%$; VPN - $702/986 = 71\%$; RPP - $605/889 + 340/1042 = 2,1$; RPN - $284/889 + 702/1042 = 0,48$.				

Fig. 3 - Análise do desempenho do teste ergométrico comparada com a cinecoronariografia. Diamond GA, Forrester JS².

tiva da sensibilidade e especificidade do teste para estabelecer o ponto de transição diagnóstica mais adequado, e cuja estratégia geral seria⁴: a) se a principal preocupação é evitar resultado falso-positivo (o resultado do teste pode indicar uma cirurgia arriscada para o paciente), então o ponto de corte deve objetivar o máximo de especificidade; b) se a preocupação maior é evitar resultado falso-negativo (o resultado do teste em suspeito de AIDS), então o ponto de corte deve objetivar o máximo de sensibilidade.

Em relação ao TE, a maioria dos autores prefere empregar o infradesnívelamento de 1,0mm do ST-T⁶. Conseqüentemente, uma diminuição deste limiar de positividade do TE para 0,5mm aumentará a sensibilidade e diminuirá significativamente a especificidade. O oposto acontecerá se houver incremento do limiar para 2,0mm. Como então garantir a melhor escolha do ponto de viragem para positivar um teste?

Curva ROC - A curva ROC (*receive operator characteristic curve*) é a melhor maneira de estabelecer o ponto de corte, otimizando a sensibilidade e especificidade do teste diagnóstico. O pesquisador deverá selecionar vários pontos ou níveis de alteração do teste e determinar a sensibilidade e especificidade em cada ponto. Ele então construirá um gráfico da sensibilidade em função da proporção dos resultados falso-positivos. O teste ideal é aquele que alcança a extremidade mais superior e esquerda do gráfico. Uma das vantagens deste método é que as curvas de diferentes testes diagnósticos podem ser comparadas; quanto melhor o teste mais perto estará sua curva do canto superior esquerdo do gráfico⁴ (fig. 4).

Prevalência, probabilidade prévia e valores preditivos

- O valor do teste diagnóstico depende não apenas da sensibilidade e especificidade, mas também, da prevalência da doença na população onde o mesmo está sendo testado. Assim, quando mais rara a doença mais específico deve ser o teste para ser útil. Ao contrário, se uma doença é frequente (doença coronária em indivíduos de meia-idade, fumante e com angina típica), o teste deve ser muito sensível para poder ser útil ao clínico. De outra maneira, o resultado negativo expressará um resultado falso-negativo.

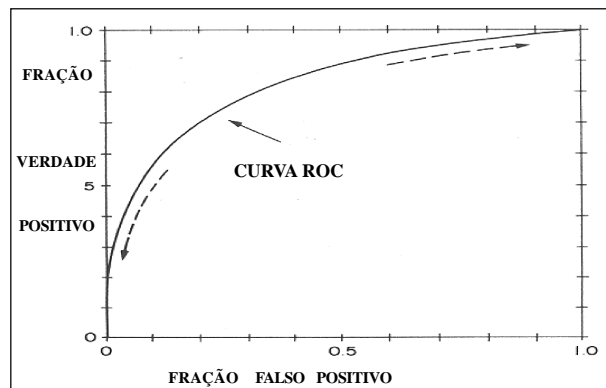


Fig. 4 - Curva ROC típica. O teste ideal é aquele que alcança o quadrante mais superior e esquerdo. (sensibilidade e especificidade de 100%)

Em cada paciente, a prevalência da doença é geralmente designada de probabilidade prévia que é baseada em dados demográficos e nas características clínicas que um doente, em particular, deve ter da doença (prevalência), e deve ser estimada antes da realização do teste diagnóstico. Por exemplo, a probabilidade prévia de DAC pode ser muito baixa, cerca de 1% em jovens recrutas do exército, mas muito alta, cerca de 90%, em pessoas de meia-idade, fumantes, hipertensos e com angina típica.

A relação entre sensibilidade e especificidade do teste diagnóstico e a probabilidade prévia de uma doença pode ser expressa formalmente pelo teorema de Bayes⁷. Após a decisão de realizar um teste, o dado mais importante para o clínico é estimar se o resultado do mesmo é verdadeiro. Esta informação, o valor preditivo positivo e negativo, é facilmente obtida através da mesma tabela em que calculamos a sensibilidade e especificidade (fig. 3).

O valor preditivo positivo (VPP) do teste diagnóstico é, portanto, a probabilidade de uma pessoa com um resultado positivo ter a doença. - $VPP = A/A+B$.

O valor preditivo negativo (VPN) do teste é a probabilidade de uma pessoa com resultado negativo não ter a doença. - $VPN = D/C+D$.

Valor preditivo é também chamado de probabilidade posterior. Como incorpora informação, tanto do teste como sobre a população em avaliação, o valor preditivo é uma boa medida da utilidade clínica global do teste. Por exemplo, assumindo uma sensibilidade e especificidade de 90% para DAC pelo TTA, o significado de um teste positivo dependerá da probabilidade prévia da DAC na pessoa em avaliação. Para probabilidade prévia de 1% (jovem recruta do exército), o VPP será de 8%. Assim, é muito provável que o teste positivo seja um resultado falso-positivo. Por outro lado, para probabilidade prévia de DAC de 90% (um paciente masculino de meia-idade com história típica de angina e hipertenso), o teste positivo resultará em probabilidade DAC de 99%. Entretanto, um teste negativo não excluirá a DAC, porque o mesmo paciente terá ainda chance de 50% de ter a doença^{2,4} (tab. I).

Assim enfatizamos que o principal determinante da estimativa do resultado de um teste é a prevalência ou pro-

babilidade prévia da doença em questão. Mesmo que o teste diagnóstico tenha altos valores de sensibilidade e especificidade, se a prevalência da doença naquele paciente for baixa, o resultado positivo do teste terá pouco valor, porque terá grande chance de ser um resultado falso-positivo. Por outro lado, se a prevalência da doença for alta e o resultado do teste vier negativo haverá grande chance de ser falso-negativo. Portanto, a grande utilidade dos testes diagnósticos é naquelas situações de prevalência intermediária (entre 25 e 65%)^{5,8}.

Razão de probabilidade (likelihood ratios) - Esta é uma outra maneira de avaliar o resultado de um teste diagnóstico. Razão de probabilidade (RP) é simplesmente a probabilidade de pessoas com uma doença ter um determinado resultado no teste, dividido pela probabilidade de pessoas sem a doença, também, terem o mesmo resultado^{4,5}.

Quando combinada com informação sobre a probabilidade prévia de uma doença, a RP pode ser utilizada para determinar o valor preditivo do resultado de um teste, o que requer expressar a probabilidade prévia em chance prévia daquela doença. As chances (*odds*) de uma doença são simplesmente a razão da proporção de pessoas com a doença pela proporção das sem doença. ($odds = p/1-p$). Por exemplo, a probabilidade prévia de 25% equivale a *odds* 1:3 (25% ÷ 75%). *Odds* de 5:1 equivalem probabilidade de 5/6 = 83%.

As chances prévias (*prior odds*) de uma doença, quando multiplicadas pela RP, determinam a chance posterior (posterior *odds*) de uma doença. Assim, a chance prévia de IAM em paciente masculino de 45 anos com queixa de dor precordial, por mais de 1h, é de 4:1 (probabilidade prévia de 80%). Considerando hipoteticamente, por exemplo, que a dosagem de asparaginase kinase (AK) foi de 150UI/dl., multiplicando a chance prévia (4:1) pela RP para IAM deste nível de AK de 5:4 (tab. II), obtemos uma chance posterior de 20:4 ou 5:1 (83% de probabilidade).

Vantagens da razão de probabilidade - 1) Permite expressar o resultado do teste diagnóstico em vários níveis de corte; 2) as RP obtidas em seqüência de vários testes podem ser multiplicadas entre si, desde que os testes sejam independentes, originando uma RP global para o conjunto dos resultados dos vários testes.

Limitações dos testes diagnósticos - Como tantos outros estudos observacionais, também os testes diagnósticos são susceptíveis a erros randômicos e sistemáticos.

Tabela I - Valor preditivo do teste com sensibilidade e especificidade de 90% variando a probabilidade prévia da doença (prevalência).

Probabilidade Prévia da Doença	Valor Preditivo Positivo	Valor Preditivo Negativo
0,001	0,01	0,9999
0,01	0,08	0,999
0,05	0,32	0,994
0,10	0,50	0,99
0,20	0,69	0,97
0,50	0,90	0,90
0,80	0,97	0,69
0,90	0,99	0,50
0,95	0,994	0,32
0,99	0,999	0,08
0,999	0,9999	0,01

Tabela II - Razão de probabilidade para o nível da asparaginase kinase sérica no diagnóstico do infarto agudo do miocárdio (IAM).

Asparaginase (UI/dl)	Kinase IAM	Pacientes Sem IAM	Razão de Probabilidade
0 - 99	15	150	15/100+150/200 = 1:5
100 - 199	25	40	25/100+40/200 = 5:4
> 200	60	10	60/100+10/200 = 12:1
Total	100	200	

Erros randômicos - Alguns pacientes com a doença terão resultados normais no teste diagnóstico, apenas por uma questão de chance. Este tipo de erro, embora inevitável, é quantificável. A maneira de estimá-lo é construir intervalos de confiança (IC) para os valores da SEN, ESP, VPP e VPN do teste diagnóstico. O IC permite ao leitor ver a variação dos valores em relação aos resultados relatados e compará-los com a variação apresentada pelos outros testes diagnósticos. Por exemplo, um teste aplicado em 100 pessoas apresenta SEN de 80% e ESP de 70%. Um novo teste diagnóstico apresentou resultado positivo em todos os cinco indivíduos doentes da amostra (SEN de 100%) e foi negativo em 9 dos 10 indivíduos sem a doença (ESP de 90%). Apesar de expressivo esses resultados foram obtidos em uma amostra pequena. Isto significa que para os IC de 95%, a SEN e ESP desse novo teste são, respectivamente, de 57% a 100% e 60% a 98%. Estes valores sobrepõem-se aos dos antigos testes. Assim não é tão claro que o novo teste represente verdadeiramente um avanço, e poderá mesmo ser pior^{5,8}.

Este problema poderia ter sido evitado, estimando-se previamente o tamanho da amostra necessária para atingir o nível desejado de sensibilidade e especificidade.

Erros sistemáticos - Os mais comuns são os vieses de amostragem, de mensuração e de publicação⁷. O **vies de amostragem** ocorre quando a amostra estudada não é representativa da população alvo no qual o teste será aplicado. Frequentemente, esses estudos são realizados em centros terciários de referência que concentram amostras de casos mais graves. Isto faz com que a sensibilidade do teste seja sobrestimada. Similarmente, se a população em que é testado a especificidade é constituída de indivíduos voluntários normais, há exagero nos valores obtidos para a especificidade. Sabe-se que indivíduos voluntários tendem a ser mais normais que o restante da população. A melhor estratégia para lidar com esses problemas é aplicar o teste em populações semelhantes àquelas em que o mesmo deverá ser empregado. Outro efeito da amostragem é que a prevalência da doença na amostra estudada geralmente é muito mais alta que a encontrada na prática clínica, o que tende a sobrestimar os valores preditos do teste em questão.

Vies de mensuração acontece quando o examinador conhece o diagnóstico do caso, quando está analisando o resultado do teste. É particularmente importante naquelas situações *bordeline*, quando há dúvida na interpretação do teste. A melhor estratégia é realizar a interpretação do teste "cego", tanto para o estado do paciente quanto para o resultado do teste de referência.

Vies de publicação - Os teste que não apresentam resultados promissores, geralmente, não são publicados. Alternativamente, os autores devem realizar testes diagnósticos com amostras representativas. Assim, mesmo negativo o resultado poderá ser significativo e ter mais chance de ser aceito para publicação.

Conclusão

Finalizando, talvez em resposta à maldição de Lorde Kelvin que afirmou não levar a sério a medicina porque não acreditava em ciência que não conseguia medir seus fenômenos, os médicos utilizam, cada vez mais, do raciocínio probabilístico na sua prática diária. Muito em breve, a continuar a velocidade de incorporação das técnicas estatísticas e metodológicas na medicina, estaremos quiçá rivalizando com a física de partículas atômicas, ramo mais avançado do conhecimento, que tem na essência do seu corpo teórico as leis da probabilidade estatística.

Descartado o exagero e independente da evolução que o futuro nos descortina, é indiscutível que a medicina vem se tornando, cada vez mais, em um empreendimento científico que exige dos seus cultores, o compromisso de integrar com eficiência as qualidades individuais dos médicos, responsáveis pela leitura adequada das queixas e manifestações dos pacientes, com a melhor informação disponível à situação em questão. Como ainda não existem regras definitivas nem leis para isto, o bom julgamento clínico continuará ainda por um longo tempo, orientando e humanizando a prática médica.

A explosão de novos e dispendiosos testes diagnósticos exige dos médicos, além do bom senso, que aprendam e utilizem de forma eficiente a estratégia sumarizada no quadro I. Prevalência, sensibilidade, especificidade, valores preditivos e razões de probabilidades são instrumentos úteis na avaliação das dezenas de testes diagnósticos que aparecem diariamente nas revistas e consultórios médicos. Ao clínico compete decidir se vale apenas usá-los ou não. Nessa análise, os benefícios clínicos devem ser cotejados com os riscos médicos, os ônus econômicos e as vantagens e desvantagens em relação aos outros exames. Embora não seja uma tarefa fácil, é dever do médico conhecer as técnicas de validação e interpretação dos testes diagnósticos para decidir, em bases científicas, sua real utilidade.

Quadro I - Planejamento do estudo do teste diagnóstico

- 1) O investigador deverá considerar a real necessidade de um novo teste diagnóstico.
- 2) Descrever a maneira como os indivíduos serão selecionados.
- 3) Escolher adequadamente o teste *golden Standard* que será utilizado do como referência pelo teste em avaliação.
- 4) Assegurar que tanto o teste *golden standard* quanto o teste em avaliação sejam avaliados de maneira padronizada e às "cegas" quanto ao resultado um do outro.
- 5) O investigador deverá estimar o tamanho da amostra necessária para atingir um intervalo de confiança de 95% com razoável precisão, tanto para a sensibilidade quanto para a especificidade.
- 6) O investigador deve avaliar as condições de recrutar o número de indivíduos estimado pelo cálculo do tamanho de amostra.
- 7) O investigador deverá planejar o resultado do estudo expressando a sensibilidade, a especificidade e os valores preditivos positivos e negativo sem relação a diferentes valores de prevalência da doença. Se possível deverá considerar o uso da curva ROC e da razão de probabilidade para descrever o desempenho do teste diagnóstico.

Referências

1. Osler W - Medical Education. In: Counsels and Ideals, 2nd ed. (cidade ???): Houghton Mifflin, 1921.
2. 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.
3. Patterson RE, Horowitz SF - Importance of epidemiology and biostatistics in deciding clinical strategies for using diagnostic tests: a simplified approach using examples from coronary artery disease. J Am Coll Cardiol 1989; 13: 1653-65.
4. Sackett DL, Haynes RB, Tugwell P - Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd ed. Boston: Little, Brown, 1991: 370.
5. Sackett DL - A primer on precision and accuracy of the clinical examination - JAMA 1992; 267: 2638-44.
6. Gianrossi R, Detrano R, Mulvihill D et al - Exercise-induced ST depression in the diagnosis of coronary artery disease: a meta-analysis. Circulation 1989; 80: 87-95.
7. Wagner HN - Bayes's theorem: an idea whose time has come? Am J Cardiol 1982; 49: 875-7.
8. Ransohoff DF, Feinstein AR - Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299: 926-30.