

## CAPÍTULO 4

# Parâmetros e Controle da Qualidade de Imunoensaios

Antonio Walter Ferreira  
Sandra do Lago Moraes  
Sandra Trevisan Beck

### PRECISÃO DIAGNÓSTICA E VALOR CLÍNICO DOS TESTES DE LABORATÓRIO

Neste capítulo vamos apresentar e discutir parâmetros intrínsecos e extrínsecos de testes de laboratório, especialmente dos imunoensaios, que serão úteis para a compreensão do real valor dos resultados obtidos quando associados a dados clínicos e epidemiológicos do paciente.

O resultado de um teste de laboratório é um resultado de probabilidade que reflete a situação clínica do paciente no momento da colheita da amostra, e quando bem avaliado representa importante ferramenta de auxílio de que o clínico dispõe para a tomada de decisão ou para a mudança de uma hipótese clínica inicial. Para que isso seja verdade, é necessário saber interpretar os resultados em função dos limites do teste utilizado. Para o clínico, interessa saber com que frequência o teste utilizado é positivo em pacientes doentes comparado com a frequência com que o teste é negativo nos indivíduos não-doentes. Nesse sentido, vamos discutir neste capítulo:

#### *Gold Standard Test* (Teste de Referência)

Sensibilidade

Especificidade

Eficiência

Reprodutibilidade

Índice Kappa

Prevalência

Valor Preditivo Positivo

Valor Preditivo Negativo

Teorema de Bayes e Probabilidade

Curva ROC (*Receiver Operating Characteristic*)

### *GOLD STANDARD TEST* (TESTE DE REFERÊNCIA)

Teste ou procedimento que é utilizado para definir o verdadeiro estado do paciente.

O teste de referência deve ser definitivo e independente do teste que está sendo avaliado. Para ser considerado definitivo, deverá atingir a real causa da doença tão diretamente quanto possível. Para ser considerado independente, não deverá ser parte de um algoritmo que inclua também o teste em avaliação e não ser tecnicamente ou biologicamente relacionado com o teste em avaliação. Por exemplo, se o teste em avaliação é um teste imunoenzimático que diagnostica a infecção chagásica pela presença de anticorpos anti-*T. cruzi* no soro, não podemos utilizar um outro teste imunoenzimático de fabricante diferente como teste de referência. O teste de referência para essa patologia deveria ser a pesquisa de *T. cruzi* no sangue dos pacientes (demonstração direta, xenodiagnóstico, cultivo) ou pesquisa de RNA/DNA por tecnologia de amplificação de ácidos nucleicos.

Também é importante lembrar que para definirmos teste de referência é preciso o conhecimento da evolução clínica da doença e dos efeitos patológicos provocados pelo patógeno. Por exemplo, na fase crônica da doença de Chagas, a sensibilidade dos métodos parasitológicos e moleculares é baixa, em função do nível de parasitas que circulam. Nesse caso, o teste de referência passa a apresentar índices de sensibilidade menores que a pesquisa de anticorpos anti *T. cruzi* presentes no sangue do paciente.

Para estudarmos os diferentes parâmetros sorológicos, vamos utilizar o **Quadro 4.1**, de dupla entrada, no qual relacionamos a probabilidade de doença com resultados obtidos



**QUADRO 4.1** Combinação binária de resultados entre o diagnóstico verdadeiro de doença e a probabilidade de resultados do teste

Teste	Doença		Total
	Presente	Ausente	
Reagente	(a) ou VP	(b) ou FP	(a + b) ou VP + FP
Não-reagente	(c) ou FN	(d) ou VN	(c + d) ou FN + VN
Total	(a + c) ou VP + FN	(b + d) ou FP + VN	(a + b + c + d) ou n

VP, FP, VN e FN são, respectivamente, os números de verdadeiro-positivos, falso-positivos, verdadeiro-negativos e falso-negativos;

VP+FN e FP+VN são o número de pacientes com e sem a doença, respectivamente;

VP+FP e FN+VN são o número de pacientes com e sem resultado positivo no teste, respectivamente;

n é o número total de pacientes estudados.

## SENSIBILIDADE

Proporção de todos os pacientes com a doença que apresentam resultados positivos quando o teste em particular é utilizado. A sensibilidade pode ser calculada como o número de verdadeiros resultados positivos dividido pelo número de todos os pacientes com a doença [verdadeiro-positivos (VP) e falso-negativos (FN)]. Isso significaria a taxa de verdadeiro-positivos ou a verossimilhança (*likelihood*) do resultado positivo em pessoa doente. O índice de sensibilidade pode ser expresso em porcentagem (%), bastando multiplicar o valor obtido por 100.

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

A definição da sensibilidade apresentada refere-se à sensibilidade clínica, que difere da **sensibilidade analítica**, que está relacionada à capacidade da técnica em **quantificar** concentrações de analitos, anticorpos ou antígenos.

## ESPECIFICIDADE

Proporção de todos os indivíduos sem a doença que apresentem resultados negativos quando o teste em particular é utilizado. A especificidade pode ser calculada como o número de verdadeiros resultados negativos dividido pelo número de todos os indivíduos sem a doença [verdadeiro-negativos (VN) e falso-positivos (FP)]. Isso significaria a taxa de verdadeiro-negativos ou a verossimilhança (*likelihood*) do resultado não-reagente em pessoa não-doente. O índice de especificidade pode ser expresso em porcentagem (%), bastando multiplicar o valor obtido por 100.

$$\text{Especificidade} = \frac{VN}{VN + FP}$$

A definição da especificidade apresentada refere-se à especificidade clínica, que difere da especificidade técnica.

A especificidade do teste pode ser influenciada por inúmeros fatores que levam a falsos resultados positivos. Anticorpos naturais e heteroanticorpos, normalmente imunoglobulinas IgM contra diferentes epítopos, são responsáveis por falsos resultados positivos no teste de hemaglutinação para a doença de Chagas. Condutas laboratoriais, como o uso de 2-mercaptoetanol em concentração adequada, eliminam anticorpos IgM sem prejuízo na detecção de anticorpos IgG. Indivíduos polinfestados por parasitas intestinais apresentam um somatório de componentes antigênicos, elicitando a formação de anticorpos que reagem com inúmeros antígenos alvo dos testes sorológicos em níveis baixos. Isso obriga os pesquisadores a adequarem limiares de reatividade ou a absorverem os soros antes de serem processados.

Atualmente, para melhorar a especificidade dos testes, antígenos de microrganismos obtidos por síntese de peptídeos ou recombinação genética têm sido utilizados (*ver Cap. 3*). Por outro lado, esse procedimento normalmente leva a uma perda na sensibilidade dos testes.

A dificuldade na obtenção de um painel de soros confiável para que sejam estabelecidos corretamente os valores de sensibilidade e especificidade leva muitos laboratórios a preparar painéis de soros a partir de resultados sorológicos de outros testes. Nesse caso, a sensibilidade e a especificidade do teste em estudo serão influenciadas pelos mesmos índices do teste escolhido como referência. Para diferenciarmos dos parâmetros verdadeiros, chamamos essa sensibilidade de co-positividade ou sensibilidade relativa e a especificidade, de co-negatividade ou especificidade relativa.

## EFICIÊNCIA, ACURÁCIA OU PRECISÃO DIAGNÓSTICA

Refere-se à relação entre o número de resultados corretos [verdadeiro-positivos (VP) e verdadeiro-negativos (VN)] no teste e o total de indivíduos testados (N), isto é, a proporção de diagnósticos corretos. É a relação entre os verdadeiro-positivos somados aos verdadeiro-negativos e o total de indivíduos testados. Usando os dados do **Quadro 4.1**, podemos estabelecer a precisão do teste sorológico pela relação expressa:

$$\text{Eficiência, Acurácia ou Precisão diagnóstica} = \frac{VP + VN}{N}$$

## REPRODUTIBILIDADE

A reprodutibilidade de um teste é definida como a obtenção de resultados iguais em testes do mesmo formato realizados com a mesma amostra biológica por diferentes técnicos em diferentes locais.

A reprodutibilidade de resultados é influenciada por diferentes fatores, que variam desde a qualidade dos reagentes às condições técnicas e operacionais dos laboratórios. Equipamentos mal calibrados, principalmente pipetas, desgaste de equipamentos e acessórios, como lâmpadas de microscópios e espectrofotômetros, são responsáveis por falhas na reprodutibilidade dos resultados. Atualmente grande parte dos laboratórios possui certificação ISO (International Organization for Standardization) de qualidade, o que garante a boa qualidade da metrologia no laboratório e critérios de compra de reagentes, previamente avaliados por sorotecia



confiável. A introdução de soros de referência como rotina diagnóstica auxilia na detecção da ocorrência de erros sistemáticos (avaliação da exatidão) ou acidentais (avaliação da precisão).

A reprodutibilidade dos resultados é fortemente influenciada por erros humanos, principalmente quando o laboratorista inventa variações técnicas que estão em desacordo com os procedimentos descritos pelos fabricantes.

Existem várias opções para avaliação de testes, reagentes e serviços. Normalmente utilizam-se a avaliação da reprodutibilidade intratestes (repetitividade), que é a obtenção do mesmo resultado por ensaios realizados, ao mesmo tempo, em duplicatas ou triplicatas, e a reprodutibilidade intertestes, que é a repetição de resultados da mesma amostra processada em dias diferentes pelo mesmo teste.

A variação da reprodutibilidade pode ser medida pelo desvio padrão (DP) ou pelo coeficiente de variação (CV), expressos pelas fórmulas:

$$DP = \frac{\sqrt{\sum (X - \mu x)^2}}{N - 1}$$

$$CV = \frac{DP \times 100}{\mu x}$$

em que:

X é o valor encontrado para cada ensaio,

$\mu x$  é o valor médio,

N é o número total de ensaios realizados.

### ÍNDICE KAPPA ( $\kappa$ )

Mede a avaliação da reprodutibilidade pelo grau e concordância entre os resultados de dois ou mais observadores. O índice Kappa leva em consideração as proporções das concordâncias observadas (PO) em relação às concordâncias esperadas (PE). O índice Kappa varia de valores negativos até 1.

O **Quadro 4.2** mostra como pode ser calculado o índice Kappa a partir dos resultados obtidos pelos observadores, e o **Quadro 4.3** apresenta a classificação do índice Kappa.

$$PO = \frac{(a + d)}{N}$$

$$PE = \frac{[(a + b)(a + c)] + [(c + d)(b + d)]}{N^2}$$

$$\kappa = \frac{(PO - PE)}{(1 - PE)}$$

**QUADRO 4.2** Combinação binária de resultados obtidos para um mesmo acontecimento entre dois observadores

Observador 1	Observador 2		Total
	Positivo	Negativo	
Positivo	a	b	a + b
Negativo	c	d	c + d
Total	a + c	b + d	N

**QUADRO 4.3** Classificação do índice Kappa

Valor de kappa	Concordância
0,00-0,20	Ruim
0,21-0,40	Fraca
0,41-0,60	Moderada
0,61-0,80	Substancial
0,81-1,00	Quase perfeita

### PREVALÊNCIA

Prevalência: número de casos de uma doença numa determinada localidade em tempo determinado. Utilizando o **Quadro 4.1**:

$$\text{Prevalência} = \frac{(VP + FN)}{N}$$

Prevalência sorológica (Ps): número de amostras de soros reagentes para anticorpos IgG para uma doença numa determinada localidade em tempo determinado.

$$Ps = \frac{(VP + FP)}{N}$$

$$\text{Prevalência verdadeira} = \frac{Ps + (E - 1)}{S + (E - 1)}$$

### VALORES PREDITIVOS

Valor preditivo é definido como a precisão de um teste prever uma condição médica.

Para respondermos a perguntas como:

- Se o resultado do teste é positivo, qual a probabilidade de que o indivíduo ensaiado tenha realmente a doença?
- Se o resultado do teste é negativo, qual a probabilidade de que o indivíduo ensaiado não tenha realmente a doença?

Para responder a essas questões, precisamos conhecer o valor preditivo positivo de um teste com resultado positivo e o valor preditivo negativo de um teste com resultado negativo.

O valor preditivo positivo (VPP) é o número de resultados verdadeiro-positivos fornecidos pelo teste dividido pelo número de todos os resultados positivos do teste. O valor preditivo negativo (VPN) é o número de resultados verdadeiro-negativos fornecidos pelo teste dividido pelo número de todos os resultados negativos do teste.

Usando o **Quadro 4.1**, podemos expressar os valores preditivos usando a fórmula simplificada do teorema de Bayes.

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{FN + VN}$$



O valor preditivo é fortemente influenciado pela prevalência da doença e pela sensibilidade e especificidade do teste.

Como exemplo, calculamos valores preditivos obtidos a partir de um mesmo teste aplicado a duas populações de área endêmica e não-endêmica para uma mesma infecção.

População 1 = prevalência de 1%

N = 10.000

S = 99%

E = 98%

P = 1%

Teste	Doença		Total
	Presente	Ausente	
Reagente	99	198	297
Não-reagente	1	9.702	9.703
Total	100	9.900	10.000

VPP =  $99/297 = 0,333 = 33,3\%$

VPN =  $9.702/9.703 = 0,999 = 99,9\%$

População 2 = prevalência de 10%

N = 10.000

S = 99%

E = 98%

P = 10%

Teste	Doença		Total
	Presente	Ausente	
Reagente	990	180	1.170
Não-reagente	10	8.820	8.830
Total	1.000	9.000	10.000

VPP =  $990/1.170 = 0,846 = 84,6\%$

VPN =  $8.820/8.830 = 0,999 = 99,9\%$

Note-se que o VPP é muito afetado pela prevalência da doença na população de estudo, enquanto o VPN não é alterado significativamente. Por isso, em imunoensaios, resultados negativos são bem aceitos, enquanto resultados positivos, freqüentemente, precisam ser confirmados.

## TEOREMA DE BAYES

Teorema de Bayes: expressões algébricas para calcular a probabilidade ou não de doença pelo resultado do teste quando são conhecidas a probabilidade do pré-teste [prevalência estimada =  $p(D)$ ], a sensibilidade e a especificidade.

$$\text{Probabilidade de doença com teste positivo} = \frac{p(D) \times VP}{p(D) \times VP + 1 - p(D) \times FP}$$

$$\text{Probabilidade de doença com teste negativo} = \frac{p(D) \times FN}{p(D) \times FN + 1 - p(D) \times VN}$$

## CURVA ROC

### (Receiver Operating Characteristic)

As características de um teste de laboratório, como sua sensibilidade, especificidade, valores preditivos positivos e nega-

tivos, não são parâmetros inalteráveis. Esses valores podem mudar, dependendo do limiar de reatividade ou ponto de corte (*cut off point*) estabelecido para o ensaio realizado, ponto a partir do qual o teste será considerado positivo. Em geral, quanto maior a sensibilidade definida para o teste, menor será sua especificidade, e vice-versa. O ponto escolhido será definido em função do objetivo do estudo.

Quando é realizada a triagem de doadores de sangue, por exemplo, faz-se necessário um teste com máxima sensibilidade (100%), uma vez que qualquer risco de doença infecciosa transmissível pelo sangue deve ser evitado, minimizando ao máximo a presença de resultados falso-negativos. Contudo, a especificidade ficará comprometida ( $< 100\%$ ), podendo ocorrer um número significativo de resultados falso-positivos que deverão ser confirmados. Se a prevalência da doença for baixa, o número de resultados falso-positivos superará o número de verdadeiro-positivos.

Já, quando a prioridade é a certeza diagnóstica, a especificidade do teste deverá ser priorizada, minimizando a chance de resultados falso-positivos, o que agrega ao resultado do teste alto valor clínico. No entanto, haverá resultados falso-negativos.

Como nenhum teste diagnóstico apresenta valores de sensibilidade e especificidade de 100%, deve-se procurar um ponto de corte que proporcione valores máximos desses parâmetros, ciente da possibilidade da ocorrência de falsos resultados positivos ou negativos, os quais serão comunicados ao usuário. Por exemplo, um teste com sensibilidade de 95% e especificidade de 98% apresentará 5% de ocorrência de resultados falso-negativos e 2% de ocorrência de falso-positivos.

É justamente para valorizar o ponto de corte que a curva ROC tem sua maior aplicação, minimizando ao máximo os eventuais falsos resultados. A curva ROC é definida como a descrição gráfica do desempenho de um teste representado pela relação entre a taxa de verdadeiro-positivos (sensibilidade) e a taxa de falso-positivos (1-especificidade). Esse procedimento visa à melhor definição da região em que existe sobreposição de dados quando são colocados em gráficos os resultados obtidos de amostras de indivíduos doentes e sadios (Fig. 4.1). Representa, portanto, a precisão intrínseca do teste, e é ideal para a comparação de testes.

Por definição, os resultados de um teste encontrados na maioria dos indivíduos saudáveis são considerados valores normais, sendo representados pela média dos valores encontrados nesse grupo acrescido de dois desvios padrões. Contudo, sempre se devem levar em conta as características do grupo considerado "normal", pois alguns valores podem sofrer variações com a idade do indivíduo, sexo, dieta etc., fazendo com que esses cortes sofram variações, dependendo da seleção feita para estudo.

Assim, o estudo de diferentes pontos de corte entre os extremos A e B (Fig. 4.1) nos permitirá determinar o valor limiar ideal de máxima sensibilidade e máxima especificidade para determinado teste.

Para tanto, podem ser elaboradas curvas ROC, construindo-se um gráfico a partir dos diferentes índices de sensibilidade e especificidade obtidos com diferentes valores de *cut off*. Para cada ponto de decisão são colocados os índices de verdadeiro-ordenadas e os índices de falso-positivos (1-especificidade) que estarão representados no eixo das abscissas, mostrando as variações ponto a ponto, para cada *cut off* possível (Fig. 4.2).<sup>6</sup>

Se um teste diagnóstico tiver acurácia perfeita, com 100% de verdadeiro-positivos, a curva ROC irá se sobrepor ao eixo y. Como na prática dificilmente encontramos tal situação, a maior



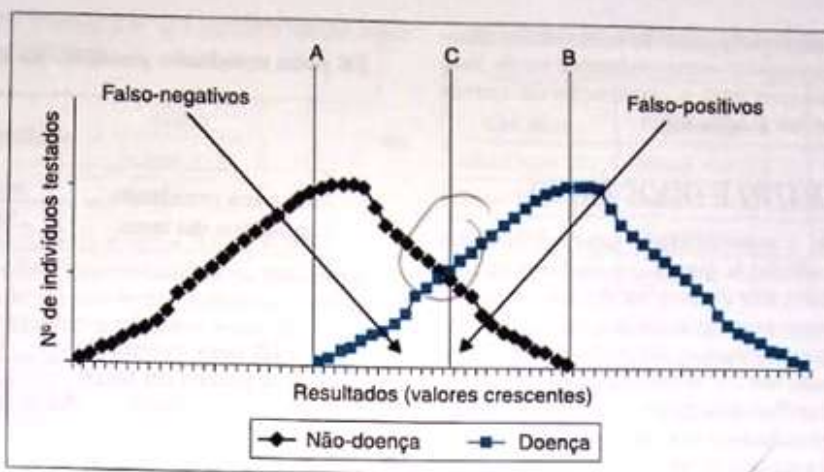


Fig. 4.1 Distribuição hipotética dos resultados obtidos em grupo de indivíduos sadios e doentes. Pontos de cortes: A = máxima sensibilidade, baixa especificidade. B = máxima especificidade, baixa sensibilidade. C = ponto de máxima sensibilidade e máxima especificidade para o teste. No *cut off* C, estão indicados os falso-positivos e os falso-negativos.

acurácia de um teste será caracterizada por uma alta proporção de resultados verdadeiro-positivos e uma baixa proporção de resultados falso-positivos, correspondendo ao ponto de corte (*cut off*) ideal na curva ROC como o ponto situado mais acima e mais à esquerda.

A acurácia global de um teste pode ser descrita como a área sob a curva ROC. Quando um modelo se torna mais perfeito (ou seja, alcançando sensibilidade e especificidade próximas a 100%), a área sob a curva aproxima-se de 1,0; quando o desempenho do modelo torna-se mais randômico, a área sob a curva aproxima-se de 0,5, representado graficamente como

uma linha de 45°; nesse caso, a capacidade de previsão do teste não é superior ao puro acaso, isto é, não possui capacidade discriminatória. Teremos então:<sup>8,9</sup>

Área sob a curva ROC	Desempenho do teste
0,7	Razoável
Superior a 0,8	Bom
Superior a 0,9	Excelente

As curvas ROC podem ainda ser utilizadas para comparar o desempenho de dois testes, analisando-se a área sob a curva ROC de cada um: quanto melhor for o teste, maior será a área sob a curva (melhor discriminação)<sup>1</sup> (Fig. 4.3). Cabe lembrar que, quando dois ou mais testes são comparados, estes devem ser avaliados sob as mesmas condições. A população estudada deve ser semelhante à que foi considerada referência na padronização do teste, em termos de estágio da doença, características e estado de saúde do grupo sem a doença. Só assim poderemos comparar a acurácia de dois testes.

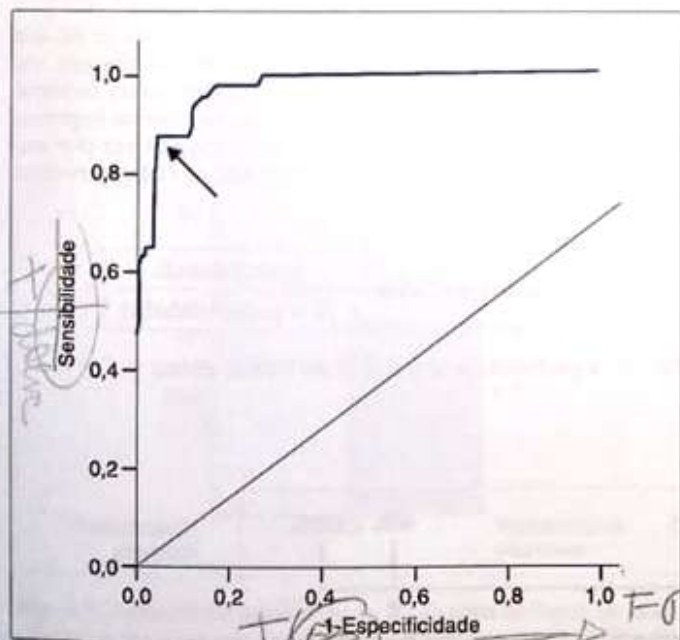


Fig. 4.2 Curva ROC – A probabilidade de resultados positivos na presença de doença está plotada no eixo y. As probabilidades de teste positivo na ausência de doença estão plotadas no eixo X. O ponto mais à esquerda e mais acima corresponde ao melhor ponto de corte para o teste em questão [maior sensibilidade (S)], menores falso-positivos (1-E), fazendo o diagnóstico com maior acurácia. A linha a 45° mostra a pior hipótese de teste.

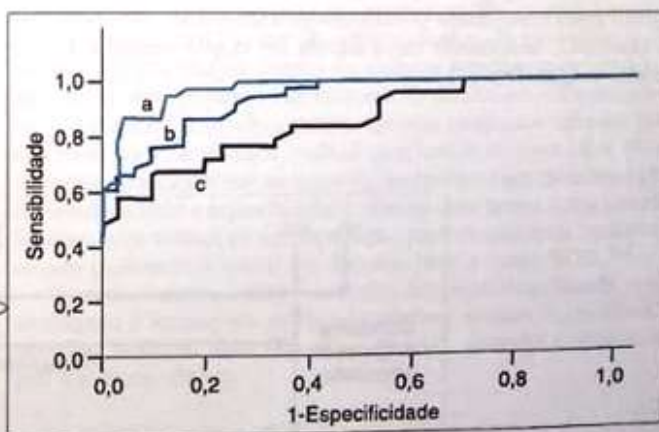


Fig. 4.3 Associação de curvas ROC: Curva "a", maior acurácia, por apresentar maior área abaixo da curva. S = Sensibilidade. 1-E = (1-Especificidade) = falso-positivos.



Atualmente existem muitos programas de informática, associados ou não ao Excel, disponíveis comercialmente ou de livre acesso na internet, que servem para a elaboração de curvas ROC para diferentes situações e aplicações.

### LIKELIHOOD RATIO E ODDS RATIO

Apresentar sensibilidade e especificidade satisfatórias é tão importante quanto saber a influência que os resultados obtidos a partir de um teste diagnóstico irão exercer na decisão clínica.

Raramente um único teste consegue excluir ou confirmar com certeza a presença de uma doença. Na maioria das vezes, um determinado exame diagnóstico é apenas uma das maneiras de aumentar ou diminuir a probabilidade de uma hipótese diagnóstica. Vários fatores são analisados conjuntamente, uma vez que diferentes doentes apresentam fatores de risco diferentes, pertencem a diferentes faixas etárias e apresentam ou não doenças associadas, com diferentes probabilidades de terem determinada doença.

Algumas vezes, o exame clínico realizado pelo médico exclui a possibilidade de uma patologia, não sendo necessária a realização de testes diagnósticos para que esta seja descartada (limiar do teste), passando então para outra hipótese. Em outra situação, os sinais e sintomas são tão evidentes, que permitem intervenção imediata (limiar de tratamento ou intervenção), dispensando qualquer exame complementar. Contudo essas são situações extremas.

Na maioria das vezes, os resultados obtidos por exames diagnósticos irão influenciar a decisão do clínico quanto à probabilidade de um determinado indivíduo apresentar ou não a doença. Um teste diagnóstico será útil apenas se os resultados obtidos modificarem a probabilidade de ocorrência ou não da doença. O teste nesse caso ajuda a transformar a probabilidade pré-teste da doença em uma nova probabilidade pós-teste, cuja magnitude e sentido de variação são determinados pela taxa de probabilidade ou *likelihood ratio* (LR) do teste, que está diretamente ligada à acurácia com a qual o teste identifica a doença em questão (Fig. 4.4).

É possível, por meio da sensibilidade e especificidade apresentadas por um determinado teste, calcular o poder que o resultado desse teste terá na mudança de opinião do clínico. O LR indicará então quantas vezes o resultado de um teste diagnóstico é capaz de mudar a probabilidade de se ter uma doença. Esse valor pode ser calculado<sup>7</sup> tanto para resultados negativos (probabilidade de excluir a doença) como para resultados positivos (probabilidade de confirmar a doença), utilizando uma tabela de dupla entrada 2 x 2, como o Quadro 4.1.

$$\text{LR para resultado positivo do teste} = \frac{[a/(a + c)]}{[b/(b + d)]}$$

ou

$$\text{LR para resultado positivo do teste} = \frac{\text{sensibilidade}}{(1 - \text{especificidade})}$$

e

$$\text{LR para resultado negativo do teste} = \frac{[c/(a + c)]}{[d/(b + d)]}$$

ou

$$\text{LR para resultado negativo do teste} = \frac{(1 - \text{sensibilidade})}{\text{especificidade}}$$

Como podemos observar, o valor de LR de um teste está extremamente ligado aos seus valores de sensibilidade e especificidade. Um teste com 50% de sensibilidade e 50% de especificidade, apresentando resultado positivo ou negativo, terá um LR = 1

$$\text{LR (+)} = 0,5/(1 - 0,5) = 0,5/0,5 = 1$$

$$\text{LR (-)} = (1 - 0,5)/0,5 = 0,5/0,5 = 1$$

Como saberemos então, a partir do valor calculado para LR de um teste, a magnitude da mudança de opinião que esse resultado provocou? Para tanto teremos que saber o conceito de *odds*. Manteremos o nome em inglês para facilitar a explicação.

*Odds* é a chance de ocorrência de um evento, traduzido literalmente como "pontos de vantagem".

Chance, ou *odds*, e probabilidade de ocorrência de um evento são relacionadas, mas não são idênticas. Chance, ou *odds*, é expressa em fração tendo o número 1 como denominador, e probabilidade de ocorrência é normalmente expressa como fração decimal, podendo ser transformada em porcentagem. Uma expressão pode ser convertida na outra conforme necessário.<sup>6</sup>

$$\text{Chance ou Odds} = \frac{\text{probabilidade}}{(1 - \text{probabilidade})}$$

Ex.: Se a probabilidade for de 0,80 (80%), então:

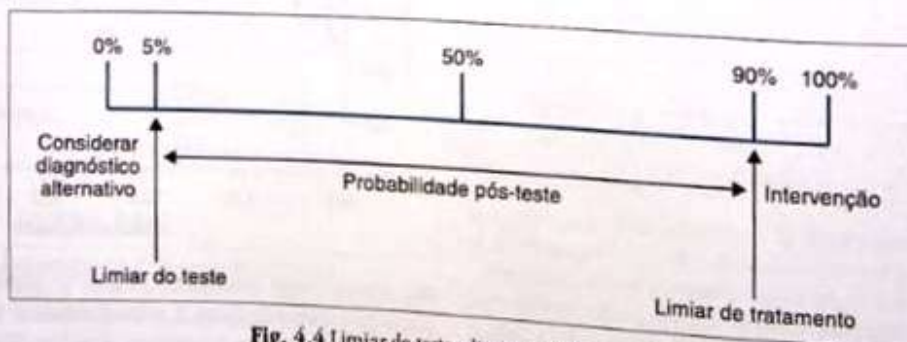


Fig. 4.4 Limiar do teste e limiar de tratamento.<sup>7</sup>



$Odds = 0,80 / (1 - 0,80) = 0,80 / 0,2 = 4/1$  (quatro vezes mais chance de ocorrência do evento)

$$\text{Probabilidade} = \frac{Odds}{(Odds + 1)}$$

Ex.: se  $Odds = 4:1$ , então:

Probabilidade =  $(4/1) / [(4/1) + 1] = 4/5 = 0,8$  (80%)

Pela *Odds* (não por valores percentuais) podemos avaliar de maneira rápida a utilidade de um teste diagnóstico para uma determinada doença, utilizando a seguinte equação:<sup>4</sup>

$$Odds \text{ Pré-teste} \times LR = Odds \text{ Pós-teste}$$

em que a *odds* pré-teste será determinada pelo clínico por meio dos dados epidemiológicos e de sinais e evidências clínicas apresentadas pelo paciente após anamnese, descrição da doença existente na literatura e experiência pessoal.<sup>5</sup>

Por meio dessa equação podemos ver que, quando o valor de LR é igual a 1, a utilidade do teste é nula, uma vez que não altera a hipótese diagnóstica inicial. Um teste diagnóstico será útil quando apresentar valores altos ou baixos de LR.

Em geral, na literatura adotam-se os seguintes intervalos:

LRs altos	LRs baixos	Efeito na probabilidade pós-teste da doença
Maior que 10	Menor que 0,1	Grande
Entre 5-10	Entre 0,1-0,2	Moderado
Entre 2-5	Entre 0,2-0,5	Mínimo
Igual a 1	Igual a 1	Nenhum

Quanto mais o valor de LR se afasta da unidade (em ambos os sentidos), mais poderoso é o teste em termos discriminativos/diagnósticos.<sup>10</sup>

Uma outra maneira de avaliarmos a utilidade de um teste diagnóstico é utilizarmos o nomograma de Fagan, o qual permite de maneira fácil, a partir do cálculo do LR e da probabilidade pré-teste, determinarmos a probabilidade pós-teste (Fig. 4.5).

No exemplo, um teste com LR = 50 altera uma probabilidade pré-teste de doença de 30% para 75% após o resultado positivo do teste.

Se quisermos comparar a acurácia de dois testes diagnósticos para a pesquisa de anticorpos anti-*T. cruzi*, numa mesma população tendo:

Teste A	Teste B
Sensibilidade de 99,9%	Sensibilidade de 98,8%
Especificidade de 99,5%	Especificidade de 99,8%

Podemos calcular o LR (+) e (-) para cada teste:

**Teste A:**

$LR (+) = 0,999 / (1 - 0,995) = 0,999 / 0,005 = 199,8$

$LR (-) = 1 - 0,999 / 0,995 = 0,001 / 0,995 = 0,001005$

**Teste B:**

$LR (+) = 0,988 / (1 - 0,998) = 0,988 / 0,002 = 494$

$LR (-) = (1 - 0,988) / 0,998 = 0,012 / 0,998 = 0,0120$

O LR(+) do teste A é de 199,8 e do teste B, 494. Esses dados significam que, após um resultado positivo no teste B, a probabilidade pós-teste da existência da doença será muito maior que a existente antes da realização do exame. O teste A também será capaz de alterar essa probabilidade, mas com um menor efeito, apesar da excelente sensibilidade do teste. Porém, se o objetivo do clínico é descartar a possibilidade da doença no mesmo paciente, o teste A seria vantajoso por apresentar um LR (-) mais baixo.

Os valores de LRs não são dependentes da prevalência da doença na população, como os valores preditivos positivos e negativos, sendo então mais estáveis. Contudo, ao compararmos os valores de LR de dois testes, apenas podemos afirmar que um teste com LR = 100 é melhor que um teste com LR = 10 se tivermos certeza de que as populações usadas para determinação da sensibilidade e especificidade desses dois testes foram semelhantes, pois variações na população podem acarretar variações nesses parâmetros, como foi descrito para a curva ROC.<sup>2,11</sup>

O conhecimento sobre o método diagnóstico utilizado, suas vantagens e limitações, pode nos ajudar a avaliar os resultados obtidos, fazendo uma diferença crucial, quando empregado por um bom clínico.

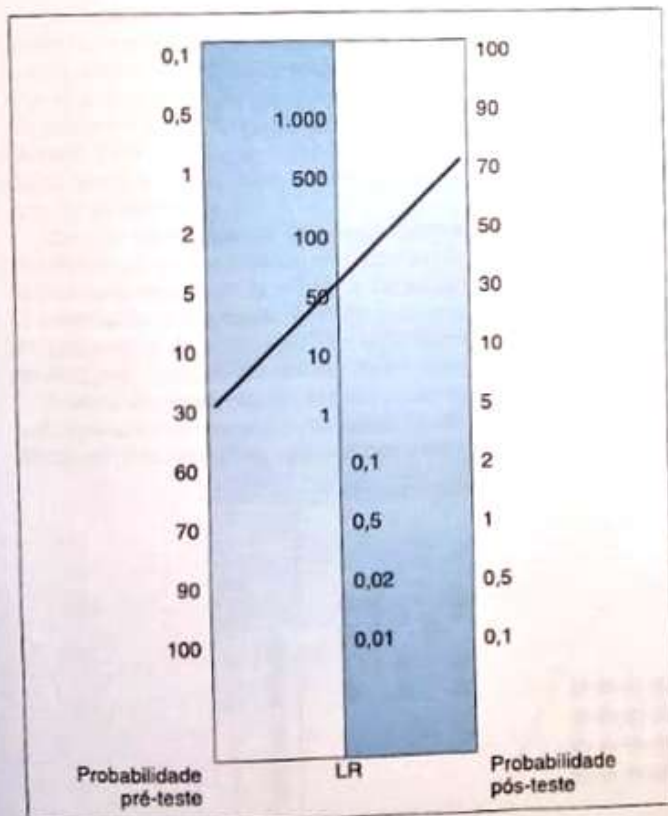


Fig. 4.5 Representação esquemática do nomograma de Fagan. Os valores plotados são apenas representativos, não correspondendo ao original. LR = *likelihood ratio* (taxa de probabilidade).

## BIBLIOGRAFIA

1. Carneiro AV. Cardiologia baseada na evidência – princípios de seleção e uso de testes diagnósticos: propriedades intrínsecas dos testes. *Rev Port Cardiol* 2001; 20:1267-74.
2. Dujardin B, Van Denende J, Van GA, Unger JP, Van der Stuyft P. Likelihood ratios: a real improvement for clinic decision? *Eur J Epidemiol* 1994; 10:29-36.
3. Ferreira AW, Ávila SLM. Sorologia: importância e parâmetros. In: *Diagnóstico laboratorial das principais doenças infecciosas e auto-imunes*, 2ª ed., 2001. Rio de Janeiro: Guanabara Koogan. p. 1-8.
4. Gambino SR. Odds, probability and likelihood ratios. *Lab Report* 1986; 8:69-71.
5. Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: getting diagnostic testing into perspective. *Q J Med* 1998; 91:247-58.
6. Hanley, JA. Receiver operating characteristic (ROC) methodology: The state of the art. *Crit Rev Diagn Imaging* 1989; 29:307-35.
7. Hayden SR, Brown MD. Likelihood ratio: A powerful tool for incorporating the results of a diagnostic test into clinical decision making. *Ann Emerg Med*. 1999; 33:575-80.
8. Kollef MH, Schuster DP. Predicting intensive care unit outcome with scoring systems. Underlying concepts and principles. *Crit Care Clin* 1994; 10:1-18.
9. Ruttimann EU. Statistical approaches to development and validation of predictive instruments. *Crit Care Clin* 1994; 10:19-35.
10. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. 2nd ed. Edinburgh: Churchill Livingstone, 2000.
11. Sox HC. The evaluation of diagnostic test. *Ann Rev Med* 1996; 47:463-71.