

AVALIAÇÃO ESTATÍSTICA DOS TESTES DIAGNÓSTICO EM MEDICINA (Um enfoque heurístico) *

Eilson Goes de Oliveira**

RESUMO

O autor apresenta uma revisão dos conceitos de sensibilidade e especificidade de testes diagnósticos na prática médica, através de um procedimento construtivista, que exige, para compreensão pelo leitor não familiarizado, um mínimo de bagagem matemática e estatística.

ABSTRACT

The author presents a review on concepts of sensitivity and specificity applied to diagnostic tests on the current medical practice towards a constructive approach; for this it is required, in the case of readers not familiar with the topic, a minimum of mathematics and statistical background.

1. INTRODUÇÃO

Os testes laboratoriais utilizados em Medicina para diagnóstico de doenças ou de condições fisiológicas (como gravidez) são possíveis de fornecerem resultados errôneos, em decorrência de, basicamente, três fatores.

- in experiência daquele que executa, lê e interpreta o dócima;
- defeitos no material e/ou instrumental utilizados;
- resposta paradoxal do organismo do paciente ao procedimento diagnóstico.

Veamos alguns exemplos ilustrativos, destinados aos ainda pouco familiarizados com a questão: um indivíduo reconhecidamente aidético, em fase terminal, com a doença manifesta na sua trágica exuberância (cânceres e infecções generalizadas) pode ser HIV negativo, quando submetido a exames laboratoriais baseados em presença de

anticorpos, pois o sistema imunológico, profundamente abalado, não mais responde a estímulos; um paciente sem nenhuma possibilidade (clínica e laboratorial) de ser portador da doença de Chagas, e capaz de responder positivamente ao teste de Machado-Guerreiro, desde que apresente uma enfermidade propiciadora de reações cruzadas (hanseníase, v.g.). No primeiro exemplo, surgiu um resultado dito **falso-negativo** (incapacidade do teste de detectar a doença); no segundo, um **falso-positivo** (falha do teste em revelar a ausência da doença).

Esquemáticamente, podemos resumir o que foi dito no esquema abaixo.

		DOENÇA	
TEST	PRESENTE	AUSENTE	
+	COERÊNCIA	FALSO-POSITIVIDADE	
-	FALSO-NEGATIVIDADE	COERÊNCIA	

* Extrato de conferência pronunciada pelo autor na Fundação Gonçalo Moniz, Bahia

** Professor organizador das disciplina Estatística Aplicada à Saúde da UNIFOR; Professor visitante de Bioestatística nos Cursos de Mestrado e Doutorado em Patologia Humana da UFBa; Professor de Bioestatística do Curso de Mestrado em Patologia Tropical da UFC.

O ideal seria que todos os resultados fossem coerentes, mas isso só é proporcionado por raros testes; além do mais, é medida de cautela o pesquisador sempre considerar a possibilidade, remota que seja, de estar diante de um resultado falso. Torna-se imperioso, contudo, estabelecer um d^ocima padrão, o chamado "gold test", aquele com menor risco de levar a falsas conclusões; no caso do câncer, por exemplo, o exame histopatológico é o teste de ouro.

O problema que se estabelece, agora, é graduar os recursos diagnóstico disponíveis para detecção de determinada patologia, em relação à sua capacidade, maior ou menor, de "escapar" às armadilhas de resultados incoerentes. É o que nos propomos apresentar, mediante um tratamento estatístico singelo, despidido do formalismo da análise bayseana, "batendo a porta da boa vontade do paciente leitor, mas nunca com muita força".

2. SENSIBILIDADE E ESPECIFICIDADE

Os dois termos que intitulam esse item referem-se às medidas estatística básicas da qualidade de um teste diagnóstico.

Sensibilidade (S) é estimada pela frequência relativa de resultados **positivos** dentre um conjunto de reais portadores da doença (ou condições); por definição, varia de 0 a 100%. Exemplificando: se a ultra-sonografia permite um diagnóstico correto do sexo da criança em 190 casos sobre um total 200 grávidas, estima-se que esse método tenha sensibilidade 95% (equivalente à fração 190/200) para caracterização intrauterina da referida característica fetal (sexo).

Outra ilustração: afirmar que o exame W-B é 99% sensível para a AIDS significa que, em média, para cada 100 pacientes efetivamente aidéticos, 99 serão diagnosticados como tal pelo W-B (sigla designando o termo Western-Blot), apenas 1 (um) caso escapando à detecção, ou seja, única presença de falso-negativo.

Evidentemente, um teste de alta sensibilidade nos protege contra a ocorrência de resultados negativos errôneos.

Especificidade (E) é estimada pela frequência relativa de resultados **negativos** dentre um conjunto de indivíduos **isentos** da doença (ou condição) estudada.

Exemplificando: se a dosagem de antígeno

carcinoembrionário resultar negativa em 180 de 300 pacientes **sem** câncer de cólon, a especificidade do referido teste em relação ao diagnóstico de câncer colônico é 60% (180/300), valor muito baixo, que limita esse recurso ao acompanhamento, nunca ao diagnóstico definitivo da doença; realmente, na questão ilustrativa, ocorreriam, em média, 40 casos (100-60) rotulados erroneamente como câncer (falsos-positivos) para cada 100 pessoas examinadas.

A especificidade elevada nos protege contra falso-positividade. É de conveniência lembrar que dois atributos considerados são independentes, isto é alta sensibilidade não implica em alta especificidade e vice-versa.

No item posterior, examinaremos uma situação com dados hipotéticos, à guisa de diretriz para uma análise mais detalhada. Os pormenores do delineamento experimental são omitidos, pois fogem ao escopo desse trabalho.

3. TESTES PARA DIAGNÓSTICO DA LEISHMANÍOSE VISCERAL (CALAZAR)

Suponhamos que 180 pacientes em área, endêmica de calazar tenham sido divididos em dois grupos: um, constando de 80 portadores de leishmaniose; outro, de 100 indivíduos não infectados. A caracterização da presença ou ausência da infecção foi estabelecida por um "gold test", o exame imunológico conhecido pela sigla DAT. Amostras de sangue foram colhidas de todos os indivíduos, neles se efetuando três testes: o DAT, o ELISA e o IFI (imunofluorescência indireta).

As tabelas I e II exibem os resultados obtidos:

Tabela I

ELISA	Leishmaniose		
	SIM	NÃO	TOTAL
+	75(a)	5(b)	80(m1)
-	5(c)	95(d)	100(m2)
TOTAL	80(n1)	100(n2)	180(n)

Tabela II

IFI	Leishmaniose		
	SIM	NÃO	TOTAL
+	75(a)	5(b)	80(m1)
-	5(c)	95(d)	100(m2)
TOTAL	80(n1)	100(n2)	180(n)

Obs.: As letras entre parênteses simbolizam um valor genérico.

De início, calculemos a sensibilidade (S) e a especificidade (E) para cada teste:

$$A) \text{ ELISA: } S = \frac{a}{n_1} = \frac{75}{80} \cong 94\%$$

$$E = \frac{d}{n_2} = \frac{95}{100} = 95\%$$

$$\text{Validade do teste*} \cdot V = S + E = 189\%.$$

$$B) \text{ IFI: } S = \frac{a}{n_1} = \frac{72}{80} = 90\%$$

$$E = \frac{d}{n_2} = \frac{98}{100} = 98\%$$

$$V = 188\%$$

A validade de cada método é analisada pelo teste usual do χ^2 (qui-quadrado), com um grau de liberdade, já que se trata de comparação de amostras independentes. A expressão do χ^2 observado é dada por:

$$\frac{\chi^2 n (ad - bc)^2}{m_1 \cdot m_2 \cdot n_1 \cdot n_2}$$

O valor crítico tabulado a nível de significância 5%, $\chi^2 = 3,84$. Analisemos, então, a significância da validade de cada recurso diagnóstico:

$$\begin{aligned} \chi^2 - \text{Para o ELISA:} &= 141,72 > 3,84 \\ & \frac{180(75 \times 95 - 5 \times 5)^2}{80 \times 100 \times 80 \times 100} \quad (p > 0,05) \end{aligned}$$

Conclusão: Validade altamente significativa.

$$\text{Para o IFI:} \quad = 142,16 > 3,84$$

$$\chi^2 = \frac{180 (72 \times 98 - 8 \times 2)^2}{80 \times 100 \times 106 \times 74} \quad p < 0,05$$

Conclusão: Validade altamente significativa.

Além da verificação global da utilidade do método, é possível examinar a significância estatística da sensibilidade e da especificidade, consideradas de per se. Para tanto, deve-se partir da consideração de que só é útil um teste de $S > 50\%$, do contrário os resultados equivaleriam aos de uma mera adivinhação ("cara ou coroa, por

exemplo); idêntico raciocínio aplica-se à especificidade. Naturalmente, para testar a significância da **sensibilidade** só se usam os pacientes com a doença; para a **especificidade**, só aqueles sem a doença (no nosso caso, leishmaniose).

Agora, o χ^2 se calcula pela expressão clássica: em que: o = valores observados

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

e = valores esperados segundo a hipótese de que $S = 50\%$ ou $E = 50\%$, conforme o caso.

Assim, para a sensibilidade do ELISA:

ELISA	0	e	
+	75	$50\% \times 80 = 40$	$\chi^2 = \frac{(75-40)^2}{40} + \frac{(5-40)^2}{40} = \frac{1225+1225}{40}$
-	5	$50\% \times 80 = 40$	
TOTAL	80	80	$\chi^2 = \frac{2450}{40} = 61,25 > 3,84 (p < 0,05)$

Conclusão: Exame com sensibilidade altamente significativa.

Agora, em relação à especificidade:

ELISA	0	e	
+	5	50	$\chi^2 = \frac{(5-50)^2}{50} + \frac{(95-50)^2}{50} = 81 > 3,84 (p < 0,05)$
-	95	50	
TOTAL	100	100	

Conclusão: Exame com sensibilidade altamente significativa.

Deixamos ao cargo do leitor a análise da significância da sensibilidade e da especificidade, isoladas, do IFI. Adiantamos que, a semelhança do ELISA, são altamente significativas.

Em resumo, ambos os testes são bastante válidos para o diagnóstico da leishmaniose visceral, segundo os dados hipotéticos apresentados; alguns trabalhos confirmam essa conclusão. Destarte, a escolha de um ou outro método depende do custo, principalmente quando da realização de exames em massa, para fins epidemiológicos.

4. PERSPECTIVAS

A análise estatística da qualidade dos testes diagnóstico em Medicina, seguramente não termina aqui, dentro dos limites de um trabalho que se propõe, fundamentalmente, a ser simples e objetivo, sujeito a uma linha construtivista de comunicação, imposta pela natureza do autor, um autodidata esforçado.

Exemplos de horizontes a serem descortinados na área estudada:

A) correção de taxas de prevalência de doenças, levando em conta a possibilidade de diagnósticos errôneos (falsos positivos e falsos negativos);

B) avaliação do ganho na capacidade de acerto diagnóstico, em função da qualidade do recurso laboratorial utilizado (valor produtivo dos testes);

C) confronto entre resultado de diferentes procedimentos diagnósticos para amostras emparelhadas;

D) sinergismo de recursos diagnósticos.

Mas aí são outros "casos"... que outras pessoas, mais peritas, haverão de contar ao paciente leitor dessas maltraçadas linhas.

BIBLIOGRAFIA RECOMENDADA

1. EL SAFI, S.H. et alii - **A comparison of the direct agglutination test and enzyme - linked immunosorbent assay (ELISA) in the sero-diagnosis of leishmaniasis in the Sudan.** Transactions of the Royal Society of Tropical Medicine and Hygiene (1989): 83, 334-337.
2. IPSEN and FEIGL - **BANCROFT'S INTRODUCTION TO BIOSTATISTICS** - 2nd. Edition - Harper and Row - New York - 1970.
3. KORTE, P. M. et alii - **Introduction of an improved direct agglutination test (DAT) for the detection of Leishmania infantum infection in Southern France.** Parasitology Research 1990, 76: 526-530.